Trans Indian Inst Met
https://doi.org/10.1007/s12666-023-02893-6

ORIGINAL ARTICLE

In-Vitro Corrosion and Wear Studies of Ceramic Layers on Additively Manufactured Zr Metal for Implant Applications

Nimu Chand Reger[1,3] · K. Buvya Devi[2] · Vamsi Krishna Balla[3] · Mitun Das[3]

Received: 11 July 2022 / Accepted: 2 February 2023

Abstract In the present study, an attempt has been made to develop *in-situ* grown ceramic layer on additively manufactured *Zr* metal by thermal oxidation (TO) treatment. Detailed characterization and testings were performed to determine the thickness of the ceramic layer, oxide phases, hardness, surface roughness, wettability *in-vitro* wear, and *in-vitro* corrosion resistance of theses oxidized specimens. The X-ray diffraction analysis confirmed the formation of ZrO_2 in the *in-situ* oxide layer and its thickness increased significantly at higher oxidation temperatures. However, among the samples, lowest *in-vitro* wear rate ($2.12 \pm 0.36 \times 10^{-6}$ mm^3/N m) was demonstrated by the samples oxidized at 600 °C for 6 h. Further this obtained wear rate was correleted with thickness of oxide layer, contact angle, surface rougness, and hardness. It is also noticed that the formation of oxide phases on Zr significantly increase the *in-vitro* corrosion resistance compared to untreated Zr substrate in Hanks Balanced Salt solution (HBSS).

Keywords Zirconium · Thermal oxidation · ZrO_2 and Zr_2O · Hardness · Contact angle · *In-vitro* wear · Corrosion resistance

✉ Nimu Chand Reger
nimu@nitt.edu

1 Department of Metallurgical and Materials Engineering, National Institute of Technology Tiruchirappalli, National Highway 67, Tanjore Road, Near BHEL, Thuvakudi, Tamil Nadu, India

2 Department of Chemistry, Thassim Beevi Abdul Kader College for Women, Kilakarai, India

3 Bioceramics and Coating Division, CSIR-Central Glass and Ceramic Research Institute, 196 Raja S.C. Mullick Road, Kolkata 700 032, India

Published online: 17 March 2023

1 Introduction

Compared to metallic implants such as 316 L stainess steel, Co–Cr–Mo alloys, Ti, Zr, and their alloys, ceramic coated metallic implants exhibit better corrosion and wear resistance because of their high hardness, low surface roughness, lower coffecient of friction, low contact angles [1]. Among various ceramic coating, zirconia-coated implants are vital for the articulating surfaces on the hip, knee, and shoulder. Over 3.5 million alumina components and over 6,00,000 zirconia femoral heads have been implanted worldwide as a result of the clinical success associated with the use of ceramics [2]. For total hip arthroplasty replacement application, early-stage wear of traditional PE with Co–Cr and zirconia femoral heads was examined, and it was discovered that the wear was lower with zirconia heads (0.08 mm/year) than with Co–Cr heads (0.17 mm/year) [3]. It was reported that the plasma electrolytic oxidation technique improves the hardness four-fold and wear resistance too by forming a well-adhered thick ceramic layer onto the Zr substrate [4].

Recently, oxidized Zr (OxZr) is manufactured using the commercial technique known as OXINIUM. In this the oxide film is *in-situ* zone of the metal and is not a coating which is enabled as a scratch-resistant counter face for a total joint replacement prosthesis [5]. Likewise, Balla et al. [6] successfully demonstrated the formation of strong and adhered *in-situ* developed ZrO_2 coating on Zr substrate by laser-processed oxidation to improve wear resistance and wettability in a physiological environment for implant applications. At present total hip and knee replacement implants are typically manufactured by conventional processing such as machining, casting, and forging. These processes still lacking due to a large amount of material wastage, not capable to produce complex shapes with micro/macrostructure,

Review

THE CHEMICAL RECORD

doi.org/10.1002/tcr.202200136
tcr.wiley-vch.de

Magnesium Phosphate Bioceramics for Bone Tissue Engineering

K. Bavya Devi,[a] V. Lalzawmliana,[b] Maktumkari Saidivya,[c] Vinod Kumar,[d] Mangal Roy,*[c] and Samit Kumar Nandi*[e]

Abstract: Magnesium phosphate (MgP) is a family of newly developed resorbable bioceramics for bone tissue engineering. Although calcium phosphates (CaP) are the most commonly used bioceramics, low solubility, and slow degradation, when implanted *in vivo*, are their main drawbacks. Magnesium (Mg) is an essential element in the human body as it plays important role in bone metabolism, DNA stabilization, and skeletal development. Recent research on magnesium phosphates has established their higher degradability, *in vitro*, and *in vivo* biocompatibility. Compared to CaP, very limited research work has been found in the area of MgP. The prime goal of this review is to bring out the importance of magnesium phosphate ceramics for biomedical applications. In this review, we have discussed the synthesis methods, mechanical properties, *in vitro* and *in vivo* biocompatibility of MgP bioceramics. Moreover, we have highlighted the recent developments in metal ion-doped MgPs and MgP scaffolds for bone tissue engineering.

Keywords: Bone regeneration, Bioceramics, magnesium phosphates, *in vitro* and *in vivo* biocompatibility, bioresorbable materials

1. Introduction

The human skeletal system along with muscle attachments provides support and protection for the body.[1] The skeleton of an adult human consists of an average of 206 bones. Bone is made up of about 70 % of inorganic calcium phosphates (CaP) reinforcement and 30 % of the organic collagen matrix.[2] Thus bone is a natural composite with collagen as matrix and calcium phosphates as reinforcement, which provides strength and hardness.[3] Bone is a dynamic living tissue, which consists of living cells that build the bone, namely osteoblasts and osteocytes, while bone also has other living cells of osteoclasts that resorb old bone during bone remodeling. Bone is constantly reconstructed, with the osteoblasts and osteoclasts cells, it takes nearly 3 years to remodel the old bone. Bone is one of the few human tissues that can regenerate themselves. Because of age factors, few diseases and other factors bone cannot regenerate by itself.[4]

To overcome bone disorders and maintain bone health, the demand for synthetic bone substitute materials with suitable physicomechanical and biological properties is increasing. Metallic, polym[illegible] and ceramics are currently being used as synthetic bone substitute materials. Among the broad area of synthetic biomaterials, ceramic materials are explored in great detail for use in orthopedic applications. These bioceramics can be bioinert (alumina, zirconia, or a composite), bioresorb-

[a] *Dr. K. Bavya Devi*
Department of Chemistry, Thassim Beevi Abdul Kader College for Women, 623517 Kilakarai, Ramanathapuram, India
[b] *Dr. V. Lalzawmliana*
Department of Veterinary Surgery and Radiology, College of Veterinary Sciences and Animal Husbandry, 799008 R. K. Nagar, Tripura West, India
[c] *M. Saidivya, Dr. M. Roy*
Department of Metallurgical and Materials Engineering, Indian Institute of Technology-Kharagpur, 721302 Kharagpur, India
E-mail: mroy@metal.iitkgp.ac.in
[d] *Dr. V. Kumar*
Department of Veterinary Clinical Complex, Faculty of Veterinary & Animal Sciences, Banaras Hindu University, pin-221005 Mirzapur, India
[e] *Prof. S. Kumar Nandi*
Department of Veterinary Surgery and Radiology, West Bengal University of Animal and Fishery Sciences, 700037 Kolkata, India
E-mail: samitnandi1967@gmail.com

Structural Chemistry
https://doi.org/10.1007/s11224-022-02085-w

ORIGINAL RESEARCH

The structure, stability, thermochemistry, and bonding in SO_3-$(H_2O)_n$ ($n = 1–7$) clusters: a computational analysis

Ambigapathy Suvitha[1,2] · Natarajan Sathiyamoorthy Venkataramanan[3] · Ryoji Sahara[4]

Received: 5 September 2022 / Accepted: 20 October 2022

Abstract

The structure, stability, and intermolecular interactions in SO_3-$(H_2O)n$ ($n = 1–7$) clusters were investigated using density functional and wave functional methods. The putative global minimum shows the SO_3 molecule tends to be on the surface water clusters. The increase in the number of water molecules chalcogen bond distance between water molecules and SO_3 decreases, while the maximum number of water molecules coordinated to the SO_3 molecule remains at three. The calculated solvation energy increases with the increase in the number of water molecules, and it does not saturate, which indicates that the addition of water molecules can add up to the existing water cluster network. The interaction energy between water molecules and SO_3 was less than the solvation energy conforming to the cluster forming of water molecules. The Gibbs free energy and entropy values decrease with the increase in cluster size, signifying the amount of water molecule decide the sequential hydration process. Thermochemistry data at various temperatures show that low-temperature regions found in the upper part of the troposphere favor hydration formation. Molecular electrostatic potentials (MESP) show reduced $V_{s,max}$ value of π-hole on sulfur atom and increased value on hydrogens of water molecules which results in the addition of water which leads to the sequential addition of water molecules to the water network. The quantum theory of atoms in molecules (QTAIM) shows the presence of S···O, O···H interactions between SO_3 and water molecules. Between water molecules O···H, H-bonding interactions were observed, and in larger clusters, O···O interaction was also noticed. QTAIM analysis shows that the water–water HBs in these clusters are weak H-bond, while the SO_3-water interaction can be classified as medium H-bonds which was further supported by the NCI and 2D RDG plots.

Keywords DFT · Hydrogen bonding · Non-covalent interactions · AIM · Chalcogen bonding

Introduction

Sulfur is one of the most abundant elements found in the universe and the earth's crust. Sulfur from land and sea enters the atmosphere through emissions and is converted to sulfur oxides [1]. These sulfur oxides particularly SO_2 and SO_3 are hazardous when emitted into the atmosphere and affect air quality, the environment, climate, and human health [2]. Once into the atmosphere, the SO gets oxidized in SO_2 and finally to SO_3 which acts as a source of H_2SO_4. The formed H_2SO_4 plays an important role in aerosol formation and is also one of the primary constituents of acid rain. In the stratosphere, UV photolysis of SO_2 generates SO_3 which rapidly reacts with water and gets converted into H_2SO_4 [3, 4]. It has been shown experimentally that the hydrolysis of SO_3 is catalyzed by the presence of a third particle and is second-order with respect to the partial pressure of water [5]. It is believed that the reaction proceeds via an intermediate complex $SO_3.H_2O$. Furthermore, the conversion of the aqueous complex to sulfuric acid is fast compared to the reversion reaction. To account for the negative temperature dependence of the above reaction, it is essential to estimate the enthalpy of activation of the above reaction. To accurately

✉ Natarajan Sathiyamoorthy Venkataramanan
nsvenkataramanan@gmail.com

Ambigapathy Suvitha
suvithaa@gmail.com

1 Department of Chemistry, Thassim Beevi Abdul Kader College for Women, Kilakarai 623 517, India

2 SHARP Substratum, Adavathur East, Trichy 620 102, India

3 Center for Computational Chemistry & Materials Science (CCC&MS), Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai 602 105, Tamil Nadu, India

4 Research Center for Structural Materials, NIMS, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

Published online: 28 October 2022

Journal of Inclusion Phenomena and Macrocyclic Chemistry
https://doi.org/10.1007/s10847-022-01144-3

ORIGINAL ARTICLE

Molecular insights into the complex formation between dodecamethylcucurbit[6]uril and phenylenediamine isomers

Venkataramanan Natarajan Sathiyamoorthy[1] · Ambigapathy Suvitha[2,3] · Ryoji Sahara[4]

Received: 4 January 2022 / Accepted: 28 April 2022

Abstract

The complexation behavior of diprotonated phenylenediamine isomers with decamethylene cucurbit[6]uril ($Me_{12}Q[6]$) in 1:1 and 1:2 stoichiometry was investigated in vacuum and solution using density functional theory (DFT). Among the isomers, *o*-phenylenediamine (*o*PDA) forms an exclusion complex in 1:1 stoichiometry, while others form an inclusion complex. In the 1:2 stoichiometry, at least one of the guest molecule lie on the surface of the $Me_{12}Q[6]$ cavity. The structural reorganization was found to depend on the mode of interaction of the guest molecule. In the gas phase, the enthalpy and Gibbs free energy are negative for all the complexes demonstrating the encapsulation process is spontaneous and thermodynamically favorable. In the solution phase, the enthalpy and entropy are negative for complexes with guest molecules *meta*- and *para*-isomers of phenylenediamine. For *o*PDA as a guest, the enthalpy and entropy are positive indicating the complex formation to be nonspontaneous. The MESP isosurface of complexes shows a higher accumulation of charge in the 1:2 stoichiometry, which reduces their stability. AIM analysis shows the interaction between *o*PDA and $Me_{12}Q[6]$ is due to hydrogen bonding with moderate strength, while for the other isomers, interactions between the benzene group and the $Me_{12}Q[6]$ cavity were noticed. EDA analysis shows the larger contribution is the electrostatic attraction and it decreases with the increase in the guest ratio. The formation of inclusion complexes by mPDA and pPDA and the surface adsorption of *o*PDA on $Me_{12}Q[6]$ and the higher accumulation of positive charge during solvation in *o*PDA@$Me_{12}Q[6]$ complexes make them labile in the solution phase.

✉ Venkataramanan Natarajan Sathiyamoorthy
nsvenkataramanan@gmail.com

1 Department of Chemistry, K.Ramakrishnan College of Engineering, Samayapuram, Trichy 621 112, India

2 SHARP Substratum, Adavathur, Trichy 620102, India

3 Department of Chemistry, Thassim Beevi Abdul Kader College for Women, Kilakarai 623 517, India

4 Research Center for Structural Materials, NIMS, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

Published online: 29 May 2022

International Journal of Research in Engineering, Science and Management
Volume 6, Issue 1, January 2023
https://www.ijresm.com | ISSN (Online): 2581-5792

Extraction of Eco-Friendly Natural Dye from Variety of Vegetables

J. Janani[1*], N. Priya[2], M. Kokila[3]
[1,2,3] Assistant Professor, Department of Chemistry, Thassim Beevi Abdul Kader College for Women, Ramanathapuram, India

Abstract: **A large number of experiments have been performed of natural dyes from plants and vegetable wastes. Natural dyes helps in the reduction of pollution level in textile dyeing, in realization of eco- friendly agriculture productions, and pollution. The main idea of extracting dyes from natural sources is to avoid the environmental pollution. Present days with global concern over the use of eco-friendly and bio degradable materials, considerable research work is being undertaken around the world on the application of natural dyes in textile industry. The present study deals with the extraction of natural dye from this species, and their application on textiles. Pre mordant technique methods for extraction of dye from collected vegetables were evaluated to determine. The cotton types of fabrics were used in the experiment to observe the strength of dye. This method following steps: vegetables are naturally dried, and are added into deionized water for soaking then baking soda is added for regulating the pH value to 6-8, the temperature is increased to 70-100 and is maintained for 60-90 min, and the obtained product is cooled and filtered, and residue is removed. The vegetable dye is obtained. The various extract were analysed phytochemical compounds. The extracts were then treated with cotton fabric using the various chemical mordants. Based on analysis, the best mordant and technique for using mordant were determined for dying fabric.**

Keywords: **Natural dye, Vegetable dye, Eco-friendly dye.**

1. Introduction

The main idea of extracting dyes from plants. Source is to avoid the environmental pollution. Present days with global concern over the use of eco- friendly and biodegradable materials, considerable research work is being undertaken around the world on the application of natural dyes in the textile industry, the effluent problem of synthetic dyes occur not only during their application in the textile industry, but also during their manufacture and possibly during the synthesis of their intermediates and other raw materials. The e of natural dyes for textile dyeing purpose, decreased to a large extent after the discovery of synthetic dyes in 1856. Dyes derived from natural source have emerged as an important alternative to synthetic dyes. Analytical studies such as IR spectrophotometry were performed on the extract. Natural dyes are dyes or colorant derived from plants, invertebrates, or minerals. The majority of natural dyes are vegetable dyes from plant sources – roots, berries, bark, leaves, and wood and other biological sources such as fungi.

*Corresponding author: janujega1995@gmail.com

Beetroot:

Betalain pigment can be used as an unnatural additive for food, cosmetics, and drugs. Beet powder can be a good source of antioxidants. To increase the acceptability of food. Beetroot dyes are less toxic. Beetroot are a great source of fiber, folate manganese, potassium, iron, and vitamin C. Beetroot juice have been associated with numerous health benefits, including improved blood flow, lower blood pressure. Beetroot are a unique source of betalains - a type of phytonutrient. These compounds are known to provide powerful antioxidant, anti-inflammatory and detoxification. It prevents fatty acids from building up and stimulates the liver, Cells. Now a days growing demand for eco- friendly / non- toxic colorants, specifically for health sensitive application such as coloration of food.

Onion:

Plant have been used for thousands of years to make natural dyes. Onion skins are one of the best natural source. In recent years, the textile industry has been severely criticized for their contribution to environmental pollution. The dry outer skins of onions can be used for coloring natural textile materials. Onion skins are simple for a few reasons, they are easy to source, they are food safe, and they do not require the aid of a mordant to achieve colorfast fabric. Onion skins do not need a mordant because they are naturally high in tannin, which binds of the color to the fabric, creating lasting, colorfast fabrics. Onion skin create an earthy range of colors. Protein fibers such as wool and silk, dye pale to medium nutmeg brown. Cellulose fibers such as cotton, hemp, with a mix out of champagne, pale, and silver pink. It is possible to extract natural dyes from byproduct such as food, wood and agricultural waste at lower costs, these extracts have other interesting properties that allow the functionalization of textiles with UV protection and antimicrobial activity.

Carrot:

In India, it was widely used for coloring of fabrics and other materials. In order of meet the growing demand for natural colorants, new pigment crops are being sought. Natural dyes from plants are of increase attention as alternate source for synthetic dye in the food. And pharmaceutical industry and they increase their added value if they possess positive effects on health and some added qualities like antifungal properties. The ancient people exclusively used dyestuffs of vegetable, minerals, and animal origin, all easily obtained in their own

Original Article

Imperative Linear Algebra for Data Science with R-Software

[1]R. Nagarathinam, [2]R. Punitha, [3]M. Nazreen Banu

[1]Associate Professor, Department of Mathematics, Dr.MGR Educational and Research Institute, Chennai, nagarathinam.hs@drmgrdu.ac.in

[2]Associate Professor, Department of Mathematics, Thassim Beevi Abdul Kader college for Women, Kilakarai, Tamilnadu, India, varshre@gmail.com

[3]Associate Professor, Department of Mathematics, Thassim Beevi Abdul Kader college for Women, Kilakarai, Tamilnadu, India, nazz_reen@yahoo.com

Abstract

Data science and machine learning are built on linear algebra. Machine learning and data science make extensive use of linear algebra, a branch of Mathematics. Machine learning relies heavily on linear algebra. Matrix representations are commonly used in machine learning models. Using linear algebra in data science means regularizing, reducing to dimensions, recognizing images, learning algorithms, and analyzing images. Many data science algorithms are based on linear algebra. This article will cover three uses of linear algebra in three different data science domains. We will discuss loss functions from the perspective of machine learning, and image convolution from the perspective of computer vision. Any prospective data scientist must learn R since it is a very strong language designed specifically for data analysis and data visualisation. With linear algebra, R is extremely useful. It has built-in data types like matrices and vectors.

Keywords: Linear Algebra, Data Science, Algorithms, R-Software.

Introduction

The area of mathematics known as "linear algebra" deals with linear equations and how they are represented in vector spaces and through matrices. The study of vectors and linear functions is what linear algebra is, to put it simply. Linear algebra involves the study of matrices, determinants, linear transformations, vector spaces, and subspaces and uses closed vectors that operate under addition and scalar multiplication. It enables us to do mathematical operations and comprehend geometric notions in greater dimensions. Nearly all branches of mathematics, including geometry and functional analysis, are based on linear algebra. Understanding its ideas is essential for comprehending the theory underlying data science.

Data science is a branch of study that combines subject-matter knowledge, programming abilities, and a working understanding of mathematics and statistics to draw out valuable insights from data.

Data scientists use machine learning algorithms to analyse data from a variety of sources, including text, images, videos, and audio, to create artificial intelligence systems that can carry out activities that often require human intellect. To study and analyse real-world phenomena using data, "data science" is a concept that combines statistics, data analysis, informatics, and their related methodologies. In the context of mathematics, statistics, computer science, information science, and domain knowledge, it makes use of methods and theories from a variety of domains. A data scientist is a person who writes programming code and uses statistical expertise in conjunction with it to derive insights from data.

Address for correspondence: R. Nagarathinam
Associate Professor, Department of Mathematics, Dr.MGR Educational and Research Institute, Chennai
Email: nagarathinam.hs@drmgrdu.ac.in

Access this article online

Quick Response Code:

Website: www.pnrjournal.com

DOI: 10.47750/pnr.2022.13.04.255

How to cite this article: R. Nagarathinam, R. Punitha, M. Nazreen Banu, Imperative Linear Algebra for Data Science with R-Software, J PHARM NEGATIVE RESULTS 2022;13(4): 1852-1857.

Data scientists may be able to avoid using linear algebra for a while, but not for very long. Here are several ways that linear algebra can help with computer vision and machine learning. Few people think of linear algebra when discussing data science in general or specific subfields like machine learning or computer vision. Because the modern tools we use to perform data science algorithms do a great job of disguising the underlying math that makes things function, linear algebra is sometimes overlooked. In data science, linear algebra is used.

Impact of Linear Algebra in Data Science

Area of mathematics called linear algebra is very helpful in machine learning and data science. The following data science fields make use of linear algebra:

- Regularization
- Reduction to dimension
- Image recognization
- Working with data sets
- Machine learning
- Computer vision

REGULARIZATION

Regularization is one of the biggest hurdles in machine learning, especially for beginners in data science. It is when a model is too close a fit for the available data, to the point that it does not perform well with any new or outside data. A concept called "regularization" is used to prevent the model from overfitting. This concept makes use of linear algebra as it uses the "norm". The norm can be defined simply as the magnitude of a vector. This magnitude can be calculated in various ways one popular way is using the Euclidean Distance. ie, using the distance from the origin. Regularization prevent overfitting as it adds the norm of the weight vector to the cost function. This makes sure that model does not become overly complex as aim is always to reduce the cost function, and therefore have to reduce this norm. This is much better understood by someone who knows the basics of linear algebra.

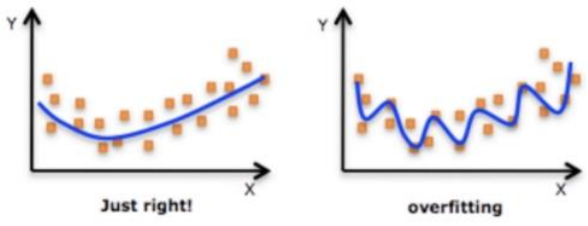


Figure 1

REDUCTION TO DIMENSION

While making Machine Learning Models, often come across data that is made up a hundreds or even thousands of variable. Our model becomes more and more complicated as these variables increase. Dimensionality reduction is the technique that reduces the number of input variables in data set. Since datasets can be easily represented as matrices, certain matrix factorization methods can be used to reduce a matrix and hence the dataset into its constituents parts. Then any operation that used to be performed on the original matrix, could be performed on the smaller matrices. Decomposition method like LU matrix decomposition and QR matrix decomposition can be easily performed using python programming.

LU Decomposition

Consider the system of equation written as a matrix equation:

$X_1+X_2-X_3=4$

$X_1-2X_2+3X_3=-6$

$2X_1+3X_2+X_3=7$

$$\begin{bmatrix}1 & 1 & 1\\ 1 & -2 & 3\\ 2 & 3 & 1\end{bmatrix}\begin{bmatrix}X1\\ X2\\ X3\end{bmatrix}=\begin{bmatrix}4\\ -6\\ 7\end{bmatrix}$$

$$AX = B$$

We can solve the system using LU Decomposition

Let A= LU and substitute into AX=B

Solve LUX=B FOR X to solve the system

Let UX=Y

LY=B and UX=Y

First solve LY=B and Y and the solve UX=Y for X

LINEAR ALGEBRA IN IMAGE RECOGNIZATION

When implementing data science models, especially in deep learning, it come across data in the form of image, however we cannot just pass an image to a model and expect it to understand it. It need to convert each image into something Mathematical or Statistical to be understood by the model. This is where linear algebra come in.

IMPACT OF LINEAR ALGEBRA IN IMAGE PROCESSING

Image processing is the manipulation of images using mathematical operations. With the introduction of computers, processing is now done using computer graphic algorithms on digital images obtained through a digitization process or directly using any digital device. Digital image processing is the use of a computer to perform image processing on digital images. Linear algebra can be used to perform computer graphics operations such as rotation, skewing, scaling, Bezier curves, reflections, dot and cross products, projections, and vector fields. Other more complex operations, such as filters, necessitate the use of linear algebra in conjunction with other mathematical tools. Linear algebra deals with matrices and all the operation to be

Performed on matrix. Any image is made of pixels, which are nothing but coloured square of varying Intensitics (for gray scale image it could be a single number with the intensity, and for coloured images, it could be the RGB value).

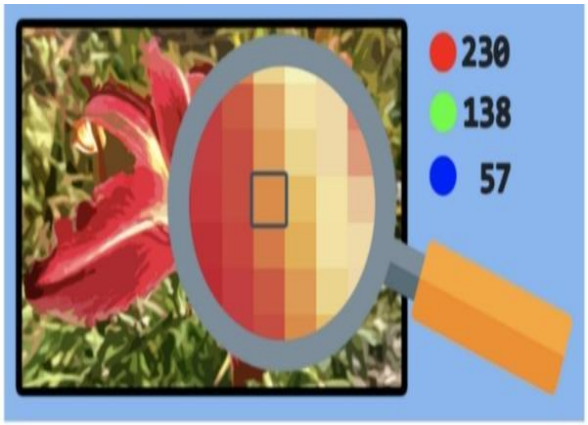


Figure 2

WORKING WITH DATA SETS

When building a machine learning model, it will most probably dealing with large data sets having multiple rows and columns. These are nothing but matrices when you spilt your dataset into training and testing data, you are performing operations on these matrices. Matrices are the key data structure in linear algebra and it deals with the various operations performed on matrix, including row and column transformations, transpose of a matrix, addition or scalar multiplication in matrices

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

This is nothing but a 5*4 matrix each record is a row and is indexed with the numbers 0,1,….4 . Each column has its name on top.

MACHINE LEARNING

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programmes can predict outcomes more accurately without having to be explicitly instructed to do so. Machine learning algorithms predict new results using historical data as input. The way in which a machine learning algorithm learns to improve its prediction accuracy is a common way to classify traditional machine learning. There are four fundamental strategies: reinforcement learning, semi-supervised learning, unsupervised learning, and supervised learning. The kind of data that scientists wish to predict determines the kind of algorithm they use.

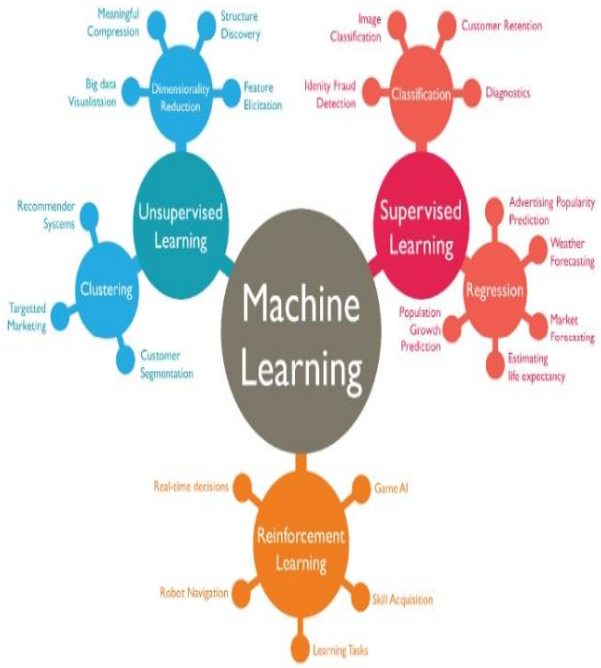


Figure.3

LINEAR ALGEBRA IN MACHINE LEARNING

Vectors, matrices, and linear transforms are the focus of the mathematics subject known as linear algebra. From the notation used to describe the operation of algorithms through the implementation of algorithms in code, it serves as a crucial basis for the field of machine learning. Despite the fact that linear algebra is essential to the subject of machine learning, the close connection is frequently ignored or described using impersonal ideas like vector spaces or certain matrix operations.

The following are significant fields of application made possible by linear algebra:

Data representation

Representation using learnt models

Eigenvectors for word embeddings

REPRESENTATION OF DATA

Data, which serves as the fuel for ML models, must first be transformed into arrays before being fed to your models. Matrix multiplication is one of the operations carried out on these arrays (dot product). Additionally, the output is returned, which can also be seen as a changed matrix or tensor of integers.

EMBEDDING WORDS

Just above it is the idea of using a lesser dimensional vector to express large-dimensional data (consider a large number of variables in your data).

EIGENVECTORS (SVD) (SVD)

Principal component analysis allows us to decrease the number of characteristics or dimensions in the data without sacrificing the quality of any of them.

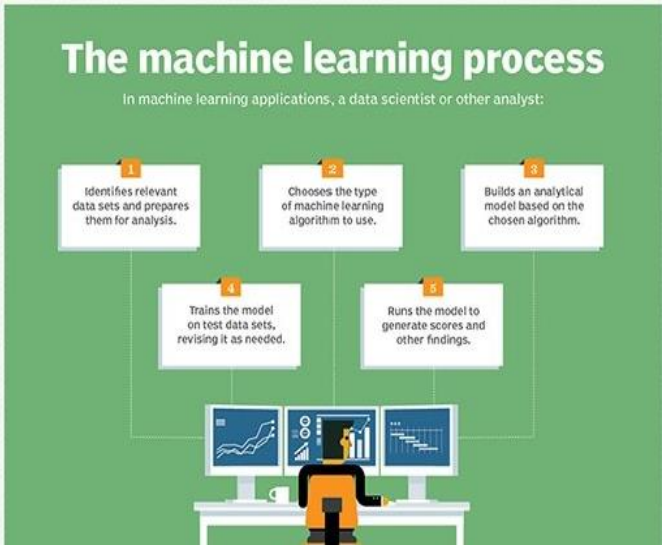


Figure 4

LINEAR ALGEBRA IN COMPUTER VISION

For machine learning, data science, and other related fields like computer vision and natural language processing, linear algebra is a potent tool. With its own implementations in the form of matrices, vectors, and tensors, linear algebra is used in computer vision. This includes basic operations, such as linear transformation, matrix operations, linear combination, and dependency of variables.

Significant contributions from linear algebra have been made to the difficult calculations used in computer vision. Large and complex matrices are necessary for the computations and operations involved in the complex matrix multiplications used in computer vision techniques. The computer vision system uses a variety of techniques to extract data from images, including compression, rotation, flip-flopping, convolution, noise reduction, object detection, facial recognition, etc.

The study of visual knowledge extraction is known as computer vision. and the fundamental idea of linear algebra is applied to the extraction. Vectors, matrices, and tensors are LA concepts.

VECTOR

A vector is a type of 1D array that is typically described with magnitude and direction.

MATRIX

It is a 2D array of numbers called a matrix. Think about an image's matrix-based pixel representation as an example. Projections, translations, rotations, scaling, and affine transformations are some of the operations of matrices.

TENSOR

It is a generalisation of matrices and vectors.

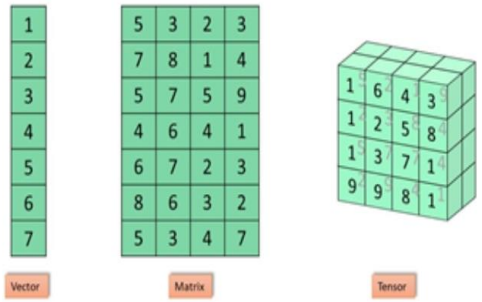


Figure. 5

VECTOR PRODUCT: MATRIX

The identity matrix is created by combining the unit vectors e1,..., em. A simpler approach to refer to random vectors a1,...an from the same vector space in a matrix is to use the letter I.

VECTORS AND MATRICE IN DATA SCIENCE

The some essential of linear algebra in the context of data science applications.

Vectors organize information that cannot be expressed as a single number and for which there exist a concept of scaling and addition. Matrices group together multiple vectors. The matrix –vector product expresses a linear combination of the column vectors of the matrix. solving a linear system

Ax=b=Ib, to find x€Rm for given b€Rm , re –expresses the linear algebra b= b1e1+…bmem ,

I=[e1e2….em] , as another linear combination b=x1a1 + x2a2+…..xnan , A=[a1a2…an] for certain problems the linear combination Ax might be more insightful, but the above transformation is information preserving, with b1x both having the same number of components

Finding the best approximation of some given b€Rm by a linear combination Ax of the column vector of A€Rm*n is known as a least squares problem and transforms the information from the m components of b into n components of x, and knowledge of the form of the column vectors. If m>n and the form of the column of A can be successfully stated, the transformation compresses information.

USAGE OF R SOFTWARE

COMMANDS TO FIND MULTIPLICATION OF MATRIX USING R

```
>
> #Multiplication of matrix
>
> A<-matrix(c(11,12,13,14,15,16,17,18,19),nrow = 3,byrow = T)
> B<-matrix(c(20,21,22,23,24,25,26,27,28),nrow = 3,byrow = T)
>
> A*B
     [,1] [,2] [,3]
[1,]  220  252  286
[2,]  322  360  400
[3,]  442  486  532
```

COMMANDS TO FIND INVERSE OF MATRIX USING R

```
> #Inverse of matrix
>
> B<-matrix(c(30,31,40,41,50,51,60,61,70),nrow = 3,byrow = T)
>
> A<-solve(B)
> A
            [,1]    [,2]        [,3]
[1,] -0.16208333 -0.1125  0.17458333
[2,] -0.07916667  0.1250 -0.04583333
[3,]  0.20791667 -0.0125 -0.09541667
>
```

COMMANDS TO FIND DETERMINANT OF A USING R

```
>
> #Determinant of A
> det(A)
[1] -0.0004166667
> #Calculating eigenvalues and eigenvectors
> A<-matrix(c(30,31,40,41,50,51,60,61,70),nrow = 3,byrow = T)
> e <- eigen(A)
> e$values
[1] 147.737576   5.317459  -3.055035
```

COMMANDS TO FIND EIGENVECTOR USING R

```
> e$vectors
           [,1]       [,2]        [,3]
[1,] -0.3948374  0.4437557 -0.74478185
[2,] -0.5497457 -0.8199420 -0.06303763
[3,] -0.7361271  0.3616296  0.66432391
> |
```

EIGENVECTORS APPLICATIONS IN DATA SCIENCE

The principal component analysis of a machine learning method makes use of the idea of an eigenvector.

Assume you have data that is very high in dimensionality and has a lot of features. It's possible that the data contains redundant features. A lot of features will also decrease efficiency and take up more disc space. In this case, PCA eliminates some of the less significant features; eigenvectors save the day. Let's go over the PCA algorithm. Let's say we wish to condense a "n"-dimensional dataset into a "k"-dimensional dataset. We'll proceed in stages.

Step 1: The data are subjected to mean normalisation and feature scaling.

Step 2: We find out the covariance matrix of our data set.

Step 3: Finding the covariance matrix's Eigenvectors is step three. It is a sophisticated statistical notion. We will discover "n" Eigenvectors matching "n" Eigenvalues because our data is in "n" dimensions.

Step 4: The fourth step entails choosing "k" Eigenvectors that correspond to the "k" biggest Eigenvalues and creating a matrix in which each Eigenvector is a column. It's time to locate the less data points now. Let's say you want to shrink a data point from the data set, "a," to dimensions of "k." Multiply the dimensions of the matrix U and transpose it.

After discussing Eigenvectors, let's move on to a more complex and valuable idea in linear algebra known as singular value decomposition, or SVD for short. To fully comprehend it, a thorough investigation of linear

SINGULAR VALUE DECOMPOSITION

Consider receiving a feature matrix A. We divide our matrix A into three constituent matrices for a specific purpose, as the name would imply. Additionally, it has been asserted on occasion that SVD is a generalisation of Eigen value decomposition. A data set's redundant characteristics are removed using SVD. Let's say you have a data set with 1000 features. There will undoubtedly be redundant features in any real data set with this many attributes.

Figure. 6

This tiger can be rendered in monochrome and seen as a matrix whose elements stand in for the pixel intensity and pertinent position. In other words, the matrix comprises data in the form of rows and columns concerning the intensity of pixels in the image. This image displays many images with various resolutions that correlate to various levels. Just assume for the time being that a higher rank indicates that there is more information about pixel intensity.

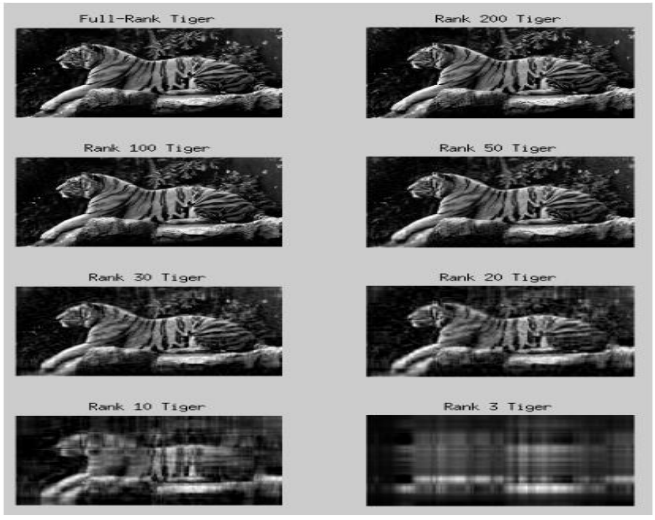


Figure. 7

It is obvious that we can achieve a reasonably good image with 20 or 30 ranks rather than 100 or 200 ranks, and in a scenario of highly redundant data, that is what we want to do. What I'm trying to say Is it true that we don't need to save every piece of data from the original dataset in order to generate a valid hypothesis? However, some of the features make it difficult to determine the optimum algorithm. For instance, multicolinearity in linear regression is caused by the presence of redundant characteristics. Additionally,

some qualities don't matter for our model. The method fits better, is more time efficient, and uses less disc space when these features are removed. The application of singular value decomposition

Conclusion

In the actual world, linear algebra is quite useful. In data science, linear algebra techniques are utilised to increase algorithm performance and produce more accurate findings. In this paper, it has gathered thatarethe uses of linear algebra in data science and provided an overview of each technique. To analyse the data sets, the data scientists can utilise linear algebra as a tool. Given the continually growing search outputs and the accessibility of the available evidence, which is a specific issue for the study sector in terms of quality improvement, machine learning algorithms are of special relevance. Regularization, dimension reduction, image identification, machine learning, and computer vision were all topics I covered. R is very useful for linear algebra.

References

[1] Sorin Mitran - Linear algebra for data science – University of North Carolina at chapel hill
[2] Herstein. L.N (2016) topic in algebra, second edition, wiley student edition
[3] Santiago M.L(2001) Modern algebra, Tata Megraw- hill publishing co ltd
[4] victor A. Bloomfield university of Minnesota Minneopolis, USA – using R for Numerical Analysis in Science and Engineering.
[5] Mike X Cohen, Practical Linear Algebra for Data Science, Released September 2022, Publisher(s): O'Reilly Media, Inc.,ISBN: 9781098120610.
[6] Kenneth Hoffman, Ray Kunze (1996) Linear Algebra,

International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 11 Issue 5, May-2022, Impact Factor: 7.957

Qualitative Analysis of Phytochemicals Screening of Extract of Papaya Seeds

K.M. Buvaneswari[1], Dr. S. Sumayaa[2], Dr. S. Muthumaresswari[3], R. Lakshmi Shree[4]

[1]PhD Scholars, Department of Home science and Research centre, Thassim Beevi College for Women (Affiliated to Alagappa University) Kilakarai, Ramanathapuram District, Tamilnadu, India
[2]Professor and Head, Department of Home science and Research centre, Thassim Beevi College for women, (Affiliated to Alagappa University) Kilkarai, Ramanathapuram District, Tamilnadu, India
[3]Associate Professor and Head, Department of Home science and Research centre, Thassim Beevi College for women, (Affiliated to Alagappa University) Kilkarai, Ramanathapuram District, Tamilnadu, India
[4]Assistant Professor and Head, Department of Home science and Research centre, Thassim Beevi College for women, (Affiliated to Alagappa University) Kilkarai, Ramanathapuram District, Tamilnadu, India

Abstract

Phytochemical analysis of fruit seeds is an important commercial interest in the food industry to produce nutritional supplements for the treatment of degenerative diseases. Papaya seeds have a high medicinal value. The present study was intended to investigate the Preliminary phytochemical studies of different extracts of Carica papaya seeds. Dried and ground papaya seeds (5g), respectively mixed with 95ml of solvent (water, methanol, and ethanol) for 24 hours were used for this study. Extraction was done by Soxhlet extraction methods. The extract was then separated and sorted through Whatman's filter paper. Three different extracts of papaya seeds were prepared and analyzed for their phytochemical Substance. The results revealed the presence of phytochemical constituents was higher in ethanol extract of papaya seeds than in methanol and aqueous products. Therefore, the results scientifically confirm the use of papaya seeds in traditional medicine and can be used to treat various diseases caused by free radicals and chemicals due to the presence of secondary metabolites.

Key Index Terms: Papaya seeds- Phytochemicals- Soxhlet extraction- Extracts-disease

INTRODUCTION

The interest in natural antioxidants has increased considerably in recent years because of their beneficial effects on prevention and risk reduction in several diseases. Phenolic compounds are biologically active molecules that serve as natural antioxidants. Preliminary screening studies may aid in the identification of bioactive principles, which could lead to drug discovery and development (1).Papaya is a member of the Caricaceae family, which includes four genera. In India, the genus Carica L. is represented by four species, the most frequently grown and well-known of which is Carica papaya L. (2)

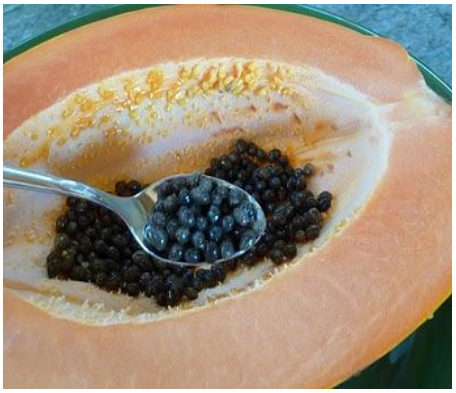
Papaya fruit is a sweet, cheap, and easily available tropical fruit. In addition to papaya, its seeds also have economic value and medicinal properties. Papaya seeds are rich in many nutrients and antioxidants. The nutrients contained in papaya seeds are folic acid, potassium, copper, magnesium, and fiber. It also contains some very important nutrients such as protein, carbohydrates, beta-carotene, fatty acids, benzyl isothiocyanate, glucosinolate benzyl, beta-sitosterol, the enzyme caricin, myrosinase, and 40% water (3,13). Medicinal plants containing phytochemicals such as flavonoids are said to have antioxidant, antibacterial, and antiallergic effects, maintain an inflammatory balance and reduce the risk of certain types of cancer. (4). Many plant compounds are found in papaya fruits and various components such as seeds, latex, leaves, and roots. Papaya seeds also contain a variety of botanical ingredients that help treat many diseases.

The pulp of seeds and skins contains a variety of antioxidant phytochemicals, including natural phenols and flavonoids. Phytochemicals are bioactive nutrient-free plant compounds found in fruits, vegetables, grains, and seeds that reduce the risk of serious chronic illness (14,15). But there is little or no information on the phytochemical properties of this nutritious seed. Hence, the present work evaluates the qualitative phytochemical analysis of poorly utilized papaya

seed. Considering all these facts, the present study was designed to investigate the presence of various phytochemicals in the three different extracts of papaya seeds, a plant that evokes various therapeutic effects.

MATERIAL AND METHODS

Collection and preparation of sample: The plant Carica papaya was collected in a sterile polythene bag, rinsed, sundried, and made into a powdery form before use. Plant materials were collected from a local fruit shop in the Ramanathapuram district. The fruit seeds were collected, washed in tap water, rinsed in sterile distilled water, and dried for 5 days at 100 C. The dried fruit seeds were blended to powder with a clean mortar and pestle and stored in airtight glass containers kept in a laboratory cupboard until required for further analysis. The fruit seeds were separately extracted with ethanol, methanol, and aqueous.



Papaya Seeds

Papaya seeds Powder

Preparation of extracts of papaya fruit seeds: Extraction of seeds powder was done by maceration and Soxhlet Extraction. In the maceration, 20g powdered seeds were soaked in 100 mL water and 20 g powdered seeds were soaked in 100 mL methanol for 5 days in a closed container at room temperature with occasional shaking or stirring. The extract was then repeated from the fruit seeds particles by straining. The process is repeated once or twice with fresh solvent. Finally, the last residue of the extract was pressed out of the plant particles using a mechanical press and then filtered using Whatman filter paper.

Qualitative Phytochemical analysis of Papaya seeds:
Phytochemicals like flavonoids and phenols are strong antioxidants and have an important role in the health care system (5). Screening of active compounds from plants has led to the discovery of new medicinal drugs which have efficient protection and treatment roles against various diseases, including cancer and Alzheimer's disease (6). Qualitative phytochemical screening was done for the evaluation of major phytochemical constituents such as Alkaloids, Flavonoids, Tannin, Saponins, Terpenoids, Phenol, Anthro quinine, Glycosides, Phytosterol, and steroids using the standard procedure of analysis (7.8,9). The reactions in this analysis revealed the presence or absence of these compounds in the seed with different extractions.

Test for Alkaloids: The concentrates were treated with 2% dilute HCl and kept in a water bath at 100ºC for 2minutes. In the wake of cooling, scarcely a few drops of 5% NaOH solution were added into the blend and were noticed for the arrangement of turbidity or yellow precipitate.

Test for Flavonoids: A couple of drops of NaOH solution were added to the extracts. Development of exceptional yellow color which vanishes on adding oil. HCl was noticed.

Test for Tannins: Uniform mixture of 0.5ml of the extract was set up in 1ml of water. To the blend, 2 drops of FeCl3 solution were added. The presence of catholic tannins is demonstrated by the appearance of a green dark tone.

Test for Saponins: Five ml of distilled water and 0.5ml of concentrate were mixed and shaken for 15 minutes. Foam development shows the presence of saponins in the seed powder extract.

Test for Terpenoids: To the 4ml of concentrate, 0.5ml of acidic anhydride, 0.5ml of chloroform, and conc. H2SO4 was added to the mixture and noted for the arrangement of red-violet tone.

Test for Glycoside Keller- Gilani test 0.5ml of concentrate was blended in with 2ml of glacial acetic acid to the combination; a few drops of conc. H2SO4 and FeCl3 were added. Reddish-brown colored shading at the intersection of two layers and presence of phenols was indicated by a blue-green or black colouring. Blue-green color in the upper layer demonstrates the presence of glycosides in the seed powder extract.

Test for the presence of phenols: The presence of phenols was determined by using the method described by Yadav & Agarwala [20]. Two grams of the sample were soaked in 10ml of methanol. After 24 hours, it was filtered using Whatman filter paper (No1). 2ml of the extract was mixed with a dilute ferric chloride solution. A blue-green or black coloration was formed indicating the presence of phenols.

Test for Anthocyanin: 2 ml of concentrate was blended in with 1ml of 2N NaOH and boiled in a water bath for 5minutes. Perseverance of somewhat blue-green shading marks the presence of anthocyanin.

Test for Phytosterol: 5 ml of concentrate was treated with 3ml conc. H2SO4 followed by 2ml of chloroform. The blend was permitted to rest briefly. The appearance of reddish-brown precipitate in the chloroform layer was noticed for the presence of phytosterol.

Test for Steroids: In a test tube, 1 ml of the sample was taken and mixed with 5 ml chloroform, at that point, an equivalent volume (5 ml) of concentrated sulphuric acid was carefully added through the sides of the test tube. If the upper layer transforms into a red tone and the sulphuric acid layer becomes a yellow tone with slight green fluorescence which shows the presence of steroids.

RESULTS AND DISCUSSION

Phytochemical constituents such as tannins, flavonoids, alkaloids, and several other aromatic compounds or secondary metabolites of plants serve as a defense mechanism and curative properties against predation by many microorganisms, insects, and herbivores (10). The analysis of phytochemical screening of papaya seed was shown in Table 1. The different extractions (methanol, ethanol, and aqueous) revealed the presence and absence of phytoconstituents of the papaya seed

Table-I: Qualitative phytochemical screening of different extracts of papaya seeds

Sample	Phytochemical Substance	Methanol	Ethanol	Aqueous
Papaya Seeds	Alkaloids	+	+	+
	Flavonoids	+	+	+
	Tannin	+	++	+
	Saponins	-	+	+
	Terpenoids	+	-	+
	Phenols	++	++	-
	Anthro quninone	+	++	-
	Glycosides	-	++	-
	Phytosterol	-	+	-
	Steroids	++	+	+

Table I shows the phytochemical screening of Papaya seed seeds where, 7 out of 10 components such as Alkaloids, Flavonoids, Tannin, Saponin, Terpenoid, Phenol, Anthro quinone, and steroids were present with methanol solvent extraction; 9 out of 10 components such as Alkaloids, Flavonoids, Tannin, Saponins, Terpenoids, Phenol, Anthro quinine, Glycosides, Phytosterol, and steroids were present with Ethanol solvent extraction and 6 out of 10 components such as Alkaloids, Flavonoids, Tannin, Saponins, Terpenoids, and steroids were present with Aqueous extraction.

The presence of phytochemicals with biological activity can be beneficial and of great medicinal value. It is said that phytochemicals such as saponins, terpenoids, flavonoids, tannins, steroids, and alkaloids have anti-inflammatory effects (11,16). Glycosides, tannins, flavonoids, and alkaloids have hyperglycemic activities. Steroids and triterpenoids showed analgesic properties (12,17).

CONCLUSION

Some of the plant seeds investigated for phytochemical constituents appeared to have the potential to improve user health as a source of useful medicines and in the presence of various compounds essential for health. From the foregoing, it can be concluded that phytochemical screening in methanolic, aqueous, and ethanolic extracts of papaya seeds revealed the presence of major phytochemical constituents such as saponins, tannins, flavonoids, alkaloids, phenolic compounds were high in the Ethanolic extract than methanolic and aqueous extracts comparatively. The phenol and flavonoid may be the potential chemo preventive and anticancer substances. It is, therefore, recommended that the cultivation of these oilseeds should be upgraded to enhance their wider utilization. Further study has to be encouraged on the effect of processing of these nutrient-rich seeds. Thus, the research focused on bringing to light their potential for commercial exploitation

REFERENCES

[1]. Rybarczyk, A. and Amarowicz, R. (2007). Silica gel column chromatography of phenolic compounds in sweet lupin seeds extract. Bromatologia i Chemia Toksykolo-giczna. 4: 375–379.
[2]. Duenas, M., Hernandez, T., Estrella, I., and Fernandez, D. (2009). Germination as a process to increase the polyphenol content and antioxidant activity of lupin seeds (*Lupinus angustifolius* L.). Food Chemistry. 117: 599–607.
[3]. Kavitha Vijayaraghavan, S.Mohamed Ali, and R.Maruthi. (2013). Studies On Phytochemical Screening And Antioxidant Activity Of *Chromolaena Odorata* And *Annona squamosa*. International Journal of Innovative Research in Science Engineering and Technology. 2 (12): 7315-7321
[4]. Edeoga, HO, Okwu , DE and Mbaebie, BO(2005) Phytochemical constituents of some Nigerian medicinal plants, African J. Biotech 4(7): 685- 688
[5]. Sood A, Uniyal PL, Prasanna R, Ahluwalia AS (2012) Phytoremediation potential of aquatic macrophyte, Azolla. Ambio 41:122–137
[6]. Mallikharjuna P B, Rajanna L N, Seetharam Y N, Sharanabasappa G K. Phytochemical studies of *Strychnos potatorum*. *Lf-A medicinal plant E-J Chem.* 2007;4:510–518.
[7]. Aravind G, Bhowmik D, Duraivel S, Harish G (2013) Traditional and medicinal uses of Carica papaya. Journal of Medicinal Plants Studies 1: 7-15.
[8]. Ajayi IA, Ajibade O, Oderinde RA (2011) Preliminary phytochemical analysis of some plant seeds. Res J Chem Sci 1(3): 58-62
[9]. Zhou K, Wang H, Mei W, Li X, Luo Y, et al. (2011) Antioxidant Activity of Papaya Seed Extracts. Molecules 16(8): 6179-6192.
[10]. Devi P S S, Kumar N (2017) The surprising health benefits of papaya seeds: A review. 6: 424-429.
[11]. Raaman N. Phytochemical Techniques. New India Publishing Agency, New Delhi, 2006, 19-24.
[12]. Tiwari P, Kumar B, Kaur M, Kaur G, Kaur H. Phytochemical screening and Extraction: A Review. Internationale Pharmaceutica Sciencia. 2011; 1(1):98-106
[13]. Auwal MS, Saka S, Mairiga IA, Sanda KA, Shuaibu A and Ibrahim A. Preliminary phytochemical and elemental analysis of aqueous and fractionated pod extracts of *Acacia nilotica* (Thorn mimosa). Veterinary Research Forum. 2014; 5(2):95-100.
[14]. Sofowra, A. 1993. Medicinal Plants And traditional Medicine In Africa. Spectrum Books Ltd., Ibadan, Nigeria, pp. 191-289.
[15]. Trease, G.E., Evans, W.C. 1989. Pharmacognosy, 11th edn., Bailliere Tindall, London, pp. 45-50.
[16]. Harborne, J.B. 1973. Phytochemicals Methods. Chapman and Hall Ltd., London, pp. 49-188.
[17]. Nwofia GE, Ojimelukwe P, Eji C. 2012. Chemical composition of leaves, fruit pulp and seeds in some Carica papaya (L) morphotypes. Int. J. Med. Arom. Plants, 2: 200-206

Review Article

Food Label and its Influence among Indian Consumers - A Review

Shireen S*, Muthumareeswari S, Sumaya, Bhuvaneswari KM, Lakshmi Shree R

Department of Home Science and Research Center, Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramanathapuram, India

***Corresponding Author:** Shireen S, Department of Home Science and Research Center, Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramanathapuram, India

Received: 25 May 2022; **Accepted:** 03 June 2022; **Published:** 24 June 2022

Citation: Shireen S, Muthumareeswari S, Sumaya, Bhuvaneswari KM, Lakshmi Shree R. Food Label and its Influence among Indian Consumers - A Review. Journal of Food Science and Nutrition Research 5 (2022): 543-551.

Abstract

Food Labels are a preface to packaged foods. There is an important component known as Nutrition Facts Panel (NFP), which precisely gives the amount of macro and micronutrients per 100g or one serving. Consumers were found to read them but do not alter their purchasing behavior accordingly. Therefore this study aims in identifying the hindering factors in food label usage. Articles done from 2013 to 2021, published on PubMed, Research Gate, Google Scholar and other independent Indian and International Journals on Indian population were reviewed. Certain factors like: education, gender, income and area of residence had an impact on reading and purchasing foods. Other major factors found to be hindering food label usage were knowledge of the buyers. This study concluded by recommending mandatory food labeling requirements in all Indian food products and to provide proper education and awareness to consumers to enable them to select healthier food choices in future. Further studies were also required to be done on a larger scale to generalize the findings and correct the narrow paths.

Keywords: Food label; Packaged foods; Food label knowledge; Consumer

1. Introduction

India is the world's second largest producer of food next to China, and has the potential of being the largest in the food and agricultural sector. The food processing industry is one of the biggest industries in India- and is ranked fifth in terms of production,

consumption, export and expected growth. The food sector has been witnessing a drastic change in consumption pattern especially in terms of food. Eventually, diet-related health issues have increased intensely from the last few years across the globe. It has become significant that the consumer must have knowledge of the product they are consuming. The best way to inform the consumer about the product and the nutrients it carries is to provide information about the product. Food label on the product plays the role of informing the consumer about the product, its ingredients and nutrients it contains. A label provides basic product information (including common name, list of ingredients, net quantity, shelf life, grade/quality, vegetarian society logo, country of origin, name and address of manufacturer, dealer or importer and food standards agency [1]. It also provides health, safety, nutrition information which includes instructions for safe storage, handling, nutrition information such as quantity of fats, protein, carbohydrate, vitamins & minerals and preservatives, colors, if used any, quantity per serving of stated size of food (in the nutrition facts table) and specific information on products for special dietary use. Moreover, It also acts as a vehicle for food marketing promotion and advertising promotional information and claims such as 'low fat', 'cholesterol free', 'high source of fiber', 'natural', 'organic', 'no preservatives added'[2]. Awareness and knowledge together with correct delivery of nutritional facts are interrelated which can form the consumer's choice of food. Therefore, this review aims at highlighting the existing literature on the socio-demographic, knowledge, attitude and preference as well as the health related factors which affect the usage of food labels and also factors influencing label usage among Indian consumers with the following objectives

- Association between socio-demographic characteristics and food label usage
- Association between knowledge, attitude and use of food labels.
- Reasons for non-usage of food labels by people.
- Most commonly checked information on food labels.

2.Methodology

The current study is a descriptive in nature and is based on secondary datas collected from various research papers, review papers, e-sources like PubMed", "ResearchGate" and "Google Scholar" using the keywords "food labels", "nutrition labels". Only Indian articles published from 2012 to 2022 were considered along with other international articles as references.

3. Background studies

Association between socio-demographic characteristics and food label usage-

a. Education

Consumer's education is one of the factors of food label usage [3]. Awareness on pre-packaged food labels was found to be associated with education of the consumers. Understanding and interpreting on nutrition labels is attributed only to education. Health-related information on the food labels is clearer and educated people are more aware of what they eat and its association with their health [4,5] People who have higher educational qualifications found it easier to interpret nutrition labels than those with lesser or no education at all. It was also proved that less educated people found difficulty in understanding the technical terms used in the food labels [6,7]. However, there are contrasting studies

which proved that educational qualification of consumers barely had any effect over knowledge of nutrition labels. People with lack of time but with adequate education were found to be purchasing without looking at the food label, which proves that education had no association with the habit of reading nutrition labels [8].

b. Age

Knowledge of food labels is significantly associated with age. Age of the consumers determines the level of importance they show towards reading and understanding food labels. The younger population was found to be more cautious in buying foods according to nutritional composition and health claims than older people.

c. Gender

Women were found to spend more time purchasing which eventually led them to read the nutritional content [9]. Women were found to be more concerned about calories, micronutrients like vitamins and minerals, fat, salt and sugar intake. Though adequate information on females using food labels more than males was not proved, it can be a controversy in some cases too [10,11].

d. Income

People with higher income have the choice to compare similar food products with differing rates than those who can afford only one particular type of food. Certain consumers were also found in wasting their income on junk foods rather than eating nutritious food which can be afforded by them. They also find it time consuming on shopping due to work to earn a good income [12,13].

e. Household Size

The structure of the family is one of the key aspects in determining the usage of food labels. Increase in members of a family might lead to poor food choices as they won't be able to afford healthy food for everyone. Another major aspect is satisfying everyone at home which is very difficult especially in large household composition. Household size was also being deemed as not associated with food purchase behavior in some contrasting studies [14].

f. Area

The locality in which people reside also determines their usage of food labels in purchasing foods. Urban residents in India were found to be more aware of the importance of food labels while those in rural areas don't know them or are not educated enough to understand or afford them. Health education and social media campaigns could help increase awareness among rural people which will help them to make healthier choices [15].

Association of Knowledge, Attitude and food label use

Nutrition knowledge and awareness has a significant impact on understanding and utilization of food and nutrition labels by the consumers. Better diet quality could be achieved with better nutritional knowledge. The foods that are familiar to the consumers prevent them from using food labels. Less nutritional awareness also leads to health related dietary issues [16,17]. Though many people were found to read food labels, the point is that they don't alter their purchase decision due to many reasons. It can be due to their inability to understand the information they

read. A few consumers were found to be making purchase decisions based on nutrition facts [18]. Many people were aware of the importance of reading a food and nutrition label, but due to the difficulty and monotony in reading labels they preferred to avoid their usage. The major reason was that the labels were not simple and it did not deliver accurate information about the product. Easy-to-understand labels must be formulated to make everyone get interested in reading them. People preferred that the nutritional information must be listed per 100g rather than serving size as this is much more relatable and enables easier calculation by anyone [19]. They also wanted a standard format to be followed while displaying nutritional content on the food label.

i. Traffic light nutrition labeling is effective as they are self-explanatory but it might be overrated by consumers and they might misinterpret the food due to the colors used [20].

ii. Consumers prefer the nutrition information to be present in front-of-pack (FOP) as this can enable easier access to the required deciding factors than making them search for the information. Though some misleading results were found as many feel that foods with FOP nutritional labeling were healthier than those with nutrition facts mentioned behind [21].

"The use of nutritional labels" has been a multidimensional issue. Under this study researchers gave various definitions related to nutritional labeling and alternatively also found out the role of nutritional labels for influencing consumers for choosing healthy food products.[22]Attitude of consumer towards nutritional evidence mentioned on food labels is positive, but most of the consumers refer food labels only for brand comparison. Some of the factors that affects consumer buying decisions were food label designs, nutritional information, health claims and quality of product etc [23]. A cross-sectional study was conducted by Vemula, et al. [24] in 2013 to determine the knowledge of consumers and their food label usage. This study involved both quantitative and qualitative forms. $1/5^{th}$ of them bought pre-packed foods daily whereas, 45% bought once a week. 90% were found to be reading food labels out of which 81% checked the date of manufacture and expiry dates alone.$1/3^{rd}$ only checked for nutrition information and the list of ingredients. Due to the lack of nutritional knowledge, many consumers did not prefer to read nutrition labels and found the terms very technical too. Quality symbols were read by 60% of them. Education levels and checking food labels were found to be positively correlated and girls as well as women were found to be more concerned about fat and sugar contents on the nutrition labels. The study concluded that, majority of the consumers found it difficult to understand and interpret nutritional information on the labels. Shekhar SK & Raveendran PT [25] in the year 2014 conducted a study among students to understand their perception on nutrition labels. They used various tests and found that their perception on nutrition labels did not change according to their age or gender. A survey was conducted by Priyadarshini V [26] in 2014 on assessing the awareness among the consumers in regard to the food label information provided. The awareness level was satisfactory regarding different aspects mentioned on the food labels but, their usage among consumers to make purchase decisions was found to be low. The information of the label was checked by 67.5%, but they were found to be checking only basic information like manufacturing and expiry dates as well as MRPs. This study concluded by suggesting certain guidelines to be adopted by manufacturers and government to help consumers to understand the nutritional facts. Further

research was also demanded to find the suitable format of nutritional facts displayed on the labels. Dunford et al. [27] in the year 2015 conducted a survey to find the food packages being labeled in accordance with Indian and International Standards. 4166 packed foods were included in the study which was classified under 14 food groups. Nutrient information consisting of calories, proteins, sugar, carbohydrates, total fats as per the Food Safety and Standards Authority of India (FSSAI) were noted among 52% of the foods. Only 27% adhered to Codex criteria which included sodium and saturated fats. A significant correlation and variation was noticed among leading brands, food groups and manufacturing countries in adhering to the standards. The study highlighted that the majority of Indian foods were found to be displaying inadequate information in regard to nutrition labels as per the Indian and International guidelines. The enforcement of these mandated criteria in India would help consumers in making wise choices especially the quantity of saturated fats and sodium guidelines. Madhvapaty H & DasGupta A [28] conducted a study in 2015 to find the requirement regarding the alternative formats for nutrition labeling to be enforced especially as Front-of-Pack display. The questionnaire included a traffic light method of nutrition facts displayed to find the preference of the consumers. Though traffic light nutrition labels were highly preferred among the consumers; it also required fine-tuning. Overloading information must be prevented. The manufacturers must enable the consumers to view the actual nutritional content in order to facilitate the nutrition label usage among them. A review was carried out by Kodali S & Telaprolu N [29] in 2016 which aimed to highlight the literature on the influence of food labels on consumer purchase behavior in order to find gaps for future research. Health claims were found to be poorly understood by consumers but their attitude towards them was positive. More research on the preferred method of label display was required to conclude their usage. It was found that the influencing factors on a food label were the nutrition information, health claims, label design, its accessibility and quality. A cross-sectional study was done by Gavaravarapu, et al [30] in the year 2016 which aimed to find the usage of food labels, the knowledge of adolescents on food labels in India. It was found that almost 95% of food labels were in tune to Indian standards. Though many thought that the information on the labels to be confusing or too cramped with information. Consumer education was recommended to help adolescents in making wise dietary choices. 70% wanted a change in the display of label components which requires research into alternative methods to ease the consumer understanding. A descriptive study was carried out by Kaur VP, Kaur N & Qumar N [31] during 2016 to find the awareness levels on different information provided on the food labels. It aimed to find the association between purchase decisions and food labels. Every participant was found to be having adequate knowledge on food labels. 56.66% of the males were found to purchase packed-foods more than one time in a week, 16.66% purchased only once a week and 6.66% bought once a month, whereas 20% bought occasionally. Awareness on different components of the food label was found to be varied among the study population. Certain information was found to be used more when compared to other aspects on the labels. Therefore, the study was used as an aid for food manufacturers to know the consumer preference as well as the researchers to conduct further awareness studies. Anitha, M.C & Devi AK [32] conducted a study in 2017 regarding

the frequency of use of nutrition labels and their awareness. Children aged 10-12 years and who belonged to low-income status were found to be unaware of nutrition facts used on the food labels. Further studies were required to educate school students to enable them to make wise choices from their young age. A cross-sectional survey study was conducted by Robert SD & Chandran in 2017 to find the food label reading frequency and to associate their knowledge, practice and attitude on reading food labels. Only 5% were noted to be aware of food label information and education status was not found to be correlated with food label knowledge. 19% acknowledged the importance of food labels and 46% read them frequently. 90% refused to buy food packs with a label and 19% were found to calculate their intake based on the provided facts. Therefore this study recommended further interventional studies to educate the consumers on food labels. A systematic review was done by Donga, G & Patel, N[33] during 2018 on the association between different socio-demographic factors on usage and altering purchasing behavior with regard to nutrition labels. Income and education had a positive correlation in influencing nutrition label usage. Married people and females were found to use them more than their counterparts. Whereas, household size, age, frequency of purchase and time were found to have negative correlation. A simple and easier version of nutrition label was demanded for better usage among all consumers. A survey study was done by Donga G & Patel N [1] during 2018 which aimed to study the different aspects of nutrition labels which were given importance by the consumers and to find the people who often read them and their influence on purchase decisions. 2192 participants were included out of which 95.8% had knowledge on nutrition labels and 88% were found to be reading them. The important information checked on nutrition labels were calories from fat, cholesterol and total fat contents. Gomathi SJR & Kar SS [34] performed a cross-sectional study in 2021 to find the number of people reading and checking food labels to decide on purchasing and those using the food label information. It also aimed to find the different categories of packed foods adhering to the labeling regulations. Many people were aware of the information on the labels whereas, the ability to understand was very less. Therefore, educating the people on various food label aspects is required in future studies to improve their utilization among them. Reasons for non-usage of labels-

One of the major issues in not using food labels are: the difficulty in understanding the terms used in relation to nutrition and other health claims. Some were not interested and others had a lack of time due to which they did not prefer to read the food labels before purchasing. Some feel that the nutrition information and claims are not what they appear to be and certain others avoid reading food labels as they are familiar with the brand they buy regularly. Those preferring attractiveness of food package, taste of food were other reasons for not giving importance to food labels [35,36].

4. Conclusion

From the above reviewed Indian articles, it was proved that food labels were one of the most underestimated sources of information that consumers are aware of but did not alter their choices based on them. They found certain terms to be ambiguous, which was one of the major key reasons for non-usage of the food labels. Some of the key findings derived from this review are:

a. Socio-demographic characteristics of the people like: education, gender, income and area of residence

had an impact on the reading and utilization of food labels and making purchase decisions.

b. Knowledge was found to be associated in understanding the seriousness of choosing healthier choices through reading food labels.

c. Their preferences and opinions also had a correlation to usage of food labels, like the format of labels and the difficulties they encountered in understanding the information.

Survey results are more specific when it comes to the use of nutritional information in the food labels. Educating consumers on different aspects of food labeling enables them to make more informed purchasing decisions on healthier food choices.

References

1. Donga G, Patel N. The Nutrition Label-which information is important to theConsumer? International Journal of New Technology and Research 4 (2011): 53-56.
2. Gautam D, Naresh P. A Review of Research Studies on Factors Affecting Consumers' use of Nutritional Labels. Nutri Food Sci Int J 7 (2018): 555713.
3. Carrillo E, Varela P, Fiszman S. Influence of Nutritional Labels. Journal of Food Science 77 (2012): 1-8.
4. Syed H, Abbas Z, Muhammad B. Awareness of Pakistani consumers towards nutritional labeling on product packaging in terms of buying behavior. International Journal of Business and Social Science 3 (2012): 97-103.
5. Petrovici D. Nutritional knowledge, nutritional labels and health claims on foods.British Food Journal 114 (2012): 768-783.
6. Cooke R, Papadaki A. Nutrition label use mediates the positive relationship between nutrition knowledge and attitudes towards healthy eating and dietary quality among university students in the UK. Appetite 83 (2014): 297-303.
7. Alexandra F, Chrysochou P, Krystallis A. Consumer response to food labels in an emerging market: the case of Romania. International Journal of Consumer Studies 38 (2014): 166-174.
8. Van derMerwe D. Consumers' knowledge of food label information: an exploratory investigation. Public Health Nutrition 16 (2012): 403-408.
9. Vemula SR. Use of food label information by urban consumers in India- a study among supermarket shoppers. Public Health Nutrition 17 (2013): 2106-2114.
10. Aygen G. Determinants of nutrition label use among Turkish Consumers. International Journal of Humanities and Social Science 2 (2012): 53-70.
11. Patricia U. Consumers' knowledge, attitude and use of Food Labels in an urban city in Southeast Nigeria. Advance Nutrition 7 (2016): 22.
12. Yong KC, Ming MF, Debbie AL. Socio-demographic and lifestyle factors associated with nutrition label use among Malaysian adults. British Food Journal 117 (2015): 2777-2787.
13. Sinne S, Anna KE, Pia KH. Who is the purchaser of nutrition-labeled products? British Food Journal 5 (2017) 12-17.
14. Zul AAL, Amizi AM. Consumption awareness of Indian Community towards food labels in Klang, Selangor. International

Academic Research Journal of Business and Technology 1 (2015):45-50.

15. Sun PC, Hsien-Long H, Fang YC. Factors instead of demographic characteristics related to nutrition label use. British Food Journal 117 (2015): 3024-3038.
16. Zul AAL. Food Labels' impact assessment on Consumer purchasing behavior in Malaysia. Journal of Food Products Marketing; 22 (2015): 137-146.
17. Zaidi SH, Bakhtiar M. Awareness of Pakistani Consumers towards Nutritional Labeling on product packaging in terms of buying behavior. International Journal of Business and Social Science 3 (2012): 97-103.
18. Geaney F. Nutrition knowledge, diet quality and hypertension in a working population. Preventive Medicine Reports 2 (2015): 105-113.
19. Montandon CA, Christopher C. Effective nutrition labels for fast food consumers. British Food Journal 118 (2016): 2534-2549.
20. Angela S, Seamus O'R, Kathleen O'S. Nutritional knowledge, nutritional labels and health claims on food. British Food Journal 114 (2012): 768-783.
21. Miller LM, Cassady DL. The effects of nutrition knowledge on food label use. A review of the literature. Appetite 92 (2015): 207-216.
22. Azman N, Sahak SZ. Nutritional Label and Consumer Buying Decision: A Preliminary Review. Procedia - Social and Behavioral Sciences 130 (2014): 490-498.
23. Kodali S, & Telaprolu N. Food Label and its influence on Consumer Buying Behavior: A Review of Research Studies. International Journal of Science and Research 7 (2018): 386-390.
24. Shekhar SK, Raveendran PT. An exploration into Indian's perception on food products nutritional labeling. Management Research and Practice 6 (2014): 65-78.
25. Dunford EK. The adherence of packaged food products in Hyderabad, India with nutritional labeling guidelines. Asia Pacific Journal of Clinical Nutrition 24 (2015): 540-545.
26. Priyadarshini V. Awareness and use of food labeling informations among consumers in Bhubaneswar City. Food Science Research Journal 5 (2014): 114-119.
27. Madhvapaty H, DasGupta A. A study of food product labeling for products aimed at children. IOSR Journal of Business and Management 17 (2015): 88-96.
28. Gavaravarapu SM, Saha S, Vemula, S. Knowledge and practices of using food label information among adolescents in Kolkata, India. National Institute of Nutrition 37 (2016): 12-25.
29. Kaur VP, Kaur N, Qumar N. Assessment of consumer awareness about usage of food labels and its impact on food buying behavior. International Journal of Research 4 (2016): 10-19.
30. Anitha MC, Devi AK. An outlook on nutrition and food labeling among selected school children aged 10-12 years in Coimbatore city. International Journal of Community Medicine and Public Health 4 (2017): 3461-3467.
31. Robert SD, Chandran A. Survey on consumer knowledge and use of food labels.

International Journal of Health Sciences and Research 7 (2017): 203-209.

32. Donga G, Patel N. A review of Research Studies on factors affecting consumers' use of nutrition labels. Nutrition and Food Science International Journal 7 (2018): 001-007.
33. Gomathi SJR, Kar SS. Consumer awareness and status of Food Labeling in selected supermarkets of Puducherry: an exploratory study. International Journal of Advanced Medical and Health Research 5 (2021): 36-40.
34. Bandara BES. Impact of food labeling information on consumer purchasing decision: with special reference to faculty of Agricultural Sciences. Procedia Food Science 6 (2016): 309-313.
35. Su D. A sex-specific analysis of nutrition label use and health, Dougles County, Nebraska. Preventing Chronic Disease Public Health Research, Practice and Policy 12 (2015): 1-14.

© May 2022 | IJIRT | Volume 8 Issue 12 | ISSN: 2349-6002

Nutrient Analysis of Indian Gooseberry Seeds

K.M.Buvaneswari[1], Dr.S.Sumayaa[2]

[1] *PhD Scholars & Assistant Professor, Department of Home Science and Research Centre, Thassim Beevi College for Women (Affiliated to Alagappa University) Kilakarai, Ramanathapuram District, Tamil Nadu, India*

[2] *Professor and Head, Department of Home Science and Research Centre, Thassim Beevi College for Women, (Affiliated to Alagappa University) Kilakarai, Ramanathapuram District, Tamil Nadu, India*

Abstract- **Global fruit consumption is increasing significantly worldwide due to its disease prevention and health benefits due to the presence of nutrients and other bioactive compounds needed for human well-being. Fruit processing waste is highly perishable and seasonal and is a problem for pollution monitoring agencies. In some fruits, seeds make up about 20% of the total weight. This study aims to investigate the nutritional composition of Indian goose fruit seeds. The seeds of Indian gooseberry seeds were extracted and analyzed for their Macro and micronutrient content. The results showed that the estimated ash and moisture content per 100 gm was 3.76% and 7.16% respectively. The estimated amount of carbohydrate content was 17.07 g, Protein 9.25 g, fat 5.45 g, and fiber content 3.18 g, respectively. Calcium and iron content was moderate at about 23mg/100gm and 7.73mg /100gm of the seed powder; whereas the seeds had higher amounts of phosphorus with an average value of 388.4mg/100gm. The selected sample proved to have a high content of potassium of 310mg/100gm. Therefore, it means that 20 g of Indian goose fruit seed powder will meet a quarter of the potassium requirement. The sodium content was found to be 4.25mg/100gm, which is quite a moderate amount compared to other fruit seeds. Current study results show that seeds can be good health supplements and food ingredients if used and handled properly. Indian gooseberry seeds is nutritionally valuable based on their nutrients.**

Index terms: Fruit Processing, Health Supplements, Indian gooseberry seeds, Nutrients

I.INTRODUCTION

Fruits provide quick ways to provide an adequate supply of vitamins, minerals, and fiber to people living in tropical areas. Many fruits and vegetables have a low energy level and it is recommended to control weight. In some cases, fruit contains about 85% water, fat, and protein in very small amounts, a good portion of the available carbohydrates such as cellulose, starch in small amounts, vitamins, and sugars [2]. Fruits are rich in vitamins, minerals, fiber, phytochemicals, and antioxidants in the branches, seeds, and peels [3]. In addition, many fruits are used to make beverages, such as fruit juice or alcoholic beverages such as wine, brandy, or vinegar. The fiber content of fruits and vegetables has been reported to have beneficial effects on blood cholesterol and is helpful in preventing intestinal infections. It has also been reported that people who eat a diet rich in fruits and vegetables have very low levels of many cancers [4].

A nutritionist has suggested that eating at least five servings of fruits and vegetables a day can help people maintain good health throughout their lives, protect them from heart disease and cancer, Type 11 diabetes, kidney stones and reduce stroke. Fruits are also useful as a nutritional supplement and are recommended worldwide as processed food [5]. However, some fruits are also known to have anti-nutrient properties such as phytates and tannins, which can reduce nutrient bioavailability, if present in high concentrations [6]. About 15% of the world's diseases are the result of malnutrition and malnutrition. However, not eating enough fruit can lead to serious health conditions such as vitamin deficiencies, serious illnesses (such as cancer, and heart disease), digestive problems, and weight gain [7]. There is therefore a need to do more research on the nutritional qualities of the fruit in order to incorporate the right fruit seeds into our daily diet.

The use of Indian gooseberry in Indian communities is very widespread. This is due to the discovery of Indian gooseberry fruit in India. Many people eat Indian gooseberries because of their flavor,

prescription medicinal and healthy foods, and their cheap availability in the region. Obviously, the most commonly used part of the Indian gooseberry is always the edible part i.e. the soft part, the layer between the skin and the seeds. Part of the fruit is usually not eaten due to its bad taste, little knowledge of its medicinal/nutritional value, and difficult processing. Therefore, different people never choose Indian gooseberry seeds mainly because they are not informed about the nutrients in the seed component. The study aimed to establish a basis for the formation of nutrients in the seed component and thus proposes the optimal exploitation of Indian gooseberry seeds. The main objective of the current study is to analyze the nutrient content of Indian gooseberry seeds.

II. MATERIALS AND METHODS

2.1. Selection of area

In Tamil Nadu, Indian gooseberries are commonly available in the local fruit markets as well as from departmental stores to small groceries. The investigator had easy access to a nearby vegetable market and was also interested in studying the nutritional composition of the gooseberry seed powder commonly available in the Indian region.

2.2. Collection of Indian gooseberry seeds

The investigator freshly collected the fruit samples from a local vendor at the vegetable market, Ramanathapuram. The fruits were consumed and deseeded. The seeds were collected and kept separately in a dry container for further processing.

2.3. Development of dry gooseberry seeds powder

The fruit seeds were collected, washed in tap water, rinsed in sterile distilled water, and dried for 5 days at 100 C. The dried fruit seeds were blended to powder with a clean mortar and pestle and stored in airtight glass containers kept in a laboratory cupboard until required for further analysis.

Figure-I-Development of Indian goose berry seeds powder

2.4. Proximate analysis of the nutrient content of Indian gooseberry seed powder

In order to estimate the micro and macronutrients such as moisture, ash, carbohydrates, protein, fat, and crude fiber of Indian gooseberry seeds powder, AOAC procedures are followed. Ash solution is prepared and used for estimating the mineral content such as calcium, iron, phosphorus, potassium, and sodium by using standard Procedure.

TABLE I METHODS FOLLOWED FOR THE ESTIMATION OF MACRO AND MICRO NUTRIENTS OF INDIAN GOOSEBERRY SEEDS

MACRO NUTRIENTS	METHODS
Ash (percent)	AOAC Method – 2000
Moisture (percent)	AOAC method – 2000
Carbohydrate (gm)	Anthrone method -2001
Protein (gm)	Lowry et al – 2001
Total fat (gm)	Freeman et al – 1957
Fiber (gm)	AOAC method – 2000
MICRO NUTRIENTS	METHOD
Calcium	Hawk et al – 1957
Iron	Wong – 1928
Phosphorous	Fiske et al – 1925
Potassium	Venkatesh et al – 2011
Sodium	Sharma et al – 2013

III. RESULTS & DISCUSSION

3.1. Nutritional quality of the selected nutrient seed powder

The seeds of the Indian gooseberry possess equally high beneficial qualities in comparison with the fruit. The seeds contain high amounts of macro and micronutrients and also being one of the richest sources of phytochemicals, they are highly recommended for therapeutically benefits and in the prevention of many ailments. The seeds can also be incorporated into many healthy food products to increase the overall nutritional value.

TABLE II ESTIMATED MACRO NUTRIENT OF THE INDIAN GOOSEBERRY SEED POWDER

S.NO	Nutrients	Amount / 100gm
1.	Ash (percent / 100gm)	3.76
2.	Moisture (percent / 100gm)	7.16
3.	Carbohydrate (gm)	17.07
4.	Protein (gm)	9.25
5.	Total fat (gm)	5.45
6.	Fiber (gm)	3.18

Table II shows the nutrient composition of the selected Indian gooseberry seed powder. The estimated amount of ash and moisture per 100gm was 3.76% and 7.16% respectively. The estimated amount of carbohydrate content was 17.07gm, the protein was 9.25gm, fat content was 5.45 gm and fibre content was 3.18 gm respectively. Carbohydrates and proteins are the highly essential nutrients for the human body and they are the main energy source, providing fuel for the central nervous system and for all the other organs. Proteins are required for the development, tissue repair/ regeneration and reconstruction of the body and they are responsible for the synthesis of antibodies, blood cells, hormones and enzymes (7).

Fat content was higher in seed powder of Emblica when compared with other fruit seeds with a mean value of 5.8%. Tetradecanoic acid was a common component in both seed and seed coat and this acid possesses antioxidative properties, anti-inflammatory and antiarthritic properties (8,9).The seed powder was also found to have a good amount of dietary fibre of 3.42%. It is known through some clinical trials that intake of dietary fibre has a positive effect in the control of diabetes and body weight (10) .

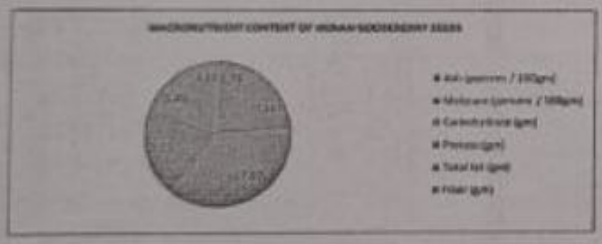


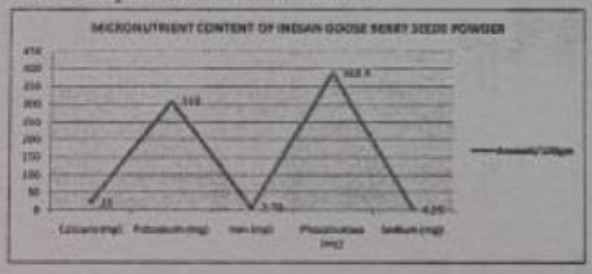
TABLE II ESTIMATED MICRO NUTRIENT OF THE INDIAN GOOSEBERRY SEED POWDER

S.no	Micro nutrients	Amount/100gm
1.	Calcium (mg)	23
2.	Potassium (mg)	310.0
3.	Iron (mg)	7.73
4.	Phosphorous (mg)	388.4
5.	Sodium (mg)	4.25

The above table III showed that Minerals are said to be inorganic nutrients that may be present as single atoms or in singlet form according to Palmer, 2001. The calcium content was moderate at about 23mg/100gm of the seed powder, whereas the seeds possessed higher amounts of phosphorus with a mean value of 388.4mg/100gm.

The iron was found to be 7.73gm/ 100gm of the seed powder. Amla seeds powder had a good amount of iron. Iron can bind a variety of ligands including cyanide, carbon monoxide, oxygen binding proteins such as hemoglobin, myoglobin and cytochrome oxidase, etc(11)

The selected sample proved the potassium content was a high content of 310mg/100gm and also potassium is the most abundant intracellular cation which is known to activate various enzymes which are involved in catalyzing the transfer of phosphoryl groups or elimination reactions (12). According to RDA, 2000mg of potassium is required (13,14). It thus implies that 20g of Indian gooseberry seed powder will satisfy one –fourth requirement of potassium. Sodium content was found to be 4.25mg/100gm which is a quite moderate amount when compared to other fruit seeds.



IV.CONCLUSION

The seeds of the Indian gooseberry were analyzed to obtain moisture, ash, protein lipids, carbohydrates, and fiber. It is found that gooseberry seeds can be considered a good source of macro and micronutrients (seeds in particular), lipid, carbohydrate (seeds in particular), protein (seeds in particular), fiber, and moisture, which benefit human health. Therefore, people can use these seeds to supplement their diet. Importantly, this study supported the use and benefits of various nutritional and therapeutic benefits of this fruit seed on the basis of an analysis of its nutrients. Further research could be done to study the benefits that gooseberry seeds may also have in the cosmetics industry. In this day and age where many degenerative diseases are present, people are trying to return to natural remedies to avoid the side effects of allopathic medicines. In this view, the research conducted can

be beneficial in preventing many degenerative diseases. The use of gooseberry seeds is virtually invisible to the public although it is rich in many bioactive compounds

REFERENCE

[1] LE Voorrips, RA Goldbohm, PG Van, F Sturmans, RJ Hermus, et al. (2000) Vegetables and Fruits consumption and risk of colon and rectal cancer in a prospective cohort study: The Netherlands cohort study on diet and cancer. Am J Epidemiol 152(11): 1081-1092.

[2] IO Alaekwe, OE Mojekwu (2013) Proximate Composition and Utilization of Napoleona imperialis Fruits. J Nat Sci Res 3(6): 160-165.

[3] AS Lim, MS Rabeta (2013) Proximate analysis, mineral content and antioxidant capacity of milk apple, malay apple and water apple. Int Food Res J 20(2): 673-679.

[4] WA Fila (2013) Comparative Proximate Compositions of Watermelon Citrullus Lanatus, Squash Cucurbita Pepol and Rambutan Nephelium Lappaceum. Int J Sci Technol 2(1): 81-88.

[5] OA Ekpete, OS Edori (2013) Proximate and mineral composition of some Nigerian Fruits. Int J Res Rev Pharm Appl Sci 3(2): 287-295.

[6] CA Edem, ID Miranda (2011) Chemical Evaluation of Proximate Composition, Ascorbic Acid and Anti-Nutrients content of African star apple (Chrysophyllum africanum) fruit. IJRRAS 9(1): 146-149.

[7] A Sheldon (2018) Problem from lack of fruits and vegetables. 8. K Claude, UA Godwin (2015) Phytochemical profile and antibacterial activity of crude extracts of the pod of Aframomum angustifolium (Sonn.) K. Schum. Eur Journals Biol Res 5(2): 36-41.

[8] Dr. Lanka.S., A review on pharmacological, medicinal and ethnobotanical important plant: phyllanthus emblica linn.(syn.emblica officinalis). World J.Pharm Res 2018; 7(4): 380-396

[9] Dr. Hassan.M., Mughal.S., Aslam.A, Emblica officinalis(Emblica): A prospective Review on distinctive properties and therapeutic applications of Emblica. Biomedicine and nursing 2020;6(2):22-30.

[10] Grover HS, Deswal H, Singh Y, Bhardwaj A. Therapeutic effects of Emblica in medicine and dentistry: A review. J Oral Res Rev 2015;7:66

[11] Habib-ur-Rehman. K.A., Studies on chemical constituents of Phyllanthus emblica, Nat, prod. Res., 2007; 20:775-781

[12] Singh, E., Sharma, S., Pareek, A., Dwivedi, J., Yadav, S., & Sharma, S., (2011). Phytochemistry, traditional uses and cancer chemopreventive ctivity of Amla (Phyllanthus emblica): The Sustainer. Journal of applied pharmaceutical science. ISSN: 2231-3354.

[13] Tripathi, V.K., Singh, M.B., & Singh, S., (1988). Studies on comparative compositional changes in different preserved products of Amla (Emblica officinalis Gaertn). var Banarasi, Indian Food Packer. 42: 62-66.

[14] Palmer, T 2001. Enzymes: Biochemistry, Biotechnology, Clinical chemistry. Horwood publication, England, ISBN:9781898563785, pages:402.

ANALYSIS OF INTRUSION DETECTION BASED ON SWARM INTELLIGENCE AND CLASSIFIER TECHNIQUES

Dr.A.Shanthisona[1] & Dr.A.Krishnaveni[2]

Assistant Professors[1&2],

PG & Research Department of Computer Science, Tiruppur Kumaran College for Women, Tirupur[1]&

Department of Computer Science & Research Centre, Thassim Beevi Abdul Kader College For Women, Kilakarai[2].

***Abstract*—**The Intrusion Detection System (IDS) plays a significant role in identifying anomalies or attacks in the network. Intrusion detection is observing and analyzing the events that occur in a network to address security problems. Network anomaly detection is an effective method for ensuring better security. Recently, several techniques have been aimed at intrusion detection in wireless networks. The conventional techniques are challenging to extract the intrusion behavior features from a high-dimensional dataset. Additionally, the conventional method resulted in minimum accuracy and higher time complexity. Swarm Intelligence (SI), a relatively new bio-inspired family of methods, seeks inspiration in the behavior of swarms of insects or other animals. A major contribution of this work is a detailed comparison of several SI-based IDS in terms of efficiency. Particle swarm optimization (PSO) is a robust and simple to implement optimization technique has been used in order to improve the performance. NSL-KDD dataset used to evaluate the performance of proposed method. The result demonstrated the proposed method has comparable performance based on detection rate.

Keywords— ***Intrusion Detection System, Swarm optimization, Detection Rate, Ensemble technique.***

1. INTRODUCTION

An intrusion detection system (IDSs) is becoming most needed in system defence. Recently, machine learning methodologies are playing an important role in detecting network intrusions (or attacks), which further helps the network administrator to take precautionary measures for preventing intrusions. Classification method plays an active role in obtaining the effective counter measure against the intrusions. Hence, Aldweesh et al. (2020) stated that the accuracy entirely depends on classification, it is important to use feature selection techniques to select an optimum subset of features. In this paper, a novel

IDS is proposed which employs the butterfly optimization algorithm (BOA), a recent meta-heuristic, to perform feature selection. For each butterfly (i.e. features), fitness is estimated by using cosine similarity index measurement. Based on measured fitness, optimal features are selected. After that, Polytomous adaptive base class ensemble technique is applied to recognize the different types of attacks. An ensemble classifier constructs number of weak learners using training data. Then, the results of weak learners are combined to form strong classification result. Here, Support Vector Machine classifier is utilized for identifying suspicious activities. With the result of strong classifier, the data is classified as normal or anomalous. Followed by, different types of attacks are correctly identified with selected features. Finally, the ensemble classification identifies the weak learner with minimum error. Thus it helps to improve the accuracy on intrusion detection with minimum time. Hence, the proposed Hybrid Jaccard Index Butterfly Optimization Feature Selection–Polytomous Adaptive base Class Ensemble method (**HJIBOFS-PABCE).**The experimental works of on the parameters such as Intrusion Identification Rate, Intrusion Detection Accuracy, Intrusion detection time. The experimental result shows that the proposed technique improves intrusion detection accuracy with reduced time and error rate than the state of art methods.

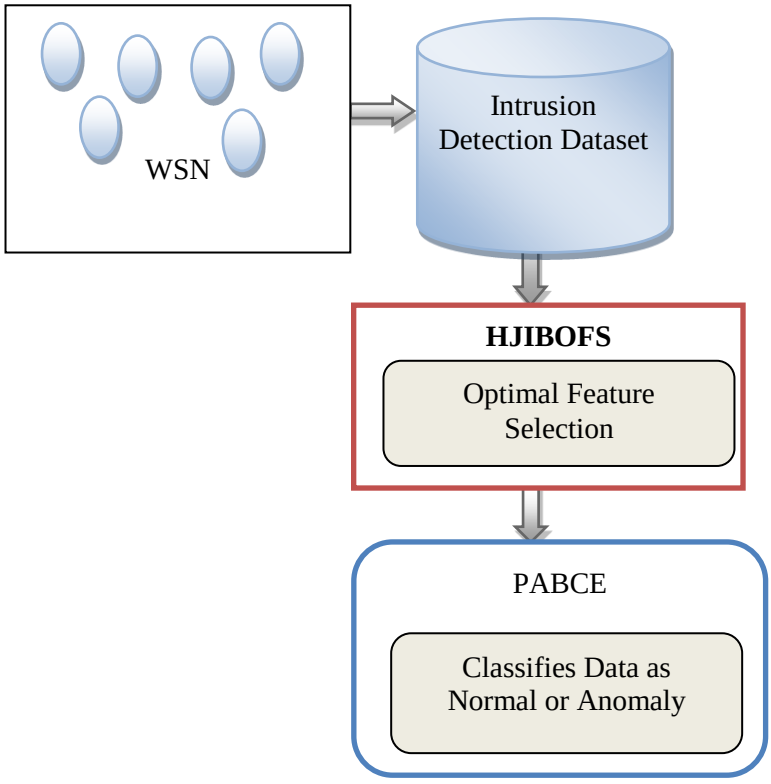


Figure 1.Architecture diagram of HJIBOFS-PABCE

1.1. Paper Outline

The rest of this paper is organized into five different sections. Section 2 elaborates the issues and challenges of intrusion detection in the related works. Section 3 describes the model for intrusion detection based on feature selection and classification. In section 4, experimental evaluation is carried out with dataset and the performance results of various metrics are discussed in section 5. Finally, section 6 provides the conclusion of the work.

2. RELATED WORKS

Manzoor and Kumar (2017) developed an artificial neural network-based intelligent intrusion detection system with fewer features. The system was to identify valuable and useless features using both information gain and correlation. The system then performs intelligently to categorize analysis data into attack and non-attack classes. The feature minimized system to achieve an identical performance as a normal system. The system extracts features using concepts of achieving information and correlation. The method uses preprocessing to eliminate redundant and irrelevant data from the dataset to develop resource exploitation and minimize time complexity. However, the system was unsuccessful in using the fast converging learning algorithms to achieve an accurate detection rate.

Shenfield et al.(2018) developed an artificial neural network to identify the intrusion with higher accuracy. The intelligent intrusion detection system has extensively developed the performance of signature-based detection methods. An artificial neural network classifier employed the detection system to recognize shell code patterns of network traffic. The designed method was unsuccessful to choose the optimal attributes to reduce the time complexity in intrusion detection.

Karami (2018) designed a modified Self-Organizing Map (SOM) with benign outliers for anomaly-based intrusion detection. This approach enables better investigations by integrating a large amount of network traffic into an easy-to-understand 2D format and an effortless user interaction. However, the designed system was unsuccessful to perform efficient intrusion detection for a large volume of network data.

Tama et al. (2020) introduced the stacked ensemble method to develop the performance of anomaly-based intrusion detection. The stacked ensemble was ensemble-based learners with other ensembles such as random forest, gradient boosting machine,

and XGBoost. However, the method was unsuccessful to consider multi-class classification. A different conventional stacking method was employed were several single weak learners. The most important issues identified from the above-said literature are overcome by introducing a novel technique. The brief explanation model is presented in the next section.

3. METHODOLOGY.

3.1Hybrid Jaccard Index Butterfly Optimization (HJIBOFS)

The HOBOFS process is performed in proposed technique for feature selection. The fitness of optimal features is carried out by cosine similarity index where the space and Existing Multi-objective algorithm CLSOBBOA developed by Adel SaadAssiri (2021**)** to perform the detection of intrusion in network. It effectively detects the intrusion by selecting features but failed to reduce error occurrences with selected feature. Hence, butterfly optimization is carried in proposed technique to choose optimal features for detecting intrusion in network.

$$\varphi_F = \mathrm{J}(\mathrm{A},\mathrm{B}) = |\mathrm{A} \cap \mathrm{B}| \,/\, |\mathrm{A} \cup \mathrm{B}| \qquad \text{-Eqn (1)}$$

Using equation (5.1), fitness value among attribute is estimated and denoted by φ_F, attribute is symbolized as 'A, B' specifies two groups with various features. As a result, the output of fitness ranges from 0 to 1.This process is repeated until the maximum iteration is reached. A novel hybrid technique was introduced by Singhet al., (2021) to detect intrusion data in network. It effectively detects intrusion data with enhanced accuracy and minimum time. But, the optimization technique failed to achieve better accuracy on detection. Therefore, Butterfly Optimization is carried in proposed technique to perform efficient optimal feature selection for attack detection.

Input: Number of attributes '$A_1, A_2, A_3, \ldots A_n$'
Output: Select relevant features
Begin
Step 1: For each attribute A_i
Step 2: Initialize the population of $A_1, A_2, A_3, \ldots A_n$
Step 3: Compute fitness function 'φ_F'
Step 4: Update the position of butterfly
Step 5: **End if**
Step 6: **If** (Iteration <max) then

Step 7: Obtain global best features
Step 8: **Else**
Step 9: Go to step 3
Step 10: **End if**
Step 11: **End for**
End

Algorithm 1: Hybrid Jaccard Index Butterfly Optimization

3.2. Polytomous Adaptive Base Class Ensemble (PABCE)

The proposed technique performs Polytomous adaptive base class ensemble process with the selected optimal features. During ensemble classifier algorithm, the different types of attacks are identified by classifying data into various classes. An ensemble classifier is a classifier built depending on mixture of the weak learners. The process of classifying data into different classes is referred as Polytomous. Thus, strong classification result is provided for attack detection.

Input: Number of data '$D_1, D_2, D_3, \ldots D_n$'
Output: Improved suspicious intrusion detection and anomaly intrusion detection accuracy
Begin
Step 1: For each training data 'D_i'
Step 2: Construct 'b' weak learners
Step 3: Construct optimal hyper plane $\vec{q}.D_i + K = 0$
Step 4: Find two marginal hyper planeM_1, M_2
Step 5: The output of weak classifier is $R = sign \sum qy_i \, P\,(D_i, D_A)$
Step 6: **If**$(R > 0)$**then**
Step 7: Data is classified as 'normal'
Step 8: **else**
Step 9: Data is classified as 'anomalous'
Step 10:End if
Step 11: Construct multiple hyper planes
Step 12: Classifies the multiple attacks
Step 13: End for

Step 14: Combine a weak learner $Z = \sum_{i=1}^{b} R_i$

Step 15: **for** each R_i

Step 16: Initialize the weight 'φ'

Step 17: Compute the classification error

Step 18:End for

Step 19: Apply totally corrective property

Step 20: Update the initial weight$Z = \sum_{i=1}^{b} R_i * \eta_t$

Step 21: Find weak learner with minimum error $\arg\min[error\ (R_i)]$

Step 22: Return strong classification results

End

Algorithm 2.Polytomous Adaptive Base Class Ensemble technique

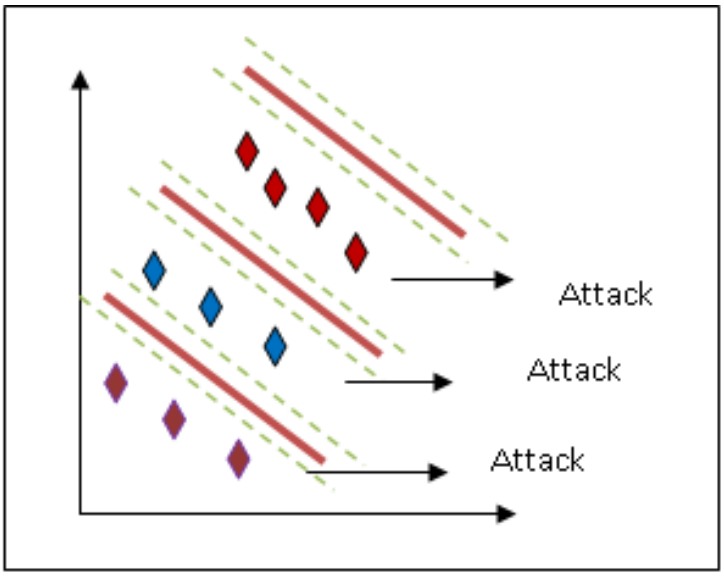


Figure 2. SVM Classifier –Attack detection

4. RESULTS AND DISCUSSIONS

Experimental evaluations of proposed **HJIBOFS -PABCE** and existing **CLSOBBOA** methods namely are implemented using Java language using NETBEANS8.2 IDE tool. For the experimental consideration, NSL-KDD dataset is used and taken from https://www.kaggle.com/hassan06/nslkdd/version/1

This dataset is an enhancement of KDD'99 dataset from which duplicate instances were removed and improved the classification results. The dataset comprises 42 attributes. In NSL-KDD dataset, 42 attributes (i.e., features) are taken as input for feature selection

process. In feature selection process, the relevant features are selected and irrelevant features are removed.

4.1 Performance analysis of intrusion identification rate

Intrusion identification rate is measured as the ratio of number of data identified as suspicious to the total number of data from dataset. It is measured in terms of percentage (%) and formulated as follows.

$$IIR = \frac{Numberofdataidentifiedassuspicious}{Numberofdata} * 100$$

…… Eqn (2)

The intrusion identification rate 'IIR' is determined using equation (2).

Sample calculation:

Existing CLSOBBOA Number of data identified as suspicious is 880 and the total number of data is 1000. Then, the intrusion identification rate is$IIR = \frac{880}{1000} * 100 = 88$ %.

Proposed HJIBOFS -PABCE: Number of data identified as suspicious is 930 and the total number of data is 1000. Then, the intrusion identification rate is$IIR = \frac{930}{1000} * 100 =$ 93%.

Instances	Intrusion identification rate (%)	
	CLSOBBOA	HJIBOFSPABCE
1000	88	93
2000	87	94
3000	90	95

4000	88	96
5000	92	93
6000	90	95
7000	88	92
8000	91	93
9000	89	95
10000	90	96

Table 1: Tabulation of Intrusion Identification Rate

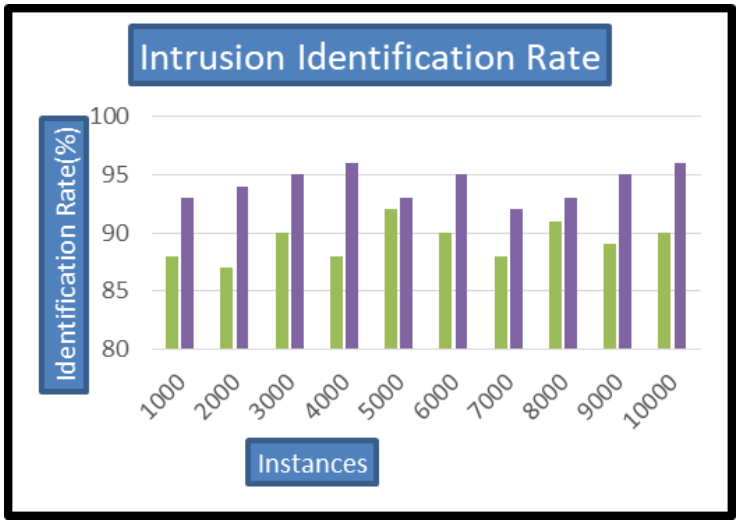


Figure 3: Intrusion Identification Rate

Table 1.shows Tabulation of Intrusion Identification Rate and Figure 3.presents the performance analysis of intrusion identification rate concerning a different number of data.

4.2 Performance analysis of intrusion detection accuracy

Anomaly intrusion detection accuracy determined as the ratio of the number of correctly identified anomalous data to the number of data considered from dataset. The intrusion detection accuracy is evaluated in terms of percentage (%) and formulated as given below.

$$IDA = \frac{N_{identified\ as\ anomalous}}{N} * 100$$

…… Eqn (3)

From equation (3), Intrusion Detection Accuracy 'IDA' is calculated based on the 'N' number of data. Here, '$N_{identified\ as\ anomalous}$' specifies number of data identified as anomalous.

Existing CLSOBBOA: Number of data correctly detected as anomaly is 850 and the total number of data is 1000. Then, the intrusion detection accuracy is measured as$IDA = \frac{850}{1000} * 100 = 85\ \%$.

Proposed HJIBOFS-PABCE: Number of data correctly detected as anomaly is 910 and the total number of data is 1000. Then, the intrusion detection accuracy is measured as$IDA = \frac{910}{1000} * 100 = 91\ \%$.

Instances	Intrusion Accuracy(%)	
	CLSOBBOA	HJIBOFSPABCE
1000	85	91
2000	86	93
3000	89	94
4000	88	95
5000	91	93
6000	90	95
7000	87	92
8000	90	93
9000	88	94
10000	89	95

Table 2: Intrusion Detection Accuracy

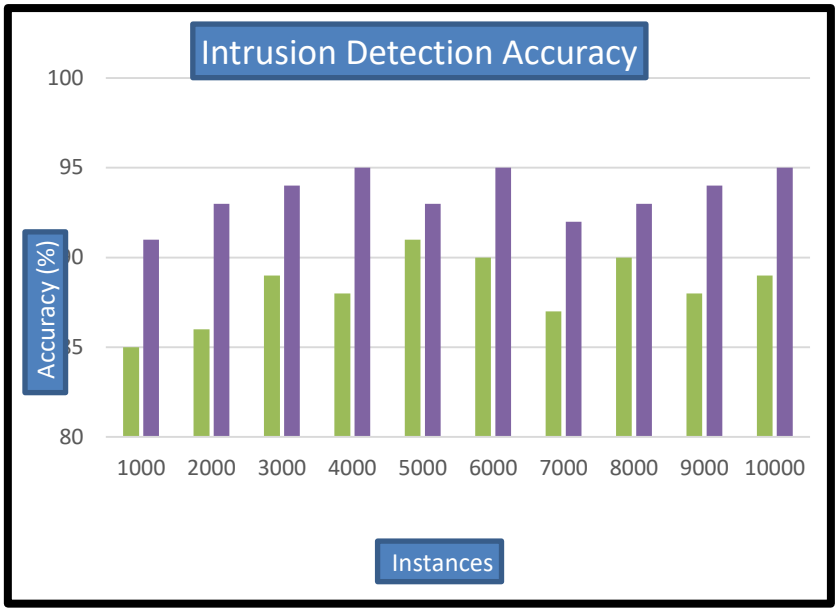


Figure 4: Intrusion Detection Accuracy.

4.3 Performance Analysis of Intrusion Detection Time

Intrusion detection time is defined as the amount of time taken by the classifier to identify data attacks as normal or anomalous. It is evaluated in terms of milliseconds (ms). The detection time is calculated in equation.

$$ID_{Time} = N * Time_{identifynormaloranomalous} \quad \text{- Eqn (4)}$$

From equation (4), time for intrusion detection is measured. Here, 'N' denotes the number of data and. '$Time_{identifynormaloranomalous}$' describes time consumed to identify single data as normal or anomalous.

Sample Calculation

Existing CLSOBBOA: Time to identify single anomaly data is 0.023ms, and the total data is 1000. Then, the intrusion detection time is calculated as$ID_{Time} = 1000 * 0.023 = 23\ ms$.

Proposed HJIBOFS -PABCE: The time to identify single anomaly data is 0.021ms, and the total number of data is 1000. Then, the intrusion detection time is calculated as$ID_{Time} = 1000 * 0.021 = 21\ ms$.

Instances	Intrusion Detection Time (ms)	
	CLSOBBOA	HJIBOFSPABCE
1000	23	21
2000	28	24
3000	32	27
4000	36	30
5000	39	34
6000	43	39
7000	47	42
8000	51	46
9000	58	50
10000	63	54

Table 3: Intrusion Detection Time

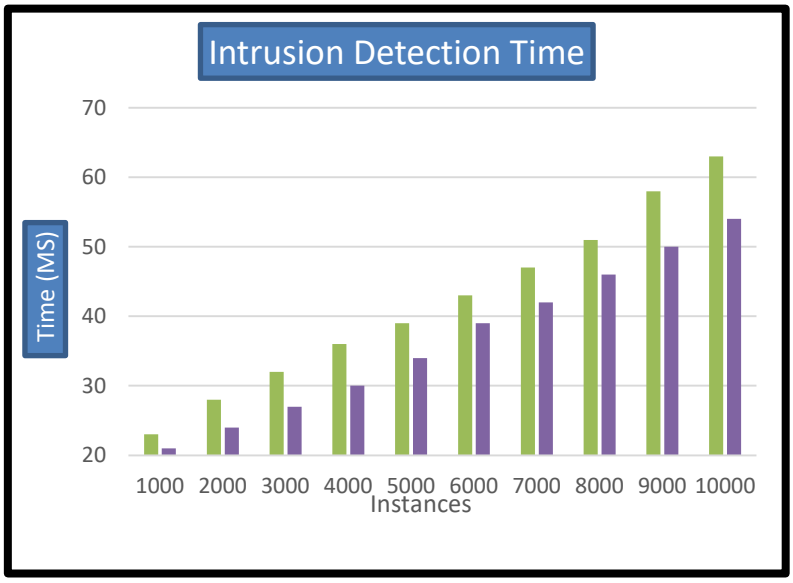


Figure 5: Intrusion Detection Time

Table 3.shows Tabulation of Intrusion Detection time and Figure 5 presents the performance analysis of intrusion detection time concerning a different number of data.

5.CONCLUSION

An efficient **HJIBOFSPABCE** is introduced to attain higher accuracy on anomaly intrusion detection in a network. An intrusion detection dataset considers the number of attributes. For each attribute, it is measuring fitness for all dragonflies using the Jaccard similarity index. Based on measured fitness for intrusion attack detection, selecting the optimal attribute helps reduce time and improves accuracy.

BIBLIOGRAPHY

Aldweesh, A., Derhab, A., &Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems, 189, 105124.

Adel SaadAssiri (2021) On the performance improvement of Butterfly Optimization approaches for global optimization and Feature Selection https://doi.org/10.1371/journal.pone.0242612

Dua, M. (2020). Attribute selection and ensemble classifier based novel approach to intrusion detection system. Procedia Computer Science, 167, 2191-2199.

Elmasry, W., Akbulut, A., &Zaim, A. H. (2020).Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic.Computer Networks, 168, 107042.

Falcão, F., Zoppi, T., Silva, C. B. V., Santos, A., Fonseca, B., Ceccarelli, A., &Bondavalli, A. (2019, April).Quantitative comparison of unsupervised anomaly detection algorithms for intrusion detection. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (pp. 318-327)

Haq, E. U., Lyu, X., Jia, Y., Hua, M., & Ahmad, F. (2020). Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach. Energy Reports, 6, 1099-1105. Ibrahim, D., &Alruhaily, N. (2020). Anomaly Detection in Wireless Sensor Networks: A Proposed Framework.

Karami, A. (2018). An anomaly-based intrusion detection system in presence of benign outliers with visualization capabilities. Expert Systems with Applications, 108, 36-60.

Kasongo, S. M., & Sun, Y. (2019). A deep learning method with filter based feature engineering for wireless intrusion detection system. IEEE Access, 7, 38597-38607.

Manzoor, I., & Kumar, N. (2017). A feature reduced intrusion detection system using ANN classifier. Expert Systems with Applications, 88, 249-257.

Singh, A., Nagar, J., Sharma, S., &Kotiyal, V. (2021). A Gaussian process regression approach to predict the k-barrier coverage probability for intrusion detection in wireless sensor networks.Expert Systems with Applications, 172, 114603.

Shenfield, A., Day, D., &Ayesh, A. (2018).Intelligent intrusion detection systems using artificial neural networks. ICT Express, 4(2), 95-99

Tama, B. A., Comuzzi, M., & Rhee, K. H. (2019). TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. IEEE Access, 7, 94497-94507.

JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CERVICAL CANCER PREDICTION USING MACHINE-LEARNING TECHNIQUES

Ms J Fathima Kaleema
Department of Computer Science,
TBAK College for Women,
Kilakarai, Tamilnadu, India.

Abstract: Cervical cancer is the fourth most common cancer among women around the world. The objective of this study is to provide a comparative study to predict the cervical cancer dataset. The extraction involved over 38 attributes here used three different machine learning algorithms (XGBoost, Decision Tree, Logistic Regression) has been applied on four different medical tests (Biopsy, Cytology, Hinselmann, and Schiller) as four different target variables. The disease cannot be identified in the early stage. The result showed that the performance of EML outperforms other classifiers after evaluation. In this paper it exposes the classifiers can effectively achieve the best performance with the least number of highly important attributes.

Keyword: Machine Learning, XGBoost , Decision Tree, Logistic Regression

I INTRODUCTION

Cervical cancer is a type of cancer that begins in the cervix which connects the uterus and vagina. This cancer can affect the deeper tissues of their cervix and may be spread to other parts of their body after breast cancer, colorectal cancer, lung cancer and skin cancer as per [1].Many young women become infected with multiple types of human papillomavirus, which then can increase their risk of getting cervical cancer in the future and it may spread to other parts of their body . Most of the women who don't have early abnormal changes who do not have regular examinations are at high risk for localized cancer by the time they're age of 40 and for invasive cancer by age of 50.The target of this paper to diagnos the four different variables hinselmann, schiller, citology, and biopsy. This paper examines to diagnose the biopsy

Cervical cancer dataset collected from kaggle dataset and used google colab the platform for the purpose of coding for dataset prediction. The methodology involves use of supervised learning algorithms and classification technique like Decision Tree, Logistic Regression and XGBoost Classifier with Dimensionality Reduction technique.

II LITERATURE REVIEW

Early detection of cervical cancer is a way to avoid the death rate because most of the women affected from cervical cancer .They didn't identify the cervical cancerous in abnormal state this is main cause of an increase the female death rate in this world[2].

WEN WU "Data-Driven Diagnosis of Cervical Cancer with Support Vector Machine- Based Approaches", Department of Blood Transfusion, Jinan Military General Hospital, Jinan, China Year: 2017 [3]. In this paper it claimed that some set of data to the way of increasing the possibility of developing cervical cancer. An advantage of this paper is to SVM(Support vector Machine) to introduce the diagnosis the cervical cancer. It supports to improve the SVM method, support vector machine-recursive feature elimination and support vector machine-principal component analysis (SVM-PCA), are proposed to diagnose the cancer samples.

H. Teame et al., "Factors associated with cervical precancerous lesions among women screened for cervical cancer in Addis Ababa, Ethiopia: A case control study," PLoS ONE 13(1) e0191506, 2018[4]. Age play the vital role in increasing the risk of developing the cervical cancer. In this paper, most of the women infected with HPV during sexual transmitted infections were associated with precancerous lesion.

DATA COLLECTION

The data was collected from the kaggle Machine Learning repository and available in online. The dataset contains 858 instances and 36 different features

No	Attribute	No	Attribute
1.	Age	19.	STDs:pelvic inflammatory disease
2.	Number of sexual partners	20.	STDs:genital herpes
3.	First sexual intercourse	21.	STDs:molluscum contagiosum
4.	Num of pregnancies	22.	STDs:AIDS
5.	Smokes	23.	STDs:HIV
6.	Smokes (years)	24.	STDs:Hepatitis B
7.	Smokes (packs/year)	25.	STDs:HPV
8.	Hormonal Contraceptives	26.	STDs: Number of diagnosis
9.	Hormonal Contraceptives (years)	27.	STDs: Time since first diagnosis
10.	IUD	28.	STDs: Time since last diagnosis
11.	IUD (years)	29.	Dx:Cancer
12.	STDs	30.	Dx:CIN
13.	STDs (number)	31.	Dx:HPV
14.	STDs:condylomatosis	32.	Dx
15.	STDs:cervical condylomatosis	33.	Hinselmann
16.	STDs:vaginal condylomatosis	34.	Schiller
17.	STDs:vulvo-perineal condylomatosis	35.	Citology
18.	STDs:syphilis	36.	Biopsy

III DATA PRE-PROCESSING

It is a technique that is used to convert raw data into numeric dataset .In this dataset some data having NaN (Not a Numeric) values are replaced with numerical values with use of digit transformation using sklearn techniques. After converting dataset feeding it to the algorithm. For getting better result from the Machine Learning algorithm, the ML algorithm not support null values so there is need to preprocess for medical dataset which has major attribtues, The dataset was split into Train dataset and Testing dataset.In this process 20% of test dataset and 80% of trained dataset

IV METHODOLOGY

i) XGBoost

Extreme Gradient Boosting or XGBoost is a decision tree based ensemble ML Algorithm that is used in the library of gradient boosting Some of the major benefits of XGBoost are that its highly scalable/parallelizable, easy to visualize, quick to execute, and typically outperforms other algorithms are used to more regularized model formalization, to control over-fitting, which gives it better performance.

ii) DECISION TREE

Decision Tree is a supervised learning technique that will be used for both classification and regression problems, but mostly it is preferred to solving classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. This algorithm compares the values of root attribute with the root (dataset) attribute and, based on the comparison, it follow the branch of the next node.

iii) LOGISTIC REGRESSION

Logistic regression is also one of the most popular Machine Learning algorithms, It's comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

V RESULTS AND DISCUSSION

The experimental test results on the proposed model are explained the predictive performance of XGBoosting, Logistic Regression and decision tree algorithm is analyzed by the performance metrics of accuracy.

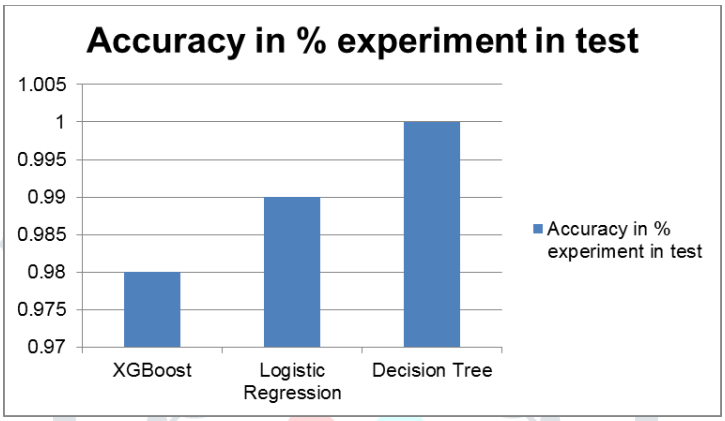
PREDICTIVE ACCURACY ANALYSIS

The predictive performance of the proposed model is experimented on the training set. The predictive accuracy of the proposed model is shown in Figure 1. Moreover, the accuracy for XGBoosting, Logistic Regression and Decision Tree for cervical cancer classification test is given in Table 1 it shows the test accuracy for comparison of three algorithms. Fig 1 shows the result of accuracy in the percentage of experiment test

Table 1: Accuracy of XGBoosting, Logistic Regression and decision tree

Learning Algorithm	Accuracy in % experiment in test
XGBoost	**0.98**
Logistic Regression	**0.99**
Decision Tree	**1.0**

Figure 1 Accuracy after pre-processing



VI CONCLUSION

In this paper, it proposed cervical cancer prediction model with XGBoosting Logistic regression and decision tree algorithm on cervical cancer dataset collected form kaggle data repository. The proposed model solves the problem of biased classification on imbalanced observation by non-ensemble algorithm through ensemble classifier namely the decision tree. The predictive performance of the proposed model is evaluated by employing different performance metrics such as accuracy on the test set. The result of performance analysis reveals that the decision tree algorithm has better performance to comparing between logistic regression and XGBoosting algorithms. Hence, the Decision tree algorithm is a better performance on prediction of the majority class and poor performance on the minority class.

REFERENCE

[1] American Cancer Society, Cancer Facts for Women, Available at: https://www.cancer.org/healthy/find-cancer-early/womens-health/cancer-facts-for-women.html(Accessed date 25.07.2020).

[2] W. J. Koh et al., "Cervical cancer, version 2.2015," JNCCN Journal of the National Comprehensive Cancer Network, vol. 13, no. 4. Harborside Press, pp. 395–404, 01-Apr-2015.

[3] W. Wu and H. Zhou, "Data-driven diagnosis of cervical cancer with support vector machine-based approaches," IEEE Access, vol. 5, pp. 25189–25195, Oct. 2017.

[4] H. Teame et al., "Factors associated with cervical precancerous lesions among women screened for cervical cancer in Addis Ababa, Ethiopia: A case control study," PLoS ONE 13(1) e0191506, 2018

[5] S. Sharma, "Cervical Cancer stage prediction using Decision Tree approach of Machine Learning",*International Journal of Advanced Research in Computer and Communication Engineering* vol. 5, Issue 4,2016.

[6]D. N. Punjani, K. H. Atkotiya, "Cervical Cancer Prediction using Data Mining",*International Journal for Research in Applied Science & Engineering Technology (IJRASET)* , vol. 5, Issue XII, 2017.

[7] S. Sharma, "Cervical Cancer stage prediction using Decision Tree approach of Machine Learning",*International Journal of Advanced Research in Computer and Communication Engineering* vol. 5, Issue 4,2016.

[8] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13–17, pp. 785–794, San Francisco, CA, USA, August 2016.

[9] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, 2013.

[10] Y. M. S. Al-Wesabi, A. Choudhury, and D. Won, "Classification of cervical cancer dataset," in *Proceedings of the 2018 IISE Annual Conference and Expo*, pp. 1456–1461, Orlando, FL, USA, May 2018.

[11] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: an ensemble approach," *Future Generation Computer Systems*, vol. 106, pp. 199–205, 2020.

[12]"Cervical cancer," 2020, https://www.nccc-online.org/hpvcervical-cancer/cervical-cancer-overview/.

[13] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: an ensemble approach," *Future Generation Computer Systems*, vol. 106, pp. 199–205, 2020.

Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776700

FRACTAL DIMENSION FEATURE BASED BREAST TUMOR CLASSIFICATION USING DATA MINING

A.Krishnaveni[1], M.S Irfan Ahmed[2]
Assistant professor1, Professor2 in Department of Computer Science 1&2
Thassim Beevi Abdul Kader College for Women, Kilakarai- 623517,
Ramanathapuram, Tamilnadu, India
E-mail: drakrishnaveni@gmail.com1,directortbakc.rir@gmail.com2

Abstract
Cancer is a disease of Deoxyribonucleic Acid (DNA). A breast tumor is a malignant growth that creates from breast tissue. It is an illness wherein dangerous (malignant growth) cells structure in the tissues of the breast. Whenever cancer is analyzed as harmless, specialists will typically let it be instead of eliminating it. Imagine a future where users can predict when any disease might occur, treat any disorder in real-time, and even prevent diseases from ever happening. This is the future of medicine, one where we will be able to create a precise roadmap for a disease-free life. As an independent, non-profit researcher focused on genomics research has a century of experience, and the foresighted vision to make this exciting future possible. In this paper, Predicting a person's tumor (normal, benign, or malignant) based upon his or her tumor features like fractal dimension, symmetry, concave points, concavity, compactness, smoothness, area, texture, perimeter, and radius are calculated. Machine learning classification techniques and data mining algorithms can be used on these types of problem statements.
Keywords: DNA, ACS, Malignant, Disorder, Researcher, Fractal dimension

I INTRODUCTION

The Indian Council of Medical Research (ICMR) measures that there will be a twelve percent ascend in disease cases in India in the following five years. The most widely recognized types of disease influencing individuals in India are inseparable malignant growth, cervical disease, and oral malignant growth. One in each ten new malignant growth analyses in India yearly, is a bosom disease, according to reports and it causes the most number of disease-related passings in ladies. Inseparable disease happens in milk-delivering channels of ladies and other bosom cells, which partition quickly and structure bumps. They may ultimately spread to lymph hubs or somewhere else. Early determination, medical procedure, prescriptions, and radiation can effectively treat the illness. [1]. Protecting human beings from cancer is needed to improve our country in a healthy way. All of us know about health. Important cells in the human body have functioned in a well-defined manner is necessary. If not functioning means some cells are broken/damaged. These are called unhealthy cells/cancer cells. Abandoned intensification of the breast area severely causes breast cancer. So screening the mammogram of every human being is necessary to control/reduce breast cancer.

In the initial stage of cancer have been some symptoms of precaution like breast pain, emotional stress, a deposit of micro classification, increasing the breast density, etc. The mammography test is the primary and also referred by many doctors for scanning cancer in the initial stage. Areas like machine learning and medical image processing are always ready to give rapid

algorithms for computing breast growth for analysis. Medical images containing confidential information about patients' disease types like whether the mammogram input image has cancer or not being secured. Input images having some disturbances like blurred, low contrast, unwanted information like background, patient id, label, etc. So pre-processing is required to remove unnecessary data in the acquired image.

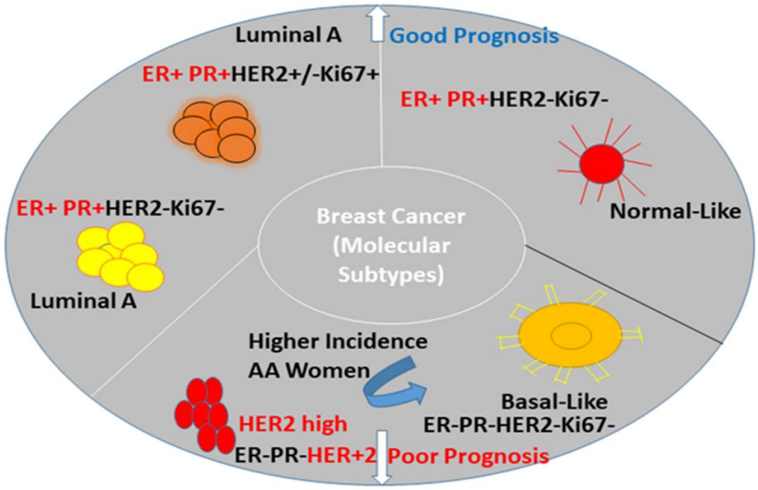


Figure 1: Breast Malignancy image [2].

In Figure 1, breast cancer and its molecular subtypes are explained. Estrogen (E+) Receptor-Positive, Progesterone (P+) Receptor-Positive, and HER-2(H+) Receptor Amplified are the basic three forms of breast cancer. Aromatase inhibitors, Tamoxifen, Herceptin, Trastuzumab, and Monoclonal Antibodies are prescribed treatments by Mark Adams. Triple-Negative Breast Cancer (TNBC) are (E-), (P-), and (H-) and recommended treatment is chemotherapy.

Medical images containing confidential information about patients' disease types like whether the mammogram input image has cancer or not being secured. Input images having some disturbances like blurred, low contrast, unwanted information like background, patient id, label, etc. So pre-processing is required to remove unnecessary data in the acquired image.

II LITERATURE REVIEW

Manish Charan et al. portrayed bosom disease as the second most normal malignant growth and the main source of malignant growth-related passing in ladies all over the planet [2]. The point of Habib Dhahri et al. denoted a hereditary user interface design strategy for selecting the appropriate elements with the help of machine learning [3]. Adam B Nover et al., assessed existing and arising innovations utilized for bosom malignant growth screening and discovery to recognize regions for potential improvement [4]. Omaer Faruq Goni et al., a review has recommended a profound neural organization with highlight choice methods to anticipate bosom disease. The examination of the recommended system is measured by exactness, precision, review, explicitness, awareness, f measure, and MCC [5]. Tanishk Thomas et al. expects to give a similar report by applying different AI calculations, for example, Support Vector Machine, K-Nearest Neighbor, Naïve Bayes, Decision Tree, K-implies, and Artificial Neural Networks on Wisconsin Diagnostic dataset to foresee bosom malignant growth at a beginning phase [6]. D. Sandeep et al., Lal Hussain et al., Dana Bazazeh et al., make a correlation of different AI calculations which considered and resolved the breast tumor [7-9].

Mamatha Sai Yarabarla et al., utilize the new advances in the improvement of CAD frameworks and related procedures. The backbone of the venture is to foresee whether or not the individual is having bosom malignant growth [10]. Nur Syahmi Ismail et al., contrasted bosom malignant growth identification and two model organizations of profound learning strategies. The general cycle includes picture preprocessing, order, and execution assessment [11]. Naresh Khuriwal et al. applied profound learning innovation to determine bosom disease to have an exactness of 99.67% [12]. Aditi Kajala et al. presents a short outline of the bosom disease conclusion utilizing AI calculations used to build the proficiency and adequacy of foreseeing malignant growth [13]. Shubham Sharma et al. introduced a correlation between the generally well-known AI calculations and strategies usually utilized for bosom disease expectation [14].

In medical research, correct predictions are needed to identify whether the person has the disease's tumor or not. Even though all the latest technologies came and were utilized for predictions of tumors but the mortality rate of a tumor in the breast is not down. To resolve this context the introduced work is needed to identify the most significant parameter involved in sample selection and classification of tumor class A or B is observed.

III MATERIALS AND METHODS

The Wisconsin Breast Cancer Dataset (WBCD) was downloaded from the UCI machine learning repository [16]. For the classification of breast cancer, real attributes in multivariate form have no missing values from five sixty-nine instances and thirty-two attributes. Based on the physiology element analysis some fluid samples were taken from the patient's breast for the creation of this dataset. Some important attributes are patient ID and category that belongs to benign or malignant. Tumor features like fractal dimension, symmetry, concave points, concavity, compactness, smoothness, area, texture, perimeter, and radius are calculated. From this feature list, the most significant feature identification is the primary task of our proposed work. Then the main aim is the classification of breast image whether the dead cells of tissues of the breast (malignant) or unrestrained growth of the cells but not affected seriously (benign) are identified.

Step1: Download WBCD Dataset from UCI Machine-learning repository
Step2: Data Discretization
Step3: Ten-fold cross-validation
Step4: Classification by different classifiers
Step5: Output data is either in Benign or in the malignant category.
The above-mentioned five steps are clearly explained in the following Figure 2.

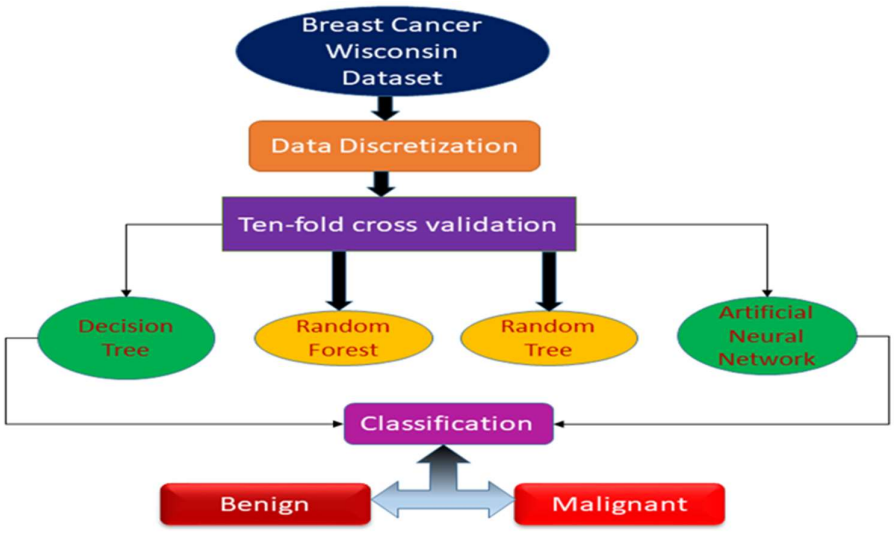


Figure 2: Framework for Classification

Early-stage tumor prediction is mainly observed and picturized in above Figure 2.

Classification:

In this research work, two classes are used for the prediction of breast tumors. CLASS A is used to represent the tumor which is not a serious case but is a benign strategy. CLASS B is involved in the prediction of malignant causes of breast cancer.

3.1 MODELS USED FOR CLASSIFICATION

Selecting the appropriate model for forecasting breast cancer in an earlier stage is a difficult process. The large volume of data set needs dimensionality reduction and type of the model like supervised or unsupervised. In supervised learning (classification) target variable has an either-or choice like in breast cancer, cancer, or non-cancer. Based on this observation and also obtained dataset having the defenseless variable CLASS A/ CLASS B. Therefore, the Decision table, Random forest, Random tree, and artificial neural network models are utilized and explained in the following.

1) Decision Table (DT)

A Decision Table (DT) is only a plain portrayal of all conditions and activities. Choice Trees are constantly utilized while the handling rationale is extremely convoluted and includes numerous circumstances. The primary parts utilized for the arrangement of the Data Table are Conditions Stubs, Action Stubs, and rules.

2) Random Forest (RForest)

It assembles choice trees on various examples and takes their greater part vote in favor of characterization and normal if there should arise an occurrence of relapse.

3) Random Tree (RT)

Random trees arbitrarily select the nodes and decide how to classify the given input randomly. It gives the solution strategy faster and also incorporates the details from the random forest.

4) Multi-Layer Perceptron (ANN)

Multi-facet perceptron characterizes the most perplexing engineering of counterfeit neural organizations. It is considerably shaped by different layers of the perceptron.

IV PERFORMANCE EVALUATION

The accuracy of introduced classifiers using ten-fold cross-validation with the performance of the presented model by Correlation Coefficient (CF), Despicable Complete Inaccuracy (DCI), Source Despicable Square off Miscalculation (DSM), Source Relation Square off Fault (RSF).

4.1 Correlation Coefficient (CC): It is utilized to quantify how solid a relationship is between two factors. It very well may be helpful in information investigation and displaying to all the more likely comprehends the connections between factors. The factual connection between the two factors is alluded to as their relationship. In the breast cancer dataset, the correlation coefficient is used to scatter the images in visualization. So how many features of the same variable made the difference in the entire data set was identified easily.

4.2 Mean Absolute Error (MAE): An average of all outrights errors are called a Despicable Complete Inaccuracy (Mean Absolute Error).

MAE = Average of All outright errors

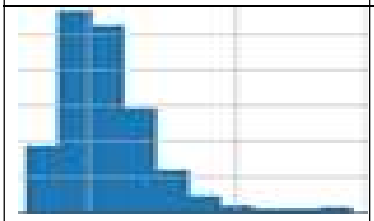
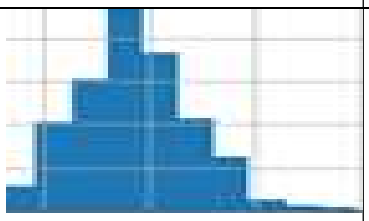
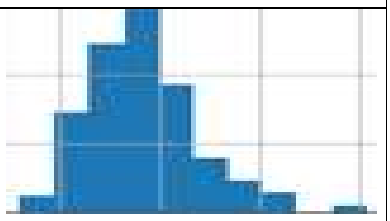
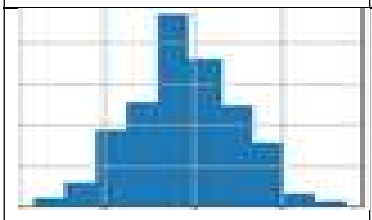
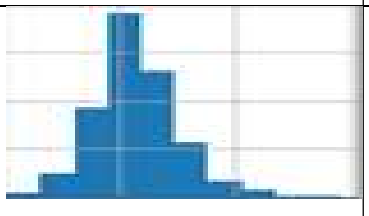
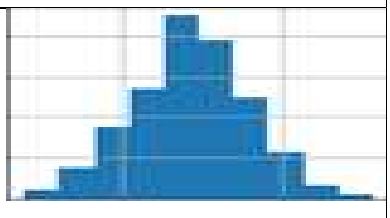
4.3 Root Mean Squared Error (RMSE): The relapse line fits the proportion of the informative elements is called root mean squared error. Root Mean Squared Error can likewise be interpreted as Standard Deviation in the residuals.

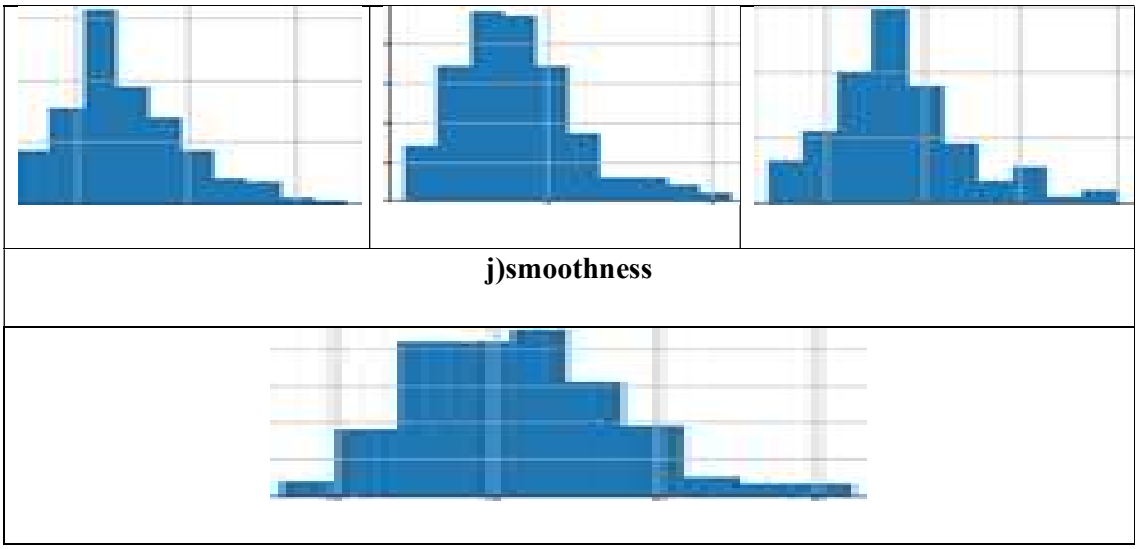
4.4 Relative Absolute Error (RAE): Overall outright blunder between two numeric vectors is known as SDSM. It is also known as a relative absolute error [27].

4.5 Relative Root Squared Error (RRSE): It is very useful to find out the error rate by using the direct pointer method [18].

From the following Table 1, visualization of all the specified attributes results is viewed by the Weka data mining tool after applying the ten-fold cross-validation. The purpose of choosing the ten-fold cross-validation is it returns the desired results in a fine-tuned manner.

Table 1 Visualization results obtained by the weka data mining tool

a) fractal dimension	**b) area**	**c) symmetry**
		
d) perimeter	**e) texture**	**f)radius**
		
g) compactness	**h) concavity**	**i)concave point**



j)smoothness

In following Table 2, describes the various machine learning models (DT, RF, RT, ANN) used for classification, and the performance of each classifier is noted with the quality metrics such as CF, MAE, RMSE, RAE, and RRSE. The results show that ANN has a high accuracy rather than other models. Here classification error occurred during the report generated with the multilayer perceptron model like MLP zero. So test dataset faced this type of classification error and sensitivity is used to measure the positive results obtained by the introduced multilayer perceptron classifier.

Table 2 Classification results obtained by DT, RF, RT, and ANN

Metrics/ Models	CF	MAE	RMSE	RAE	RRSE
DT	**0.7588**	**0.0082**	**0.0118**	**60.7786**	**65.1356**
RF	**0.9157**	**0.0053**	**0.0077**	**39.1832**	**42.4013**
RT	**0.7726**	**0.0085**	**0.012**	**63.0825**	**66.6229**
ANN	**0.9666**	**0.0028**	**0.0047**	**21.2167**	**25.92**

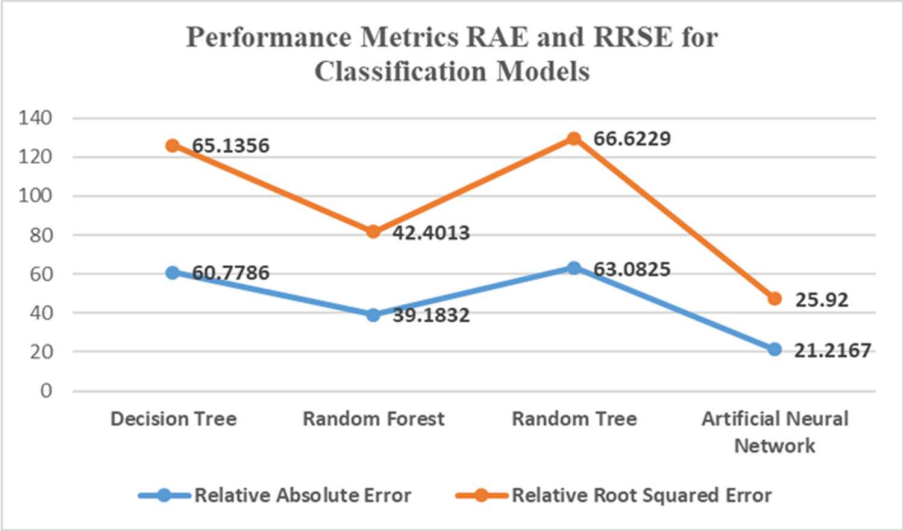


Figure 3: Classification Results

Figure 3, shows that Artificial Neural Network (ANN) has good performance. Then Random Forest has taken a place. Therefore ANN and Random Forest Classifiers give better classification results.

V CONCLUSION

Globe contains many objects which perform certain operations. Human perception is eminent to find out those objects are called images. We can't imagine our world without objects. Image capturing is the primary task for performing several operations. But the captured image should be in a clear manner such as high contrast, high quality, without blurred, without noise are required. Biomedical imaging is necessary to synthesize and analyze the internal regions of the human body without any changes. To give extra stamina to the classification is done by using machine learning models is addressed in this work. In this research work, DT, RF, RT, and Artificial Neural Network (ANN) with Ten-fold cross-validation are applied and accuracy was calculated by using error rates. Correlation Coefficient (CC), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Root Relative Squared Error (RRSE) are used to evaluate the performance of the presented model. Finally, ANN has a minimum error rate and high accuracy than other introduced models. In the future, machine learning models applied to various real-time applications of social media like Facebook, Twitter, and so on.

References

[1] News18.com

[2] Manish Charan, Ajeet K. Verma, Shahid Hussain, Swati Misri, Sanjay Mishra, Sarmila Majumder, Bhuvaneswari Ramaswamy, Dinesh Ahirwar, Ramesh K. Ganju, "Molecular and Cellular Factors Associated with Racial Disparity in Breast Cancer", Int. J. Mol. Sci. 2020, 21, 5936; doi:10.3390/ijms21165936

[3] Habib Dhahri, Eslam Al Maghayreh, Awais Mahmood, Wail Elkilani, Mohammed Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms Hindawi Journal of Healthcare Engineering", Volume 2019, Article ID 4253641,11pp. https://doi.org/10.1155/2019/4253641

[4] Adam B. Nover, Shami Jagtap, Waqas Anjum, Hakki Yegingil, Wan Y. Shih, Wei-Heng Shih, Ari D. Brooks, "Modern Breast Cancer Detection: A Technological Review" Hindawi Publishing Corporation, International Journal of Biomedical Imaging Volume 2009, Article ID 902326, 14 pages doi:10.1155/2009/902326
[5] Omaer Faruq Goni, Fahim Sifnatul Hasnain, Abu Ismail Siddique, Oishi Jyoti, Habibur Rahaman, "Breast Cancer Detection using Deep Neural Network", 2020 23rd International Conference on Computer and Information Technology (ICCIT), 19-21 December 2020 IEEE
[6] Tanishk Thomas, Nitesh Pradhan, Vijaypal Singh Dhaka, "Comparative Analysis to Predict Breast Cancer using Machine Learning Algorithms: A Survey", Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT-2020) IEEE
[7] D. Sandeep, G.N. Beena Bethel, "Accurate Breast Cancer Detection and Classification by Machine Learning Approach" 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC) DOI: 10.1109/I-SMAC52330.2021.9640710 IEEE
[8] Dana Bazazeh, Raed Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis", 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA) DOI: 10.1109/ICEDSA.2016.7818560 IEEE
[9] Lal Hussain, Wajid Aziz, Sharjil Saeed, Saima Rathore, Muhammad Rafique, "Automated Breast Cancer Detection Using Machine Learning Techniques by Extracting Different Feature Extracting Strategies", 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) IEEE
DOI: 10.1109/TrustCom/BigDataSE.2018.00057
[10] Mamatha Sai Yarabarla, Lakshmi Kavya Ravi, A. Sivasangari, "Breast Cancer Prediction via Machine Learning", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), DOI: 10.1109/ICOEI.2019.8862533 IEEE
[11] Nur Syahmi Ismail, Cheab Sovuthy, "Breast Cancer Detection Based on Deep Learning Technique", 2019 International UNIMAS STEM 12th Engineering Conference (EnCon), DOI: 10.1109/EnCon.2019.8861256 IEEE
[12] Naresh Khuriwal, Nidhi Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm", 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) DOI: 10.1109/ICACCCN.2018.8748777 IEEE
[13] Aditi Kajala, V K Jain, "Diagnosis of Breast Cancer using Machine Learning Algorithms-A Review", 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)"DOI: 10.1109/ICONC345789.2020.9117320 IEEE
[14] Shubham Sharma, Archit Aggarwal, Tanupriya Choudhury, "Breast Cancer Detection Using Machine Learning Algorithms" 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS) 10.1109/CTEMS.2018.8769187 IEEE
[15] William H Wolberg, W Nick Street, and Olvi L Mangasarian. 1992. Breast cancer Wisconsin (diagnostic) data set. UCI Machine Learning Repository [http: //archive. ics. uci. edu/ml/] (1992).
[16] https://archive.ics.uci.edu/ml/datasets.php

[17] Arumugam K, Ramasamy S, Subramani D, "Binary Duck Travel Optimization Algorithm for Feature Selection in Breast Cancer Dataset Problem", Vol.251, pp-157-167 Springer (2022).
[18] https://www.gepsoft.com/gxpt4kb/Chapter10/Section1/SS07.htm
[19] Krishnaveni Arumugam, Shankar Ramasamy, Duraisamy Subramani, "Chaotic Duck Traveler Optimization (cDTO) Algorithm for Feature Selection in Breast Cancer Dataset Problem", Turkish Journal of Computer and Mathematics Education, Vol.12 No.4 (2021), pp-250-262.
[20] A.Krishnaveni and Shankar, R. and Duraisamy, S., Versatile Duck Traveler Optimization (VDTO) Algorithm Using Triple Segmentation Methods for Mammogram Image Segmentation to Improving Accuracy (March 13, 2021). Available at SSRN: https://ssrn.com/abstract=380 3814 or http://dx.doi.org/10.2139/ssrn.3803814
[21] A.Krishnaveni, R.Shankar, S.Duraisamy, "Duck Cluster Optimization Algorithm with K-Means Clustering for Mammogram Image Segmentation", Solid State Technology, SCOPUS Indexed ISSN NO: 0038-111X Volume 63, Issue 6, (2020).
[22] A.Krishnaveni, R.Shankar, S.Duraisamy, "An Efficient Methodology for Breast Tumor Segmentation using Duck Traveler Optimization Algorithm", PalArch's Journal of Archaelogy of Egypt/Egyptology, Scopus Indexed ISSN NO: 1567- 214X Volume 17, Issue 9, (2020).
[23] A. Krishnaveni, S.Duraisamy, R.Shankar, T.Latha Maheswari, "Heightened biased optimized duck traveler collaborative feature selection and classification for breast cancer dataset problem", IJSRC, 3(2), pp-19-24 (2021).
[24] Krishnaveni A, Shankar R, Duraisamy S, "A Survey on Nature Inspired Computing (NIC): Algorithms and Challenges", Global journal of computer science and technology: D Neural & Artificial Intelligence volume 19 issue 3 version 1.0 Year (2019).
[25] A.Krishnaveni , R. Shankar and S. Duraisamy, "A Review on various Image Thresholding Methods for Mammogram Image Segmentation" Compliance Engineering Journal ISSN NO: 0898-3577 Volume 11, Issue 2, (2020).
[26] Krishnaveni Arumugam, Shankar Ramasamy, Duraisamy Subramani "Improved Duck and Traveler Optimization (IDTO) Algorithm: A Two-way efficient approach for breast tumor segmentation using multilevel thresholding", European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 7, Issue 10, (2020).
[27] https://byjus.com/physics/errors-absolute-error-relative-error/
[28] M.S.Irfan Ahmed, A.Krishnaveni, "Hybrid Data Mining Method for Early Detection of Breast Cancer", Webology, Volume 18, No.6, pp-959-978, (2021).

Authors Profile

Dr. Irfan Ahmed is currently working as the Director-Research, Industry-Institute Relations at Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramnad, Tamilnadu. He received his MCA from Bharathidasan University in 1994. After that, he obtained his Ph.D in Computer Science with specialization in Trusted Networks from Algappa University, Karaikudi. Dr. Irfan's research expertise includes Trustworthy Computing, Networking, and Data Mining, etc. He has authored 4 books and published more than 58 international and national journals along with several conference publications. Professor Ahmed has 27 years of experience including teaching, research, and industry. He has supervised 12 Ph. D students and currently, 3 more are under his supervision. He is the reviewer of a number of journals all over the globe. He has

given several invited and contributed presentations in multiple national and international conferences and symposia. He is the life member in IST (Indian Society for Technology Education), ACM (Association for Computing Machinery), Associate life member in IACSIT (International Association of Computer Science and Information Technology), a member in IAENG (International Association of Engineers).

Uttar Pradesh Journal of Zoology

Volume 44, Issue 11, Page 61-73, 2023; Article no.UPJOZ.2578
ISSN: 0256-971X (P)

An Efficient Fuzzy based MRI Brain Tumor Segmentation and Classification

A. Krishnaveni [a++*] and M. S. Irfan Ahmed [a#]

[a] Department of Computer Science, Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramanathapuram, Tamil Nadu, India.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.56557/UPJOZ/2023/v44i113521

Editor(s):
(1) Prof. Aurora Martínez Romero, Juarez University, Mexico.
Reviewers:
(1) Jaideep Gera, ST Marys Group of Institutions Guntur, India.
(2) Xinxing Wu, Midway University, United States.

Original Research Article

Received: 01/04/2023
Accepted: 01/06/2023
Published: 16/06/2023

ABSTRACT

Picture division is a significant testing factor in clinical picture division. This paper portrays the division strategy comprising two stages. In the initial step, the MRI cerebrum picture is gained from the patients' information base, in that film, relics and commotion are taken out after the Binarization strategy is applied for picture division. The Binaraization works with the assistance of the Fuzzy C Means Clustering calculation, hence the calculation assumes the principle part in the framework, in this most minimal level of the weight vector, higher worth of cancer pixels, calculation speed is accomplished by the Fuzzy C Mean with vector quantization. The point of this exploration work is to give an assortment of fluffy c-implies (FCM) calculation that gives picture bunching utilizing the MRI Brain Tumor data set. The proposed calculation joins the neighborhood spatial data and dark level data in an original fluffy manner. The new calculation is called Fuzziness Confined Message C-Means (FCM^2). FCM^2 can defeat the drawbacks of the known fluffy c-implies calculations and simultaneously improves the grouping execution. The significant trait of FCM^2 is the utilization of a

---

[++] *Assistant Professor;*
[#] *Professor;*
**Corresponding author: Email: drakrishnaveni@gmail.com;*

fluffy neighborhood (both spatial and dark level) closeness measure, expecting to ensure clamor heartlessness and picture detail conservation. Besides, the proposed calculation Experiments performed on manufactured and true pictures show that the FCM^2 calculation is powerful and proficient, giving strength to uproarious pictures.

Keywords: Cell; fuzzy; Binarization; clustering; segmentation; brain tumor.

1. INTRODUCTION

One of the significant errands of picture examination is picture division. Picture division parcels picture into groups that are not difficult to examine [1-4]. A portion of the uses of picture division incorporate example acknowledgment and object ID, including extraction and clinical imaging [5-9]. For clinical pictures, picture division assumes an indispensable part. In this paper, a strategy is introduced which is applied to identify the various tissues present in MRI cerebrum pictures. Subsequent to identifying the mind tissues, a few sicknesses can be analyzed without any problem [10-14].

Brain tumor cell recognition is important and needed for finding the tumor in the earlier stage to reduce the mortality rate of a human being. This paper involves finding pattern images from the given source image [15,16]. In medical images, if the user needs to find the cell structure or bone crack information, we need to manually verify the area. Instead of manual analysis if a system that highlights all the crack patterns, then it would be useful [17-19]. This paper helps to identify the search pattern image from the given large image. In addition, the pattern image can be cropped from the source image itself or can be selected from another file. The rectangular pixel area of the pattern is collected in an array and checked with source image pixel data row-wise and column-wise. The pixel percentage can be adjusted so that nearly matching images can also be recognized.

2. LITERATURE REVIEW

Sandeep Kumar et al. there is a variety of the fluffy c-implies (FCM) calculation is introduced that gives picture bunching. The proposed calculation consolidates nearby spatial data and dark level data in a clever fluffy manner. The new calculation is called fluffy nearby data C-Means (FLICM). By utilizing this calculation they can beat the inconveniences of the past calculations and simultaneously improves the bunching execution [20].

Qiuyu Song et al. raise a better self-learning weighted fluffy calculation, which straightforwardly acquires various loads in distance estimation through consistent iterative self-learning, then, at that point, the distance metric with the loads got from self-learning is implanted in the goal capacity of the fluffy grouping calculation to further develop the division execution and vigor of the calculation. An enormous number of investigations on various kinds of pictures show that the calculation can smother the clamor as well as hold the subtleties in the picture, the impact of portioning complex commotion pictures is better, and it gives preferable picture division results over the current most recent fluffy grouping calculations [21].

Sucharitha M et al., suggested Fuzzy Local Information C-Means calculation (FLICM) is presented which incorporates a boundary-free clever fluffy variable. This calculation can be applied for clinical picture division for characterizing mind MRI into various locales, for example, Gray matter (GM), White matter (WM), and Cerebro-Spinal liquid (CSF). Subsequent to arranging into various tissue locales, a few illnesses can be dissected without any problem. The proposed strategy shows the viability of calculation in portioning cerebrum pictures into various tissue types [22].

Pooja B et al. two calculations are thought of. One is level set division utilizing fluffy c means by utilizing exceptional highlights (SFCM) and another is a division of mind MRI pictures utilizing DWT and head part investigation (PCA) are additionally handled utilizing support vector machine (SVM) for characterization. The execution assessment is finished by processing mean square mistake, top sign to commotion proportion (PSNR), most extreme contrast, outright mean mistake and so forth Here DWT utilizes k-implies bunching and level set utilizes fluffy c-implies grouping. The spatial limitations are named with various records, for example, the client can pick on a specific district of interest and emphasize the form ventures until a more precise outcome is acquired [23].

Standard FCM, Fuzzy Local data C-implies grouping calculation [FLICM], Reformulated Fuzzy Local data C-implies bunching calculation [RFLICM] are contrasted with investigating the exactness of our proposed approach. Grouping results show that the RFLICM division technique is fitting for ordering tissues in cerebrum MR picture [24].

Consequently, to beat the vagueness brought about by the above unique effects, an upgraded fluffy unwinding approach called fluffy relaxation-based modified fluffy c-implies is introduced to group calculation by Magudeeswaran et al.. In the proposed strategy, openness-based sub-picture fluffy brilliance variation calculation is executed for the improvement of cerebrum tissues, and it is trailed by a modified fluffy c-implies grouping calculation to section the upgraded mind attractive reverberation picture into white matter, dim matter, and cerebrospinal fluid tissues. The proposed technique adjusts its prosperity on mind tissue division and offers broad help to radiologists and clinical focuses [25].

The diagram in light of pixel esteem is drawn taking the different focuses from the expanded cells lies in the first situation from the impacted locale. Here the impacted district is considered an ellipsoid shape and the volumes have been determined from it. A fluffy level set calculation is proposed by Pritam R et al. to work with clinical picture division and in this exhibition of assessment of the proposed calculation was conveyed [26].

3. MATERIALS AND METHODS

3.1 Source Image Selection

In the source image selection module, the jpg, gif, or bitmap image is selected. The image is previewed in a picture box with the size mode property set to auto-size. The image property is used to create a bitmap class in .Net so that the width and height of the image are obtained.

3.2 Pattern Image(s) Selection

In the pattern image(s) selection module, the jpg, gif, or bitmap image is selected. Optionally the second image is selected if to be recognized in the second image.

3.3 Pattern Recognition

In the pattern recognition module, both the source and pattern images are displayed. When the check button is clicked, the source and pattern images are copied into separate arrays. During recognition, for faster comparison, the source image's color information for each pixel is copied and compared immediately with the corresponding pattern image pixel. If the pixel is not matched, then the function returns a false value so that further pixels in the rectangular area are not compared. The threshold value can be set so that only some percent of the image pixels are compared.

3.4 Problem Direction

The existing system is analyzing images manually. We cannot recognize all the patterns in the images. In the previous, we have to split the large image into a small image after that only we can identify the part of an image that we need. This is not an effective method to get a clear part of an image. The existing system has the following disadvantages,

1 The manual approach reduces the accuracy of pattern identification.
2 More time-consuming.
3 Efficiency is less.

3.5 Proposed System

The proposed system is analyzing images through software. We can recognize all the patterns in the images easily. This method simplifies our work to get a clear part of an image with a simple process. This method is very helpful in various fields like medical, engineering, textiles, etc.

3.6 Advantages of Proposed System

The proposed system has the following advantages,

- The proposed approach increases the pattern identification accuracy.
- Less time consuming
- Efficiency is more.
- More patterns can be identified in a single large image at a time.

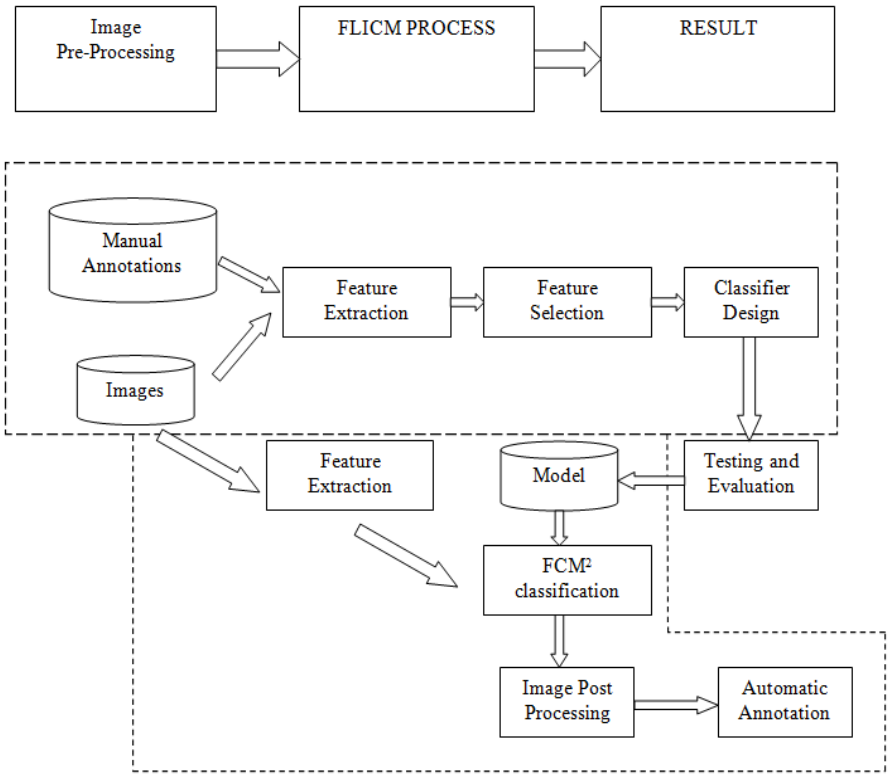


Fig. 1. Systematic Diagram for the Introduced method (FCM^2)

In introduced method, new fuzziness feature F_{ki} is incorporated over the objective function of predictable FCM, to overcome the interruption of any value and is shown below,

$$F_{cm} = \sum_{mgN, n_i \neq m} \frac{1}{d_m + |1} (1 - u_{cn})^k ||y - u_k||^2 \quad (1)$$

Where c is reference cluster, spatial Euclidean distance between pixels m and n is , unit of relationship is resolute as u_{cn} which is the relative of nth pixel in k^{th} cluster, fuzziness relationship has a increment supporter and s denoted as m and the midpoint of cluster k has the sample as v_k .

FCM^2 method is projected by combining both three-dimensional particulars and gray smooth particulars into predictable function and is stated in way of F_{ki} as,

$$J_m = \sum_{m=1}^{N} \sum_{k=1}^{c} [u^m ||y_m - v_k||^2 + F_{ki}] \quad (2)$$

FCM^2 algorithm is described below as:

Step 1: Set the quantity of cluster as c, fuzzification limit l and the stop of process s.

Step 2: Take up fuzziness divider conditions with experimental and miscalculation.

Step 3: Allocate the significance of twist security to be a = 0.

Step 4: Constellation examples is resolute created on the formulation in the following Eq.3

$$v = \frac{\sum_{i=1}^{N} u^l xi}{\sum_{i=1}^{N} u_{ki}^l} \quad (3)$$

Step 5: Find the fuzziness divider environment through the support of the following equation as

$$\beta_{ki} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|x_i - v_k\|^2 +}{\|x_i - v_j\|^2 + Gj_i} \right)^{1/m-1}} \quad (4)$$

Step 6: When $\{\beta^{(m)} - \beta^{(m+1)}\} < \varepsilon$, then termination comes to the procedure, otherwise,

set $b = b + 1$ and transfer to Step 4.

Fuzziness divider environment β is improved into a crisp divider when the comprehensive

repetition is ended. To achieve the upstairs procedure defuzzification procedure is experienced and is stated below.

$$DF_i = argument\{\{\beta_{ki}\}\}\ldots\ldots\ldots\ldots \quad (5)$$

Fuzziness picture accomplished subsequent to doing the interaction is then changed over into a fresh sectioned picture. The above idea is carried out to portion mind MRI pictures into three tissue types and is liberated from any anomalies and furthermore autonomous of boundaries. The division execution is significantly more upgraded than the current methods.

At the point when exceptions are available in the picture, the great characteristics of the calculation is made sense of by two fundamental cases. They are,

Event (i): Pixels present in the nearby window are undermined by commotion however pixel situated at the Center isn't a clamor pixel.

In such case, dark degrees of boisterous pixels to that of different pixels inside the window are unique and henceforth their enrollment values are adjusted by factorGki. Thusly, the connection of loud pixels is stifled by the factor Gki. Clamor is taken out by including the blend of spatial and dark-level limitations in work. The calculation is likewise more worthwhile in the event of anomalies.

Event (ii): Pixels present in the nearby window are homogenous, not debased by clamor however pixel situated at the Center is a commotion pixel.

In such a case, focal pixel enrollment esteem is adjusted by factor Gki by considering both the spatial and the dark level of the without commotion adjoining pixels in a fluffy way. In spite of the fact that enrollment worth of focal pixel doesn't relate to the commotion, the proposed strategy turns out to be more powerful to exceptions.

4. RESULTS

Trial and error are gone through for the proposed strategy. The calculation is executed utilizing MATLAB and tried on mind MRI pictures to investigate the division precision of the proposed approach. The examination is made and the nature of the division of the proposed calculation can be determined by utilizing SVM Classification.

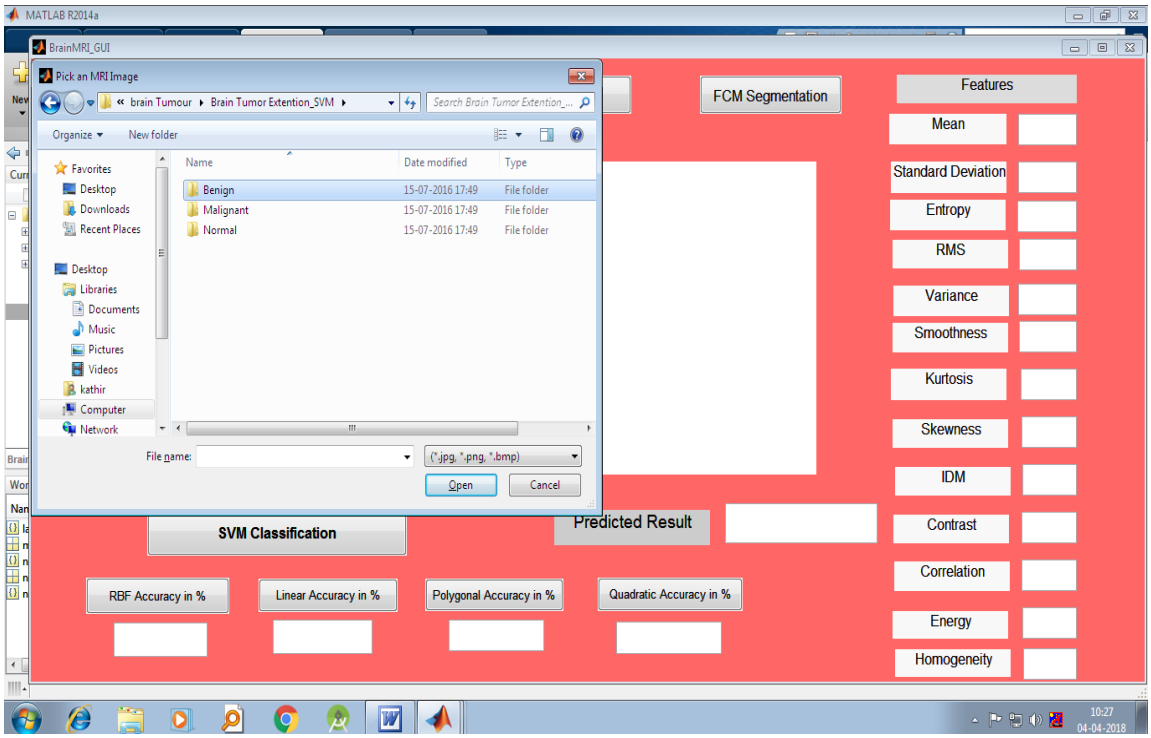


Fig. 2. Input MRI Image

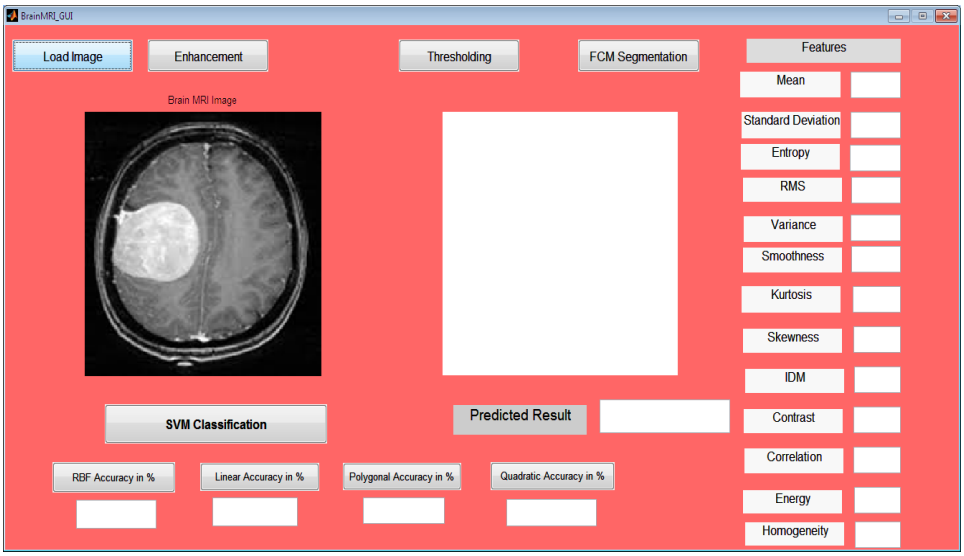


Fig. 3. Uploading Image

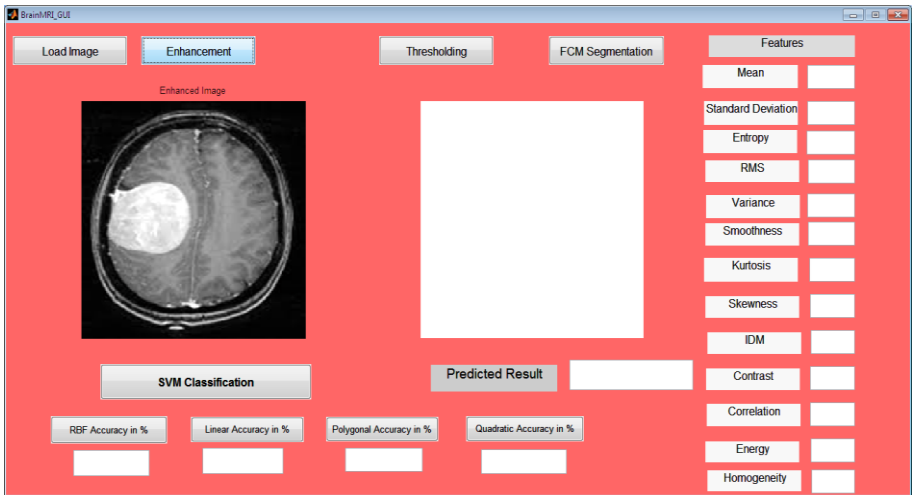


Fig. 4. Enhanced Image

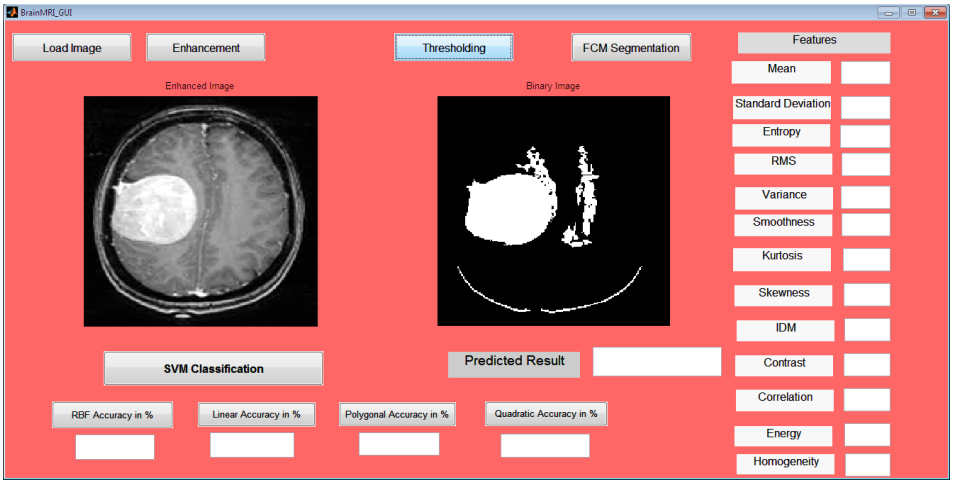


Fig. 5. Thresholding Binary Image

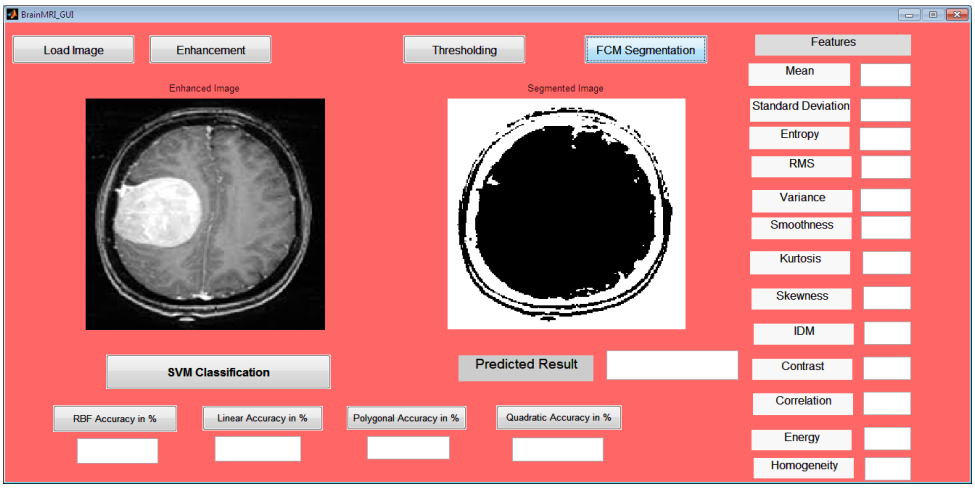


Fig. 6. Segmented Image

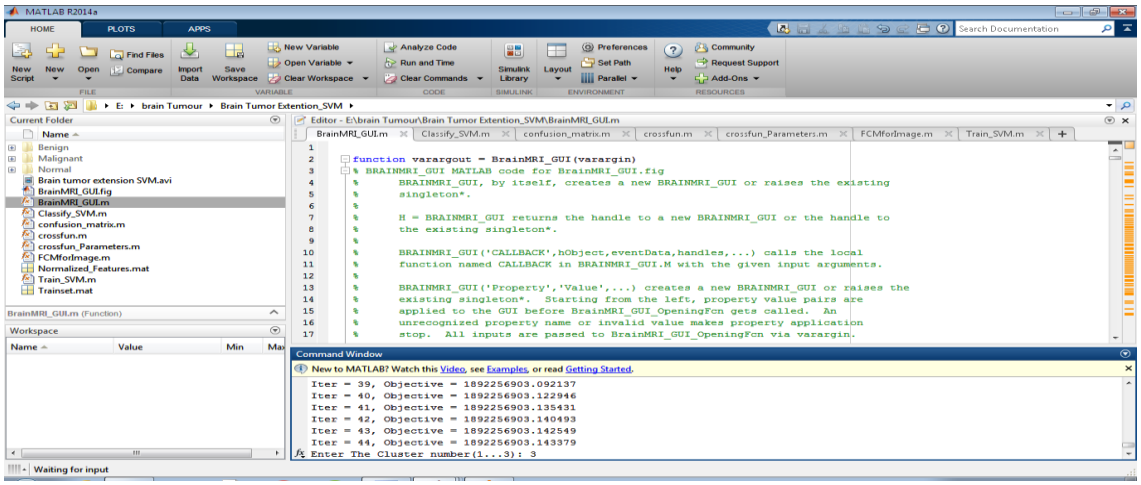


Fig. 7. Input Cluster Count Image

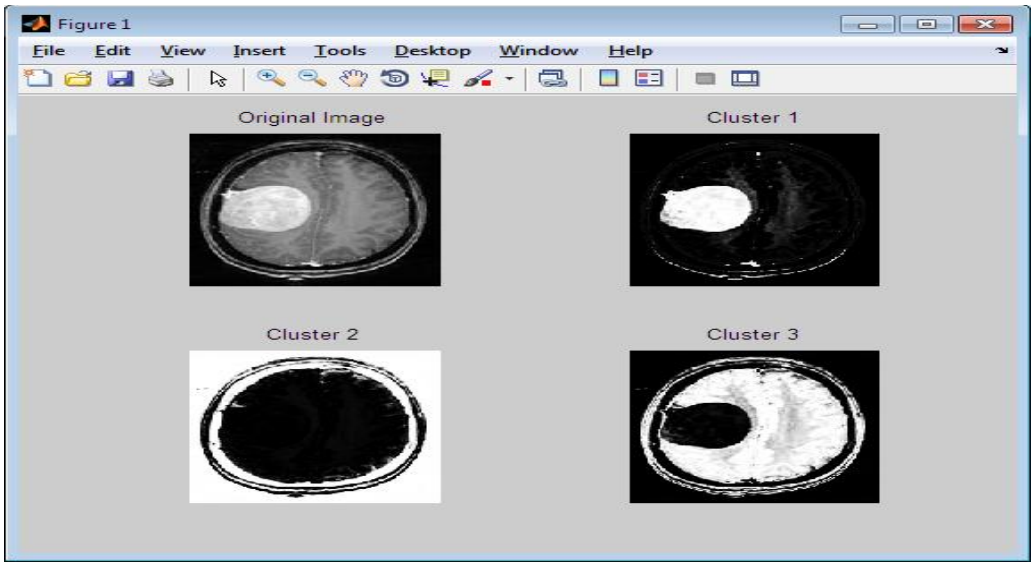
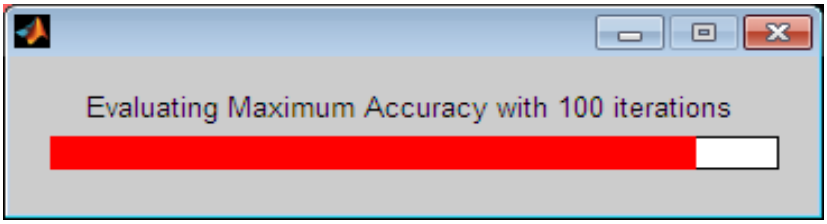


Fig. 8. Clustered Image

RBF Accuracy Finding



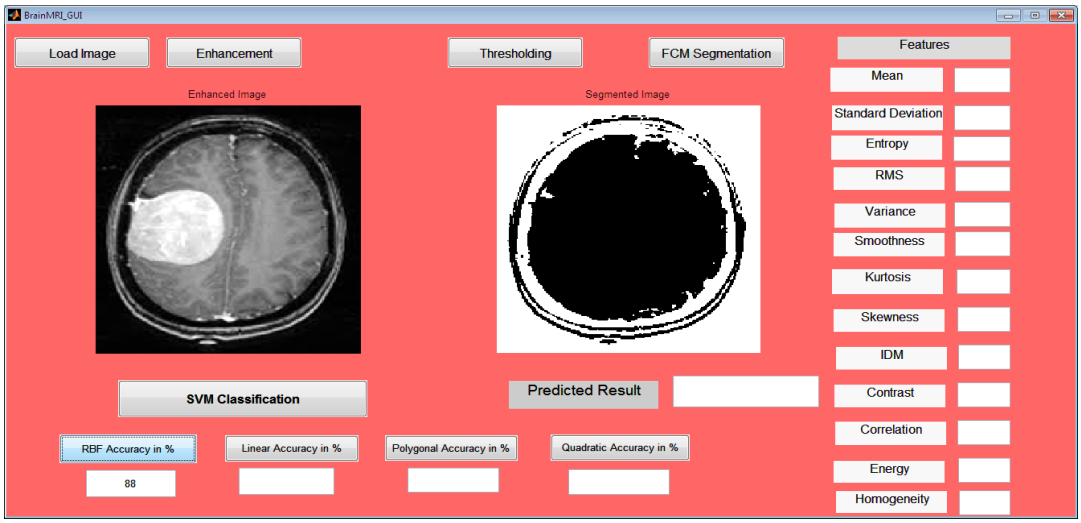
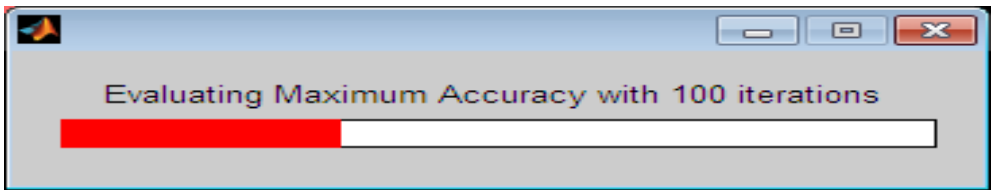


Fig. 9. RBF Accuracy Result

Linear Accuracy Calculation Processing



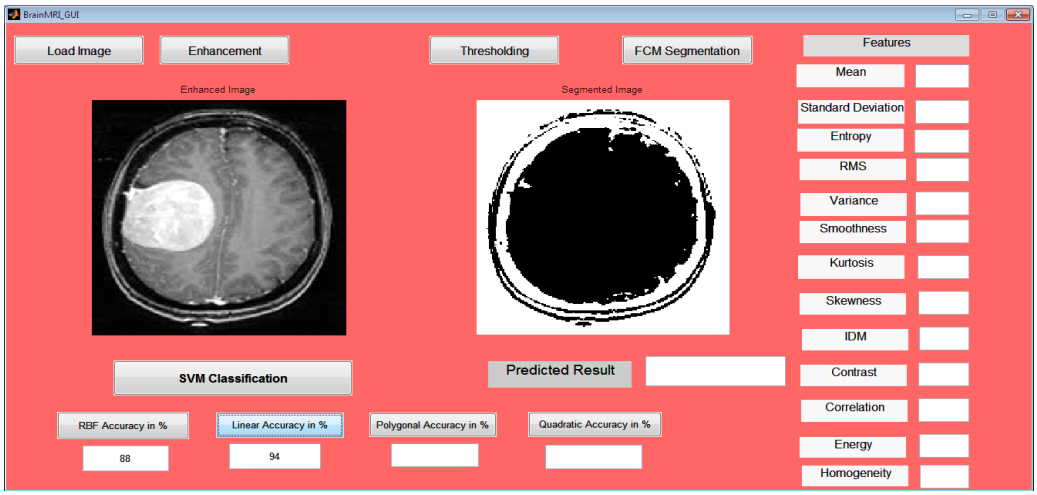


Fig. 10. Linear Accuracy in Percentage

Polygonal Accuracy Finding

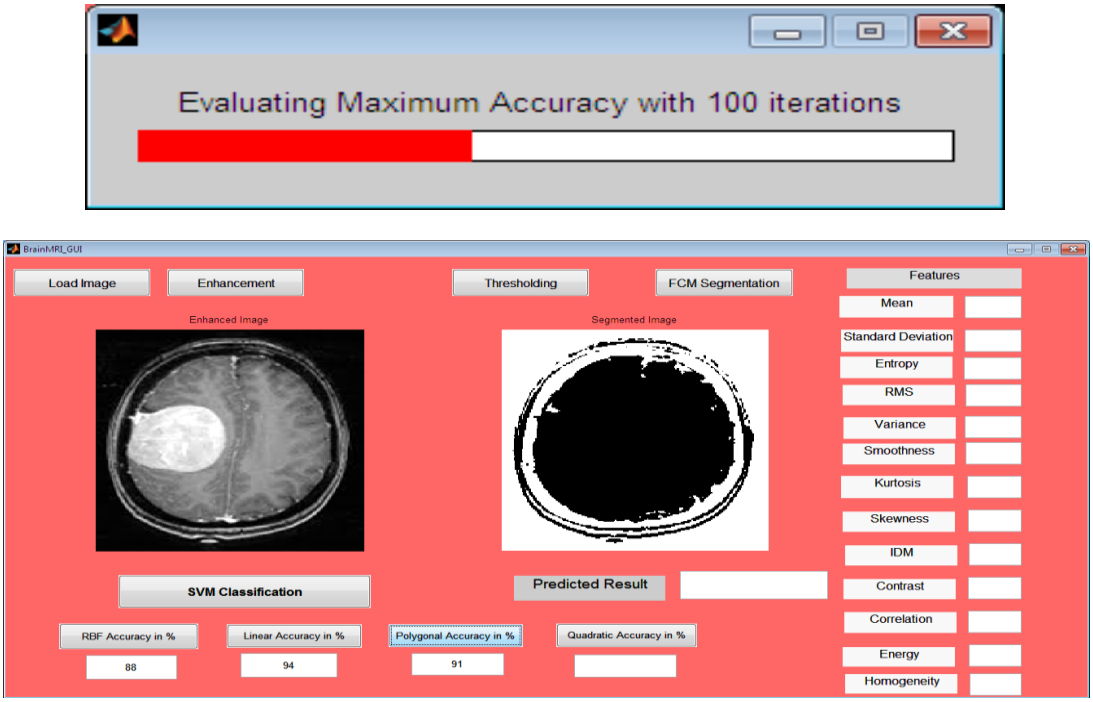


Fig. 11. Polygonal Accuracy Result

Quadratic Accuracy Finding Result

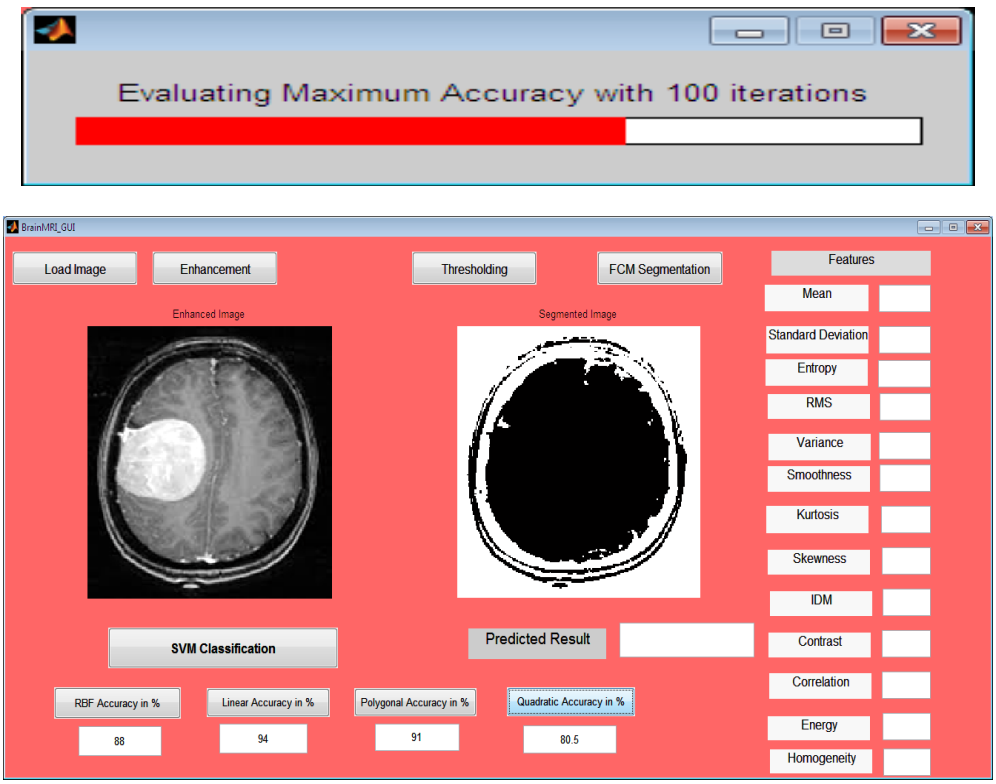
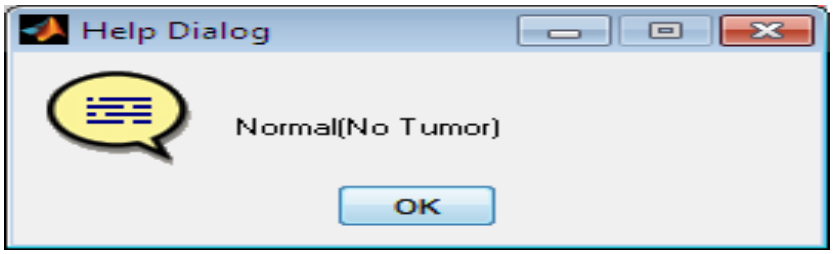
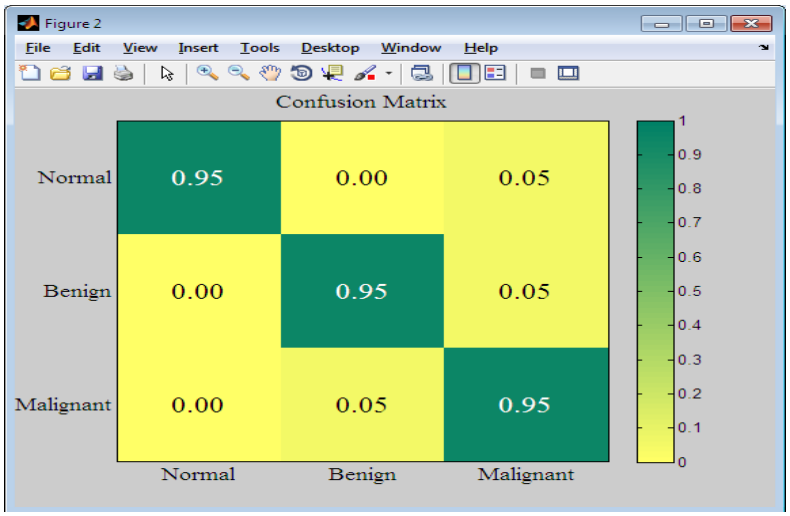
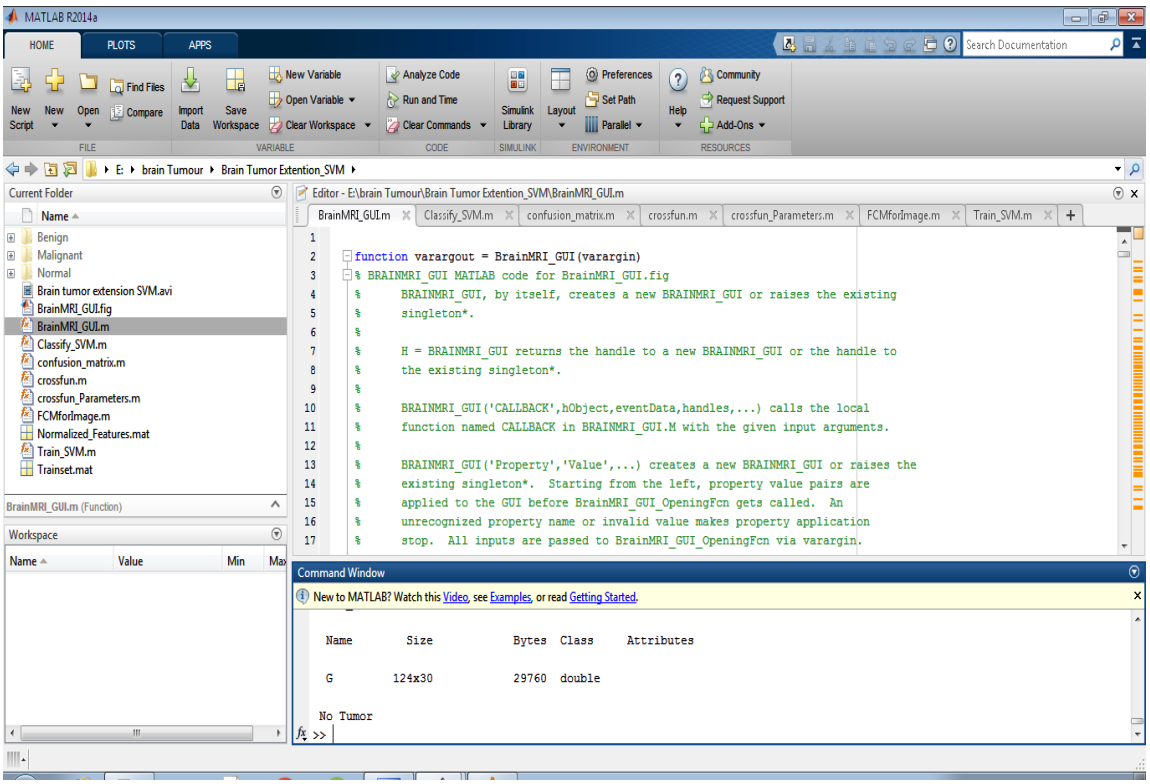


Fig. 12. Quadratic Accuracy Result





OUTPUT IN COMMAND WINDOW



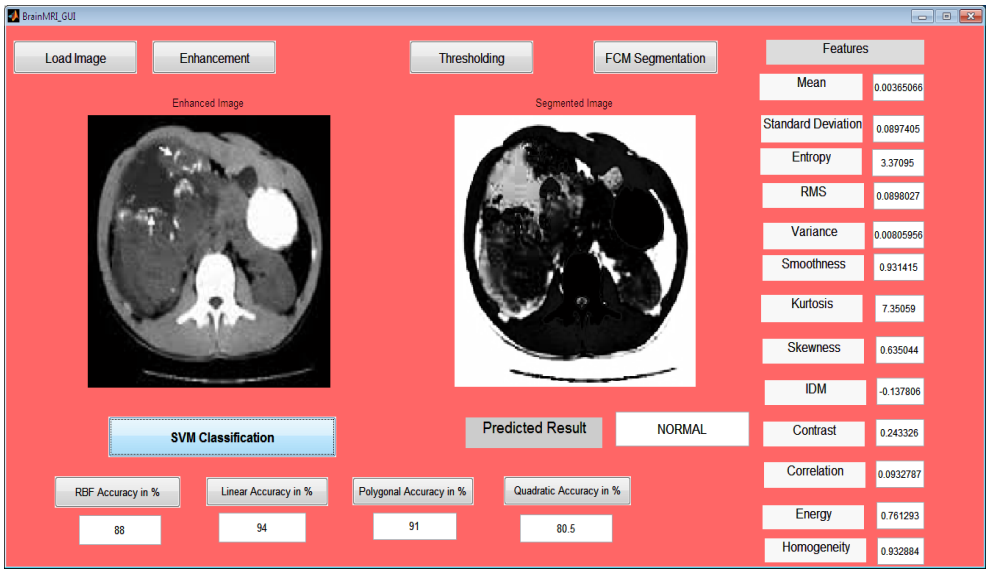


Fig. 13. Features Classification

Table 1. Quality Metrics

Quality Metrics	Accuracy in (%)
RBF Accuracy	88
Linear Accuracy	94
Polygonal Accuracy	91
Quadratic Accuracy	80.5

The segmentation accuracy achieved in the proposed method with respect to noise is 94%. Linear accuracy with SVM feature classification has the prior place and Polygonal accuracy have the 91%. Then RBF accuracy is 88% and Quadratic accuracy is 80.5%.

5. DISCUSSION

The performance of the proposed method is described in the result sections. Input MRI images are used and downloaded from Whole Brain Atlas (WBA). In the present work, more than 500 brain MR images with different sizes have been used for evaluation. For pre-processing, noise removal and enhancement methods are applied to the original MRI images. Then Thresholding technique was applied and binary image can be used for segmentation. Segmented image is used for clustering. Only tumor part is clearly segmented and result showed in the clustered image. Accuracy is the performance metric used to evaluate the results obtained by SVM feature classification. Linear accuracy shows the proposed method has a high accuracy for brain tumor segmentation.

6. CONCLUSION AND FUTURE WORK

Through this paper, the issue of the manual example is wiped out. Since exceptionally less information is given and any individual can utilize the application. When the pixel esteem is viewed as wrong in a given rectangular region, the whole region is disregarded for additional pixel examination. This outcome in quick work and their general acknowledgment time is decreased. The end clients are supposed to have the most un-working inclusion with structures to run this item. The application diminishes acknowledgment time and helps in further developing mistake-free and effective examples distinguishing proof. The application is tried well so the end clients utilize this product for their entire example acknowledgment-related activities. The accompanying upgrades ought to be done from now on. The multi-stringing choice ought to be added to distinguish the pixels at the same time in more regions of the picture. The picture can be changed over to a grayscale design and the example can be perceived as paying little heed to a variety of data. This will bring about quick acknowledgment. The brain

network idea in the event this at presented from now on will help in acknowledgment far better. The application is planned with the end goal that more example picture determination should be possible and at the same time thought about. For instance, coarse grain, as well as fine-grain rice pictures, can be chosen and minded a rice plate for quality. The application is grown to such an extent that these improvements will be incorporated effectively with modules present in the examination work.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Sarmad M, Robertas D, Faisal MS, An Efficient Approach for the Detection of Brain Tumor Using Fuzzy Logic and U-NET CNN Classification, ICCSA 2021, LNCS. 2021;12953:105–118.
2. Abhishek B, Minakshi B, Amlan C, Punit S. MRI Brain Tumor Segmentation and Analysis using Rough-Fuzzy C-Means and Shape Based Properties, Journal of King Saud University - Computer and Information Sciences. 2022;34(2):115-133.
3. Roqaie K, Roqaie M, Nazila L, Armin A, Mohammad BS, Arash ZS, Brain Tumor Segmentation using Hierarchical Combination of Fuzzy Logic and Cellular Automata, J Med Signals Sens. 2022;12(3):263–268.
4. Wadhwa A, Bhardwaj A, Singh Verma V. A review on brain tumor segmentation of MRI images. Magn Reson Imaging. 2019; 61:247–59.
5. Sharif, Muhammad. An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor. Pattern Recognition Letters. 2020;129:150-157.
6. Sahu, Satya Prakash. Segmentation of Lungs in Thoracic CTs Using K-means Clustering and Morphological Operations. Advances in Biomedical Engineering and Technology. Springer, Singapore. 2021;331-343.
7. Moeskops, Pim. Adversarial training and dilated convolutions for brain MRI segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham. 2017;56-64.
8. Hamad YA, Konstantin VS, Mohammad BN. Detection of brain tumor in MRI images, using a combination of fuzzy C-means and thresholding. International Journal of Advanced Pervasive and Ubiquitous Computing (IJAPUC). 2019;1:45-60.
9. Mittal, Mamta. Deep learning based enhanced tumor segmentation approach for MR brain images. Applied Soft Computing. 2019;78:346-354.
10. Mandle AK, Sahu SP, Gupta G. Brain Tumor Segmentation and Classification in MRI using Clustering and Kernel-Based SVM. Biomed Pharmacol J. 2022;15(2).
11. Hamad, Yousif Ahmed, Konstantin Vasilievich Simonov, and Mohammad B. Naeem, Detection of brain tumor in MRI images, using a combination of fuzzy C-means and thresholding. International Journal of Advanced Pervasive and Ubiquitous Computing (IJAPUC). 2019; 1:45-60.
12. Li, Ming. Brain tumor detection based on multimodal information fusion and convolutional neural network, IEEE Access. 2019;7:180134-180146.
13. Huang, Hong. Brain image segmentation based on FCM clustering algorithm and rough set. IEEE Access. 2019;7:12386-12396.
14. Yogalakshmi G, Sheela Rani B. A Review On The Techniques Of Brain Tumor: Segmentation, Feature Extraction And Classification. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE; 2020.
15. Yahya MA. Mohammed, Said El G, Ismail J. A survey of methods for brain tumor segmentation-based MRI images, Journal of Computational Design and Engineering. 2023;10:266–293.
16. Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation, Med Image Anal. 2018;43:98–111.
17. Karayegen G, Aksahin MF. Brain tumor prediction on MR images with semantic segmentation by using deep learning network and 3D imaging of tumor region, Biomedical Signal Processing and Control. 2021;66:102458.
18. Sujan M, Alam N, Noman SA, Islam MJ. A segmentation based automated system for

brain tumor detection. Int J Comput Appl. 2016;153:41–9.

19. Sathies T, Kumar K, Rashmi, Sreevidhya R, Sandhya LK, Sangeetha TJ. Brain tumor detection using SVM classifier, Third International Conference on Sensing, Signal Processing and Security (ICSSS), IEEE; 2017.
20. Sandeep Kumar M, Jai Kumar V. An Efficient Algorithm for Segmentation Using Fuzzy Local Information C-Means Clustering, International Journal of Scientific and Engineering Research. 2011;2(12).
21. Qiuyu S, Chengmao W, Xiaoping T, Yue S, Xiaokang G. A novel self-learning weighted fuzzy local information clustering algorithm integrating local and non-local spatial information for noise image segmentation, Applied Intelligence; 2021.
Available: https://doi.org/10.1007/s10489-021-02722-7
22. Sucharitha M, Parimala Geetha K. Segmentation of Brain Tissue using Fuzzy Local Information C-Means Technique, JCPS. 2016;9(1).
23. Pooja BM, Goudar RH. Segmentation of Brain MRI Images using Fuzzy C-Means and DWT, IJSTE - International Journal of Science Technology & Engineering. 2016;2(12).
24. Sucharithaa M, Parimala Geethab K. Brain tissue segmentation using fuzzy clustering techniques, Technology and Health Care. 2015;23:571–580.
25. Magudeeswaran V, Bharath S. Brain tissue segmentation for medical decision support systems, Journal of Ambient Intelligence and Humanized Computing; 2020.
Available: https://doi.org/10.1007/s12652-020-02257-8
26. Pritam RD, Dinesh DP. An Ameliorate Technique for Brain Lumps Detection Using Fuzzy C-Means Clustering, IJMTER. 2017;4(1).

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

PREDICTIVE ANALYTICS USING MACHINE LEARNING TECHNIQUES IN REAL TIME APPLICATIONS

S Habeeb Mohamed Sathak Amina
Assistant Professor, Department of Computer Science
Thassim Beevi Abdul Kader College For Women, Kilakarai

ABSTRACT

The developments in the field of artificial intelligence and machine learning have changed the world of computation where intelligent computation techniques and algorithms are introduced. Machine Learning models have a very well track record of being used as predictive models. Predictive analytics is the branch of advanced analytics, mainly used in statistical and analytics techniques to make predictions about future outcomes. It uses many techniques from data mining, statistics, modeling, and machine learning to analyze present and past data to make future predictions. Machine Learning has grown rapidly in the context of data analysis and computing which allows the applications to function intelligently. It is the subset of Artificial Intelligence. Machine Learning and predictive analytics is the key to effectively performing data analytics. Machine Learning can learn from past data sets, while predictive analytics focus is on forecasting specific target variables. This paper mainly focuses on Predictive analytics using machine learning in different real-time applications such as Medical, Banking, Security, etc., to predict future outcomes effectively.

Keywords: Predictive Analytics, Predictive Models, Machine Learning, Supervised Learning

I. INTRODUCTION

In this technologically emerging world, humans are in need to accomplish their tasks simply. Machine Learning made this possible by teaching machines how to handle data more efficiently from the experience of data. Machine learning relies on processing a large amount of data to find common patterns. It uses algorithms to identify and analyze the patterns in the data to predict future outcomes without being explicitly programmed. Predictive analytics is the future of Machine Learning. Historical and transactional data patterns are exploited by these models are used to identify the risks and opportunities for the future. Machine learning and predictive analytics are used together to make better decisions based on the effectiveness of past behavior. Machine learning is an appropriate technique to build predictive models due to their outstanding performance in managing large-scale datasets for their uniform characteristics. Predictive analytics uses many techniques from machine learning to analyze current data to make predictions to forecast future activities, behavior, and trend by extracting patterns from large datasets. Machine Learning algorithms optimize and uncover new statistical patterns to analyze data which form the backbone of predictive analytics.

II. LITERATURE REVIEW

Batta Mahesh et al. [2019] explained various machine learning algorithms. Machine Learning can be Supervised or Unsupervised. If you have a lesser amount of data and clearly labeled data for training, opt for Supervised Learning. Unsupervised Learning would generally give better performance and results for large data sets. This paper also gives an introduction to most of the popular machine learning algorithms.

Vaibhav Kumar, M. L. Garg, et. al. [2018] explained the scope of development of new models for the task of predictive analytics. And also discussed an

opportunity to add additional features to the existing models to improve their performance in the task.

Iqbal H. Sarker et. al., [2021] give an overview of machine learning algorithms for intelligent data analysis and applications. The author briefly discussed how various types of machine learning methods can be used for making solutions to various real-world issues

B. Nithya, Dr. V. Ilango et. al. [2017] discussed Machine Learning and predictive analytics techniques could revolutionize the entire healthcare industry by providing accurate insights and predictions related to symptoms, diagnoses, procedures, and medications. The author also explained different machine learning tools and techniques in health care provinces and their exclusive use in the diagnosis and predictions of various types of cancers.

III. PREDICTIVE ANALYTICS PROCESS

Predictive analytics involves several steps for data analysis to predict the future based on present and past data. The process of predictive analytics is represented in figure 1 given below.

Figure 1: Predictive Analytics Process

A. Problem Definition:

To develop a predictive model, it must be clear what is the objective and scope of prediction. Identify the data sets which will be used by implementing predictive analytics methodology.

B. Data Collection:

It involves collecting the necessary details required for the analysis. It uses the historical or current data which predictive analysis is to be performed. This data may be in the form of structured, semi-structured, or unstructured based on the source of data.

C. Data Analysis:

It is the process of cleaning and removing unnecessary, erroneous, and redundant data to refine our data sets. It involves the exploration of data and analysis to identify some patterns or new outcomes from the data set. This phase helps to discover useful information and concludes by identifying some new patterns or trends.

D. Statistics:

This phase is used to validate the findings, assumptions, and hypotheses by using statistical methods and machine learning techniques. Probability theory and regression analysis are widely used analytics techniques.

E. Predictive Modelling:

This stage involves various algorithms to build predictive models based on statistical methods and the patterns observed. It provides an accurate result of predictive models about the future outcome. In many applications, multi-model evaluation is used to find a better solution.

F. Deployment:

This phase provides an option to deploy the results, reports, or other metrics created by statistical analysis into the real world to perform an effective decision-making process based on modeling.

G. Monitoring:

The results generated by the model are periodically monitored and check the performance to ensure that the desired output is obtained to make accurate predictions.

Predictive analytics is a step-by-step process to make accurate predictions to identify success rates or reduce risk at the early stage

IV. PREDICTIVE ANALYTICS MODELS

Machine learning can increase the speed of processing and analyzing the data for predictive analytics programs. Predictive analytics is driven by predictive modeling. Predictive analytics and machine learning together form predictive models which include

machine learning algorithms. Predictive analytics models are divided into five categories.

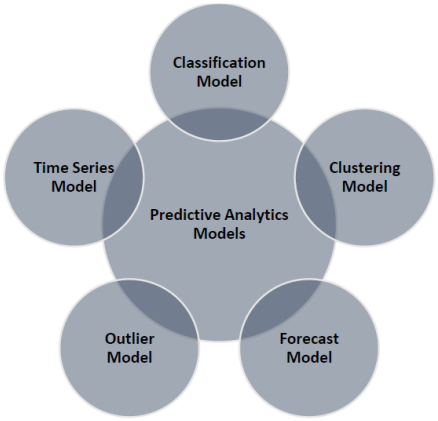


Figure 2: Predictive Analytics Models

A. Classification Model:

It is considered one of the common and simplest models. It categorizes the data based on what it learns from past events.

B. Clustering Model:

This model separates data into different groups based on common or similar attributes.

C. Forecast Model:

It is the most commonly used predictive analytics model. It deals with metric value prediction to estimate the value for new data from the past venture.

D. Outlier Model:

This model works by analyzing abnormal or outlying data entries within datasets. This model helps to identify anomalous data either by themselves or about different categories.

E. Time Series Model:

This model mainly focuses on the sequence of data points where time is the input parameter. This model works by using different data points of historical data to develop a numerical metric that will predict future trends of a specified period.

V. MACHINE LEARNING TECHNIQUES FOR PREDICTIVE MODELING

Predictive analytics uses machine learning for data modeling due to its performance to accurately process large data sets and perform deeper analysis to identify new patterns. Machine learning algorithms are broadly divided into two groups: Supervised learning which is used to construct predictive models, and unsupervised learning that is used to build descriptive models.

Supervised, machine learning algorithms are used to build a model that makes predictions based on past uncertain evidence. There are different machine learning algorithms that can be applied in predictive modeling are listed below.

A. Decision Tree Techniques:

It is one of the predictive modeling approaches. A decision tree is a graph to represent choices and their results in form of a tree to relate the decisions and their possible consequences. This technique is used to identify the ways to separate a data set based on different conditions to make an effective decision

B. Bayesian Methods:

This method is used to construct a statistical model based on Bayes' theorem which terms the events prior and posterior. It takes random variables as parameters to define the probability of occurrence of an event.

C. Artificial Neural Network:

Artificial Neural networks (ANN) or neural networks replicates the human nervous systems intended to simulate the behavior of biological neuron composed of artificial neurons. It is a nonlinear modeling technique where they learn the relationship between the inputs and the outputs through training

D. Ensemble Methods:

This method creates multiple models and then combines them to produce improved results to solve the computational problem. These models are developed by training several similar types of models and finally combining their results for prediction.

E. Instance-Based Learning:

This technique operates by comparing the current instances with the previously trained instances, which have been stored in memory. Its name is derived from the fact that it creates assumptions from the training data instances.

VI. APPLICATIONS OF PREDICTIVE ANALYTICS AND MACHINE LEARNING

Machine learning techniques automate predictive modeling by generating algorithms for training data to recognize new patterns and behaviors. There are many applications of predictive analytics in a variety of domains. From medical analysis to the stock market, prediction can be done by historical data. Some of the popular applications are listed below.

A. Banking and Financial Services:

In the banking and financial services industries, predictive analytics and machine learning are used in conjunction to detect fraudulent behavior and reduce suspicious transactions. It helps to measure market risk and identify profit rates. It helps to scan historical datasets and identify risk areas to prevent risks by making better decisions.

B. Healthcare:

Machine learning predictive analytics has a large scope of application in the field of healthcare and medicine. By training algorithms with large and varied data sets, patient symptoms can be observed. This can further help in providing a faster and more accurate diagnosis at the early stage.

C. Security:

Machine learning predictive analytics play a vital role in security. It helps to improve services and performance to understand consumer behavior and enhance data security. It also helps to detect anomalies, fraud activities. This makes the professionals focus on designing different strategies to protect the system.

VII. CONCLUSION AND FUTURE SCOPE

Machine learning with predictive analytics is becoming key to every organization. Now with the advancements in the field of artificial intelligence and the development of machine learning techniques, there is a trend nowadays of using machine learning techniques in predictive analytics. Predictive Analytics helps not only for the growth of an organization but also prevents the degradation through analysis of fraudulent activities. The process of selecting a machine learning algorithm involves matching the features of the data to be learned in the existing approaches. This paper gives a better understanding and the impact of each algorithm on the prediction. It clearly explains various machine learning techniques for prediction. A literature study is set to identify a different machine learning algorithm to assess the features that impact the prediction model.

The future work of this paper will be comparing and adding additional features to the existing models to improve their performance and get better prediction result in real-time applications

REFERENCES

[1] Batta Mahesh, “Machine Learning Algorithms - A Review”, International Journal of Science and Research (IJSR) ISSN: 2319-7064 Research Gate Impact Factor (2018): 0.28

[2] Vaibhav Kumar, M. L. Garg, “Predictive Analytics: A Review of Trends and Techniques”, International Journal of Computer Applications (0975 – 8887) Volume 182 – No.1, July 2018

[3] Iqbal H. Sarker,” Machine Learning: Algorithms, Real-World Applications, and Research Directions”, SN Computer Science (2021) 2:1604

[4] Nishchol Mishra, Dr. Sanjay Silakari, “Predictive Analytics: A Survey, Trends, Applications, Opportunities & Challenges” International Journal of

Computer Science and Information Technologies, Vol. 3 (3), 2012, 4434- 4438

[5] B. Nithya, Dr. V. Ilango, "Predictive Analytics in Health Care Using Machine Learning Tools and Techniques", International Conference on Intelligent Computing and Control Systems ICICCS 2017, 978-1-5386-2745-7

[6] Eric Siegel, 2016, "Predictive Analytics", John Willey and Sons Ltd.

BRAIN TUMOR SEGMENTATION AND CLASSIFICATION USING FUZZINESS CONFINED MESSAGE C-MEANS (FCM2) ALGORITHM

DR.A.KRISHNAVENI
Assistant Professor
Department of Computer Science
Thassim Beevi Abdul Kader College For Women
Kilakarai, Ramanathapuram

M.S IRFAN AHMED
Professor
Department of Computer Science
Thassim Beevi Abdul Kader College For Women
Kilakarai, Ramanathapuram

***Abstract*—Picture division is a significant testing factor in clinical picture division. This paper portrays the division strategy comprising two stages. In the initial step, the MRI cerebrum picture is gained from the patients' information base, in that film, relics and commotion are taken out after the Binarization strategy is applied for picture division. The Binaraization works with the assistance of the Fuzzy C Means Clustering calculation, hence the calculation assumes the principal part in the framework, in this most minimal level of the weight vector, higher worth of cancer pixels, calculation speed is accomplished by the Fuzzy C Mean with vector quantization. The point of this exploration work is to give an assortment of fluffy c-implies (FCM) calculation that gives picture bunching utilizing the MRI Brain Tumor data set. The proposed calculation joins the neighborhood spatial data and dark level data in an original fluffy manner. The new calculation is called Fuzziness Confined Message C-Means (FCM2). FCM2 can defeat the drawbacks of the known fluffy c-implies calculations and simultaneously improves the grouping execution. The significant trait of FCM2 is the utilization of a fluffy neighborhood (both spatial and dark level) closeness measure, expecting to ensure clamor heartlessness and picture detail conservation. Besides, the proposed calculation Experiments performed on manufactured and true pictures show that the FCM2 calculation is powerful and proficient, giving strength to uproarious pictures.**

Keywords —*Cell, Fuzzy, Binarization, Clustering, Segmentation, Brain Tumor.*

I. INTRODUCTION

One of the significant errands of picture examination is picture division. Picture division parcels picture into groups that are not difficult to examine. A portion of the uses of picture division incorporate example acknowledgment and object ID, including extraction and clinical imaging. For clinical pictures, picture division assumes an indispensable part. In this paper, a strategy is introduced which is applied to identify the various tissues present in MRI cerebrum pictures. Subsequent to identifying the mind tissues, a few sicknesses can be analyzed without any problem.

Brain tumor cell recognition is important and needed for finding the tumor in the earlier stage to reduce the mortality rate of a human being. This paper involves finding pattern images from the given source image. In medical images, if the user needs to find the cell structure or bone crack information, we need to manually verify the area. Instead of manual analysis if a system that highlights all the crack patterns, then it would be useful. This paper helps to identify the search pattern image from the given large image. In addition, the pattern image can be cropped from the source image itself or can be selected from another file. The rectangular pixel area of the pattern is collected in an array and checked with source image pixel data row-wise and column-wise. The pixel percentage can be adjusted so that nearly matching images can also be recognized.

II. LITERATURE REVIEW

Sandeep Kumar et al. there is a variety of the fluffy c-implies (FCM) calculation is introduced that gives picture bunching. The proposed calculation consolidates nearby spatial data and dark level data in a clever fluffy manner. The new calculation is called fluffy nearby data C-Means (FLICM). By utilizing this calculation they can beat the inconveniences of the past calculations and simultaneously improves the bunching execution [1].

Qiuyu Song et al. raise a better self-learning weighted fluffy calculation, which straightforwardly acquires various loads in distance estimation through consistent iterative self-learning, then, at that point, the distance metric with the loads got from self-learning is implanted in the goal capacity of the fluffy grouping calculation to further develop the division execution and vigor of the calculation. An enormous number of investigations on various kinds of pictures show that the calculation can smother the clamor as well as hold the subtleties in the picture, the impact of portioning complex commotion pictures is better, and it gives preferable picture division results over the current most recent fluffy grouping calculations [2].

Sucharitha M et al., suggested Fuzzy Local Information C-Means calculation (FLICM) is presented which incorporates a boundary-free clever fluffy variable. This calculation can be applied for clinical picture division for characterizing mind MRI into various locales, for example, Gray matter (GM), White matter (WM), and Cerebro-Spinal liquid (CSF). Subsequent to arranging into various tissue locales, a few illnesses can be dissected without any problem. The proposed strategy shows the viability of

calculation in portioning cerebrum pictures into various tissue types [3].

Pooja B et al. two calculations are thought of. One is level set division utilizing fluffy c means by utilizing exceptional highlights (SFCM) and another is a division of mind MRI pictures utilizing DWT and head part investigation (PCA) are additionally handled utilizing support vector machine (SVM) for characterization. The execution assessment is finished by processing mean square mistake, top sign to commotion proportion (PSNR), most extreme contrast, outright mean mistake and so forth Here DWT utilizes k-implies bunching and level set utilizes fluffy c-implies grouping. The spatial limitations are named with various records, for example, the client can pick on a specific district of interest and emphasize the form ventures until a more precise outcome is acquired [4].

Standard FCM, Fuzzy Local data C-implies grouping calculation [FLICM], Reformulated Fuzzy Local data C-implies bunching calculation [RFLICM] are contrasted with investigating the exactness of our proposed approach. Grouping results show that the RFLICM division technique is fitting for ordering tissues in cerebrum MR picture [5].

Consequently, to beat the vagueness brought about by the above unique effects, an upgraded fluffy unwinding approach called fluffy relaxation-based modified fluffy c-implies is introduced to group calculation by Magudeeswaran et al.. In the proposed strategy, openness-based sub-picture fluffy brilliance variation calculation is executed for the improvement of cerebrum tissues, and it is trailed by a modified fluffy c-implies grouping calculation to section the upgraded mind attractive reverberation picture into white matter, dim matter, and cerebrospinal fluid tissues. The proposed technique adjusts its prosperity on mind tissue division and offers broad help to radiologists and clinical focuses [6].

The diagram in light of pixel esteem is drawn taking the different focuses from the expanded cells lies in the first situation from the impacted locale. Here the impacted district is considered an ellipsoid shape and the volumes have been determined from it. A fluffy level set calculation is proposed by Pritam R et al. to work with clinical picture division and in this exhibition of assessment of the proposed calculation was conveyed [7].

III. MATERIALS AND METHODS

Source Image Selection

In the source image selection module, the jpg, gif, or bitmap image is selected. The image is previewed in a picture box with the size mode property set to auto-size. The image property is used to create a bitmap class in .Net so that the width and height of the image are obtained.

Pattern Image(s) Selection

In the pattern image(s) selection module, the jpg, gif, or bitmap image is selected. Optionally the second image is selected if to be recognized in the second image.

Pattern Recognition

In the pattern recognition module, both the source and pattern images are displayed. When the check button is clicked, the source and pattern images are copied into separate arrays. During recognition, for faster comparison, the source image's color information for each pixel is copied and compared immediately with the corresponding pattern image pixel. If the pixel is not matched, then the function returns a false value so that further pixels in the rectangular area are not compared. The threshold value can be set so that only some percent of the image pixels are compared.

3.1. Problem Direction

The existing system is analyzing images manually. We cannot recognize all the patterns in the images. In the previous, we have to split the large image into a small image after that only we can identify the part of an image that we need. This is not an effective method to get a clear part of an image. The existing system has the following disadvantages

1. The manual approach reduces the accuracy of pattern identification.
2. More time-consuming.
3. Efficiency is less.

3.2. Proposed System

The proposed system is analyzing images through software. We can recognize all the patterns in the images easily. This method simplifies our work to get a clear part of an image with a simple process. This method is very helpful in various fields like medical, engineering, textiles, etc.

Advantages of Proposed System

The proposed system has the following advantages,

- The proposed approach increases the pattern identification accuracy.
- Less time consuming
- Efficiency is more.
- More patterns can be identified in a single large image at a time.

In introduced method, new fuzziness feature Fki is incorporated over the objective function of predictable FCM, to overcome the interruption of any value and is shown below

$$F_{cm} = \sum_{mgN,\ n_i \neq m} \frac{1}{d_m + 1}(1-u_{cn})^k ||y - u||^2 \ldots\ldots (1)$$

Where c is reference cluster, spatial Euclidean distance between pixels m and n is, unit of relationship is resolute as $u$cn which is the relative of nth pixel in kth cluster, fuzziness relationship has a increment supporter and s denoted as m and the midpoint of cluster k has the sample as vk.

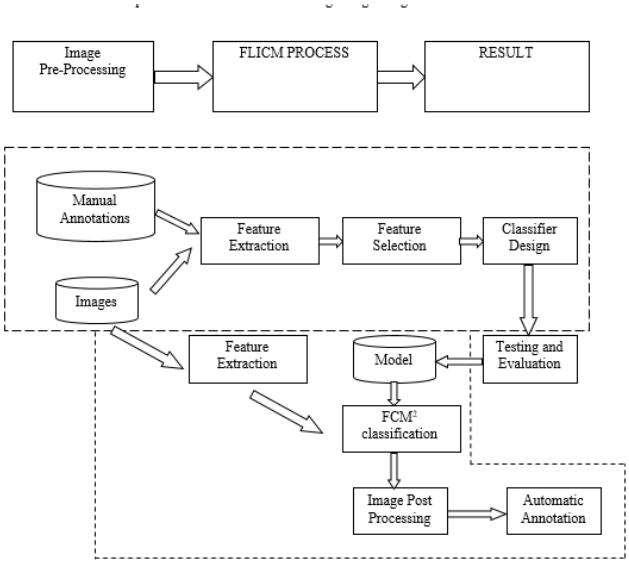


Fig. 1. Systematic Diagram for the Introduced method (FCM2)

FCM2 method is projected by combining both three-dimensional particulars and gray smooth particulars into predictable function and is stated in way of Fki as,

$$Im = \sum_{m=1}^{N} \sum_{k=1}^{c} [u^{m}_{ki} \|y_m - v_k\|^2 + F_{ki}] \dots \dots \dots (2)$$

FCM2 algorithm is described below as:

Step 1: Set the quantity of cluster as c, fuzzification limit l and the stop of process s.

Step 2: Take up fuzziness divider conditions with experimental and miscalculation.

Step 3: Allocate the significance of twist security to be a = 0.

Step 4: Constellation examples is resolute created on the formulation in the following Eq.3

$$v = \frac{\sum_{i=1}^{N} u^{l} x_i}{\sum_{i=1}^{N} u^{l}_{ki}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

Step 5: Find the fuzziness divider environment through the support of the following equation as

$$\beta_{ki} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|X_j - v_k\|^2 + G}{\|X_i - v_j\|^2 + G_{ji}} \right)^{1/m-1}} \dots\dots (4)$$

Step 6: When {β (m) −β (m+1)} < ε, then termination comes to the procedure, otherwise, set $b = b + 1$ and transfer to Step 4.

Fuzziness divider environment β is improved into a crisp divider when the comprehensive repetition is ended. To achieve the upstairs procedure defuzzification procedure is experienced and is stated below.

$$DF_i = argument\{max\{\beta_{ki}\}\} \dots\dots\dots\dots\dots (5)$$

Fuzziness picture accomplished subsequent to doing the interaction is then changed over into a fresh sectioned picture. The above idea is carried out to portion mind MRI pictures into three tissue types and is liberated from any anomalies and furthermore autonomous of boundaries. The division execution is significantly more upgraded than the current methods.

At the point when exceptions are available in the picture, the great characteristics of the calculation is made sense of by two fundamental cases. They are,

Event (i): Pixels present in the nearby window are undermined by commotion however pixel situated at the Center isn't a clamor pixel.

In such case, dark degrees of boisterous pixels to that of different pixels inside the window are unique and henceforth their enrollment values are adjusted by factorGki. Thusly, the connection of loud pixels is stifled by the factor Gki. Clamor is taken out by including the blend of spatial and dark-level limitations in work. The calculation is likewise more worthwhile in the event of anomalies.

Event (ii): Pixels present in the nearby window are homogenous, not debased by clamor however pixel situated at the Center is a commotion pixel.

In such a case, focal pixel enrollment esteem is adjusted by factor Gki by considering both the spatial and the dark level of the without commotion adjoining pixels in a fluffy way. In spite of the fact that enrollment worth of focal pixel doesn't relate to the commotion, the proposed strategy turns out to be more powerful to exceptions.

IV. EXPERIMENTAL RESULTS

Trial and error are gone through for the proposed strategy. The calculation is executed utilizing MATLAB and tried on mind MRI pictures to investigate the division precision of the proposed approach. The examination is made and the nature of the division of the proposed calculation can be determined by utilizing SVM Classification

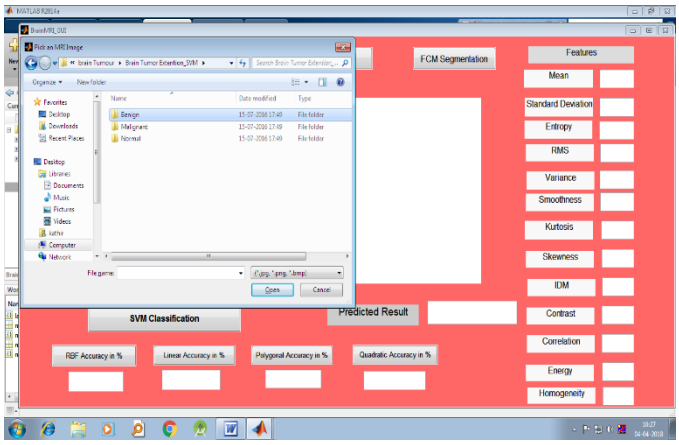


Fig. 2. Uploading Image

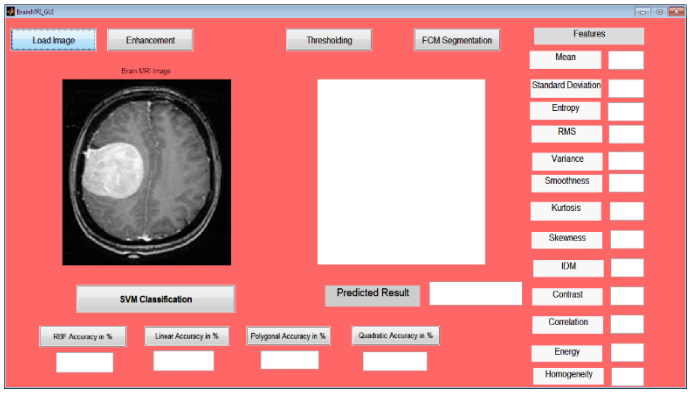


Fig. 3. Uploading Image

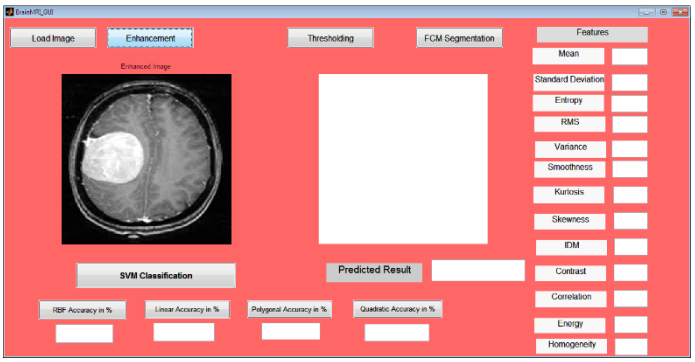


Fig. 4. Enhanced Image

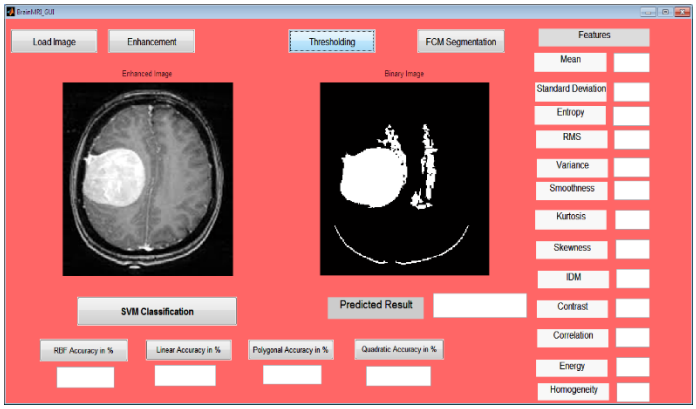


Fig. 5. Thresholding Binary Image

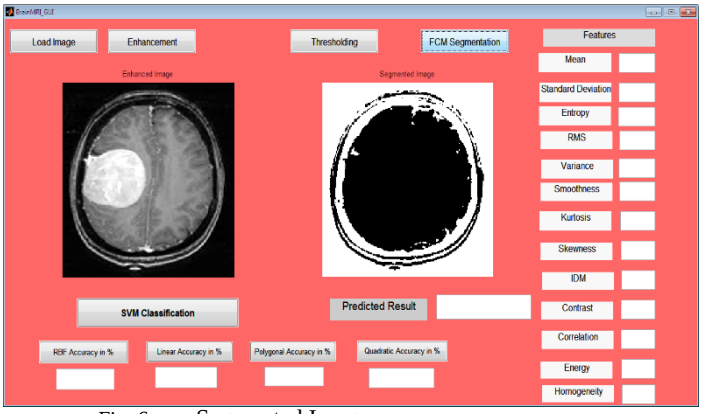


Fig. 6. Segmented Image

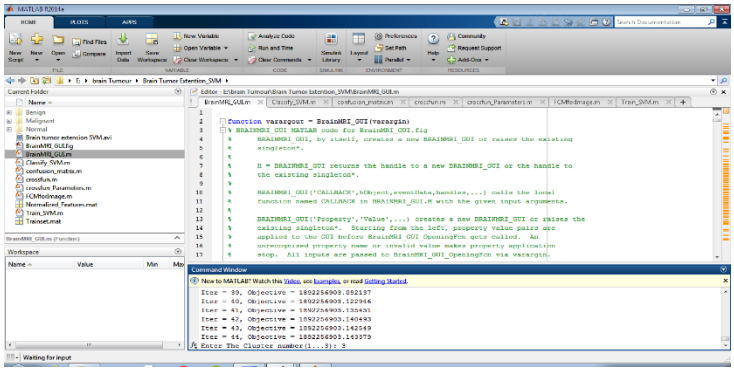


Fig. 7. Input Cluster Count Image

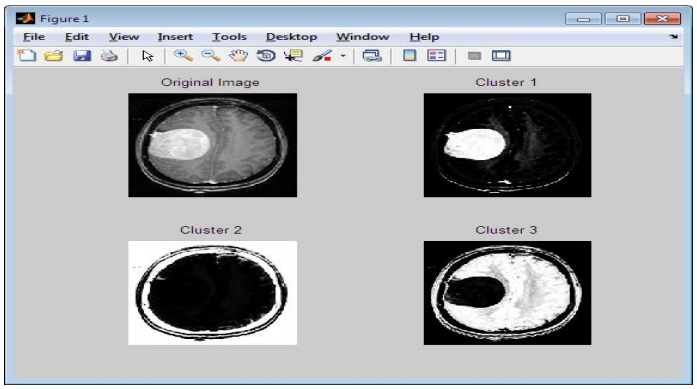


Fig. 8. Clustered Image

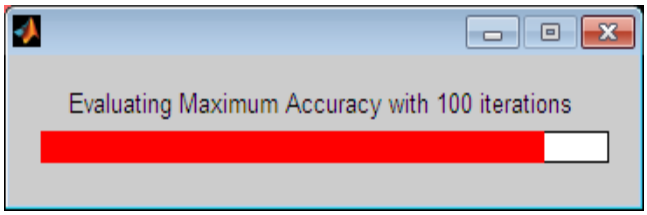


Fig. 9. RBF Accuracy Finding

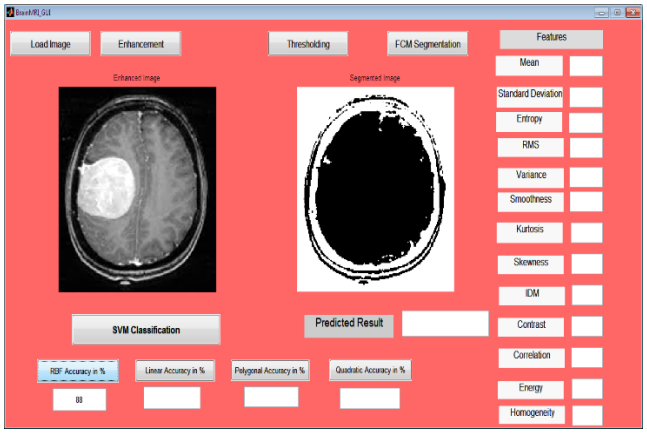


Fig. 10. RBF Accuracy Result

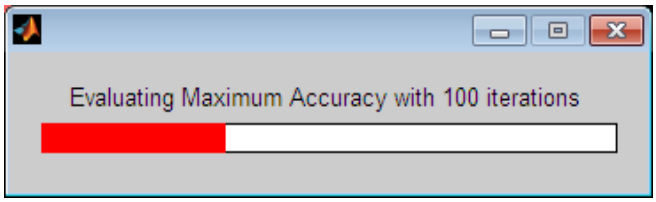


Fig. 11. Linear Accuracy Calculation Processing

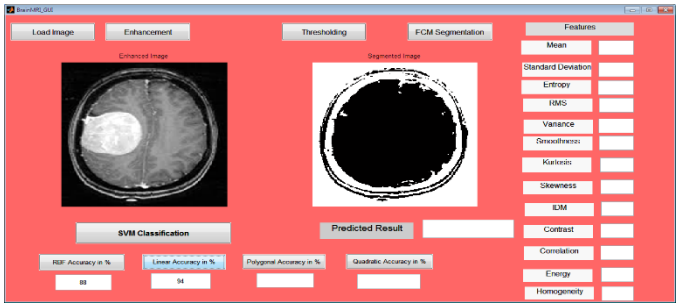


Fig. 12. Linear Accuracy in Percentage

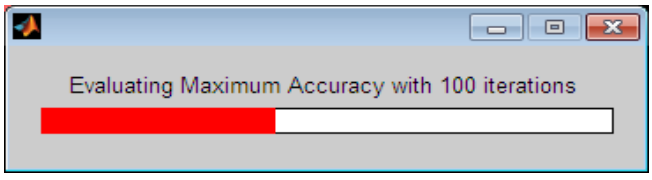


Fig. 13. Polygonal Accuracy Finding

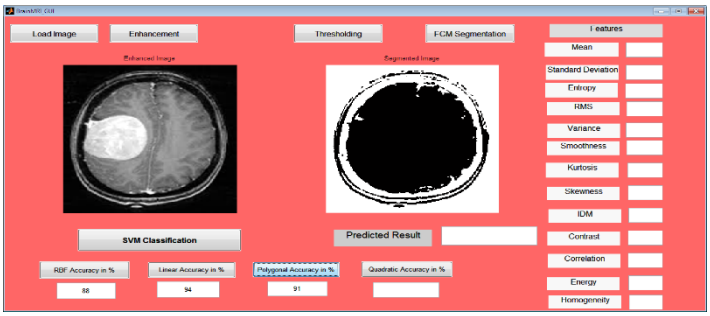


Fig. 14. Polygonal Accuracy Result

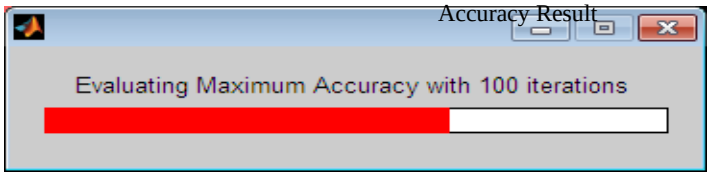


Fig. 15. Quadratic Accuracy Finding Result

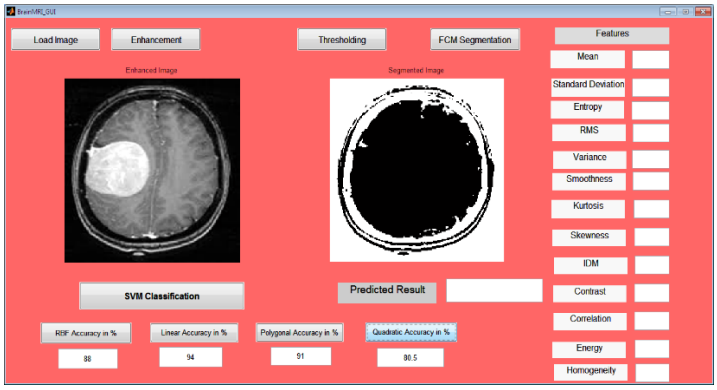


Fig. 16. Quadratic Accuracy Result

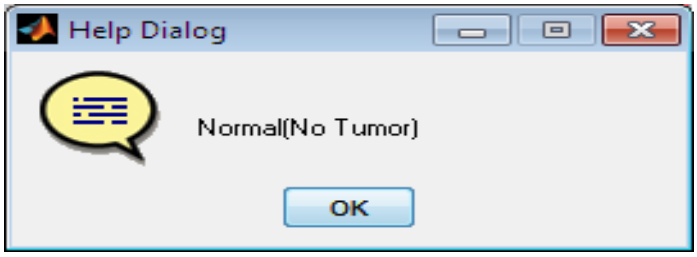


Fig. 17. Result

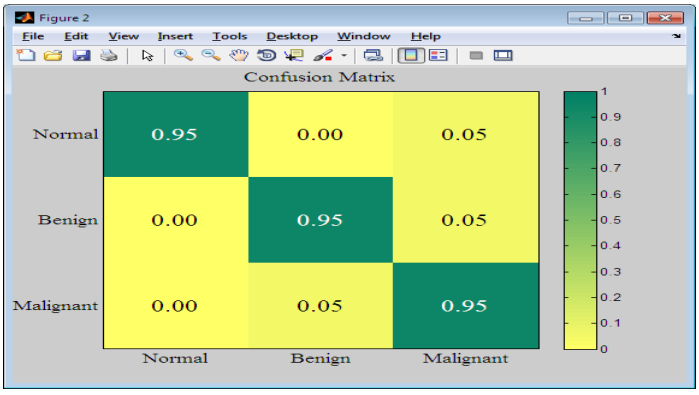


Fig. 18. Result

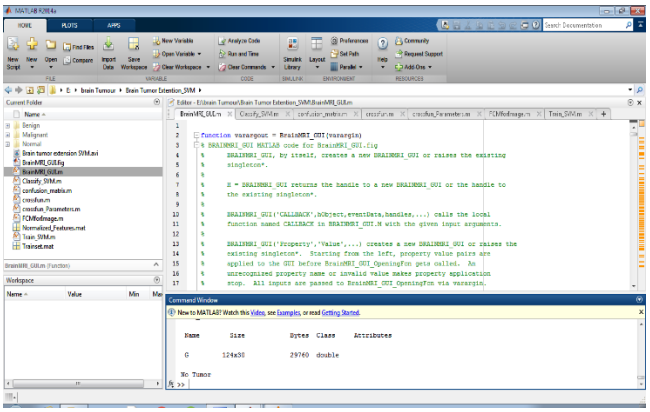


Fig. 19. Output in Command Window

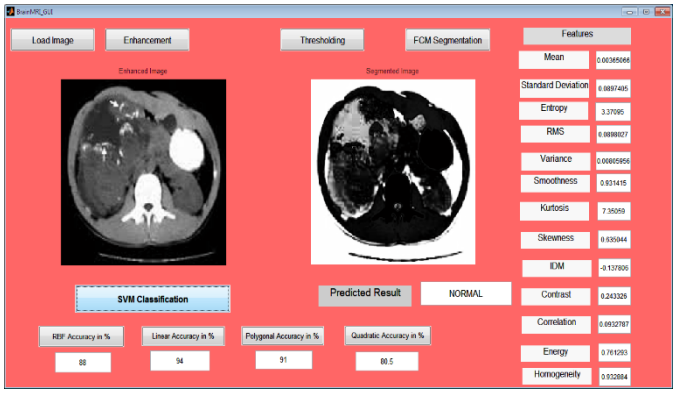


Fig. 20. Features Classification

Table 1: Quality Metrics

Quality Metrics	Accuracy in (%)
RBF Accuracy	88
Linear Accuracy	94
Polygonal Accuracy	91
Quadratic Accuracy	80.5

The segmentation accuracy achieved in the proposed method with respect to noise is 94%.

V. CONCLUSION AND FUTURE WORK

Through this paper, the issue of the manual example is wiped out. Since exceptionally less info is given; any individual can utilize the application. When the pixel esteem is viewed as wrong in a given rectangular region, the whole region is disregarded for additional pixel examination. This outcome in quick work and their general acknowledgment time is decreased. The end clients are supposed to have the most un-working inclusion with structures to run this item

The application diminishes acknowledgment time and helps in further developing mistake-free and effective examples distinguishing proof.The application is tried well so the end clients utilize this product for their entire example acknowledgment-related activities. The accompanying upgrades ought to be done from now on. The multi-stringing choice ought to be added to distinguish the pixels at the same time in more regions of the picture. The picture can be changed over to a grayscale design and the example can be perceived as paying little heed to a variety of data. This will bring about quick acknowledgment. The brain network idea in the event this at presented from now on will help in acknowledgment far better. The application is planned with the end goal that more example picture determination should be possible and at the same time thought about. For instance, coarse grain, as well as fine-grain rice pictures, can be chosen and minded a rice plate for quality. The application is grown to such an extent that these improvements will be incorporated effectively with modules present in the examination work.

References

[1] R. Agrawal, T. Imielinski and A. Swami, Mining Association Rules between Sets of Items in large Databases, Proc. ACM SIGMOD, New York, NY, USA:ACM, pp. 207-216, 2005.

[2] G. Smyth, Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments, Statistical Applications in Genetics and Molecular Biology, vol. 3, no. 1, pp. 3, 2004.

[3] S. Mallik and et al., Integrated Analysis of Gene Expression and Genome wide DNA Methylation for Tumor Prediction: An Association Rule Mining based Approach, CIBCB, IEEE SSCI, Singapore, pp. 120-127, 2013.

[4] S. Bandyopadhyay, U. Maulik and J.T.L. Wang, Analysis of Biological Data: A Soft Computing Approach, World Scientific, Singapore, 2007.

[5] M. Anandhavalli, M.K.Ghose and K.Gauthaman, Association Rule Mining in Genomics, International Journal of Computer Theory and Engineering, vol.2, no. 2, pp. 1793-8201,2010.

[6] D. Arthur and S. Vassilvitskii, k-means++: the advantages of careful seeding, In Proc. of ACM-SIAM SODA 2007, Society for Industrial and Applied Mathematics Philadelphia, PA, USA, pp. 1027-1035, 2007.

[7] S. Bandyopadhyay and et al., A Survey and Comparative Study of Statistical Tests for Identifying Differential Expression from Microarray Data, IEEE/ACM TCBB, 2014, DOI: 10.1109/TCBB.2013.147.

[8] [8] A. Thomas and et al., Expression profiling of cervical cancers in Indian women at different stages to identify gene signatures during progression of the disease, Cancer Medicine, 2013, DOI: 10.1002/cam4.152.

[9] J. Liu and et al., Identifying differentially expressed genes and pathways in two types of non-small cell lung cancer: adenocarcinoma and squamous cell carcinoma, Genet Mol Res, 2014, DOI: http://dx.doi.org/10.4238/2014.January.8.8.

[10] W. Wei and et al., The potassium-chloride co transporter 2 promotes cervical cancer cell migration and invasion by an ion transport-independent mechanism, J Physiol., 2011, DOI: 10.1113/jphysiol.2011.214635.

J Physiol., 2011, DOI: 10.1113/jphysiol.2011.214635.

Hybrid Classifier System for Wheat Classification

Dr.Shweta K R[1], Dr.Krishnaveni A [2]

Assistant Professor[2], Department of Computer Science
Thassim Beevi Abdul Kader College for Women, Kilakarai- 623517, Ramanathapuram, Tamilnadu, India

Abstract:

The economy has always been based on agriculture. Due to a variety of causes, this agricultural activity has not developed. The majority of tasks are completed without the use of contemporary technologies. At the moment, seed classification is done using human knowledge. The existing analysis of seed classification is ineffective and lacks a validation system. This paper presents a hybrid classifier-based approach for classifying wheat. The outcome of the experiment shows that the proposed hybrid technique's accuracy is superior to the algorithms currently in use for classifying wheat seeds. In this work, two supervised machine learning algorithms are combined to produce a hybrid model which consists of Naïve Bayes and BayesNet algorithms. The result shows that the hybrid model gave 97% accuracy as against the 92.45% and 92.21% of the Naïve Bayes and BayesNet models respectively. The purpose of this study is to improve the prediction accuracy on seed classification by hybridizing machine learning algorithms of Naïve Bayes and BayesNet. Also, a comparison between the proposed hybrid algorithm and individual algorithms is conducted to demonstrate the increase in classification prediction accuracy. Comparing the suggested algorithm to the individual classifiers, the performance is improved.

Key Words: Naive Bayes, BayesNet, Hybrid classifier, Agriculture

I.INTRODUCTION

Wheat is an essential cereal crop for human being that counts about 50 species around the world. Wheat species are of different sizes, colors, shapes, and textures. Yellow, red, and white Wheat seeds are the most common categories.

Classification of wheat seeds has importance for reducing the cost and for increasing the yield of the corn crop. Humans can use photos of seeds to evaluate different types thanks to technology. This study will help to develop a system for the identification of wheat varieties. The identification process is conducted by analysing digital images of wheat seeds through ML techniques. In order to classify Wheat seed varieties using hybrid features, this study set out to provide an improved hybrid features classification framework.

II.REVIEW OF LITERATURE

Further details on seed quality, impurity identification, and outliers are revealed through seed analysis and categorization. Moreover, before planting, it is very necessary to check the corn variety, its condition for achieving a good yield of each variety [8]. Mostly, conventional identification methods are used in Pakistan. The scarcity of human resources stands out as a major negative. It is difficult to find these types using human base perception [9]. Classifying maize seeds is crucial for lowering costs and raising yields of the maize crop. Technology helps the human being to evaluate varieties using images of seeds [10].

This work will contribute to the creation of a system for classifying different maize types. By using Machine Learning approaches to analyse digital photographs of maize seeds, the identification procedure is carried out. During the last few years, ML approaches have motivated the researchers to overcome the given below issues, which are related to human visual perception. According to the above literature survey, such an approach has already been working for classification of crops, land covers, fruit grading and for the solution of medical problems effectively [11, 12].

Many researchers have been used the ML approaches for the classification of different seed varieties such as Pandey and his fellows proposed technique based on content-based image recovery (CBIR) to identify seeds, such as grain, wheat, or rice, based on their features. The accuracy rate of the system through the use of a support vector machine (SVM) was 84.4% while the accuracy rate of the artificial neural network (ANN) was 95% [13]. Pourreza and his co-researchers identified nine varieties of common wheat, Iranian seeds using linear discrimination analysis (LDA) classifier, and achieved accuracy 96.15% when the best 50 features of the classifier were used [14]. Silva and Sonndara categorised rice varietals with 92%

accuracy using the ANN technique. [15] The multilayer perceptron (MLP) achieved 99.46% accuracy for the classification of rice grain types, and the neuro-fuzzy classifier reached 95.73% accuracy when Pazoki and his colleagues classified five central varieties of rice grains grew up in various Iranian climates. The accuracy was 94.40% of MLP and 96.73% of the neuro-fuzzy network (NFN) after feature selections [16].

The automated system for identifying diseased maize seed was given by Daskalov and his research team. They employed an SVM classifier to classify an image taken using a CCD camera, and their accuracy rate was 93.05% [22]. Discriminant Analysis (DA) was used by Qiu and his associates to present the Cultivar Classification Corn Seed. They used wavelength features and the SVM classifier to obtain 98.78% accuracy [23]. The automatic classification framework for classifying maize seeds using computer vision was put forth by Li and his research group. To distinguish between typical and aberrant maize seeds, they employed shape and colour characteristics. Deep learning was used in the suggested framework to obtain 96.67% accuracy [24]. With the aid of a remote sensing dataset, Zhang and his co-researchers proposed the system that categorises the different types of maize seeds.

i)Naïve Bayes:

The Naive Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem.

- It is primarily employed in text categorization with a large training set.
- The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms, aiding in the development of quick machine learning models capable of making accurate predictions.
- Because it is a probabilistic classifier, it makes predictions based on the likelihood that an object will occur.

The words Naive and Bayes, which make up the Nave Bayes algorithm, are as follows:

- Naive: It takes the position that the existence of one feature is unrelated to the occurrence of other features, which is why it is termed naive.
- Bayes: It is so named because it relies on the Bayes' Theorem premise.

Bayes' Theorem:

- Bayes' theorem, often known as Bayes' rule or Bayes' law, is a method for calculating the likelihood of a hypothesis given some previous information. The conditional probability determines this.
- The Bayes theorem's formula is as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

Naive Bayes' Classifier Operation:

1. Create frequency tables from the provided dataset.

2. Create a Likelihood table by calculating the odds of the provided attributes.

3. Next, determine the posterior probability using Bayes' theory.

ii)BayesNet:

A Bayesian network, sometimes referred to as a Bayes network, Bayes net, belief network, or judgement network, is a probabilistic graphical model that uses a directed acyclic graph (DAG) to describe a set of variables and their conditional relationships. When determining the chance that any one of a number of potential known causes

contributed to an event that already happened, Bayesian networks excel.

A Bayesian network is a directed acyclic graph where each node represents a distinct random variable and each edge represents a conditional dependency. Formally, the presence of an edge (A, B) in the graph connecting the random variables A and B indicates that P(B|A) is a component of the joint probability distribution. As a result, in order to conduct inference, we must be aware of P(B|A) for all values of A and B. Given that WetGrass has an edge into Rain in the example above, P(WetGrass|Rain) will be a factor in the conditional probability table, with its probability values listed next to the WetGrass node.

The local Markov property, which asserts that a node is conditionally independent of its non-descendants supplied, is satisfied by Bayesian networks. Given that Sprinkler is conditionally independent of its non-descendant, Rain, given Cloudy, this indicates that P(Sprinkler|Cloudy, Rain) = P(Sprinkler|Cloudy). We may reduce the joint distribution we generated using the chain rule in the previous section to a more manageable size thanks to this characteristic. The joint distribution for a Bayesian network is, after simplification, equal to the product of P(node|parents(node)) for all nodes, as follows:

$$P(X_1,\ldots,X_n)=\prod_{i=1}^{n}P(X_i \mid X_1,\ldots,X_{i-1})=\prod_{i=1}^{n}P(X_i \mid Parents(X_i))$$

Due to the fact that most nodes typically have few parents in comparison to the size of the network as a whole, this aspect enables us to significantly minimise the amount of processing needed in bigger networks.

III. METHODOLOGY

Dataset:

The proposed method for seed categorization is examined in this section. The UCI Machine Learning Repository provided the dataset. With 210 instances and 7 attributes, the seed collection contains three different types of wheat: Kama, Rosa, and Canadian. Different seed parameters are taken into consideration for the experimental purpose. Seven geometric characteristics of wheat kernels were measured to provide the data:

1. area A,
2. perimeter P,
3. compactness C = 4*pi*A/P^2,
4. length of kernel,
5. width of kernel,
6. asymmetry coefficient
7. Length of kernel groove.

The dataset including these parameters and their values is used in the suggested classification approach for wheat seed. Training and testing datasets are created from the data set. 30% of the data were used to test the model's performance, while 70% of the data were used to train the model.

Proposed Model:

The proposed model for seed classification includes following steps:

Step 1: Import the dataset

Step 2: Preprocessing of data

Step 3: Applying Hybrid classifier (Naïve Bayes and BayesNet) on preprocessed dataset

Step 4: classification of wheat seeds using hybrid classifier

IV.RESULT AND DISCUSSION:

This study set out to enhance the prediction accuracy of hybrid machine learning algorithm. Hence, this section shows the analysis results and discussion about individual and proposed hybrid machine learning classifiers prediction accuracy on classification of wheat. Here we have selected two supervised machine learning algorithms; they are Naïve Bayes algorithm and BayesNet algorithm. By combining two selected algorithms, the proposed hybrid algorithm was produced.

The Naive Bayes algorithm is combined with BayesNet algorithm to increase accuracy. The Naive Bayes provides the accuracy of 93.45.The BayesNet provides the accuracy of 94.90.The proposed hybrid algorithm provides the accuracy of 95.35, which is better than NaiveBayes and BayesNet**.** The results prove that hybrid algorithm provides more accuracy than NaiveBayes and BayesNet and this is better for seed classification base.

The following Table 1, shows the comparison of prediction accuracy of Bayesnet, Naïve Bayes algorithm and the proposed Hybrid algorithm.

The overall results show higher values of accuracy for most datasets when we compare the individual and hybrid algorithm according. It can be

observed that in many cases, the hybrid algorithm outperforms over the individual classifiers.

Table 1: Accuracy of Proposed hybrid classifier

Classifier	Accuracy
Naïve Bayes	92.90
BayesNet	92.75
Hybrid classifier	97.14

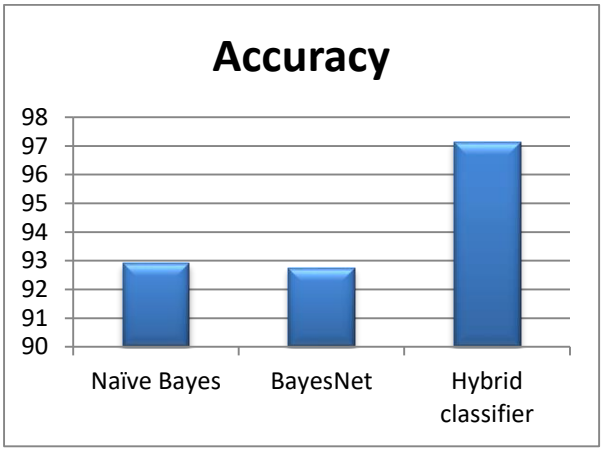


Figure 1: Accuracy of hybrid classifier

The above graph in Figure 1 shows the prediction accuracy of selected individual algorithms and proposed hybrid algorithm on seed classification. The proposed hybrid algorithm has better prediction accuracy than the individual algorithms.

V.CONCLUSION

In this study, the proposed hybrid algorithm demonstrates that hybrid machine learning techniques perform better than the individual algorithms on classification of seeds. The proposed hybrid algorithm composed of BayesNet and Naïve Bayes algorithm. The proposed hybrid algorithm outperform over the selected individual algorithms. The outcome demonstrates that the hybrid machine-learning algorithm is essential for increasing the classification precision of individual machine learning algorithms.

REFERENCES

[1]Flint-Garcia, S. A. 1 Kernel Evolution: From Teosinte to Maize. In Maize Kernel Development, 2017; pp 1.

[2] Chopin, F.; Roucolle, J. U.S. Patent No. 9,661,820; U.S. Patent and Trademark Office: Washington, DC, 2017.

[3] Saeed, M. A.; Irshad, A.; Sattar, H.; Andrews, G. E.; Phylaktou, H. N.; Gibbs, B. M. Agricultural Waste Biomass Energy Potential in Pakistan. Proceedings of the International Conference held in Shanghai, PR China, Leeds, October, 2015.

[4] Lynch, P. J.; Stangland, G. R.; Holland, G. J. U.S. Patent No. 9,072,266; U.S. Patent and Trademark Office: Washington, DC, 2015.

[5] Sastry, K. S. Methods of Combating Seed-transmitted Virus Diseases. In Seed-borne Plant Virus Diseases; Springer: India, 2013; pp 185–284.

[6] Liu, K. Soybeans: Chemistry, Technology, and Utilization; Springer, 2012.

[7] Kamilaris, A.; Prenafeta-Boldú, F. X. Deep Learning in Agriculture: A Survey. Comput. Electron. Agric. 2018, 147, 70–90. DOI: 10.1016/j.compag.2018.02.016.

[8] Tarabalka, Y.; Benediktsson, J. A.; Chanussot, J.; Tilton, J. C. Multiple Spectral-spatial Classification Approach for Hyperspectral Data. IEEE Transactions on Geoscience and Remote Sensing. 2010. DOI: 10.1109/ TGRS.2010.2062526.

[9] Asif, S.; Ahmad, M.; Zafar, M.; Ali, N. Prospects and Potential of Fatty Acid Methyl Esters of Some Non-edible Seed Oils for Use as Biodiesel in Pakistan. Renewable Sustainable Energy Rev. 2017, 74, 687–702. DOI: 10.1016/j. rser.2017.02.036.

[10] Altieri, M. A. Agroecology: The Science of Sustainable Agriculture; CRC Press, 2018.

[11] Mccabe, D.; Kadir, T.; Rodrigues, E. U.S. Patent Application No. 13/304,512, 2018.

[12] Qadri, S.; Khan, D. M.; Qadri, S. F.; Razzaq, A.; Ahmad, N.; Jamil, M.; Awan, S. A. Multisource Data Fusion Framework for Land Use/land Cover Classification Using Machine Vision. J. Sens. 2017, 2017, 1–8. DOI: 10.1155/ 2017/3515418.

[13] Pandey, R.; Naik, S.; Marfatia, R. Image Processing and Machine Learning for Automated Fruit Grading System: A Technical Review. Int. J. Comput. Appl. 2013, 81(16), 29–39. DOI: 10.5120/14209-2455.

[14] Pourreza, A.; Pourreza, H.; Abbaspour-Fard, M. H.; Sadrnia, H. Identification of Nine Iranian Wheat Seed Varieties by Textural Analysis with Image

Processing. Comput. Electron. Agric. 2012, 83, 102–108. DOI: 10.1016/ j.compag.2012.02.005.

[15] Silva, C. S.; Sonnadara, U. Classification of rice grains using neural networks; 2013.

[16] Pazoki, A. R.; Farokhi, F.; Pazoki, Z. Classification of Rice Grain Varieties Using Two Artificial Neural Networks (MLP and Neuro-Fuzzy). J. Anim. Plant Sci. 2014, 24(1), 336–343.

[17] Khoshroo, A. L. I. R. E. Z. A.; Arefi, A. R. M. A. N.; Masoumiasl, A. S. A. D.; Jowkar, G. H. Classification of Wheat Cultivars Using Image Processing and Artificial Neural Networks. Agric. Commun. 2014, 2(1), 17–22.

[18] Shahid, M.; Naweed, M.; Qadri, S. Mutiullah, and EA Rehmani,"Varietal Discrimination of Wheat Seeds by Machine Vision Approach,". Life Sci. J. 2014, 11(6), 245–252.

[19] Dubey, B.; Bhagwat, S.; Shouche, S.; Sainis, J. Potential of Artificial Neural Networks in Varietal Identification Using Morphometry of Wheat Grains. Biosyst. Eng. 2006, 95(1), 61–67. DOI: 10.1016/j. biosystemseng.2006.06.001.

[20] Guevara-Hernandez, F.; Gomez-Gil, J. A Machine Vision System for Classification of Wheat and Barley Grain Kernels. Spanish J. Agric. Res. 2011, 9(3), 9. DOI: 10.5424/sjar/20110903-140-10.

[21] Chen, X.; Xun, Y.; Li, W.; Zhang, J. Combining Discriminant Analysis and Neural Networks for Corn Variety Identification. Comput. Electron. Agric. 2010, 71, S48–S53. DOI: 10.1016/j.compag.2009.09.003.

[22] Daskalov, P.; Kirilova, E.; Georgieva, T. Performance of an Automatic Inspection System for Classification of Fusarium Moniliforme Damaged Corn Seeds by Image Analysis. MATEC Web of Conferences. 210, 02014. EDP Sciences, 2018.

[23] Qiu, G.; Lü, E.; Wang, N.; Lu, H.; Wang, F.; Zeng, F. Cultivar Classification of Single Sweet Corn Seed Using Fourier Transform Near-Infrared Spectroscopy Combined with Discriminant Analysis. Appl. Sci. 2019, 9(8), 1530. DOI: 10.3390/app9081530.

[24] Li, X.; Dai, B.; Sun, H.; Li, W. Corn Classification System Based on Computer Vision. Symmetry. 2019, 11(4), 591. DOI: 10.3390/sym11040591.

[25] Zhang, L.; Liu, Z.; Ren, T.; Liu, D.; Ma, Z.; Tong, L.; Zhang, C.; Zhou, T.; Zhang, X.; Li, S. Identification of Seed Maize Fields with High Spatial Resolution and Multiple Spectral Remote Sensing Using Random Forest Classifier. Remote Sens.. 2020, 12(3), 362. DOI: 10.3390/rs12030362.

Face Mask Detection using improved Convolutional Neural Network

Ms M.Nagavalli
Assistant Professor, Department of Computer Science,
Thassim Beevi Abdul Kader College for Women, Kilakarai

Abstract

During the CoVID 19 pandemic, people must wearing face masks in public areas to prevent and reduce the probability of virus transmission and dissemination. The healthcare sector is in a terrible condition. There have been several measures attempted to prevent the spread of this disease, such as the usage of masks, which the World Health Organization strongly recommends (WHO). In this study, the CNN algorithm is improved using the ReLU activation function. Three activation functions—ReLU, Sigmoid, and Softmax—were used to search for face masks. The Softmax was able to achieve 90% accuracy. The accuracy of the Sigmoid was 95.19% as well. ReLU architecture attained an accuracy rate of 99.72%.

Keywords: Facemask detection, Convolutional Neural Networks, Activation Function, ReLU Function

Introduction

Most of the world's population has been harmed as a result of the COVID-19 (Coronavirus) pandemic. COVID-19 is a respiratory disease that triggers severe pneumonia in those who are affected. The virus enters an airspace through coughing, sneezing, or direct contact with an infected person. It can also enter through salivation beads, respiratory droplets, or nasal droplets emitted during these actions. Globally, the COVID-19 virus claims thousands of lives every day. [4] According to a World Health Organization (WHO) assessment on the coronavirus (COVID-19), as of November 22, 2021, there.

5 148 221 deaths and 258 million diagnosed COVID-19 cases worldwide.[5] In order to prevent the viral spread of disease, people should wear face masks and maintain a social distance. Whether people are wearing face masks or not, a real-time application that watches people in public should be created using an effective and efficient computer vision method. Identifying the face is the first step in determining whether a mask is present on it. This separates the entire

process into two phases: face detection and face mask detection.

Wear a protective facemask to prevent the spread of the coronavirus as part of the survival advice. The most efficient preventive measure against COVID-19 must be adopted by wearing a facemask. It is challenging to physically check to see if people are correctly wearing face masks and to alert the victim in crowded and public places [1]. This work takes a more straightforward approach to facemask recognition and alerting the person if it is not being worn. A system that automatically identifies whether or not a person is wearing a face mask and notify the higher authorities if not wearing a mask. [2] In this work is trained using convolutional neural network with mobilenet. The dataset contains 3801 images. The model is trained with two different classes-with and without masks. The category with masks includes faces with masks of different angles and different poses and different color masks. The category without masks contains faces without masks, hands used as masks and improper wearing of facemasks. 80% of the dataset are used for training and 20% for testing. [3] The use of face masks called for the need to have facial detection machines which can be developed from Keras, Scikit-learn and OpenCV. This paper examines the Convolutional Neural Network (CNN), which is utilized for detecting and classifying an individual wearing a mask.

Activation Functions

The final value produced by a neuron is provided via activation functions. In essence, an activation function is a simplified function that converts inputs into outputs with a defined range. Different types of activation functions carry out this role in various ways. For instance, the sigmoid activation function accepts input and maps the output values to ranges between 0 and 1.

This function is included in an artificial neural network as one of the ways to aid in the network's ability to recognise intricate patterns in data. Artificial neural networks are given nonlinear real-world features through these functions. Basically, x is defined as inputs, w weights, and we send f (x), which is the value passed to the output of the network in a simple neural network. This will then serve as the last output or the first layer's input.

If the activation function is not applied, the output signal becomes a simple linear function. A neural network without activation function will act as a linear regression with limited learning power. But we also want our neural network to learn non-linear states as we give it

complex real-world information such as image, video, text, and sound.

Softmax

Softmax function calculates the probabilities distribution of the event over 'n' different events. In general way of saying, this function will calculate the probabilities of each target class over all possible target classes. Later the calculated probabilities will be helpful for determining the target class for the given inputs. Softmax is an activation function that outputs the probability for each class and these probabilities will sum up to one. Cross Entropy loss is just the sum of the negative logarithm of the probabilities

The main advantage of using Softmax is the output probabilities range. The range will 0 to 1, and the sum of all the probabilities will be equal to one. If the softmax function used for multi-classification model it returns the probabilities of each class and the target class will have the high probability.

Sigmoid

A Sigmoid function is a mathematical function which has a characteristic S-shaped curve. There are a number of common sigmoid functions, such as the logistic function, the hyperbolic tangent, and the arctangent. In machine learning, the term sigmoid function is normally used to refer specifically to the logistic function, also called the logistic sigmoid function. All sigmoid functions have the property that they map the entire number line into a small range such as between 0 and 1, or -1 and 1, so one use of a sigmoid function is to convert a real value into one that can be interpreted as a probability. One of the most widely used sigmoid functions is the logistic function, which maps any real value to the range (0, 1). Note the characteristic S-shape which gave sigmoid functions their name (from the Greek letter sigma).

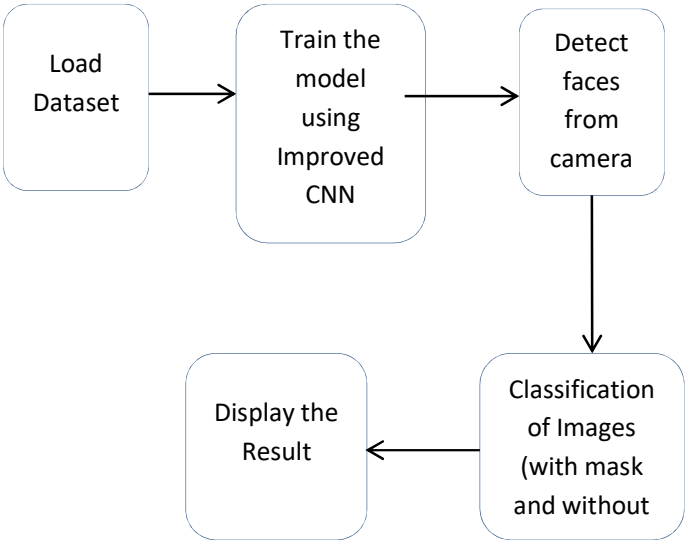
ReLU

The ReLU is the most used activation function in the world right now. Since, it is used in almost all the convolutional neural networks or deep learning. Perhaps the most often utilised function for hidden layers is the rectified linear activation function, or ReLU activation function. It is widespread because it is easy to use and successful in overcoming the drawbacks of other formerly well-liked activation functions, such as Sigmoid and Softmax.

ReLU is half rectified (from bottom). f(z) is zero when z is less than zero and f(z) is equal to z when z is above or equal to zero. The function and its derivative both are monotonic. ReLU's use thereby aids in preventing the exponential development of the compute needed to run the neural network.

More computationally efficient to compute than Sigmoid like functions since Relu just needs to pick max(0,x) and not perform expensive exponential operations as in Sigmoids. Networks with Relu tend to show better convergence performance than sigmoid. The ReLU function's primary benefit over other activation functions is that it does not concurrently activate all of the neurons. Relu activation function gives the best train accuracy and validation accuracy.

Methodology



Architecture of Proposed System

f(X) = max (0,X)

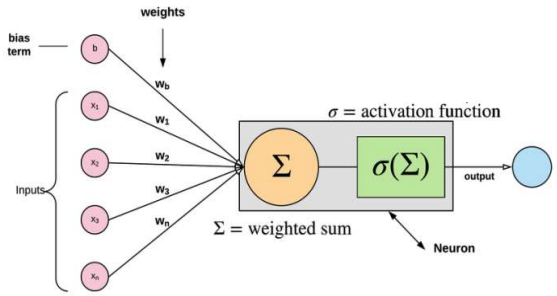


Diagram of a simple convolution neural network

Dataset

A dataset containing images of face with mask and without mask. The training dataset contains100 images,which is used to train. The testing dataset contains 300 images, which is used to test.

Result and Discussion

When relu activation function is applied,it provides highest accuracy of 90 and lowest loss value of 25.For Sigmoid activation function,it yields less accuracy of 59 and loss value of 27.With Softmax function,it gives accuracy of 59 and loss value of 68.Thus Relu activation function provides better performance when compared with sigmoid and softmax function.

S.No	Activation Function	Accuracy
1	ReLU	90
2	Sigmoid	59
3	Softmax	59

Table 1 : Accuracy of various activation functions

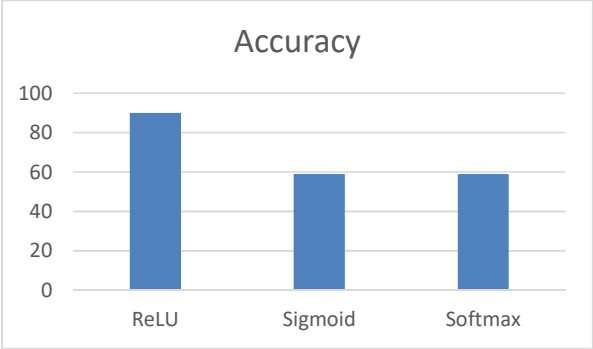


Fig 1.Accuracy of various activation functions

S.No	Activation Function	Loss
1.	ReLU	25
2.	Sigmoid	27
3.	Softmax	68

Table 2.Loss value of various activation functions

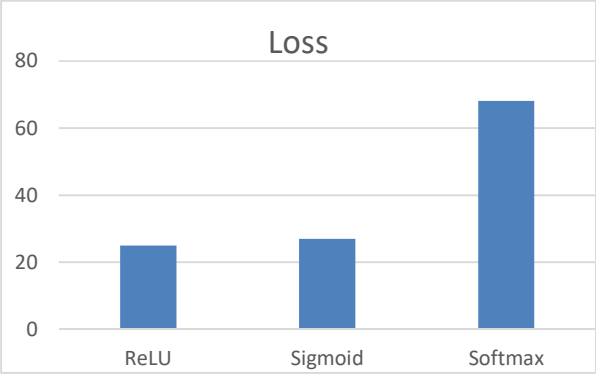


Fig 2 Loss value of various activation functions

Conclusion

In our research we have proposed a system that automatically identifies whether or not a person is wearing a face mask. Using Convolutional Neural Networks and motion learning strategies in neural networks, A dataset containing images of face with mask and without mask. The training dataset contains100 images, which is used to train. The testing dataset contains 300 images, which is used to test. In this study, the CNN algorithm is improved using the ReLU activation function. Three activation functions—ReLU, Sigmoid, and Softmax—were used to search for face masks. The Softmax was able to achieve 90% accuracy. The accuracy of the Sigmoid was 95.19% as well. ReLU architecture attained an accuracy rate of 99.72%.

References:

[1] Suresh K, Palangappa MB, Bhuvan S "Face Mask Detection by using Optimistic Convolutional Neural Network", IEEE Xplore Part Number: CFP21F70-ART ISBN: 978-1-7281-8501-9

[2] Rahinidevi V, Vijay Prasanth K, Aravind Losan S, Sandhya Devi R S " Face Mask Detection using CNN and OpenCV" , International Research Journal of Engineering and Technology (IRJET)

e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 09 Issue: 01 | Jan 2022

[3] Riya Chiragkumar Shah, Rutva Jignesh Shah "Detection of Face Mask using Convolutional Neural Network "

[4] Militante SV, Dionisio NV (2020) Real-time facemask recognition with alarm system using deep learning. In: 2020 11Th IEEE control and system graduate research colloquium (ICSGRC). IEEE, pp 106–110. https://doi.org/10.1109/ICSGRC49013.2020.9232610

[5] Who coronavirus disease (covid-19) dashboard. https://www.worldometers.info/coronavirus/. Accessed: 22-Nov-2021

[6] Modes of transmission of virus causing covid-19. https://bit.ly/3DAyqRu. Accessed: 22-Nov-2021

JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FISH DISEASE IDENTIFICATION: COMPARISON OF CLASSIFICATION ALGORITHMS USING DEEP LEARNING METHODS

[1]Dr M S Irfan Ahmed,
Director, Research and Industry –Institute Relations,

[2] Ms B Thillai Eswari,
Assistant Professor

Thassim Beevi Abdul Kader College for Women,
Kilakarai,Tamilnadu, India

Abstract: Aquaculture plays a very vital role in the world economy. In our country, Tamil Nadu has the second longest coastal area . Agriculture and fishing provide the majority of people's livelihoods in Ramanathapuram district. Aquaculture has many activities, including chank farming , pearl farming, dry fish farming, shrimp farming, etc. Fish disease is one of the important factors affecting the rural economy. The fish disease is examined as the root cause of mass production and economic deprivation by fish farmers. Species disease detection and behaviour monitoring are the main tasks for human visualization. The objective of this study is to provide a comparative study of the classification algorithms to identify fish diseases. A more advanced system with more computing power can facilitate a deep learning approach. The features of the Epizootic Ulcerative Syndrome (EUS) in fish are extracted in this study.

Key words: Fish Disease, Classification algorithm, image processing, Deep Learning, R-CNN Algorithms

I.INTRODUCTION

Routine diagnosis is an important part of fish health management. Epizootic Ulcerative Syndrome (EUS) is a fungal pathogen, a serious fish disease problem in India and in many countries.It is noted that temperature is a critical factor for fungal growth and fish mortality. There are some other diseases caused by humans, industry wastage, and population. The Existing system demonstrated a comparison between K-means clustering and C-means fuzzy logic. The proposed system has experimented with the Random forest algorithm, lazy and J48 algorithm on infected fish images of Mandapam Regional center of CMFRI. We found that the deep learning technique achieved higher accuracy and efficiency than other methods in most studies.

Deep learning is an important element of data science, which is statistics and predictive modeling. It is extremely beneficial to those who are tasked with collecting, analyzing, and interpreting large amounts of data. Deep learning makes this process faster and easier. Deep learning can be thought of as a way to automate predictive analysis, in which traditional machine learning algorithms are linear, deep learning algorithms are stacked in a hierarchy of increasing complexity and abstraction.

II. LITERATURE REVIEW

Yang X et al. have reported on applications of Deep learning including live fish identification, fish classification, behavioral analysis, size or biomass estimation, and prediction of water quality. When performing fish identification tasks, CNN models show an accuracy 18.5% higher than that of SVM models [1]. Zhoa et al. has reported that the machine learning algorithms and techniques adopted in intelligent fish aquaculture in the past five years are expounded, and the application of machine learning in aquaculture is explored in detail, including the information evaluation of fish biomass, the identification, and classification of fish, behavioral analysis and prediction of water quality parameters [2]. Ahmed MS et al. divide their work into two portions. In this section image pre-processing and image, segmentation has been applied to reduce noise and extract the image. In the portion, we extract the features to classify the diseases with the help of the Support Vector Machine (SVM) algorithm. SVM performs notably with 91.42&94.12 of accuracy [3]. Pauzi SN et al. have reported the objective of this paper is to briefly review the work established in the fish disease detection field with the use of numerous classification techniques of image processing, including

rule-based expert system, machine learning, deep learning, statistical method, and hybrid method. The review involves the improvement in image processing techniques that would be valuable for further advancement in terms of performance [4]. Knausgård KMet al. have proposed a two-step deep learning approach for the detection and classification of temperate fishes without pre-filtering. The first step is to detect every single fish in an image(YOLO)object detection technique. In the second step a Convolutional Neural Network (CNN) with the Squeeze-and-Excitation (SE) architecture for classifying each fish in the image. The existing solution achieves the state-of-the-art accuracy of 99.27% using the pre-training model. These post-training model results are also high; 83.68% and 87.74% with and without image augmentation [5]. Cui S, Zhou Y, and others are experienced that, a data augmentation approach was conducted. In this paper, the overfitting problem is solved by the Dropout algorithm. They conclude the following things (1) Establish the data set to include real blur ocean water condition; (2) CNN to explore an applicable solution for fish detection referred by revising loss function and other parameters; (3) The system is targeted at an embedded system for AUV design with all possible optimizations [6]. Rekha BS, Srinivasan et al. proposed that 3 phase methodology 1. Detection 2. Data augmentation 3. Classification. The CNNs used in the detection module and the classification module achieved a validation accuracy of about 90% and 92% respectively. The future scope includes working on R-CNN for better results in the detection and classification phases [7]. Waleed A.Medhat et al. reported that Identification of diseased fish at early stages is important to prevent spreading fish diseases. The system diagnoses three different types of fish diseases automatically. Different CNN architectures were applied to our collected data-set images in different color spaces. The Alexnet architecture achieved superior results in the XYZ color space [8].

Tseng SLet al. reported that apply the method for underwater fish disease detection using a restricted amount of training data set. This work is based on the U-Net model, modified with different encoders to achieve high accuracy detection and segmentation results [9]. Kappeler A et al.'s authors introduced a video SR algorithm using convolutional neural nets. This paper explains CNN makes use of spatial information as well as temporal information. We introduced an adaptive motion compensation scheme to deal with motion blur and fast-moving objects [10]. Waleed A Medhat et.al authors discussed architecture in the RGB color space and discuss measurements of water's temperature and pH value, and for sending a notice to users' mobile phones as their future work [11].

III. PROPOSED METHODS

In this section the fish species are collected from various coastal area in Ramanathapuram District and CMFRI.we can classify the fishes and also identified the diseased (EUS) fish non(EUS) diseased fish in scales

Figure 1:a.Non-Disease

Figure 2.Diseased fish(EUS)

A. Dataset

An openly available dataset is downloaded from the UCI Machine Learning Repository.

Table 1: Fish Disease Dataset

Fish Disease	Disease Dataset	Fish Disease	Disease Dataset	Fish Disease	Disease Dataset	Fish Disease	Disease Dataset	Fish Disease	Disease Dataset
EUS 1	Yes	EUS 11	YES	EUS 21	Yes	EUS 31	Yes	EUS 41	Yes
EUS 2	No	EUS 12	No	EUS 22	No	EUS 32	No	EUS 42	No
EUS 3	Yes	EUS 13	YES	EUS 23	Yes	EUS 33	Yes	EUS 43	Yes
EUS 4	No	EUS 14	No	EUS 24	No	EUS 34	No	EUS 44	No
EUS 5	YES	EUS 15	YES	EUS 25	Yes	EUS 35	Yes	EUS 45	Yes
EUS 6	No	EUS 16	No	EUS 26	No	EUS 36	No	EUS 46	No
EUS 7	YES	EUS 17	YES	EUS 27	Yes	EUS 37	Yes	EUS 47	Yes
EUS 8	No	EUS 18	No	EUS 28	No	EUS 38	No	EUS 48	No
EUS 9	YES	EUS 19	YES	EUS 29	Yes	EUS 39	Yes	EUS 49	Yes
EUS 10	No	EUS 20	No	EUS 30	No	EUS 40	No	EUS 50	No

B. Classification

The Proposed Methodology to detect the EUS disease in the Fish scale is present or absent in the fish. Various Classification Algorithms are analysed and extract the Scale region and classify with training data set. The KStar is suitable method for detecting and classification of the EUS fish disease images. The current method is given better Accuracy and minimum processing time taken for performance of the new fish classification.

1. J48

J48 algorithm is one of the preeminent machine learning algorithms to scrutinize the data categorically and continuously. When used for this purpose, for example, it takes up more memory and reduces classification performance and accuracy for all types of data.The J48 algorithm is used to classify the data set and improve the performance for accurate results . J48 algorithm is suitable for deep learning methods to analyse the data and identify the dataset and to increase the accuracy of algorithm.

2. K-star

K Star is an instance-based classifier, which means that a test instance's class is determined by the class of training instances that are comparable to it. Its use of an entropy-based distance function sets it apart from other instance-based learners.The K* algorithm is used to measure distance, as a and define the constant attributes and missing values. This algorithm mainly concern with instance-based learner which uses such as classification an evaluation. In this experiment, K* algorithm given the maximum accuracy compare to other algorithms.

3. Decision Table

A clear visual representation for indicating the actions to take in response to certain conditions is a decision table. They are programmes whose results are a collection of deeds. The structural format can be represented using a decision table. We obtained lower accuracy using this approach. Although the decision table is essentially identical, this technique requires that each condition be considered.

4. Ramdom Forest Tree

This algorithm is one of the famous supervised algorithms. It is used for classification and regression problems. The Random Algorithm is suitable for a large range of data items. It is predicated on the idea of ensemble learning, which is the procedure of merging various classifiers to address a complicated issue and enhance the model's performance. It is very flexible. Through this algorithm, the weighted average is 100, but it will take more time for execution.

C. R-Convolutional Neural Network

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. A neural network is a system of hardware and/or software patterned after the operation of neurons in the human brain and also worked with all the animals, birds, and spices. R-CNN is one of the first large and successful applications of convolutional neural networks to the problem of object localization, detection, and segmentation.

Region based CNN consists of three modules — Region Proposal, Feature Extractor and Classifier.

Region Proposal. The region proposal tries to detect the species in different regions of different sizes and aspect ratios.

Feature Extractor: Each proposed region will be trained by a CNN network and the last layer will be extracted as features.

Classifier: The features are extracted and to classify the diseased images for each regions.

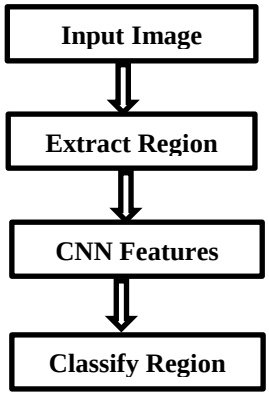


Figure: 3 Architecture of R-CNN

IV SPECIES PROCESSING TECHNIQUES

Image processing is among the most promising technologies for enhancing raw images gathered from external sources such as cameras, satellite sensors, space probes, aircraft, and others .Image processing technique is used to improve the quality of the image and give the technical assurance. The process for fish disease detection involves a fundamental procedure to extract and classify features from infected fish that is implemented in several steps.

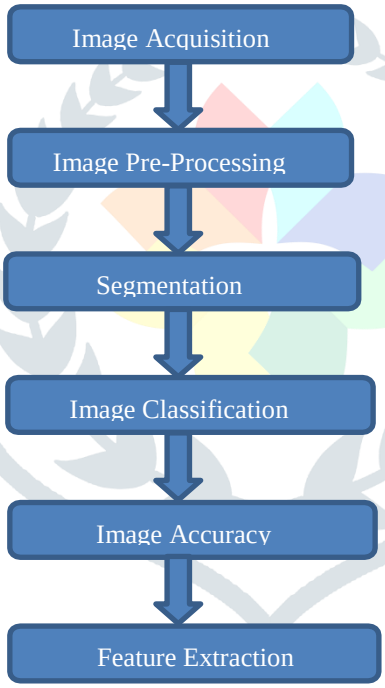


Figure 4: Steps of Methodology

Species Acquisition: The collection or acquiring of images from various coastal areas through the cameras and image sensors. The database was collected from the fishermen of some coastal areas of Ramanathapuram District and CMFRI Mandapam.

Species classification: Separate the disease-infected fish from the healthy fish. The various classification algorithms, such as J48,lazy Kstar ,Decision Tables, and Random Forest has been applied after extracting the feature.

Methods and techniques for classification are evaluated. The convention strategy for segmentation and classification of the object is resolved. Various classification algorithms are applied to calculate accuracy. In particular, the Kstar Algorithm is given great accuracy and efficient time for execution compared to other algorithms.

Table2: Comparision of K means and Kstar Algorithm

S.No	Algorithm	Weighted Avg.Accuracy	Time
1	K-Means	64.74%	0.2
2	KStar	100%	0.1

IV. Result and Discussion:

In this research, we have to apply four algorithms for the classification process. The algorithms are J48, Lazy KStar, Decision Table and Random Forest. The result is the following:

In Table 2, we compare the four different types of classification algorithms based on weighted accuracy and time**.** The results vary in their time and weighted average accuracy. The two algorithms take the same amount of time.but the weighted average accuracy is different. We consider both the weighted average accuracy and the time .

Table 2: Comparision of Classification Algorithms

S.No	Algorithm	Weighted Avg.Accuracy	Time
1	J48	67.74%	0.1
2	Kstar	100%	0.1
3	Decision Table	67%	0.01
4	RandomForest	100%	0.2

Table3: Comparision of K means and Kstar Algorithm

S.No	Algorithm	Weighted Avg.Accuracy	Time
1	K-Means	64.74%	0.2
2	KStar	100%	0.1

This figure compares the four types of algorithm J48,KStar, Decision Table and Random Forest

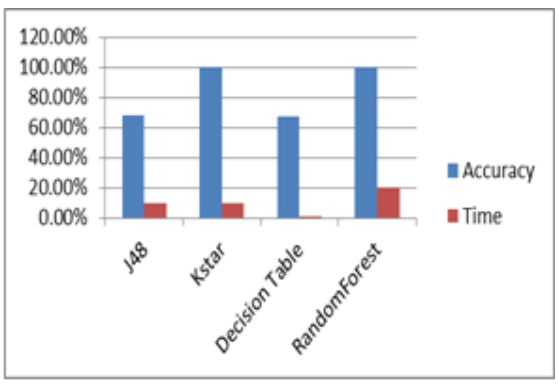


Figure 5: Training Accuracy of Algorithms

This figure compares the proposed algorithm (K Star) and the existing algorithm(K Means).

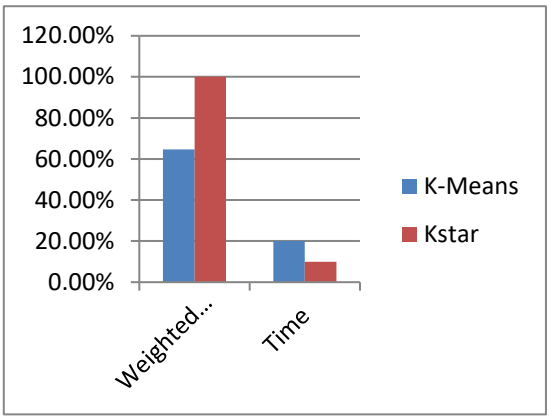


Figure:6 Comparision ofAlgorithm

V. Conclusion

This proposed method was designed to distinguish between the classification of diseased and non-diseased fish in various areas. Many people are economically dependent on fish farming and production in Ramanathapuram District. It will be concluded that many efforts have been made for fish disease detection with the help of image classification studies, which have been conducted. The existing method uses the K-Means classification algorithm for better accuracy and minimum execution time. The proposed method using the Lazy K-Star algorithm takes the minimum execution time for classification and also provides a weighted average accuracy of 100% compared with the existing method. The Lazy K-Star algorithm is exemplary for fish disease classification.

VII. References

[1]. Yang X, Zhang S, Liu J, Gao Q, Dong S, Zhou C. Deep learning for smart fish farming: applications, opportunities and challenges. Reviews in Aquaculture. 2021 Jan;13(1):66-90.

[2]. Zhao S, Zhang S, Liu J, Wang H, Zhu J, Li D, Zhao R. Application of machinelearning in intelligent fish aquaculture: A review. Aquaculture. 2021 Apr 3:736724.

[3]. Ahmed MS, Aurpa TT, Azad MA. Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture. Journal of King Saud University-Computer and Information Sciences. 2021 May 19.

[4]. Pauzi SN, Hassan MG, Yusoff N, Harun NH, Bakar AA, Kua BC. A review on image processing for fish disease detection. InJournal of Physics: Conference Series 2021 Aug 1 (Vol. 1997, No. 1, p. 012042). IOP Publishing

[5]. Knausgård KM, Wiklund A, Sørdalen TK, Halvorsen KT, Kleiven AR, Jiao L, Goodwin M. Temperate fish detection and classification: A deep learning based approach. Applied Intelligence. 2021 Mar 22:1-4.

[6]. Cui S, Zhou Y, Wang Y, Zhai L. Fish detection using deep learning. Applied Computational Intelligence and Soft Computing. 2020 Jan 23;2020

[7] Rekha BS, Srinivasan GN, Reddy SK, Kakwani D, Bhattad N. Fish detection and classification using convolutional neural networks. InInternational Conference On Computational Vision and Bio InspiredComputing 2019 Sep 25 (pp. 1221-1231). Springer, Cham.

[8]. Waleed A, Medhat H, Esmail M, Osama K, Samy R, Ghanim TM. Automatic Recognition of Fish Diseases in Fish Farms. In2019 14th International Conference on Computer Engineering and Systems (ICCES) 2019 Dec 17 (pp. 201-206). IEEE.

[9]. Tseng SL, Lin HY. Fish Detection Using Convolutional Neural Networks with Limited Training Data. InAsian Conference on Pattern Recognition 2019 Nov 26 (pp. 735-748). Springer, C

[10]. Kappeler A, Yoo S, Dai Q, Katsaggelos AK. Video super-resolution with convolutional neural networks. IEEE transactions on computational imaging. 2016 Mar 30;2(2):109-22.

[11]. Waleed A Medhat H Esmail M Osama K Samy R and Ghanim T M 2019 Automatic recognition of fish diseases in fish farms Proc. Int. Conf. Comput. Eng. Syst. pp. 201–6 DOI:0.1109/ICCES48960.2019.9068141

[12]. Han C Lee S Han S and Park J 2011 Two-stage fish disease diagnosis system based on clinical signs and microscopic images ICCSA 2 pp. 635–47

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 08 | August - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Heart Attack Prediction in Machine Learning Environment

Dr. P. Senthil Kumari[1], S. Vinitha[2]*
,
[1] Head In charge, Dept. of Computer Science, Thassim Beevi Abdul Kader College for Women, Kilakarai,
[2] M.Phil. Research Scholar, Tamil Nadu, India.
senthilmathimca@gmail.com

Abstract. Healthcare aims to impart more personal, preventive and predictive clinical treatments to patients nowadays. Since digitized data acquisition is growing more, Artificial Intelligence (AI) applications such as Machine Learning (ML) using Data Mining (DM) are used in the areas which are misunderstood as the province of human expert's fields. AI performs mimicry of human cognitive functions. Patient care requires interaction between clinicians, caretakers, managers and executives to decide on a clinical method (treatment, diagnosis and therapy) and a non-clinical method (technology chosen, budget, etc.). Due to advancements in deep neural networks, researchers and medical practitioners focus on developing an automated system to anticipate the occurrence of heart attacks. In the proposed research work, an extreme learning machine using AI and ML based diagnosis system using R-tool for heart attack prediction using classifiers (Decision Tree (DT) and Random Forest Algorithm (RFA)) is proposed. The objective of this research work is to find the most suitable classification algorithm for predicting the heart disease. The performance of classifiers is evaluated with performance metrics such as classification accuracy and error rate. The performance of the proposed system has been validated on 13 features and a reduced set of features. The feature reduction improved the performance of classifiers in the system performance metrics of accuracy and error rate. The proposed ML-based decision support system will assist the doctors in diagnosing the prediction of heart attack for the patients efficiently.

Keywords - Machine Learning, Data Mining, Decision Tree and Random Forest Algorithm

I. INTRODUCTION

Netflix attracts people by giving people's favorite movies and series. Amazon attracts people by providing people's favorite items. Google attracts people by giving the terms and conditions they search for. Detailed personal profiling can be created with all this data used to understand and predict healthcare trends. Artificial Intelligence can be used in all areas of healthcare, from admitting the patient to discharge [1].

AI is a collection of technology. AI is implemented in diagnostics, therapeutic and population health management. Specific AI technologies relevant to healthcare are Machine Learning (ML), AI based Deep Neural Networks (DNN), Deep Learning (DL) and Natural Language Processing (NLP) [2].

Precision medicine is the existing application of ML in healthcare to predict the treatment protocol. Supervised learning is used for this purpose. Digital healthcare applications, Complex algorithms and genes, proteins and metabolites-based tests are three clinical areas where innovation in precision medicine is used. Reduced healthcare costs, reduced drug response side effects, and improving drug action effectiveness are the advantages of precision medicine [3].

Google collaborates with health delivery networks to provide prediction models from big data to alarm clinicians of heart failure conditions, etc. In addition, ML has complex dialogue management and conversational flexibility to handle translation and dissemination of information to help patients [4].

Healthcare organizations use ANN to provide care delivery at a minimum cost by informing management decisions at the micro, meso and macro levels. Individual patient decisions come under the micro-level. Decisions made at the group (for example. Organization or department) are meso-level. Finally, large groups or public agencies decide the allocation of resources whose decisions reflect society's macro level. Information for decision making (in this paper, thirteen attributes of a patient) are captured and entered into Excel sheet for further analysis of categorization [5].

NLP comprises speech recognition, text analysis, translation, etc. Statistical and semantic NLP are classifications o NLP. Statistical NLP is a combination of DL and ANN. NLP includes the interpretation of clinical documents, reports and conversational AI. Clinical notes in cardiology help derive metadata in the development of downstream ML applications [6].

Empathy and ethical dilemma lack between AI and patients. Streamlining and creative solutions are effective and preferred by humans. During the SARS-COV-2 epidemic, there is a need to promote remote healthcare technologies to save lives and reduce the risk to both healthcare workers and patients [7].

More than 500 people are affected by stroke all over the world. In China, stroke is the first cause of death, and in North America, stroke is the fifth cause of death. US$ 689 billion medical expenses are spent on stroke. To prevent the death rate by stroke and reduce the medical expenses on stroke, this research paper proposes an optimal framework for heart attack prediction in balanced high dimensional big data.

The rest of the paper is structured as follows: Section II includes the existing works related to various big data analytics in healthcare, data mining in healthcare, machine learning in healthcare and artificial neural networks in healthcare. Section III describes the implementation details of the proposed heart attack prediction model in big data environment. Finally, section IV discusses the results and discussions of the proposed work, and section V presents the conclusion of this paper.

II. RELATED WORKS

Big Data Analytics in Healthcare

Core providers of healthcare services (such as nurses, doctors, lab technicians, etc.), critical healthcare services (such as health insurance policies and research based on medical field) and the recipients of healthcare services (public people and patient) are the three entities of medical healthcare system. Applications of Big Data Analytics (BDA) are five-fold: health awareness among common people, interactions among stakeholders in the healthcare environment, hospital administration practices, treatment of particular diseases and technology used in healthcare [8]. Healthcare defined three levels of big data analytics with enhanced functionality and values: descriptive, predictive and prescriptive. Descriptive level reports current situations and problems. The predictive level simulates and models the AI techniques. Finally, the prescriptive level optimizes clinical, financial and other outcomes [9].

MEDLINE, Web of Science, Embase, Scopus, Cochrane Database of Systematic Reviews via Cochrane Library and Epistemonikos are six databases used for big data analytics. In addition, patients' data are collected from Electronic Health Records (EHR), private hospital's patient databases and general public hospital's information systems. "Probability of dying from cardiovascular disease was the most frequently assessed health indicator and core priority within the WHO (World Health Organization) General Programme of Work 2019/2023 [10].

Hadoop, MapReduce, Google Big Query, Microsoft Windows Azure and Jaql are some of the tools used for analyzing big data. Frequency matching and pattern matching are the diagnostic methods used for heart disease with the help of ECG patterns [11]. Clinical prediction models of supervised learning methods are classified into three types: statistical methods (for example. Linear regression and Bayesian model, etc.), sophisticated methods in ML and Data Mining (DM) (for example. Decision trees and ANN) and survival methods to predict survival outcomes. These prediction models discover the relationship between several attributes and an outcome variable [12].

Data Mining in Healthcare

Data mining is an approach that efficiently mines data from a large database to get necessary data. Predictive and descriptive models are the two types of data mining tasks. Constructive diagnosis and treatment of the disease, detection of misuse and deceit, inpatient records, etc., are examples of data mining [13].

The Data pyramid shows the architecture and illustrates data management in figure 1.a. It demonstrates the data flow from raw data through an analytical engine in the Hadoop environment to get the required output. Data lifecycle
shows the different processes in data such as data collection, data pre-processing, data reduction and transformation, data analytics and data output, as in figure 1.b [14].

A multinomial Naïve Bayes algorithm was proposed by Saqlin et al. to diagnose heart failure. Thirty variables data set is used. Various classification algorithms (Neural networks, Decision Tree, Logistic Regression, et.) were used for comparison. An accuracy of 86.7% and a curvature area of 92.4% are achieved [15].

Disease diversity, heterogeneity of treatment and outcome, etc., are the characteristics of medical data. Big data mining I medical field are more difficult to obtain. They are based on several formats and organized data. Therefore, professional knowledge is important to analyze and interpret results. Surveillance Epidemiology and End Results (SEER), Medical Information Mart for Intensive Care (MIMIC), China Health and Nutrition Survey (CHNS), Health and Retirement Research (HRS) and Gene Expression Profiling Interactive Analysis (GEPIA) are some of the public medical database [16].

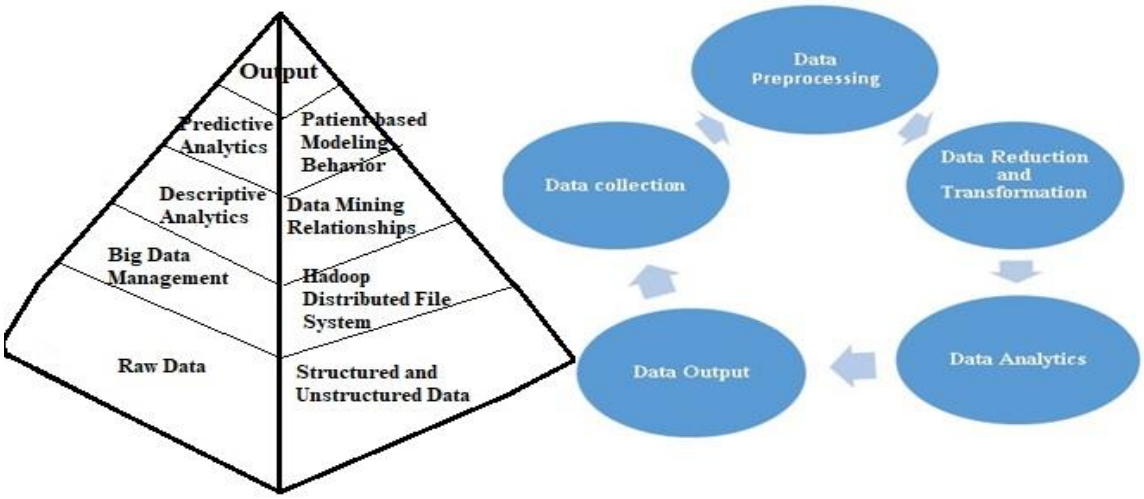


FIGURE 1. a) *Data Pyramid* b) *Data Lifecycle*

WEKA (Waikato Environment for Knowledge Analysis), KEEL (Knowledge Extraction Based on Evolutionary Learning), KNIME (Konstanz Information Miner), R programming language, Rapid miner and Orange are some of the data mining tools used in healthcare. In this proposed work, the R programming language is used for data mining purposes [17].

Machine Learning in Healthcare

People use natural human language to interact with a chatbot. ELIZA is the premier chatbot. Patient care chatbots improve the communication between clinic-patient and doctor-patient through remote testing, monitoring and telephone consultation. The recent chatbot can answer any query based on the Frequently Asked Questions (FAQ) using latent semantic analysis of NLP and Artificial Intelligence Markup Language (AIML) [4].

Supervised learning has classification and regression as two types. Random Forest Algorithm (RFA) can perform in both these categories. The ground method of RFA is recursion. In this algorithm, decision trees are created. RFA is used for imbalanced datasets since it is insensitive to noise. Moreover, RFA eliminates overfitting [18]. In the proposed work for heart attack prediction, a random forest algorithm is used.

DL finds very complex correlations. ANN with more than ten layers is used for this purpose. Simulation of artificial neurons is triggered. Watson of IBM AI machine and Deep Mind of Google AI machine are used for healthcare applications [1].

Artificial Neural Networks in Healthcare

Trained ANNs behave similarly to the biological neural cluster. Digital model of the human brain and can find the nonlinear relationship between dependent and independent variables. As with first and second-level headings, all words except prepositions. ANNs are used in cardiology, diagnosis, electronic signal analysis and medical image analysis [19]. For example, Senthilkumar et al. proposed a novel method to improve accuracy in predicting a heart

attack. A hybrid of Random Forest and Linear classification techniques is used to get improved accuracy of 88.7% [20].

A real-time heart attack prediction based on apache Spark is used to stream data against ML through in-memory calculations. Streaming processing and data storage visualization are two subparts. Spark MLib with Spark streaming is used for the first part. Apache Cassandra is used for the second part. UCI (University of California, Irvine C. A) is the data set used [21].

DL is an advanced version of NNs. DL explores more complex non-linear patterns in data mining. Arterys – one of the medical imaging clouds AI obtained clearance from the Food and Drug Administration (FDA) of the United States of America to publicize its Arterys Cardio DL package legally. The application implements AI to present editable and automated ventricle segmentation [23].

The computing of the brain-inspired Bioinspired computing. Magnetic Nano-devices display behaviors of neurons and synapses. Fully integrated spintronic Complementary Metal Oxide Semiconductor (CMOS) bioinspired hardware is proposed by Julie Grollier et al. [24].

III. HEART ATTACK PREDICTION IN MACHINE LEARNING ENVIRONMENT

The proposed work is implemented in a Pentium IV processor, 2GB RAM, 1 GB disk space and R 2.0.0 and R tool. Compiled RPMs (Red hat Package Manager) for Linux. Party package plot tree is used to show the plotting of prediction of decision tree classification algorithm. Data collection, feature selection and classification algorithms are the three modules used in this proposed work. Random forest algorithm is implemented with caret package prediction. R tool is used for classification. Misclassification error is analyzed in the decision tree. Error rate and histogram are shown in caret package prediction of random forest.

Conventional clinical practice depends on the manual patients' records. Conventional Decision Support Systems (DSS) use EHR, and other digital medical data use AI-based ML algorithms for decision making. Integrative DSSs use data mining and DNN techniques to make a decision. Fully automated clinical systems behave as e-doctors [22]. Deep Neural networks are the combination of the input layer, output layer and hundreds of hidden layers. Neural networks with hundreds of layers model complex relationships between the input, output and complex mathematical
operations for the neurons are designed. The convolutional layer extracts spatial relations. Modern neural networks work with millions of parameters and consume huge amounts of resources to train.

Supervised learning of training the UCI data set is performed systematically. First, data cleaning and preprocessing are performed. Then, missing values (no recorded values for several attributes), outliers (inconsistent data) and scaling transformation of data (range of numeric attributes) are performed. In features selection, numeric variables that are restricted to have continuous values are clustered together. The reduced features set is given for any classification algorithms (in this proposed work, DT and RFA) to predict heart attack. The output of the classification algorithm is given as input to Extreme Learning Machine (ELM) for further process. Feedforward neural networks are used in ELM. Hidden nodes are assigned randomly with nonlinear transformations. ELM, with ML techniques' help, is obtained whether the patient has a heart attack or not. The data mining process is shown in figure 2.

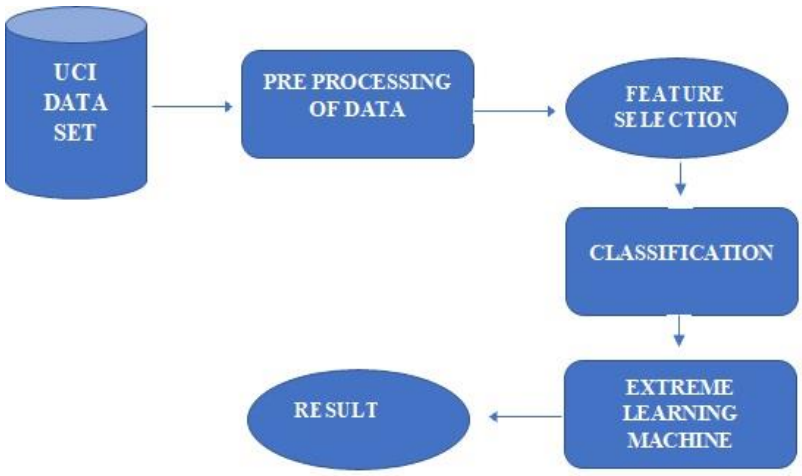


FIGURE 2. *Data Mining in Classification Algorithm*

UCI Cleveland heart diseases data set includes 304 instances with no missing values. The number of major vessels (CA), SLOPe of the peak exercise ST segment (SLOP), EXercise induced rate Achieved (EXCHANGE), RESTing Electro Cardio

Graphic (RESTECG), Cholesterol (CHOL), Resting Blood Pressure (RBP), Age, Sex, Fasting Blood Sugar (FBS), Chest Pain (CP), Maximum heart rate Achieved (THAL), ST depression induced by exercise relative to rest (OLPEAK) and reversible defect are the thirteen attributes taken for consideration to predict the heart attack. All the attributes are numerical. CP determines the pain type, ranging from 1 to 4. RBP ranges from 92 to 100. FBS is a Boolean value true of false with values 1 or 0. RESTECG varies from 0 to 2. THAL ranges from82 to 185. EXANGE is a Boolean value with 1 or 0. Part of UCI Data set is shown in figure 3.

Data cleaning, preprocessing and filtering of data set are performed. Attribute dependency is verified to find the hidden patterns and relationships. After analyzing the data and applying advanced data mining techniques, data is given as classification algorithms. The classification algorithm has two steps: the learning step and the prediction step. In the learning step, the model learns from the trained data. In the prediction step, the model predicts the response from the given input data.

A decision tree creates a training model that can predict the value of the target variable by using decision variables inferred from training data. Categorical variable decision tree and continuous variable decision tree are the two varieties of the decision tree. The decision tree follows disjunctive normal form. Algorithm selection used in the decision tree is based on the target variable. Iterative Dichotomiser 3 (ID3), C4.5 (successor of ID3), Classification and Regression Tree (CART), CHi-square Automatic Interaction Detection (CHAID) and Multivariate Adaptive Regression Splines (MARS) are the algorithms used in the decision tree. The following algorithm 1 shows the steps involved indecision tree based clustering.

Algorithm 1: Decision tree based clustering
Input: D – UCI Data set with thirteen features taken for consideration

1. For all thirteen features do
a. For each sample do
Execute the Decision tree algorithm
b. End for
2. End for
3. Obtain the total number of leaf nodes $l_1, l_2, l_3, \ldots l_n$.
4. Split the data set D into $d_1, d_2, d_3, \ldots d_n$ based on leaf nodes.

Output: Partition data sets $d_1, d_2, d_3, \ldots d_n$

Random forest algorithm creates a forest and makes it random. The larger the number of trees in the forest, the results are more accurate. Random forest uses a bagging algorithm and ensemble learning technique. The random
forest can be used for classification and regression tasks. If there are enough trees in the forest, the classifier will not over-fit the model. Random forest creation and prediction from random forest classifier are the two stages in the random forest. It needs a longer training period. The following algorithm 2 shows the random forest based clustering algorithm.

Algorithm 2: Random forest based clustering
Input: D – UCI data set with thirteen features taken for consideration

1. For all thirteen features do
a. For each sample do
Execute the random forest algorithm
b. End for
2. End for
3. Obtain the total number of leaf nodes $l_1, l_2, l_3, \ldots l_n$.
4. Split the data set D into $d_1, d_2, d_3, \ldots d_n$ based on leaf nodes.

Output: Partition data sets $d_1, d_2, d_3, \ldots d_n$

"Read CSV" operator of rapid miner tool is first loaded for secondary data retained in Comma Separated Values (CSV) file. Then, only a subset of data is selected from the loaded data. "The operator uses select Attributes" for selecting the subset. Finally, it is given to the "X-Validation" operator. The AdaBoost technique is implemented for classification by the R tool. Results of the decision tree and random forest are compared. The accuracy of a random forest is greater than a decision tree. The error rate of the random forest classifier is lesser than the decision tree. The proposed system architecture is shown in figure 4.

Age	Sex	chesp pain type	resting bp	cholestrol	fasting blood sugar	electrocardiographic	max heart rate	exercise induced angina	oldpeak	slope of peak	exercise ca	thal	nsp
63	1	1	145	233	1	2	150	0	2.3	3	0	6	1
67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	3
41	0	2	130	204	0	2	172	0	1.4	1	0	3	1
56	1	2	120	236	0	0	178	0	0.8	1	0	3	2
62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3	2
63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	3
56	0	2	140	294	0	2	153	0	1.3	2	0	3	1
56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
44	1	2	120	263	0	0	173	0	0	1	0	7	2
52	1	3	172	199	1	0	162	0	0.5	1	0	7	1
57	1	3	150	168	0	0	174	0	1.6	1	0	3	

FIGURE 3. *Part of UCI Data Set*

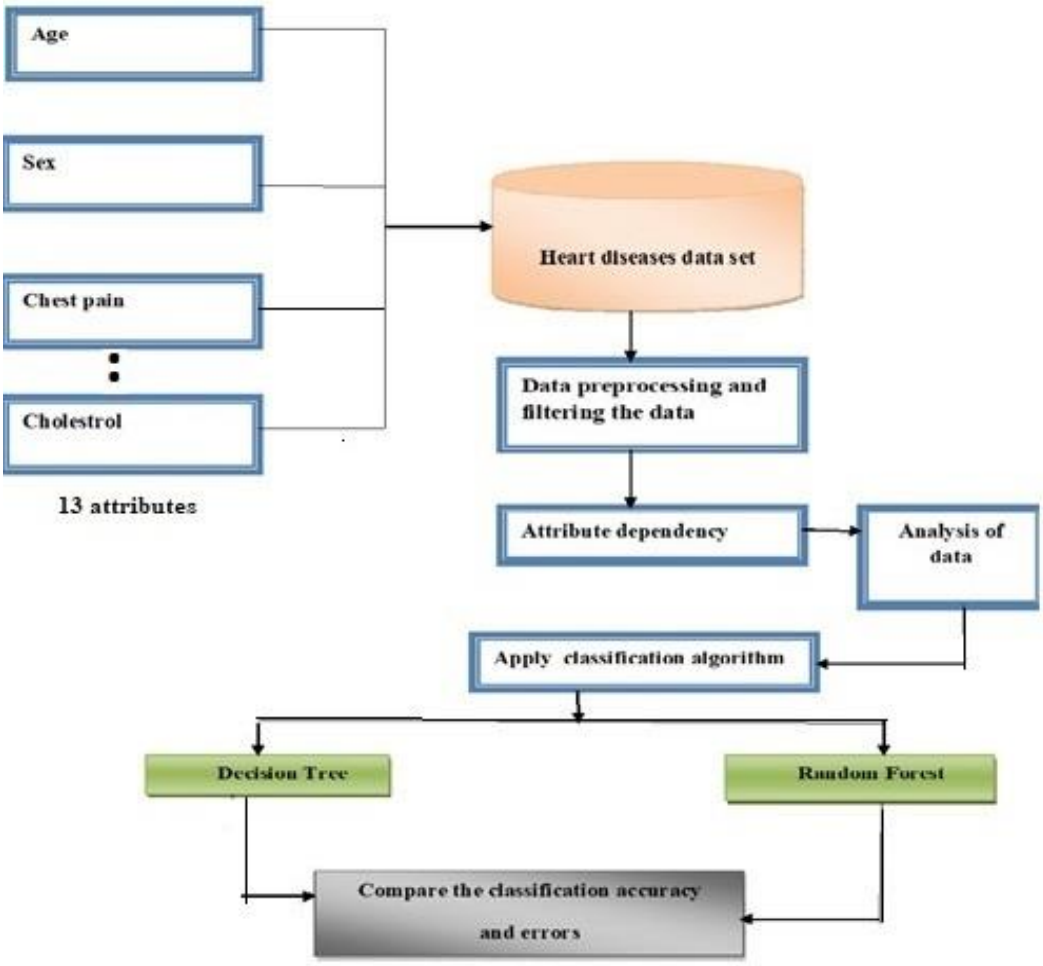


FIGURE 4. *Proposed Framework for Heart Attack Prediction*

IV. RESULTS AND DISCUSSION

R tool displays the predictions such as n-normal, s-suspect and p-pathological to depict the results. The performance of the classification algorithm is analyzed with thirteen parameters. If a classification algorithm is trained, the predictive model will present the best accuracy with the help of the following classifications:

- TP (True Positive) – Number of patients who are correctly classified as having the disease
- TN (True Negative) – Number of patients who are correctly classified as healthy
- FP (False Positive) – Number of patients who are misclassified as healthy
- FN (False Negative) – Number of patients who are misclassified as having the disease

The data set is prepared and partitioned. A decision tree with a party package of the R tool is created. It is shown in figure 5. Prediction is applied to the trained data. Partitioning of the predicted data set is performed. A decision tree is created with rpart package. Data of tree are validated. Misclassification errors are found and validated.

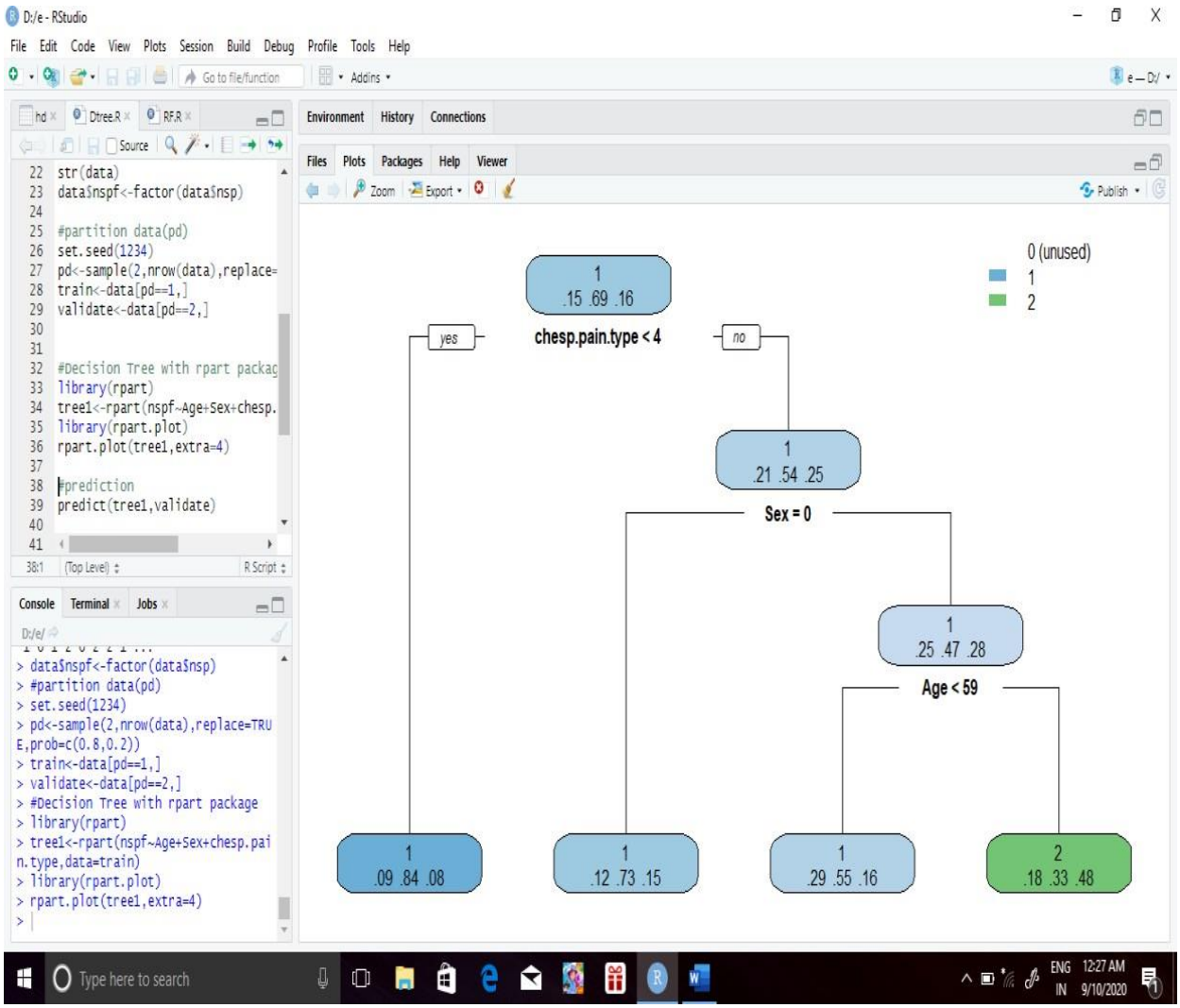


FIGURE 5. *Decision Tree Created using Party Package*

The data set is prepared and partitioned for the random forest. Random forest is created using a caret package. To train the data, prediction is performed, and a confusion matrix is created. The confusion matrix created in the random forest classifier is shown in Figure 6. The error rate of random forest is plotted in the graph. Misclassification errors are validated. The accuracy of the two classifier algorithms is shown in table 1 and figure 7.a. Error rate of two classifier algorithms is shown in table 2 and figure 7.b.

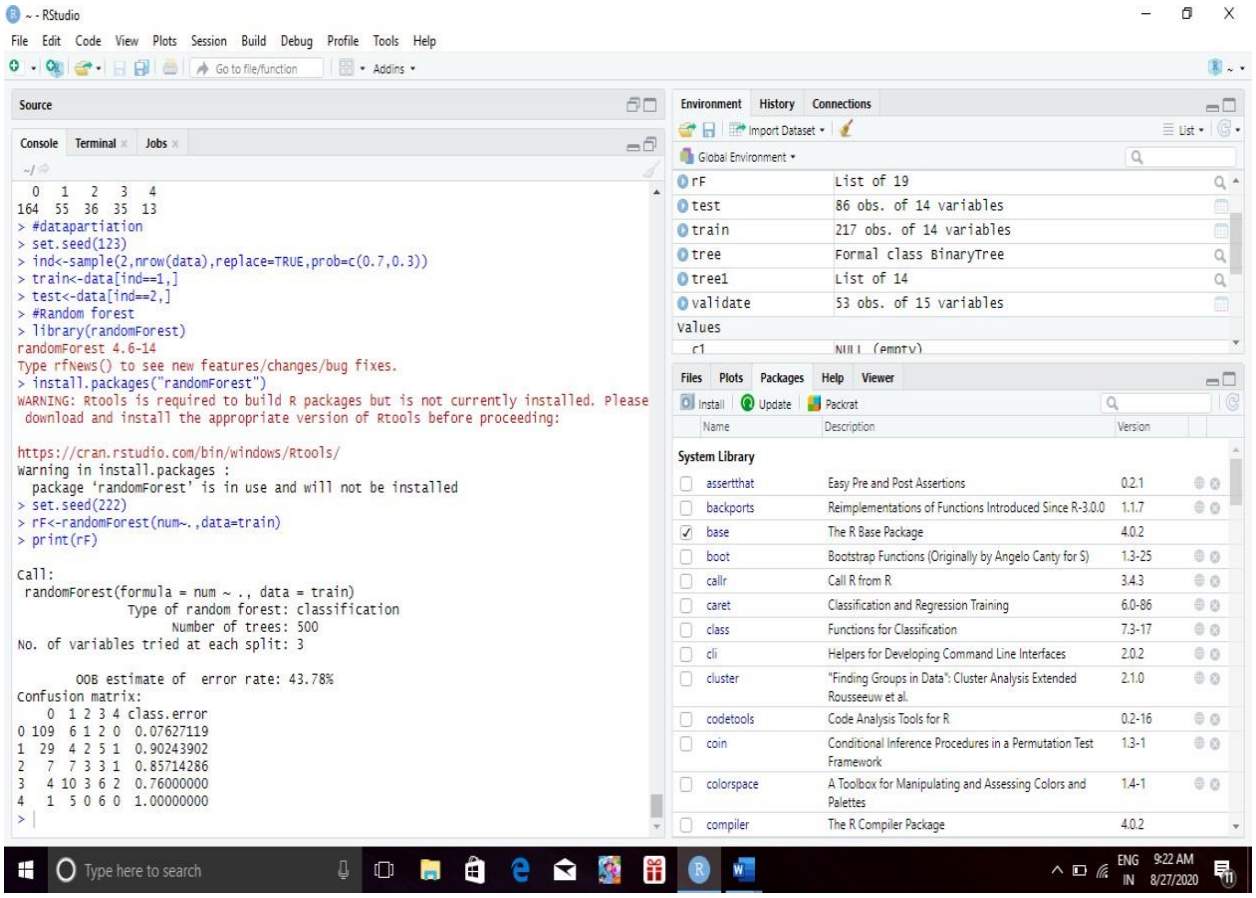


FIGURE 6. *Random Forest Confusion Matrix*

Accuracy is the fraction of many correct predictions over the total number of predictions.

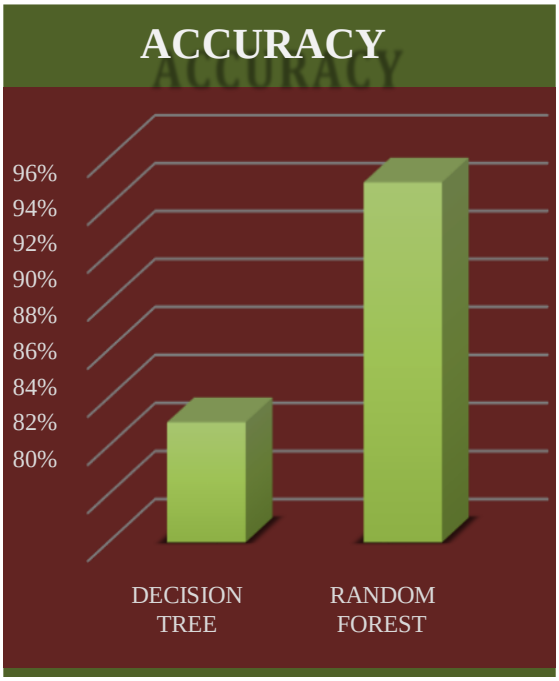
TABLE 1. *Accuracy of two classifiers*

Algorithm	Accuracy
Decision Tree	85%
Random Forest	95%

The error rate is based on the wrong predictions in the testing process. Therefore, error Rate is calculated as (1 – (Number of wrong predictions / Total number of predictions)).

TABLE 2. *Error Rate of two classifiers*

Algorithm	Error Rate
Decision Tree	47%
Random Forest	43%



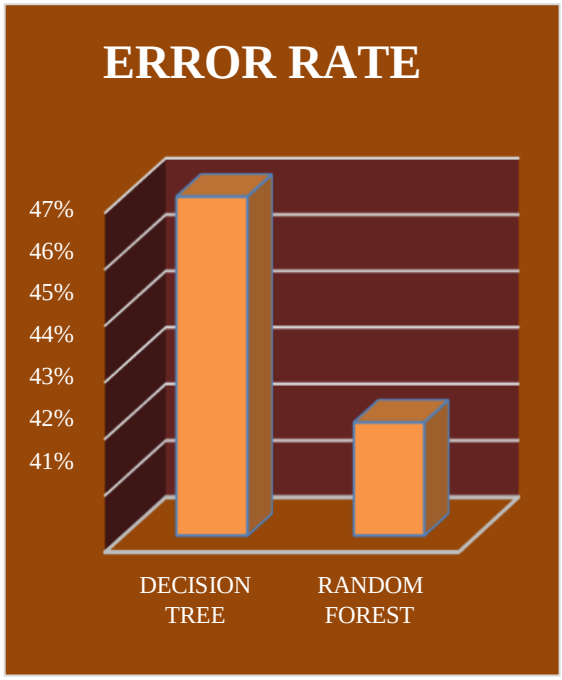


FIGURE 7. a) *Accuracy* of Classifiers b) *Error Rate of Classifiers*

Results are compared with the existing work [20]. Senthilkumar Mohan et al. proved that their research work provides 88.4% with the hybrid random forest with a linear model (HRFLM) when compared with Naïve Bayes, generalized linear model, logistic regression, deep learning, etc. The proposed research work provides 95% accuracy rate which is better than exiting work.

V. CONCLUSION

Machine Learning is a flourishing field of medicine to fuse statistics and computer science into the medical field, and the process is semi-automated. ML is the future of biomedical research, computer-aided diagnosis and personalized medicine to enhance global healthcare. Public illiteracy about AI increases the burden of trust in AI. According to a South Korean survey, only 5.9% of physicians know AI techniques from a professional perspective. Some people will not trust the government to safeguard personal EHR in the research field. There are many ethical values and hazards to be overridden to implement AI techniques in the more sensitive medical field. The predictive model will predict the accuracy of heart attack based on four classifications: True Positive (TP), True Negative, False Positive and False Negative. Sensitivity and specificity denote the persons' ability to identify TP rate and TN rate. To conclude the work, the Random forest algorithm has the highest classification accuracy performance with the lowest error rate compared with the decision tree classification algorithm to predict the heart attack.

REFERENCES

1. Adam Bohr and Kaveh Memarzadeh,"Chapter 2: The Rise of Artificial Intelligence in Healthcare Applications", Artificial Intelligence in Healthcare, DOI: https: //doi.org/10.1016/B978-0-12-818438- 7.00002-2, 2020, pp. 25-60.
2. Thomas Davenport and Ravi Kalakota, "The Potential for Artificial Intelligence in Healthcare", *Royal College of Physicians Future Healthcare Journal*, June 2019, Vol. 6, Issue 2, pp. 94 – 98, DOI: 10.7861/futurehosp.6-2-94.
3. Konieczny L and Roterman I., "Personalized precision medicine", Bio-Algorithms Med-Syst 2019, 15.
4. Sarkar Siddique and James C.L. Chow, "Machine Learning in Healthcare Communications", MDPI Encyclopedia 2021, 1, pp.220-239, https://doi.org/10.3390/encyclopedia1010021.
5. Guler S, Hurton S, C winn M and Molinari M, "Levels in Decision Making and Techniques for Clinician", International Journal of Digestive Diseases. 2015; 1(1:2).

6. Weng, W. H., Wagholikar, K. B, McCray, A. T, Szolovits, P. and Chueh, H.C, "Medical Subdomain Classification of Clinical Notes Using A Machine Learning Based Natural Language Processing Approach", BMC Med. Inform. Decis. Mak, 2017, 17, pp.1-13.
7. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD and Desroches CM, "Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views", J. Med Internet Res 2019, 21:e12802.
8. Sayantan Khanra, Amandeep Dhir, A. K. M. Najmul Islam and Matti Mäntymäki, "Big Data Analytics in Healthcare: A Systematic Literature Review", Enterprise Information Systems, 14:7, 878-912, DOI: 10.1080/17517575.2020, 1812005, Oct 2020.
9. William R. Hersh, "Healthcare Data Analytics", Hoyt, RE, Yoshihashi, A, Eds.(2014), Health Informatics: Practical Guide for Healthcare and Information Technology Professionals, Sixth Edition, Pensacola, FL, Lulu.com., pp. 1-13, Chapter 3.
10. Israel Júnior Borges do Nascimento , Milena Soriano Marcolino, Hebatullah Mohamed Abdulazeem, Ishanka Weerasekara, Natasha Azzopardi-Muscat, Marcos André Gonçalves and David Novillo-Ortiz, "Impact Of Big Data Analytics On People's Health: Overview of Systematic Reviews and Recommendations for Future Studies", J Med Internet Res. 2021 Apr 13;23(4):e27275, DOI: 10.2196/27275.
11. Lidong Wang and Cheryl Ann Alexander, "Big Data Analytics in Healthcare Systems", International Journal of Mathematical, Engineering and Management Sciences Vol. 4, No. 1, pp. 17–26, 2019.
12. Chandan K. Reddy and Charu C. Aggarwal, "Healthcare Data Analytics", Chapman and Hall/CRC Press, June 2020, pp. 1-35.
13. Rakhi Ray, "Advances in Data Mining: Healthcare Applications", International Research Journal of Engineering and Technology (IRJET), Volume: 05, Issue: 03, Mar-2018, pp. 3738-3742.
14. Revanth Sonnatti, "Improving Healthcare Using Big Data Analytics", International Journal of Scientific & Technology Research, Volume 6, Issue 03, March 2017, pp. 142-146.
15. M. Saqlain, W. Hussain, N. A. Saqib, Nazar, and M.A. Kha,"Identification of Heart Failure by Using Unstructured Data of Cardiac Patients," in Proc. 45th International Conference on Parallel Processing Workshops, pp.426-431, 2016.
16. Jin Yang, Yuanjie Li, Qingqing Liu, Li Li, Aozi Feng, TianyiWang, Shuai Zheng, Anding Xu and Jun Lyu, "Brief introduction of medical database and datamining technology in big data era", *J Evid BasedMed.* 2020, Vol. 13, pp. 57–69, wileyonlinelibrary.com/journal/jebm, DOI: 10.1111/jebm.12373.
17. R. Pallavi Reddy, Ch. Mandakini and Ch. Radhika, "A Review on Data Mining Techniques and Challenges in Medical Field", International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 08, August-2020, pp. 329 – 333.
18. Senerath Mudalige Don Alexis Chinthaka Jayatilake and Gamage Upeksha Ganegoda1, "Involvement of Machine Learning Tools in Healthcare Decision Making", Computational Decision-Making Tools for Healthcare, Volume 2021 | Article ID 6679512 | https://doi.org/10.1155/2021/6679512, pp. 1-34.
19. Jignesh Patel and Ramesh K Goyal, "Applications of Artificial Neural Networks in Medical Science", Artificial Intelligence in Drug Discovery and Development Project, Current Clinical Pharmacology, Vol. 2(3), pp. 217-26, October 2007, DOI: 10.2174/157488407781668811, PubMed Source.
20. Senthilkumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", DOI: 10.1109/ACCESS.2019.2923707, Vol. 7, July 2019, pp. 81542 – 81554.
21. Abderrahmane Ed-daoudy and Khalil Maalmi, "Real-Time Machine Learning for Early Detection of Heart Disease using Big Data Approach", IEEE Xplore, International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 2019.
22. Kun-Hsing Yu, Andrew L. Beam and Isaac S. Kohane, "Artificial Intelligence in Healthcare", Nature Biomedical Engineering, Vol. 2, October 2018, www.nature.com/natbiomedeng, pp. 719-731.
23. Bernard Marr, "First FDA approval for clinical Cloud-Based Deep Learning in Healthcare", 2017, https://www. forbes.com/ sites/ bernardmarr/ 2017/01/ 20/ first- fda- approval- for- clinical- cloud- based- deep- learning- in healthcare/#7a0ed8dc161c (accessed 1 Jun 2017).
24. Julie Grollier, Damien Querlioz and Mark D. Stiles, "Spintronic Nano devices for Bioinspired Computing", Journals & Magazines, Proceedings of the IEEE, DOI: 10.1109/JPROC.2016.2597152, Oct 2016, Vol. 104, Issue 10, pp. 2024-2039.

JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MUSHROOM CLASSIFICATION: A COMPARISON OF CLASSIFICATION ALGORITHMS USING MACHINE LEARNING TECHNIQUES

Ms D Usha Rani
Department of Computer Science,
TBAK College for Women,
Kilakarai, Tamilnadu, India

Abstract : Mushrooms are used in the medical field to cure diseases like anemia, increase body immunity, diabetes, and cancer treatment. Some of the mushroom varieties are edible, but others are highly poisonous. The identification of mushrooms whether edible or poisonous is a difficult process because of the large number of mushrooms have similar characteristics. The principle of this paper is to classify the mushrooms by using Machine learning classification algorithms through a data mining tool. For the best classification, five classification algorithms were compared. The results showed that JRip classification is the best based on the difference between before and after applying data reduction in time taken.

Keywords: Data Mining, Classification, Data Pre-processing, Data Reduction, CFS subset evaluation, JRip, Decision Table, SGD, SMO, Logistic Regression.

I.INTRODUCTION

Data mining is also usually mentioned as knowledge discovery from Data (KDD). The aim of data mining is to mine useful and relevant information from huge databases or data warehouses. Knowledge discovery is a collaborative process, comprising of developing an understanding of the application domain, choosing and making a data set, pre-processing, data transformation.
Mushrooms are live in several habitats like above the ground, on the ground, or even on the plants such as deceased wood. It contains high protein, vitamins, minerals, and antioxidants. In the science field, it is one type of fungus. Mushrooms are the most sustainably produced food is not only taste but also have an excessive nutritional value. Mushrooms are also said to be a mediator to fight cancer cells and kill some types of viruses that carry infectious diseases such as polysaccharides, glycoproteins, and proteoglycans.

II. LITERATURE REVIEW

Narumol Chumuang, et al., in their paper compared seven classification algorithms for getting the highest accuracy rate. The algorithms are Naïve Bayes Multinomial Text,Naïve Bayes Updateable,Naïve Bayes,SGD Text,LEL,K-NN and Stacking. Finally, the K-NN classification algorithm shows a 100% accuracy rate.

Agung Wibowo,et al., in their paper they compared three classification algorithms for getting the highest accuracy rate. The algorithms are c 4.5, Naïve Bayes and SVM. Finally, c 4.5 classification algorithm shows 100% accuracy rate.

Kanchi Tank ., in their paper based on supervised learning algorithms such as Support Vector Machines(SVM), K-Nearest Neighbour (KNN), Naïve Bayes,Decision Trees (DT), Random Forest (RF) and Logistic Regression (LR). The result shows that Random Forest and K-Nearest Neighbour gave the highest test accuracy of 92.21% and 92.04%.

Shuhaida Ismail., in their paper presented the methodology and results for mushroom classification experiment based on their behavioral features such as characteristics, population, and habitat. Three classification algorithms were experimented with and compared. The algorithms are Decision Trees, K-Nearest Neighbour, and Naive Bayes. Principal Component Analysis (PCA) is identifying which attributes are important to classify the mushrooms. The decision Tree algorithm shows a 100% accuracy rate.

III. PROPOSED METHODS

Data Mining tools and Machine Learning techniques are used to convert raw data into some actionable, meaningful information. Three important phases are involved which are Data Pre-processing, Data Reduction and classification.

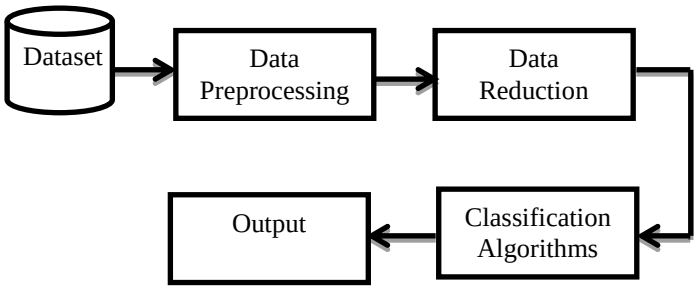


Figure 1: Methodology for mushroom classification

A. Data set

An openly available dataset is downloaded from the UCI Machine Learning Repository. It contains 23 attributes with 8124 instances of mushrooms.

Table 1: Mushroom Dataset (Attributes and Features)

S. No	Name of the Attributes	Name of the Features
	Cap Shape	Bell, Conical, Convex, Flat, Knobbed, Sunken
	Cap Surface	Fibrous, Grooves, Scaly, Smooth
	Cap Color	Brown, Buff, Cinnamon, Gray, Green, Pink, Purple, Red, White, Yellow
	Bruises	Yes, No
	Odor	Almond, Anise, Creosote, Fishy, Foul, Musty, None, Pungent, Spicy
	Gill Attachment	Attached, Descending, Free, Not attached
	Gill Spacing	Close, Crowded, Distant
	Gill Size	Broad, Narrow
	Gill Color	Black, Brown, Buff, Chocolate, Gray, Green, Orange, Pink, Purple, Red, White, Yellow
	Stalk Shape	Enlarging, Tapering
	Stalk Root	Bulbous, Club, Cup, Equal, Rhizomorphs, Rooted, Missing
	Stalk Surface above Ring	Fibrous, Scaly, Silky, Smooth
	Stalk Surface below Ring	Fibrous, Scaly, Silky, Smooth
	Stalk Color above Ring	Brown, Buff, Cinnamon, Gray, Orange, Pink, Red, White, Yellow
	Stalk Color below Ring	Brown, Buff, Cinnamon, Gray, Orange, Pink, Red, White, Yellow
	Veil Type	Partial, Universal
	Veil Color	Brown, Orange, White, Yellow
	Ring Number	None, One, Two
	Ring Type	Cobwebby, Evanescent, Flaring, Large, None, Pendant, Sheathing, Zone
	Spore Print Color	Black, Brown, Buff, Chocolate, Green, Orange Purple, White, Yellow
	Population	Abundant, Clustered, Numerous, Scattered, Several, Solitary
	Habitat	Grasses, Leaves, Meadows, Paths, Urban, Waste, Woods
	Class	Non-toxic, Toxic

B. Data Pre-processing

Data cleaning is one of the pre-processing steps that is used to prepare raw data for analysis by removing unwanted data or bad data and filling in the null values. ignore the row, Fill the missing value manually, Global constant value, Central tendency for the attribute, attribute mean or median value, most probable value is the steps to fill the missing values. Any one of the steps is enough to remove unwanted data. Smoothing by bin means, bin medians, bin boundaries, Regression, and Outlier analysis are the steps to remove the noisy data. In this dataset stalk- root has 31% of missing values. Maximum missing percentage value will affect the performance. so applied unsupervised Replace missing values to this attribute. After applying this filter, it will show 0% of the missing value.

C. Data Reduction

Data reduction is a process that reduced the original data set that is much smaller in volume and also maintains the integrity of the original data. It does not affect the result obtained from data mining before data reduction and after data reduction is the same. Data reduction increases the efficiency of data mining.

Correlation-based Feature Selection (CFS) is one of the data reduction techniques. It is used to reduce the number of variables or attributes for analysis in data set by extracting needed attributes from a large pool. It is used to improve the performance of an algorithm and improve visualization. After applying this feature selection, it will select Odor, Gill-Spacing, Stalk-surface-above-ring, and Veil-color attributes.

Table 2: Mushroom Dataset (After Applying Data Reduction)

S. No	Attributes
	Odor
	Gill-Spacing
	Stalk-surface-above-ring
	Veil-color

D. Classification Algorithms

i) Decision Table

A Decision Table is a supervised learning method that can be used for both classification and Regression problems, but mostly it is mostly chosen for solving Classification problems. It is a tree-structured classifier, where internal nodes denote the features of a dataset, branches denote the conclusion rules and each leaf node denotes the outcome. It is also called a decision tree, because looks like a tree, it begins with the root node, which enlarges on further branches, and finally constructs a tree-like structure. In a Decision tree, there are two nodes are used, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have numerous branches, whereas Leaf nodes are the output of the decisions and do not contain any extra branches. The decisions or the test are performed on the source of features of the given dataset. It is a graphical representation for getting all the probable solutions to a problem or decision based on specified conditions.

ii) JRip

The RIPPER (repeated incremental pruning to produce error reduction) algorithm is upgraded upon IREP to generate rules that match or exceed the performance of decision trees. Having progressed from numerous iterations of the rule learning algorithm, the RIPPER algorithm has a three-step process. Grow, Prune and Optimize. The first step uses a ‘separate and conquers method to include conditions to a rule until it correctly classifies as a subset of data. Like decision trees, the information gain condition is used to identify the next splitting attribute. When increasing a rule’s specificity no longer decreases entropy, the rule is instantly pruned. Until reaching the stopping condition step one and two are repeated at which point the entire set of rules is optimized using a variety of heuristics.

iii) Logistic Regression

Logistic regression is one of the Supervised Learning Methods in Machine Learning algorithms. It is mostly used to predict the dependent variable using a given set of independent variables. The values may be either Yes or No, 0 or 1, true or False, etc. it provides the probabilistic values between 0 and 1. Linear Regression is used to solve Regression problems, whereas Logistic regression is used to solve the classification problems. The sigmoid function is a mathematical function used to map the predicted values to probabilities. It plots any real value into alternative value within a range from 0 to 1. so it makes a curve like the "S" form. The S-form curve is also called the sigmoid function or the logistic function.

iv) SGD

The word ‘stochastic‘ defines a system or a process that is connected with a random probability. In Stochastic Gradient Descent methods, in each iteration, some of the samples are carefully chosen randomly instead of considering the whole data set. Although using the whole dataset is actually useful for getting to the minima in a less noisy and less random manner, the problem arises when our dataset gets big. It is computationally much less expensive than typical Gradient Descent. Hence, in most situations, SGD is chosen over Batch Gradient Descent for optimizing a learning algorithm.

v) SMO: Sequential Minimal Optimization

SMO breaks the large quadratic programming problem into a sequence of the smallest possible quadratic programming problems. These small quadratic programming problems are solved systematically, which avoids by means of a time-consuming numerical quadratic programming optimization as an inner loop. The total of memory required for SMO is direct in the training set size, which allows SMO to handle very large training sets. Because matrix calculation is omitted, SMO procedures between linear and quadratic in the training set size for several test problems. SMO's calculation time is controlled by SVM calculation; SMO is fastest for linear SVMs and sparse data sets.

IV. RESULT AND DISCUSSION

The classification of this dataset was performed to classify the mushrooms whether edible or poisonous. The results are taken from the confusion matrix.

Table 3: Comparisons of Classification Algorithms

Total Number of Instance:8124				Difference Between Before and After applying Data Reduction
S.No	Classification Algorithms	Time Taken (in Seconds)		
		Before Data Reduction Process	After Data Reduction Process	
	Decision Table	0.52	0.03	0.49
	JRip	0.30	0.17	**0.13**
	Logistic Regression	1.41	0.05	1.36
	SGD	1.06	0.20	0.86
	SMO	0.93	0.27	0.66

Before applying the process of data reduction, the JRip algorithm took less time and after the data reduction process, Decision Table algorithm took less time. The results from the above mentioned table JRip classification algorithm is the best based on the difference between data reduction process in time taken.

Finally, I conclude from Table 1, based on the difference between before and after applying data reduction JRip (0.13) is the best classification algorithm.

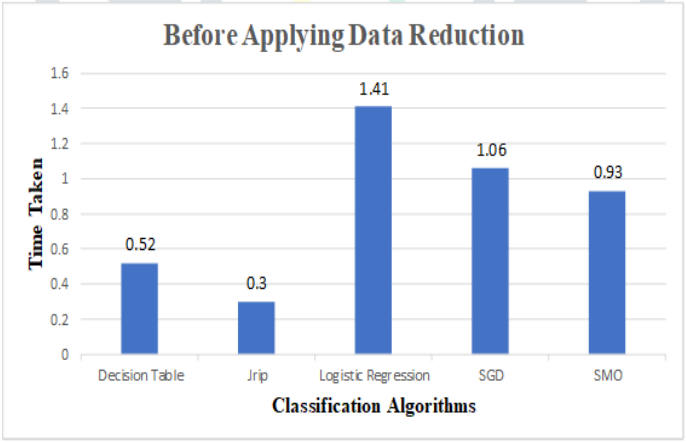


Figure 2: Bar chart representation of classification algorithms (Before Applying Data Reduction)

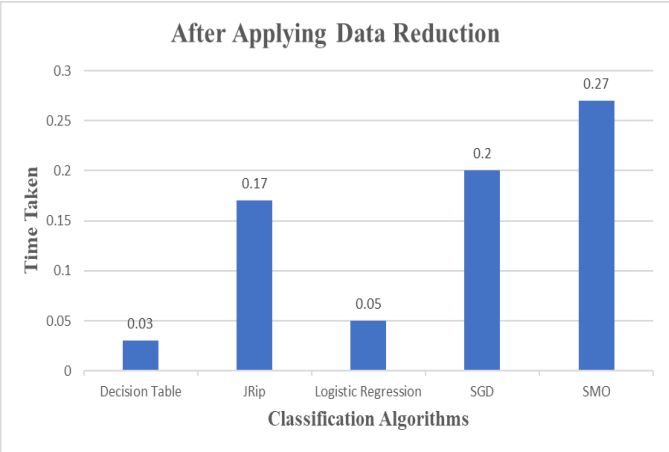


Figure 3: Bar chart representation of classification algorithms (After Applying Data Reduction)

V. CONCLUSION

This paperwork particularly focuses to classify the mushroom whether edible or poisonous because it contains fiber, protein, and antioxidants. The mushroom dataset consists of 23 attributes and 8124 instances. All of them are divided into two classes edible and poisonous. So, the classification of mushrooms is very important. Finally, I conclude JRip is the best algorithm based on processing time for this dataset. In the future, my research work concentrates on developing a mobile application with image processing.

REFERENCES

[1] Rial Adity and Setia Hadi Purwono, Jamur – Info Lengkap dan Kiat Sukses Agribisnis. Depok, Indonesia/West Java: Agriflo, 2012.
[2] Kristianus Sunarjon Dasa, "Pemanfaatan bagas sebagai campuran media pertumbuhan jamur tiram putih," vol. 11, pp. 195-201, 2011.
[3] Anna Rahkmawati, "Keanekaragaman jamur," Universitas Negeri Yogyakarta, Yogyakarta, Tech. rep. 2010. [Online].staffnew.uny.ac.id/upload/132296143/pengabdian/ppm-2010-kehati.pdf
[4] Bayu Mahardika Putra, "Klasifikasi Jamur ke Dalam Kelas Dapat Dikonsumsi Atau Beracun Menggunakan Algoritma VFI 5 (Studi Kasus: Famili Agaricus dan Lepiota)," IPB, Bogor, Laporan Akhir 2008.
[5] F.Y, Osisanwo, Akinsola J.E.T, Awodele O,Hinmikaiye J. O, Olakanmi O, and Akinjobi J."Supervised Machine Learning Algorithms:Classification and Comparison." International Journal of Computer Trends and Technology 48, no. 3 (June 25, 2017): 128–38. https://doi.org/10.14445/22312803/ijctt-v48p126.
[6] Firdous, DrxHina. "Health Benefits Of Mushroom, Uses And Its Side Effects."Lybrate, 2020. https://www.lybrate.com/topic/benefits-of-mushroom-and-its-side-effects (accessed Sep.24, 2020).
[7] E. Yukselturk, S. Ozekes, Y. K. Turel, "Prediction Dropout Student: An Application of Data Mining Methods in An Online Education Program", European Journal of Open, Distance and e-Learning, vol. 17, no. 1, pp.118 – 133, 2014.
[8] X. Y. Ren, Y. Zhao, W. Zeiler, G. Boxem, T. Li, "A Data Mining approach to analyze Occupant Behaviour Motivation", Procedia Engineering, vol 205, pp 2442-2448, 2017.
[9] Zhao, C. and Luan, J. (2006). Data mining: Going beyond traditional statistics. In New Directions for Institutional Research, 131(2) , (pp. 7 16).
[10] L.Arockiam, S.Charles,Arulkumar et.al(2010), "Deriving Association between Urban and Rural Students Programming Skills", International Journal on Computer Science and Engineering Vol. 02, No. 03, pp 687 - 690.

Integrating Seaweed Culture into Shrimp and Fish Farming Systems with Autonomous Underwater Vehicles (AUV)

Ms.M.Nagavalli*& Dr.A.Krishnaveni#
Assistant Professors*#
Department of Computer Science and Research Centre
Thassim Beevi Abdul Kader College For women, Kilakarai.

Abstract

Aquaculture is the fastest-growing food production sector in the world and has come under increasing focus and censure because of coastal pollution. Aquaculture involves major environmental impacts: waste discharge, nutrient pollution, and disease outbreaks. Seaweed aquaculture fulfils a vital role in protecting the marine ecosystem as well as preventing climate change. The Marine Algal Research Station at Mandapam was founded specifically to advance seaweed farming. The centre is well situated on the coastline of the Gulf of Mannar, where seaweed is widespread. Integrating seaweed culture into shrimp and fish farming systems is considered an achievable aquaculture practise. This system integrates the trophic level-specific productivity of a variety of aquatic species. This system's goal is to create a closed loop to minimise waste and decrease harmful impacts on the environment by using one species' waste as nutrients for another. The trophic levels in this system are based on the feeding habits of the aquatic organisms. In order to secure the sustainable production and growth of the coastal environment, it is highly important to protect it. Remote offshore subsea monitoring systems have the potential to become more safe, sustainable, and cost-effective with the help of Autonomous Underwater Vehicle (AUV) technology.

Keywords: Aquaculture, shrimp, coastal, Autonomous Underwater Vehicle, Marine Algal

Introduction

Mariculture is a specialized form of aquaculture that involves growing marine organisms for food and other goods in the open ocean or in seawater-filled tanks, ponds, or raceways. As an illustration, consider the farming of marine fish in salt water resources, including finfish and shellfish like cobia, pompano, sea bass, lobster, oysters, and seaweed. Mariculture also produces cosmetics, nutritional agar, jewelry (like cultivated pearls), and

fish meal, which are non-food items. When compared to fish raised in ponds or tanks, mariculture fish are said to be of greater quality and come in a wider variety of varieties.

Seaweed farming can gain from combined seaweed and shrimp production's positive effects on the environment by becoming more commercially viable. Nutrient bioremediation, mutual advantages for the co-cultured species, economic diversification, and enhanced profitability are all benefits of integrated aquaculture.

Aquaculture is the world's fastest expanding food production business, yet it has come under increased scrutiny and criticism as a result of coastal pollution. Effluents from intensive farming contain a high concentration of organic matter, nitrogen compounds, phosphorus, and other nutrients, rendering the water unsuited for aquaculture and contributing to eutrophication. Macroalgae plays an important function in reducing toxic waste to manageable levels and improving water quality. Aquaculture management can be improved by incorporating seaweeds into aquaculture systems. This procedure can be accomplished by either stocking seaweeds alongside shrimp at optimal stocking density or by recycling water through a pond augmented with seaweeds.

Since fish and seaweed are heterotrophic and autotrophic (i.e., consumers and producers, respectively) species, they complement one another and utilise each other's "waste." This method reduces any environmental impact while enhancing resource use, water quality, and the sustainability of the enterprise. It can also be used with other aquaculture species. Recirculation lessens the introduction of pathogens and fouling organisms, and the cost of biofiltration is recovered by turning surplus nutrients into high-quality seaweed biomass. The selling of the seaweed for abalone production or human consumption can bring in additional revenue.

Some animals fed diets containing halophytes showed poor development and food conversion efficiency, according to research. It has, however, been utilized in agriculture for soil fertility, vermicomposting, pharmaceutical as an anticancer, antibiotic, anti-hyperglycemia, anti-diabetes, antihypertension, and antihypotension, and cosmetic sectors. As a result, halophytes plants have the potential to be an ingredient and a strong source of protein in fish feed formulations.

Integrated multitrophic aquaculture techniques are promising options for overcoming the aforementioned issues. Coculture is a new word for a type of polyculture in which two or

more complementing organisms are cultivated in the same culture medium (Neoriet al., 2004).

At the chosen location, a circular cage with a minimum diameter of 6 m and a minimum depth of 4 m (113 m3 volume) or a rectangular cage measuring 6 x 4 x 4 m (96 m3 volume) can be installed. A larger outer net (predator net) surrounds them with a one-meter gap on all sides, and a bird net is affixed to the cage frame. Cages can be fixed or free-floating. Using mooring lines, chains, and anchors, the mooring system or assembly maintains the cage in the appropriate location and at the specified depth.

This idea is bound to help the local fishermen in coastal hamlets. Beneficiaries include SHGs, fishermen, fishing societies, farmers, and business people; people will be selected based on their awareness and passion.

- Seaweed **,** shrimp and fish growth Monitoring
- Predict optimal time of harvesting
- Water Quality management
- Eradication of predator Fishes
- Aeration
- Fertilization
- Monitor stocking

Literature Survey

The recently revealed IMTA concept offers an alternate method to long-term aquaculture. Many sites currently undertake intensive fish or shrimp cultivation as integrated units with seaweed and mollusc culture. The extractive components (seaweed and mollusks) in these IMTA systems collect their nutrients from the effluents of the fed components (fish or shrimp).

Farming integrated white shrimp with seaweed

Integrating seaweed culture into shrimp farming systems is seen to be a promising aquaculture strategy since seaweeds can convert dissolved inorganic nitrogen into biomass, which can be easily harvested.

Hong-xing Ge et al. 2019 investigated the effect of integrating green seaweed (Ulva prolifera) with Pacific white shrimp (Litopenaeus vannamei) at different water exchange

levels at a stocking density of 500 shrimp / m3. to analyze water quality and shrimp growth performance. This study showed that Pacific white shrimp culture (500 shrimp / m3) incorporates U. prolifera seaweed (800 mg / L) in culture systems with a 10% water exchange rate can control water quality and shrimp growth.

Seaweed planted in shallow water near the surface (30 to 50 centimeters) receives ample sunshine and grows rapidly. Water with a pH of 7.5 to 9.0 is generally considered good for shrimp farming. Ulva prolifera, a seaweed known for its rapid growth and reproduction, effectively eliminates nitrogen quickly. Ulva prolifera, a seaweed, claimed that the powerful may develop in a nutrient-rich environment and gather nutrients efficiently in the water.

Methodology

Biofloc Technology (BFT)

To keep high water quality in intensive aquaculture, the biofloc technology (BTF) offers an alternative to using water exchange. Biofloc uses the microorganisms already present in the culture water to purge the harmful substances that shrimp and fish emit. By doing this, the issue of excessive water use and effluent outflow is resolved. However, given that aquatic creatures only recover a small portion of the nutrients that are fed to them, there are two significant problems with this technology:

- The build up of dissolved, often non-toxic inorganic chemicals (like nitrate) and suspended organic materials.
- This is a waste of resources that the cultured animals could have used to grow.

Multitrophic Integrated Aquaculture (IMTA)

When regarded in the context of the Integrated Multi-trophic Aquaculture (IMTA) strategy, the accumulation of unrecovered nutrients that occurs in BFT in the form of suspended organic debris and inorganic dissolved substances can also be seen as an opportunity. This strategy involves using the leftovers from one species' culture as a source of nutrients for different species.

Different species are cultivated together in the same area to create a more balanced and sustainable ecosystem.

- Seaweed can be cultivated together with herbivorous fish such as tilapia
- Tilapia waste-Nutrient for salmon fish
- Salmon waste- Nutrients for Seaweed.
- Minimizes waste and maximizes the use of resources.

IMTA promotes a more balanced eco system, lowering the likelihood of disease outbreaks and other environmental issues. IMTA is a potential aquaculture strategy that can help to reduce the industry's environmental effect while enhancing efficiency and diversifying farmers' income. IMTA promotes a more balanced and sustainable ecology, making it an important tool for satisfying the increased demand for seafood while remaining environmentally responsible.

Underwater Optical Image Processing: a Comprehensive Review

The purpose of this study is to provide an overview of several underwater image-processing procedures, such as de-scattering, color restoration, and quality assessments. Finally, they summarize the future trends and challenges in underwater image sensor design and processing.

Underwater Optical-Sonar Image Fusion Systems:

Unmanned underwater operations employing remotely operated vehicles or unmanned surface vehicles have grown in popularity in recent years, ensuring human safety and labor efficiency. The improved optical and sonar pictures were fused with calibrated transformation matrices, and the underwater image quality measure (UIQM) and underwater color image quality evaluation (UCIQE) were employed as metrics to assess the proposed system's performance. The fused image in sonar image coordinates demonstrated qualitatively strong spatial agreement, with the average IoU between optical and sonar pixels in the fused images being 75%. The optical-sonar fusion technology will aid in the visualization and comprehension of underwater environments by providing color and distance information for unmanned operations.

A System for Autonomous Seaweed Farm Inspection with an Underwater Robot

The study describes a system and techniques for operating an AUV in a seaweed farm, including initial farm localization based on a prior estimate and dead-reckoning navigation, as well as ongoing farm scanning. Critical data from side scan sonars is acquired

from real surroundings at a test location in the ocean for algorithm development, and the results are demonstrated in a simulated seaweed farm configuration.

A sustainable integrated system for culture of fish, seaweed and abalone

Using an integrated design, a 3.3 m2 experimental system for intensive land-based growth of abalone, seaweed, and fish was constructed. The objectives were to achieve nutrient recycling, low water consumption, low nutrient outflow, and high yields.

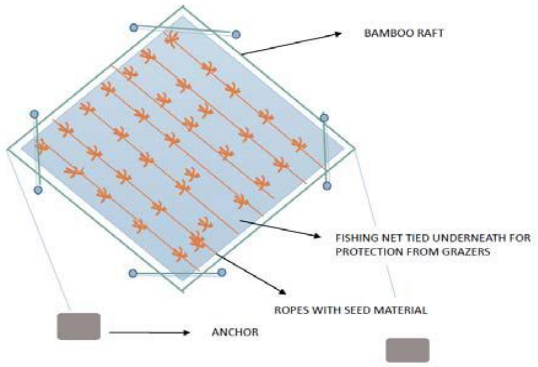


Figure 1: Bamboo raft method

A cobia cage can be connected with a total of 16 bamboo rafts, each measuring 12 feet by 12 feet and weighing 75 kg of seaweed. In a 45-day crop, the seaweed rafts that were combined with cobia cage produced an average output of 260 kg per raft as opposed to 150 kg for the unintegrated rafts. The use of cobia cage farming led to an increase of 110 kg of seaweed/raft. Additionally, the number of freshly emerged apical portions or tips in a bunch of seaweed obtained from the rafts that were integrated with the cobia cages was higher (on average 90–100 nos.,) whereas the same was lower (on average 30-40 nos.) from the rafts that were not integrated.

Figure 2: Integrating seaweed culture into shrimp and fish farming systems

An example of a multi-trophic ecosystem sharing an environment and utilizing the organic and inorganic nutrients made available by numerous organisms is shown by an integrated multi-trophic aquaculture operation in the above Figure 2.

Still, substantial and semi-intensive systems that are somewhat sustainable are where the majority of aquaculture production is produced. Concerns about the potential environmental, economic, and social effects of these frequently monospecific practices are however linked to the rapid global development of intensive marine-fed aquaculture of carnivorous finfish and shrimp, as well as to a lesser extent some shellfish aquaculture, particularly where activities are highly geographically concentrated or located in suboptimal sites whose assimilative capacities may be poorly understood and, consequently, lead to unintended consequences.

Figure 3:Shrimp farming process

The aquaculture industry needs to develop more cutting-edge, ethical, profitable, sustainable techniques that are also ecologically effective, environmentally benign, product-diversified, and societally advantageous if it is to expand. Due to the rising knowledge of picky customers regarding quality, traceability, and working conditions, maintaining sustainability has emerged as a crucial issue from not only an environmental standpoint but also from an economic, social, and technical one.

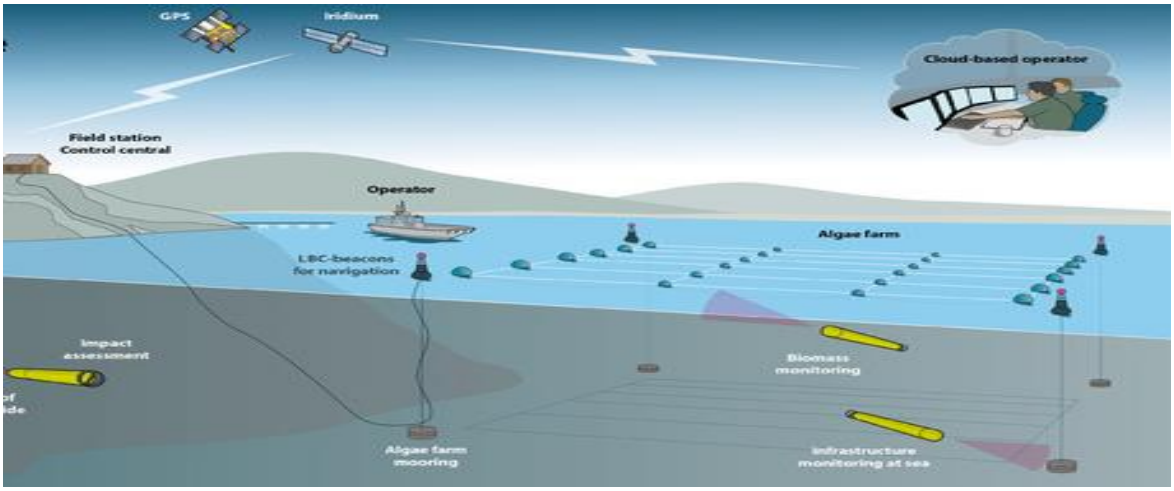


Figure 4: Monitoring System with robot

Numerous jobs that required robotic assistance have been completed by lone agents. There is more emphasis on merging numerous agents in robotic organizations in the maritime space as well, while some jobs are too complex for single agents to handle alone. This may enable these agents to perform a wider variety of activities, with the potential to develop effective systems for mapping and monitoring the maritime environment. Efficiency can be increased through integrated operations with robotic organizations, as demonstrated by Nilssen et al., even in the marine environment. Later, a large number of researchers joined the call for marine robotic organizations as a means of advancement that can lower prices and raise the complexity of jobs that can be solved.

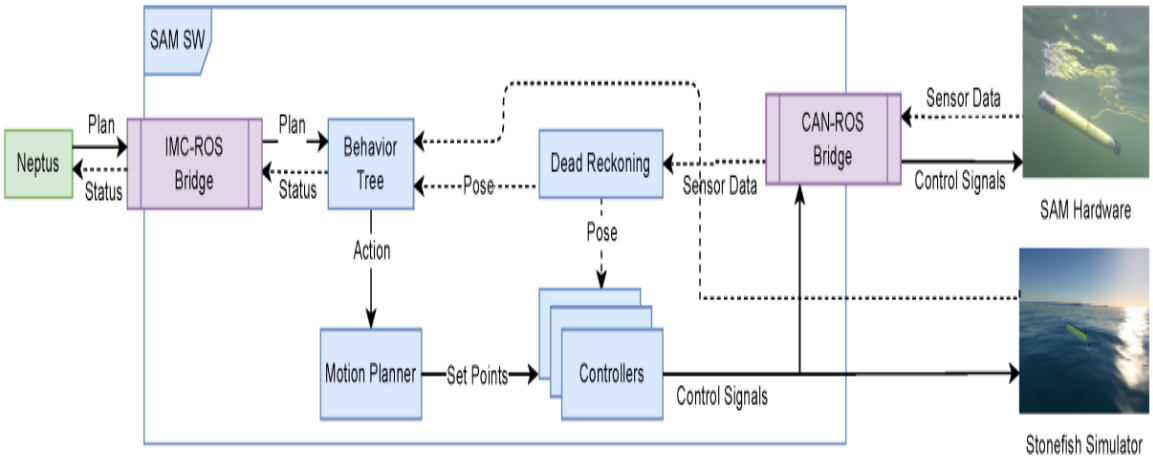


Figure 5:System architecture for shrimp farming

Superior Sensors

HUGIN Superior carries the most sensors of any commercially available AUV including the components such as HISAS 1032 Dual receiver, EM2040 Mk2, EdgeTech SBP, Cathx Ocean Camera & laser, OFG Magnetometer, Methene sensor, CTD, ADCP.

Autonomous Underwater Vehicles (AUV)

AUVs are platforms that can operate both below and above the surface and carry sensors. These agents are the vehicle itself, a control system, and maybe additional types of navigation as well as aural navigation. There are launch and recovery mechanisms for larger AUVs as well. The vehicles are available in a variety of shapes and configurations, each with a unique depth and thrust rating as well as payload sensors. There are many different types of sensors that can be used for navigation, but some of the more popular ones include cameras, GNSS, underwater acoustic location, compass, INS, and Doppler velocity log (DVL).

Around the world, the aquaculture business has been expanding quickly, with the result that waste water sources are having a substantial negative influence on the environment. As a result of this issue, numerous novel aquaculture techniques are being employed all over the nation to recycle water while reducing the amount of waste water discharged into the environment and resolving the issue of nitrogen buildup in the ecosystem. A model of aquaculture combined with seaweed or algae can be evaluated as having potential in the field of aquaculture.

From an ecological perspective, diversification also entails cultivating organisms from different and lower trophic levels, such as seaweeds, shellfish, crustaceans, echinoderms, worms, and bacteria selected based on their established or potential commercial value. This is different from simply raising different species of fish in polyculture. This strategy resembles ecosystems in nature. It is crucial to understand that there are few ways to remove nutrients from coastal ecosystems after they have done so. One of the few practical and affordable possibilities is the employment of extractive species. Therefore, when evaluating IMTA components, the economic values of extractive species' environmental services should be taken into account.

Conclusion

Technologies for seaweed aquaculture have advanced significantly over the past few decades, yet there are still obstacles to be solved. The most pressing difficulty is creating new strains using sophisticated breeding techniques. The growers will be able to extend the growing season and increase productivity thanks to superior strains. The creation of thermotolerant strains might be required in light of the problems posed by the global climate. The development of strains with disease resistance, rapid growth, high concentrations of desirable chemicals, and a decrease in fouling organisms is also necessary. It is crucial to create farm systems with advanced cultivation methods that are more reliable and economical. By transforming leftovers and uneaten feed from fed organisms into harvestable crops, IMTA promotes economic and environmental sustainability while minimizing eutrophication and enhancing economic diversification. Adopting integrated polytrophic methods should increase the aquaculture industry's environmental, economic, and social acceptability, allowing it to become a full and sustainable partner in the creation of integrated coastal management frameworks. To estimate AUV states in between position fixes, a state

estimator can also be used. A Kalman filter is one illustration of such an estimator. This might help to enhance the effectiveness of ASV tracking.

References

[1] Conchur Hughes, Jonathan W. King. Habitat suitability modelling for an integrated multi-trophic aquaculture (IMTA) system along Europe's Atlantic coast. Ecological Modelling Volume 484, October 2023, 110459.

[2]Ackefors, H. & Enell, M. 1994. The release of nutrients and organic matter from aquaculture systems in Nordic countries. *J. Appl. Ichthyol*. **10**: 225–41.

[3] Coastal Zone Canada Association 2001. *Beyond 2000: An Agenda for Integrated Coastal Management Development*. Coastal Zone Canada Association, Darmouth, 21 pp.

[4] D. Soto (Ed.), Integrated mariculture: a Global review. FAO Fisheries and Aquaculture Technical Paper. No. 529, FAO, Rome (2009), pp. 7-46

[5] Harrison, P. J. & Hurd, C. L. 2001. Nutrient physiology of seaweeds: application of concepts to aquaculture. *Cah. Biol. Mar*. **42**: 71–82.

[6] B.H. Buck, M.F. Troell, G. Krause, D.L. Angel, B. Grote, T. Chopin State of the art and challenges for offshore integrated multi-trophic aquaculture (IMTA) Front. Mar. Sci., 5 (2018), p. 165.

[7] Buschmann, A. H., López, D. A., Medina, A. 1996a. A review of the environmental effects and alternative production strategies of marine aquaculture in Chile. *Aquacult. Eng*. **15**: 397–421.

[8] Hopin, T., Yarish, C. & Sharp, G. 2001. Beyond the monospecific approach to animal aquaculture … the light of integrated aquaculture. *In* Bert, T. [Ed.]*Ecological and Genetic Implications of Aquaculture Activities*. Kluwer Academic Publishers, Dordrecht (in press).

[9] Chopin, T. Chopin Integrated Multi-Trophic Aquaculture Advancing the Aquaculture Agenda: Workshop Proceedings, OECD Publishing, Paris (2010)

[10] C. Collins, E. Bresnan, L. Brown, L. Falconer, J. Guilder, L. Jones, A. Kennerley, S. Malham, A. Murray, M. Stanley Impacts of climate change on aquaculture MCCIP Sci. Rev., (2020)

[11] Jang K. Kim, Charles Yarish3, Eun Kyoung Hwang, Miseon Park, Youngdae Kim Seaweed aquaculture: cultivation technologies, challenges and its ecosystem services. Algae 2017, 32(1): 1-13.

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Seaweeds and its Applications: A Review

By Dr. M. S Irfan Ahmed & A. Krishnaveni

Thassim Beevi Abdul Kader College For Women

Abstract- Kelp is one of the main living assets of the sea. Despite their wide applications in food and feed businesses, they have acquired significance as restorative sources due to their high mending, antimicrobial and antioxidative properties. As a rich wellspring of important compound parts, ocean growth is utilized in different businesses like beauty care products, Fuel, water treatment, and so on. Being a plant of remarkable construction and biochemical arrangement, ocean growth could be utilized profoundly for its multi-useful properties as food, energy, medication, and beauty care products. The dispersion, properties, and wide use of kelp are examined exhaustively in this paper.

Keywords: kelp, seaweeds, compound properties, living assets, business.

GJCST-D Classification: DDC Code: 338.1738 LCC Code: QK567

Strictly as per the compliance and regulations of:

Seaweeds and its Applications: A Review

Dr. M. S Irfan Ahmed [α] & A. Krishnaveni [σ]

Abstract- Kelp is one of the main living assets of the sea. Despite their wide applications in food and feed businesses, they have acquired significance as restorative sources due to their high mending, antimicrobial and antioxidative properties. As a rich wellspring of important compound parts, ocean growth is utilized in different businesses like beauty care products, Fuel, water treatment, and so on. Being a plant of remarkable construction and biochemical arrangement, ocean growth could be utilized profoundly for its multi-useful properties as food, energy, medication, and beauty care products. The dispersion, properties, and wide use of kelp are examined exhaustively in this paper.

Keywords: kelp, seaweeds, compound properties, living assets, business.

I. Introduction

There are many ways that seaweeds are like plants in that they are multicellular, they are photosynthetic, and they are important primary producers. But, the reason why seaweeds aren't plants is that they lack the vascular system of plants. They lack the structures of plants. The plants that everyone knows have to transport liquids internally and seaweeds just don't need to do that. Plants have different sides to their leaves. The top half is different from the bottom half and that's just not true for seaweeds. And the roots, while seaweeds have something that looks like roots, they function very differently. So seaweeds, are their own group. They are their own thing. Seaweeds are called marine algae or macroalgae, and this is to distinguish them from phytoplankton. They are eukaryotic and multicellular and have two important sections. There's the holdfast, which connects the seaweed to the seafloor in that area, and then the thallus, which is the part that is extended up above the holdfast. Now the thallus itself is made out of blades and stipes. So when a person looks at these different structures, the thallus is the main body of the seaweed, and the user has the holdfast, which again sort of look like roots, but behave very differently than roots. The whole purpose or goal of the holdfast is to simply hold the algae in place. It's not bringing anything up from the rocks below, it's not transporting liquid or moisture, things that roots of terrestrial plants would do. So it appears like a root but it doesn't function like a root. The leaf-like structures are called blades. Unlike earth plants, the blades of seaweed are not going to be different from one side that's the top, and to the other side, which is the bottom. These blades on these algae are just going to be swept back and forth by the seawater and so it's not known which side is going to be facing the sun to receive most of the solar input for photosynthesis. The stipe is going to be the body or the connecting portion of that algae. Now, the blades, the stipe, and in fact even the holdfast, all have chloroplasts, and they all have photosynthetic pigment and so really every surface of seaweed is able to photosynthesize. The primary seaweed is either in the haploid or the diploid generation [1-10].

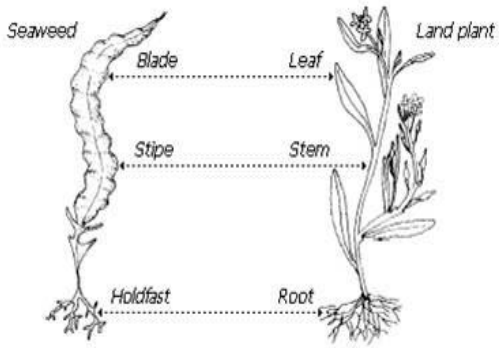


Figure 1: Seaweed structure general biology [9]

Photosynthesis: Sun+CO2+H2O= Nutrients for the seaweed +Oxygen for living persons.

Three Types of Seaweeds:

Overall there are 7000 assessed kelp, 4000 microalgae, and 50 seagrasses (Harbo, 1999). It falls into three general gatherings in view of pigmentation; brown, red, and green kelp [10].

1) Green seaweeds (Chlorophyta)

Typically radiant green in shading there is around 90 types of green kelp in the Pacific Northwest. They have sensitive-looking and clear-cutting edges going from wide to fine hair-like strands which commonly are not all the more than a cell or two thick. Green kelp is normally found in an assortment of territories going from new water streams and springs to saltwater tide pools. They can be open-minded toward a wide scope of water salinities and new water weakening [3].

Author α: Professor & Head, Director, Research-Industry and Institute Relations, Department of Computer Science, Thassim Beevi Abdul Kader College for Women Kilakarai, Ramanathapuram.
e-mail: directortbakc.rir@gmail.com
Author σ: Assistant professor, Department of Computer Science Thassim Beevi Abdul Kader College for Women Kilakarai, Ramanathapuram. e-mail: drakrishnaveni@gmail.com

Figure 2: Green Seaweed (Ulva) [9]

2) Red seaweeds (Rhodophyta)

Red Seaweeds can be in many tones including, yet not restricted to, dazzling red, pink, purple, dark red, dim, and dark. They are the most bountiful of the three divisions of kelp in our Pacific Northwest waters with around 265 species. They don't need as much light to develop as the green or earthy colored kelp so they are found from the intertidal regions to profundities of 25m. Red kelp is regularly collected for its carrageenan, which is added to food to give it a smooth, velvety surface. Porphyra is collected for nori utilized in Asian foods [4].

Figure 3: Red Seaweed (Mazzella) [9]

3) Brown seaweeds (Phaeophyta)

Typically gold, brown, olive green, or olive-brown in shading, there are around 140 types of earthy colored ocean growth in the Pacific Northwest. For the most part, they are the biggest and the quickest developing of the kelp however a couple of animal varieties stay little. Bull kelp, Nereocystis, can grow one to two feet each day, developing to be 100 feet in length in a season. Earthy colored kelp is found from the intertidal region to around 15m profound relying upon sun infiltration in the water. Earthy colored kelp (kelps) structure the submerged timberlands of the seas giving an environment to a great many species. They are critical nourishment for some spineless creatures, and kelps have been utilized by local societies for apparatuses and food [5].

Figure 4: Brown seaweed (Nereocystis) [9]

II. Literature Review

Alejandro H. Buschmann et al., [11] reviewed a few parts of kelp creation, for example, a report on the volumes of ocean growth delivered universally by both extractions from normal beds and development. They examine utilizes, creation patterns, and financial examination. Significantly, the authors give a layout to future necessities in the expectation that phycologists all over the planet will adapt to the situation, with the end goal that the possibility to be gotten from ocean growth biomass turns into a reality.

Meenakshisundaram Ganesan et al., [12] momentarily assessed the momentum status of Indian kelp assets and their use, as well as improvements in ocean growth cultivating innovations, the situation with the kelp industry, and ongoing endeavors to change ocean weed cultivating into a social undertaking. It likewise features the difficulties experienced in mainstreaming these assets in order to advance into the marine industry.

Georgia M. Ward et al., [13] surveyed sicknesses that have been accounted for in the logical writing for types of red and earthy colored kelp. The authors have zeroed in on the significant ocean growth crops filled in Asia, where a lot of this creation is trotted. They likewise give data on sickness the executives and biosecurity and a few perceptions on future bearings.

With future increments expected in the force, span, and frequencies of MHWs, Straub SC et al., [14] hope to see more substitutions of enormous seemingly perpetual living space framing kelp with more modest transient ocean growth, lessening the environment structure and successful administrations ocean growth overwhelmed reefs can give.

Kelp is a sustainable marine asset and stands out in the field of scientific classification in India when contrasted with their earthly partners, basically because of the absence of attention to their financial potential. In the current correspondence, Vaibhav A. Mantri wt al., [15] expected to address the situation with kelp biodiversity in India alongside the holes, difficulties, and open doors.

J. L. Banach et al., [16] study planned to survey the accessible information on the presence of food handling dangers in kelp, including factors impacting their presence, and to focus on the perils that might represent a gamble to human wellbeing.

Through creating quality descriptors for both steamed and dried results of the objective species, Annelise S. Chapman et al., [17] concentrate on establishing the framework for future deliberate tactile examinations. Primer tests uncovered a solid effect of animal groups on tangible insight, with P. palmata having a tactile profile unmistakable from the kelp species.

Nicholas Paul et al., [18] described, "Information is realizing that kelp contains fiber, minerals, protein, and omega-3 oils. Astuteness is realizing kelp are not high in these"? Presumably we can all imagine many models where: "kelp shrewdness" (ethno-phycology) is a higher priority than information; culture development rates and biomass productivities come into view.

Yogesh Kumar et al., [19] presented features of the nourishing and bioactive mixtures happening in various classes of ocean growth while zeroing in on their helpful exercises including but not restricted to platelet conglomeration, antiviral, antitumor, mitigating, and anticancer properties. Their survey likewise investigates the current and possible use of ocean growth as a wellspring of regular cancer prevention agents in food items.

The motivation behind Sami F.J et al., 's [20] study was to decide the bioactivity profile of ocean growth remove Eucheuma cottoni, Sargassum polycystum, and Caulerpa racemosa with the goal that it very well may be created as a restorative unrefined substance. The three sorts of kelp were removed by maceration and bioactivity testing was completed as a cancer prevention agent utilizing the ABTS revolutionary decrease strategy and bioactivity testing as UV-security with the boundaries of the rate transmission of erythema, pigmentation, and Sun Protective Factor (SPF).

Abdul Bakrudeen Ali Ahmed et al., [21] in the long run, developing kelp was a high worth and futile work. Then, at that point, after the states discovered that this thought is a cash making, low-worth, and high-volume item, they begin to put resources into this industry, and imports and products were a lucrative issue. While the ascent of fish utilization on the planet, the development of fish turned into a significant piece of each industry. Fish cultivating and fish creation rely upon freshwater or lake refinement. Kelp helps these societies and increment the yield and quality by creating sustenance in their current circumstance.

Berna Kılınc et al., [22] clarified today China, Japan, and the Republic of Korea are the biggest buyers of kelp as food. Nonetheless, as nationals from these nations have moved to different regions of the planet, the interest in ocean growth for food has followed them, as, for instance, in certain pieces of the United States of America and South America. Expanding requests in the course of the most recent fifty years exceeded the capacity to supply necessities from normal (wild) stocks. Investigation into the existing patterns of these kelp has prompted the advancement of development enterprises that presently produce in excess of 90% of the market's interest.

Cristina Garcia Sartal et al., [23] listed the utilization of alginate in the food business depends on its capacity to expand the thickness of watery arrangements and to frame gels that don't dissolve when warmed. Alginate is utilized as a stabilizer in frozen yogurt and other dairy items, in the brew to create foam, and in sodas. It is utilized as a thickener in beverages and cleansers, as an explaining specialist in wines and juice, and to expand the consistency of organic product juices and mustard (Gonzalez et al., 1998).

Pranav Nakhate et al., [24] expressed ocean growth's ability and likely usefulness to help the bio-based economy are efficiently talked about. The conceivable bioprocessing plant draws near, alongside its natural and financial parts of supportability, which are additionally managed. At last, the formative interaction, side-effect advancement, monetary help, and social acknowledgment approach are summed up, which is fundamental while considering kelp-based items' possibility.

Eduarda M. Cabral et al., [25] expect to give an outline of the overall strategies and novel mixtures with antimicrobial properties as of late disconnected and portrayed from macroalgae, accentuating the sub-atomic pathways of their antimicrobial systems of activity. The current logical proof of the utilization of macroalgae or macroalgal concentrates to expand the timeframe of realistic usability of food varieties and forestall the advancement of foodborne microorganisms in genuine food items and their impact on the tangible properties of different food varieties (i.e., meat, dairy, refreshments, fish, and pastry shop items) will likewise be examined, along with the fundamental difficulties and future patterns of the utilization of marine normal items as antimicrobials.

Adriana Leandro et al., [26] focal point of their review is in the presentation of kelp as a likely other option and as a protected food source. Here portrayed are the health benefit and concerns connecting with ocean growth utilization, and furthermore how kelp inferred compounds are now financially investigated and accessible in the food business, and the use limitations to shield them as protected food added substances for human utilization.

Simone E. M. Olsthoorn et al., [27] deliberate survey sums up data on the effect of earthy colored kelp or parts on aggravation, and irritation-related

pathologies, for example, sensitivities, diabetes mellitus, and heftiness. They center on oral supplementation hence meaning the utilization of earthy colored ocean growth as food added substances.

Joao Cotas et al., [28] concentrate on features of the three polymers, alongside their known limits, at which they can have positive as well as bad wellbeing impacts. Such information is critical to perceiving the worldview administering their effective sending and related valuable applications in people.

Maria Eggertsen et al., [29] give a synopsis of the flow of logical information on potential immediate and roundabout negative ecological impacts connected to eucheumatoid ocean growth cultivating, for example, modifications of benthic macrophyte natural surroundings and loss of local biodiversity. Moreover, we feature information holes that are of significance to address sooner rather than later, e.g., enormous scope environment impacts and ranches as likely vectors of microorganisms. We additionally give various possible administration suggestions to be executed for the proceeded with the improvement of the earth's reasonable ocean growth by cultivating rehearses in the WIO area.

A biorefinery approach Gabriela S. Matos et al., [30] shows an important thought of tackling financial and natural disadvantages, empowering fewer deposits creation near the much suggested zero-squander framework. The point of this work is to report on the recently evolved techniques for ocean growth extractions and the likely utilization of the parts extricated.

Kelp's recorded use in everyday food diet, partnered to explore discoveries, showed that macroalgae are a wellspring of supplements and bioactive mixtures with nutraceutical properties. The principle objective of Diana Pacheco et al., [31] review is to assess the records of NIS ocean growth in the Iberian Peninsula and basically investigate the capability of intrusive kelp application in the food business.

Aroa Lopez-Santamarina et al., [32] review offers a true point of view of the ebb and flows information encompassing the effects of kelp and their inferred polysaccharides on the human microbiome and the significant requirement for additional inside and out examinations concerning this subject. Creature tests and in vitro colonic-mimicking preliminaries researching the impacts of kelp ingestion on human stomach microbiota are talked about.

Ghislain Moussavou et al., [33] review centers around colorectal and bosom diseases, which are significant reasons for malignant growth-related mortality in people. It likewise depicts different mixtures separated from a scope of kelp that has been displayed to annihilate or slow the movement of malignant growth. Fucoidan removed from the earthy colored green growth Fucus has shown movement against both colorectal and bosom diseases. Moreover, we audit the instruments through which these mixtures can instigate apoptosis in vitro and in vivo.

Bahare Salehi et al., [34] present audit gives the most exceptional bits of knowledge into ocean growth research, explicitly tending to its substance synthesis, phytopharmacology, and corrective applications. Kelp is an everyday term for perceptible, multicellular benthic marine green growth. Kelp is probably the biggest maker of biomass in the marine climate and establish a significant piece of the eating routine and conventional medication in many pieces of Asia since ancient times.

Diane Purcell-Meyerink et al., [35] foresee the ocean growth industry requires huge amounts and great kelp unrefined substance that applies tension on the current normal kelp assets. Hydroponics development of ocean growth has developed impressively starting around 2009 to fulfill item needs and to safeguard wild kelp beds that are as yet in danger from over-abuse and environmental change, which has caused expansions in seawater temperatures.

Bruno Moreira Leiteet al., [36] predicted Kelp are notable for their health benefit. As of now, youth weight is an overall developing general medical issue. Three of the central point that adds to this pandemic are unfortunate dietary patterns/inactive, the absence of data on the sustenance worth of food, and the accessibility of caloric thick food with poor healthful substance. In this unique circumstance, well-being advancement through dietary instruction for young kids is of significant importance. The target of this review was to assess the effect of instruction put together mediation with respect to food naming in offspring of five distinct schools in the district of Lisbon.

III. Applications and Methods of Seaweed Cultivation

All through mankind's set of experiences, kelp has been utilized as food, people cures, colors, and mineral-rich composts. Anti-pathogenic activity by green seaweeds, seaweed chocolate, animal food, fertilizers, biotechnology, beauty care products, drug industry, and health benefits is accountable information. Many the women in coastal areas are benefited from seaweed-based cultivation [37-41].The green growth innovation wipes out difficulties related to existing carbon catch techniques.

Figure 5: Bamboo raft method [42].

Figure 6: Bamboo raft method seaweeds for harvesting [42].

Figure 5 and Figure 6 show the bamboo raft method of seaweed cultivation.

IV. Conclusion

This paper presented seaweed applications and their cultivation methods. Minority people, economic growth depends on seaweed harvesting in coastal areas like Ramanathapuram district. Women empowerment, self-employment, development, and many foods, medicine, and cosmetic products are manufactured from seaweeds. In the future, seaweed image color detection and classification are planned to be introduced.

References Références Referencias

1. Abbott, I.A. & Hollenberg, G.J. 1976. Marine Algae of California. California: Stanford University Press.
2. Castro, P. 2008. Marine Biology, Seventh Edition. New York: McGraw-Hill Inc.
3. Druehl, L. 2000. Pacific Seaweeds: A Guide to Common Seaweeds of the West Coast. BC Canada: Harbour Publishing.
4. Department of Biology Penn State University. April, 8th, 2008. http://courses.bio.psu.edu/fall2005/biol-110/tutorials/tutorial30.htm
5. Friends of the San Juans. April 8th, 2008. http://www.sanjuans.org/EelgrassHealth.htm
6. Harbo, R.M.1999. Whelks to Whales: Coastal Marine Life of the Pacific Northwest. BC Canada: Harbour Publishing.
7. Mattila, J., G. Chaplin, M. Eilers, K. Heck, J. O'Neal and J. Valentine. 1999. Spatial and diurnal distribution of invertebrate and fish fauna of a *Zostera marina* bed and nearby unvegetated sediments in Damariscotta River, Maine (USA). Journal of Sea Research. 41:321-32.
8. Sept, J.D. 1999. The Beachcombers Guide to Seashore Life in the Pacific Northwest. BC Canada: Harbour Publishing.
9. Shaw, L. May 2008. All photos not otherwise credited.
10. Waaland, R. 1977. Common Seaweeds: of the Pacific Coast. Washington: Pacific Search Press.
11. Alejandro H. Buschmann, Carolina Camus, Javier Infante, Amir Neoric, Alvaro Israele, María C. Hernández-González, Sandra V. Pereda, Juan Luis Gomez-Pinchetti, Alexander Golberg, Niva Tadmor-Shalev, Alan T. Critchley, "Seaweed production: overview of the global state of exploitation, farming and emerging research activity", European Journal Of Phycology, (2017) VOL. 52, NO. 4, 391–406, Taylor and Francis https://doi.org/10.1080/09670262.2017.1365175
12. Meenakshisundaram Ganesan, Nitin Trivedi, Vishal Gupta, S. Venu Madhav, Chennur Radhakrishna Reddy, Ira A. Levine, "Seaweed resources in India – current status of diversity and cultivation: prospects and challenges", Botanica Marina (2019) 62(5): 463–482 https://doi.org/10.1515/bot-2018-0056
13. Georgia M. Ward, Joseph P. Faisan Jr, Elizabeth J. Cottier-Cook, Claire Gachon, Anicia Q. Hurtado, Phaik E. Lim, Ivy Matoju, Flower E. Msuya, David Bass, Juliet Brodie, "A review of reported seaweed diseases and pests in aquaculture in Asia", J World Aquacult Soc. (2020) 51:815–828. DOI: 10.1111/jwas.12649
14. Straub SC, Wernberg T, Thomsen MS, Moore PJ, Burrows M, Harvey BP, Smale DA, "Resistance to obliteration; responses of seaweeds to marine heatwaves", Front. Mar. Sci. (2019) 6:763.doi:10.3389/fmars.2019.00763
15. Vaibhav A. Mantri, Monica Gajanan Kavale, Mudassar Anisoddin Kazi, "Seaweed Biodiversity of India: Reviewing Current Knowledge to Identify Gaps, Challenges, and Opportunities", Diversity (2020) 12, 13; doi:10.3390/d12010013
16. J. L. Banach, E. F. Hoek-van den Hil, H. J. van der Fels-Klerx, "Food safety hazards in the European seaweed chain", Compr Rev Food Sci Food Saf. (2020) 19:332–364. DOI: 10.1111/1541-4337.12523
17. Annelise S. Chapman, Pierrick Stévant, Wenche Emblem Larssen, "Food or fad? Challenges and opportunities for including seaweeds in a Nordic diet", Botanica Marina (2015) 58(6): 423–433DOI 10.1515/bot-2015-0044
18. Nicholas Paul, Anicia Hurtado, Leonel Pereira, Melania Cornish, Danilo Largo, Alan Critchley,

"Seaweed resources of the world: a 2020 vision. Part 2", Botanica Marina (2019) 62(5): 391–393 https://doi.org/10.1515/bot-2019-0061

19. Yogesh Kumar, Ayon Tarafdar, Prarabdh C. Badgujar, "Seaweed as a Source of Natural Antioxidants: Therapeutic Activity and Food Applications", Hindawi Journal of Food Quality Volume (2021) Article ID 5753391, 17 pages https://doi.org/10.1155/2021/5753391
20. Sami F.J, Soekamto N.H, Firdaus, Latip J, "Bioactivity profile of three types of seaweed as an antioxidant, UV-protection as sunscreen and their correlation activity", Food Research 5 (1): 441 - 447 February(2021) DOI:https://doi.org/10.26656/fr.2017.5(1).389
21. Abdul Bakrudeen Ali Ahmed, Mohaddeseh Adel, Ali Talati, Muthusamy S. Kumar, Kamaludeen Abdulrahim, Meera M. Abdulhameed, "Seaweed Polysaccharides and Their Production and Applications", Seaweed Polysaccharides. http://dx.doi.org/10.1016/B978-0-12-809816-5.00020-7 (2017) Elsevier Inc.
22. Berna Kılınç, Semra Cirik, Gamze Turan, Hatice Tekogul and Edis Koru, "Seaweeds for Food and Industrial Applications" InTech. (2017) 735-748 http://dx.doi.org/10.5772/53172
23. Cristina Garcia Sartal, Maria Carmen Barciela Alonso, Pilar Bermejo Barrera, "Application of Seaweeds in the Food Industry", Handbook of Marine Macroalgae: Biotechnology and Applied Phycology, First Edition. Edited by Se-Kwon Kim. (2012) 523-534 John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd.
24. Pranav Nakhate, Yvonne van der Meer, "A Systematic Review on Seaweed Functionality: A Sustainable Bio-Based Material", Sustainability (2021), 13, 6174. https://doi.org/10.3390/su13116174
25. Eduarda M. Cabral, Márcia Oliveira, Julie R. M. Mondala, James Curtin, Brijesh K. Tiwari, Marco Garcia-Vaquero, "Antimicrobials from Seaweeds for Food Applications", Mar. Drugs (2021), 19, 211. https://doi.org/10.3390/md19040211
26. Adriana Leandro, Diana Pacheco, Joao Cotas, Joao C. Marques, Leonel Pereira, Ana M. M. Gonçalves, "Seaweed's Bioactive Candidate Compounds to Food Industry and Global Food Security", Life (2020), 10, 140; doi:10.3390/life10080140
27. Simone E. M. Olsthoorn, Xi Wang, Berend Tillema, Tim Vanmierlo, Stefan Kraan, Pieter J. M. Leenen, Monique T. Mulder, "Brown Seaweed Food Supplementation: Effects on Allergy and Inflammation and Its Consequences" Nutrients (2021), 13, 2613. https://doi.org/10.3390/nu13082613
28. Joao Cotas, Diana Pacheco, Glacio Souza Araujo, Ana Valado, Alan T. Critchley, Leonel Pereira, "On the Health Benefits vs. Risks of Seaweeds and Their Constituents: The Curious Case of the Polymer Paradigm", Mar. Drugs (2021), 19, 164. https://doi.org/10.3390/md19030164
29. Maria Eggertsen, Christina Halling, "Knowledge gaps and management recommendations for future paths of sustainable seaweed farming in the Western Indian Ocean", Ambio (2021), 50:60–73 Springer https://doi.org/10.1007/s13280-020-01319-7
30. Gabriela S. Matos, Sara G. Pereira, Zlatina A. Genisheva, Ana Maria Gomes, Jose A. Teixeira, Cristina M. R. Rocha, "Advances in Extraction Methods to Recover Added-Value Compounds from Seaweeds: Sustainability and Functionality", Foods (2021),10,516. https://doi.org/10.3390/foods10030516
31. Diana Pacheco, Glacio Souza Araujo, Joao Cotas, Rui Gaspar, Joao M. Neto, Leonel Pereira, "Invasive Seaweeds in the Iberian Peninsula: A Contribution for Food Supply", Mar. Drugs (2020), 18, 560; doi:10.3390/md18110560
32. Aroa Lopez-Santamarina, Jose Manuel Miranda, Alicia del Carmen Mondragon, Alexandre Lamas , Alejandra Cardelle-Cobas, Carlos Manuel Franco and Alberto Cepeda, "Potential Use of Marine Seaweeds as Prebiotics: A Review", Molecules (2020), 25, 1004; doi:10.3390/molecules25041004
33. Ghislain Moussavou, Dong Hoon Kwak, Brice Wilfried Obiang-Obonou, Cyr Abel Ogandaga Maranguy, Sylvatrie-Danne Dinzouna-Boutamba, Dae Hoon Lee, Ordelia Gwenaelle Manvoudou Pissibanganga, Kisung Ko, Jae In Seo, Young Kug Choo, "Anticancer Effects of Different Seaweeds on Human Colon and Breast Cancers", Mar. Drugs (2014), 12, 4898-4911; doi:10.3390/md12094898
34. Bahare Salehi, Javad Sharifi-Rad, Ana M. L. Seca, Diana C. G. A. Pinto, Izabela Michalak, Antonio Trincone, Abhay Prakash Mishra, Manisha Nigam, Wissam Zam, Natália Martins, "Current Trends on Seaweeds: Looking at Chemical Composition, Phytopharmacology, and Cosmetic Applications", Molecules,(2019),24,4182; doi:10.3390/molecules24224182
35. Diane Purcell-Meyerink, Michael A. Packer, Thomas T. Wheeler, Maria Hayes, "Aquaculture Production of the Brown Seaweeds Laminaria digitata and Macrocystis pyrifera: Applications in Food and Pharmaceuticals", Molecules (2021), 26, 1306. https://doi.org/10.3390/molecules26051306
36. Bruno Moreira Leite, Bruno Campos, Paulina Mata, Joao Paulo Noronha & Mario Diniz, "Are seaweeds the food of the future? Challenges for its conservationand introduction in the Portuguese diet", Annals of Medicine, (2019) 51:sup1, 169-169, DOI: 10.1080/07853890.2018.1562018

37. Emer Shannon & Nissreen Abu-Ghannam (2019) Seaweeds as nutraceuticals for health and nutriation, Phycologia,58:5,563-577, DOI: 10.1080/00 3188 84.2019.1640533
38. Arumugam Maruthupandian, Panneerselvam Munirasu, Lakshmanan Archana, "Anti-Pathogenic Activity of Green Seaweeds (Chlorophyceae) From The Coast Of Ramanathapuram, Tamilnadu", International Journal of Current Innovation Research, Vol. 1, Issue 2, pp 58-62, April (2015)
39. A. Thahira Banu, S. Uma Mageswari, "Nutritional status and effect of seaweed chocolate on anemic adolescent girls", Food Science and Human Wellness4(2015)28 34 http://dx.doi.org/10.1016 /j.fshw.2015.03.001
40. Silvia Lomartire, Joao Carlos Marques, Ana M. M. Gonçalves, "An Overview to the Health Benefits of Seaweeds Consumption", Mar. Drugs 2021,19,341. https://doi.org/10.3390/md19060341
41. Radhika Rajasree SR, and S Gayathri, "Women Enterprising in Seaweed Farming With Special References Fisherwomen Widows in Kanyakumari District Tamilnadu India", J Coast Dev 2014, 17:1 DOI: 10.4172/1410-5217.1000383
42. B. Johnson, R. Narayanakumar, A. K. Abdul Nazar, P. Kaladharan, G. Gopakumar, "Economic analysis of farming and wild collection of seaweeds in Ramanathapuram District, Tamil Nadu", Indian J. Fish., 64(4): 94-99, 2017 94 DOI: 10.21077 /ijf.2017.64.4.61828-13

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Secure Online Transactions using Blockchain

Ms. K Annsheela
Assistant Professor,
Department of Computer College
Thassim Beevi Abdul Kader College for Women

ABSTRACT:

Blockchain is the most commonly used distributed database or ledger for digital currency exchanges and secure transactions. Each network participant has access to a ledger that is updated with each new transaction. The blockchain ledger is a collection of all transactions made in the past. A blockchain ledger is an ever-evolving tamper-proof data structure that contains blocks containing individual batches of transactions. This work will help understand the various security algorithms used for electronic transactions and payments. Hashing methodology is used to secure the transaction in Blockchain. Blockchain technology and its impact on companies and industries. The blockchain supports a decentralized, invariant, consistent, and secure hashing algorithm in which Proof-of-Wok is used. This systematic study helps you understand how is secure a transaction.

I. Introduction:

Advances in digitalization have revolutionized people's lives. As technology advances, we use internet-based banks, and various e-commerce payment systems are growing when shopping. It is designed to increase, improve and deliver secure electronic payment transactions. People are also interested in electronic payment systems because they reduce paperwork and transaction costs.

The system is easy for business people to use and takes less time than manual processing and latency to help businesses expand their reach to the market. The ePayment system is a method of processing or paying goods and invoices via electronic media without using cash or checks. This is called an online payment system. E-Payment uses electronic devices and various types of trading cards such as customers, productions, financial companies, etc. as a medium of computer technology to issue payment instructions directly or indirectly via a network to make payments and remittances. , Refers to the parties to electronic transactions.

II.Literature Review

Prince will Aigbe et, al[3],2014, In their study, a complete review of the different categories of online payment systems in terms of online payment processes, authentication mechanisms, and authentication types analysis

reveals that online payment systems with authentication mechanisms involving two or more authentication factors such as token encryption, digital signature, PIN tend to be more secured, reduced fraud vulnerability, and boost users' confidence in using electronic payment systems. Their work reveals that e-payment systems with authentication mechanisms involving two or more authentication factors have to be secured and reduced fraud and boost users' confidence e - payment systems

Satoshi Nakamoto et.al [4] 2016**.**Mentioned their work Online Payments or transactions were directly sent from one party to another without going through a financial institution that undergoes peer-to-peer communication. Digital signatures play a role in protection at a limit. The proposed system uses the verification of data and secures the transmission of money through bank validation.

NiturkarPallaviPravinet.al[5],2020 mentioned Blockchain technology for protecting the banking transaction without using tokens. Blockchain technology is a distributed system that works on total verification and validation of data without consideration of Miners or Tokens. Eliminating the use of miners or tokens may lead to creating a transparent and load-free network, which increase in survivability of transaction. By adopting blockchain in the distribution of databases on banking systems one can reduce attacks on the system. Blockchain without tokens plays an avital role in building a system that is more reliable for banking to perform transactions that have to be secure at a very high level.

Joseph Gualdoniet.al[4],2017, mentioned the Secure Online Transaction Algorithm: Securing Online Transaction Using Two-Factor Authentication. The Secure Online Transaction Algorithm (SOTA) would benefit not only the account holders but also the credit card companies. Our model works by using two-factor authentications and a random code that the consumer would be generated and supplied on an application. The Secure Online Transaction Algorithm creates a new level of security that has not been implemented into credit cards yet. This algorithm could instrumental in reducing the number of fraudulent purchases made with stolen credit cards.

III.Blockchain

A blockchain is a distributed database shared between nodes in a computer network. As a database, blockchain stores information electronically in digital format. Blockchain is best known for its important role in cryptocurrency systems like Bitcoin for maintainingsecure and decentralized records of transactions. Blockchain innovationis to ensure the fidelity and security of data records and to create trust without the need for a trusted third party. The main difference between a typical database andblockchain isthe structure of the data. Blockchain collects informationin groups called blocks that

contain a set of information. A block has a certain amount of storage capacity, is closed when it is full and is linked to a previously full block to create a chain of data called a blockchain. The new information that follows this newly added block is compiled into the newly formed block and added to the chain when it is full.

A hash pointer is a pointer that has an encrypted hash of the data it points to. If you are new to hash functions, think of a hash of your data as a short code that represents your data. We cannot undo the original data from this code, but we can verify that the new data matches the original data by hashing the new data and comparing this new code with the stored code. Each node in the linked list is called a block, stores some data, and references the previous block in the blockchain. The first block on the blockchain is called a genesis block because it does not point to any other block. The latest block added is called Head

Blockchain is the technical backbone of "Bitcoin" and was introduced in 2009 as a secure and fast connection technology that enables exchange. In the blockchain, each user transaction creates a block and adds it to the chain of transactions. This creates a trail of interconnected blocks. Each block is sent and added to your ledger. Application of blockchain technology in payment of Blockchain functions corresponds to the requirements of the payment base.

- **Safe.**
- **Processing speed**
- **Traceability**
- **Global Registry (Ledger)**

The distributed ledger makes it possible to connect all the parties to real-time financial transactions for quick processing while maintaining the audit trail. Processing is distributed over the network, it is almost impossible to change or manipulate data, or manipulate fraud prevention and safety stops. Blockchain, we can ship money without using banks. Describes the simplicity and elegance of such systems and understands the process of creating ourown blockchain in the Python programming language to explain nuances. Working from a blockchain:

Block or ledger and this updated block are displayed to wait for the verification node at a specific time interval. Block Verification: When an Inner work node receives an updated block validation request, the node searches other nodes in the network in network to process to repetitive process to check the block. Blockchain: If all transactions are approved in the block, the new block is connected to the current block and closes the latest status of the block into the remaining blocks on the network.

IV.Methodology:

SSL:

SSL, or Secure Sockets Layer (SSL), is the most widely used security measure to secure Internet communications. SSL provides a secure communication layer between computers on a network.SSL supports security protocols such as encryption, and authentication, and ensures that requested and transmitted data is actually forwarded to the SSL provides security with 128-bit encryption, so sensitive information transmitted over the Internet during online transactions remains private.

SET:

The SET is a system that ensures the security and integrity of scripted electronic transactions made with credit cards. Secures online credit card payments using a variety of encryption and hashing methods. The SET protocol supports development by major organizations such as Visa, MasterCard, Microsoft providing Secure Transaction Technology (STT), and Netscape providing Secure Sockets Layer (SSL) technology.

Hashing in Proof-of-Work Consensus Algorithm:

Blockchain is an ever-evolving ledger that continues an everlasting report of all of the transactions which have taken place, in a stable, chronological, and immutable manner. To stable information, blockchain makes use of a hash characteristic. The hash characteristic is one of the maximum broadly used cryptographic algorithms in blockchain technology. These are cryptographic (however now no longer encryption) algorithms designed to defend the integrity of information. In a nutshell, a hash set of rules is a mathematical characteristic that turns any enter into an output of a set length. To be cryptographically stable - and usable in blockchain technology - a hash should be collision-resistant, this means that it`s very hard to locate inputs that produce the identical output.

Types of cryptographic hash features:

- **Secure Hash Algorithms (SHA2 and SHA3)**
- **RACE primitive integrity evaluation message summary (RIPEMD)**
- **Message Algorithm 5 (MD5)**
- **BLAKE2**

To do this, a hash characteristic should have the subsequent properties:

• **One-manner**: It is feasible to move from entering to output in a hash characteristic, however now no longer vice versa. This makes it not possible to opposite engineer a collision from the preferred hash output

• **Large output area:** The handiest manner to locate hash collisions is thru brute pressure seek. According to the Pigeon Cage principle, this calls for checking as many inputs because the hash has feasible outputs. This quantity

should be huge and sufficient to make a brute-pressure seek not possible

• A hash set of rules is taken into consideration stable tilla collision may be located for it. Once this happens, it`s formally obsolete, much like MD5 and SHA1. Using hash features inside the blockchain Hash features are frequently used to defend the integrity of information. With a dependable information hash, it's far feasible to compute the hash of the information and evaluate values. If they match, the information won't have been modified for the reason that a unique hash turned into generated. The blockchain`s virtual ledger is designed to keep treasured records that might advantage an attacker if modified in their favor. Furthermore, this ledger is saved and transmitted with the aid of using a community of nodes that might be suspicious of every other.

As a result, blockchains have some exclusive makes use of for the hash features and integrity safety they provide. Some of the maximum not unusual place that makes use of hash features in blockchain include:

Uses of Hash:

• **Digital signature**: Hash features are a vital part of virtual signature algorithms, summarizing information right into a compact price even as retaining its integrity. Digital signatures are used to hold information integrity and authenticate transactions and blockchains

• **Merkle Tree**: The Merkle tree summarizes the listing of transactions contained in a block right into inside the block header. They use hash features to make certain that no Merkle timber with the identical root hash can't be located. In this manner, with the aid of using storing the unique hash inside the block header and retaining the integrity of the block header, the integrity of the transactions contained inside the block frame is likewise protected.

• **Proof of Work consensus**: The Proof of Work consensus set of rules determines that a legitimate block is a block whose header has a hash price under a positive threshold. The collision resistance of the hash characteristic is vital right here because it guarantees that it's far very hard to discover a legitimate block.

Hash function security for blockchain

The safety of hash features for hash features inside the chain is vital to defend the immutability of the virtual ledger. If the hash utilized by a blockchain is broken, an attacker can locate conflicts for crucial hashes (consisting of the blockchain chain or the Merkle tree price). This will make it simpler for malicious nodes to rewrite the blockchain community records and motivate the blockchain gadget to crash. For this reason, the safety of hash features is vital to blockchain safety. The safety of a hash characteristic may be threatened in of the ways:

• **Algorithm strength**: Hash features are designed to face up to collisions, however cryptographic algorithms occasionally break.

If the vulnerability is observed in a hash characteristic, this may correctly locate collisions inside the hash characteristic.

• **Hash output period**: Hash features are designed to make the nice manner to discover a collision is a brute pressure seek, with the hunt area being the identical length as the gap of feasible has outputs. Might also additionally have. If the sort of area does emerge as searchable - because of using a hash whose output period is just too short - then the hash is now no longer collision-resistant and liable to attack.

Conclusion:

Online payment transactions use blockchain, the purpose of which is to ensure the security of the entire process. It uses a one-way hashing algorithm and Proof-of-Work in a consensus Algorithm to securely transmit data to miners. Miners also use a proof-of-work algorithm to verify transactions using the submitted hash value. Therefore, this paper aims to provide a secure process for online transactions by overcoming attacks such as man in the middle attack and eliminating third-party ports that speed up the whole process of the Hash function using proof-of-Work. Compared to Previous algorithms like SSL, and SET, the Proof-of-Work and hashing algorithms using blockchain got the better results. In the future optimizing the hash value in blockchain and increasing the security of transactions using blockchain algorithms

References:

(1)What is Blockchain Technology? A Step-by-Step Guide For Beginners. Available Online: https://blockgeeks.com/guides/ is-blockchain-technology/

(2)Nakamoto, Satoshi. ”Bitcoin: A peer-to-peer electronic cash system.” (2008). Available Online: https://bitcoin.org/bitcoin

(3) ParitoshBanchhor, DurgeshSahu, Ankit Mishra, Mohammed BakhtawarAhmed,”A Systematic Review on Blockchain Security Attacks, Challenges, and Issues,(2021)

(4)KarthikeyaThanapal, DhirajMehta, KarthikMudaliar, and BushraShaikh,” Online Payment Using Blockchain(2020)

(5)Fan Yang, Wei Zhou, Qingqing Wu, Rui Long, Neal N. Xiong, (Senior Member, IEEE), And Meiqi Zhou, Delegated Proof of Stake With Downgrade: A Secure and Efficient Blockchain Consensus Algorithm With Downgrade Mechanism,IEEE access, august 2019.

(6)Applicationofblockchain,Availableonline(https://www.businessinsider.in/finance/news/the-growing-list-of-applications-and-use-cases-of-blockchain-technology-in-business-and-life/articleshow/74447275.cms) (accessed on march 2020)

(7)Prasanth Varma Kakarlapudi and Qusay H. Mahmoud, A Systematic Review of Blockchain for Consent Management, in mdpi journal, Healthcare 2021, 9, 137.

(8)Masihuddin, M.; Islam Khan, B.U.; Islam Mattoo, M.M.U.; Olanrewaju, R.F. A Survey on E-Payment Systems: Elements, Adoption, Architecture, Challenges, and Security Concepts. Indian J. Sci. Technol. 2017, 10, 1–19. [CrossRef]

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Optimized Weather Forecasting using Enhanced Multilayer Perceptron for Seaweed Cultivation

Ms.M.Nagavalli
Assistant Professor, Department of Computer Science,
Thassim Beevi Abdul Kader College for Women, Kilakarai
nagavallim1015@gmail.com

ABSTRACT

Seaweed is considered a high-value source of protein for humans and cattle. In this work proposed classification model used to predict the production of seaweed K.alvarezii cultivation based on seasons by weather forecasting using Enhanced MultiLayer Perceptron (MLP) Neural Network. In this MLP neural network model was built to evaluate future predictions based on attributes that are outlook, temperature, and humidity from end of training data. The dataset was collected from Kaggle for classification and classified using Enhanced Multilayer Perceptron is implemented and its performance are evaluated. The MLP function has time taken to test a model on training data 0.42 seconds, enhanced MLP gives 0.09 seconds by increasing number of neurons in hidden layer.

Keywords: weather forecasting, seaweed cultivation, neural network, ANN, Multilayer Perceptron.

1. INTRODUCTION

Inspite of one of the most developing and industrialised states, Tamil Nadu is facing a climate changing problems of coastal erosion, severe weather events and rising temperatures manifesting more recently. As per the 2019 draft action plan on climate change, the State government planned to carry out 199 projects in seven key sectors by the year 2030. The proposed budget for executing this plan was a whopping Rs 3.24 lakh crore. Rising sea levels, hotter temperatures and stronger currents along this coast, considered one of the best for commercial seaweed cultivation. Scientists say they are caused by climate change. “With the rise in sea temperature and salinity, seaweed growth has declined in the last decade,” said K Eswaran, a scientist who heads the field research unit of the Central Salt and Marine Chemicals Research Institute in Ramanathapuram district. “Women who harvest seaweed have definitely been impacted, with their incomes coming down by at least 20%.”Seaweed is leading commodity in aquaculture in Tamilnadu. Acclimatization and large scale cultivation of this alga was achieved on the coast of mandapam, south east coast of India, during 1995 to 1997. The development of seaweed cultivation areas can be caused by the biophysical environmental conditions of the waters and

climatic conditions. Planting season is one of the important factors in seaweed farming because it greatly influences the production or the end result produced. The response of seaweed growth varies over time and seasons of the year. A cultivation approach based on change the season and the environmental conditions of its waters optimal for seaweed growth. Seaweed growth has increasing by finding the weather forecasting based on whether rainy season or summer season.

Machine Learning build a model based on dataset that is used to make predictions. some implementations of machine learning use data set and neural networks in a way that imitate the working of a biological brain. Supervised learning is a type machine learning algorithm that use training data set to make predictions. Mutilayer Perceptron make use of a supervised learning technique which is called back propagation for training. MLP are the classical type of neural network, MLPs are apt for classification prediction problems where inputs are set aside a class or label.

2. REVIEW OF LITERATURE

In recent years there have been great strides in building classifiers for predictions on various datasets using varies machine learning algorithms. Machine learning, in particular, has shown improvement in accuracy on various datasets. Some of the works have been described below:

Abinaya P L Janani N [1] evaluated the performance of on forecast rainfall using WEKA data mining tool with three classifiers: 1.Navive Bayes 2. Decision Tree algorithm 3. K Nearest Neighbour 4. SVM (Support Vector Machine) and the outputs were compared based on the accuracy achieved.

Lemuel Clerk P Velasco, Ruth P.Serquina,etc [2] The need to explore different methods in selecting ANN parameters as this will help establish a reliable ANN model for rainfall forecasting.

Irfan M Malan S Muchdar F Abdullah N Juharni [3] The purpose of this research was to determine the production of seaweed K.alvarezii cultivation based on seasons. Production data collected is wet and dry production data by season.

Vaibhav. A.Mantri ,Ramalingam Dineshkumar, Mudassar Anisoddin Kazi, M. Vignesh [4] Neural-network approach in seaweed research: An emerging field for prediction and modelling critical parameters In this studies pertaining to artificial neural network (ANN) model used in studies performed to predict the thermophysical properties of Saccharina latissima, early detection of Sargassum bloom, cadmium–zinc ions biosorption by the Sargassum filipendula, optimization of different physiochemical parameters for seedling production in Gracilaria dura.

Meenakshisundaram Ganesan*, Nitin Trivedi, Vishal Gupta, S. Venu Madhav, Chennur Radhakrishna Reddy* and Ira A. Levine [5]

Seaweed resources in India – current status of diversity and cultivation: prospects and challenges In this research, the current status of Indian seaweed resources and their utilization, as well as developments in seaweed farming technologies, the status of seaweed industry and recent efforts to transform seaweed farming into a social enterprise. It also highlights the challenges encountered for mainstreaming these resources so as to evolve into a marine industry.

3. METHODOLOGY

In this section, the proposed work for weather forecasting data set is predicted by Weka tool. Collected weather and climate data for the experiment from the online repository of Kaggle. The variables on the data set of each weather measured values of climate features such as Temperature, Humidity, Wind Speed, Pressure, Daily summary. The dataset consists of 2709 instances and 12 attributes. The dataset with these attributes and its values are used in Enhanced MLP neural network method to weather forecasting for Seaweed Cultivation. MLP neural network is the classical type of neural network. They were comprised of one or more layers of neurons. Input data is give into the input layer, there may be one or more hidden layers thus providing levels of abstraction and predictions are made on the visible layer, known as output layer.

Algorithm (forward pass):

Require: pattern ~x, MLP, enumeration of all neurons in topological order
Ensure: calculate output of MLP

1: for all input neurons i do

2: set a i ← x i 3: end for

4: for all hidden and output neurons i in topological order do

5: set net i ← w i 0 + P j ∈ Pred(i) wij a j

6: set a i ← flog (net i)

7: end for

8: for all output neurons i do

9: assemble a i in output vector ~y

10: end for

11: return ~

MLPNN model has twelve input neurons consisting of the average temperature, minimum temperature, maximum temperature, average wind speed, maximum wind speed, relative humidity, total rainfall, visibility, the date, the month, and the year. Two MLPNN models were implemented with the first MLPNN model having small number of hidden neurons along with training algorithm, a supervised learning algorithm for feed-forward neural networks with super linear convergence and Hyperbolic-Tangent activation function. The Enhanced MLPNN model had more than hidden neurons along with SCG training algorithm and sigmoid activation function.

Comparative study of classification algorithm

Classifier	Time Taken to test Model	Time Taken build a Model
Multilayer Perceptron	0.42	145.15
Enhanced Multilayer Perceptron	0.09	37.37

Table -1: Sample Table format

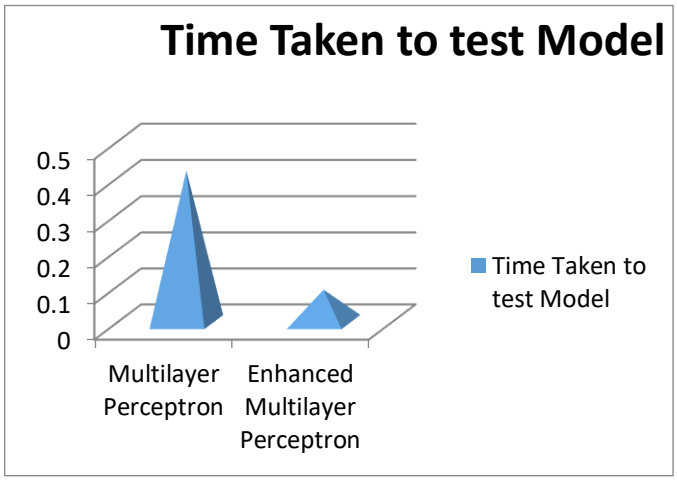


Fig-2 Time taken to build a model

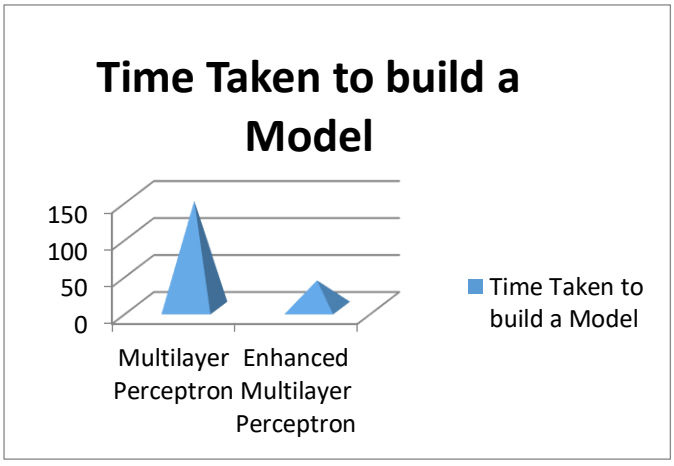


Fig-1 Time taken to test model

CONCLUSION AND FUTURE WORK

The main aim of this work is to forecasting weather for seaweed cultivation and production based on season. MLP neural network model was built to evaluate future predictions based on attributes that are outlook, temperature, humidity, windy and play from end of training data. This paper presents the analysis of seaweed cultivation and production increased during the rainy season or dry season. The MLP function has time taken to test a model on training data 0.42 seconds, enhanced MLP gives 0.09 seconds by increasing number of neurons in hidden layer. In the future, using image recognition model used to predict the better shoot tips properties of K .alvarezii, production of seaweed K.alvarezii cultivation and optimization of different shoot tips parameters for seedling production in micro propagation of tissue cultured plants.

REFERENCES

[1] Abinaya P L, Janani N " RAINFALL FORECASTING USING WEKA DATA MINING TOOL" International Research Journal of Engineering and Technology (IRJET),2020,ISSN: 2395-0072

[2] Lemuel Clerk P Velasco, Ruth P. Serquina, Mohammad Shahin A.Abdul Zamad,Bryan F.Juanico.Junneil C.Lomocso"Week-ahead Rainfall Forecasting Using multilayer Perceptron Neural Network" Procedia Computer Science 161 (2019) 386–397

[3] M Irfan1, S Malan1, F Muchdar1, N Abdullah1, Juharni11Department of Aquaculture Faculty of Fisheries and Marine Science, Khairun University,Ternate – Indonesia" Seaweed *Kappaphycus alvarezii* cultivation production based on season in Indari waters, West Bacan Subdistrict, Regency of South Halmahera, Province of North Mollucas, Indonesia"IOP Conf.Series: Earth and Environmental Science 584(2020)012043 doi: 10.1088/1755-1315/584/1/012043

[4] Vaibhav. A. Mantri, Ramalingam Dineshkumar, Mudassar Anisoddin Kazi, M. Vignesh" Neural-network approach in seaweed research: An emerging field for prediction and modelling critical parameters"

[5] Meenakshisundaram Ganesan*, Nitin Trivedi, Vishal Gupta, S. Venu Madhav, Chennur Radhakrishna Reddy* and Ira A. Levine" Seaweed resources in India – current status of diversity and cultivation: prospects and challenges"

Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022)
DVD Part Number: [illegible]; ISBN: [illegible]

A Geometric Method for Extracting Images of PDF Files

Ms. N Nagapathi
Ph.D Research Scholar, Department of Computer Science
Alagappa University
Karaikudi, Tamil Nadu, India
[illegible]@gmail.com

Dr. S S Dhenakaran
Professor, Department of CSE
Alagappa University
Karaikudi, Tamil Nadu, India
[illegible]@yahoo.com

Ms. S Uma Maheswari
Ph.D Research Scholar, Department of CSE
Alagappa University
Karaikudi, Tamil Nadu, India
[illegible]@gmail.com

Abstract— PDF is a format of the file that is used for showing research articles mostly. PDFs are actually easy to create, but extracting the data from PDF files is a challenging task. Images contain important as well as useful information which cannot be represented by text. The difficult concepts are better explained by an image than the text. In this paper, a feasible way to extract images from the PDF file is explored. This work is attempted to extract images from scholarly research publications of PDF files. The structure of a PDF file is entirely different from other file formats. A geometric method is proposed to get the position of an image in the PDF document and extracted all the images in that document. These images are saved in JPG or PNG or GIF or BMP format as a separate file. The resolution and size of extracted image are the same as the original image and are suitable for printing.

Keywords— PDF, Image Extraction, Cartesian Coordinate System, User Space, Device Space

I. INTRODUCTION

PDF expansion is the Portable Document Format, which is used to display files in an electronic format in order to prevent unauthorized alterations to the file. It is software, hardware and operating system agnostic. Text, photos and vector graphics, movies, animations, audio files, 3D models, interactive fields, hyperlinks and buttons can all be included in the PDF format. These many components can be included in a PDF document to create an article, report, presentation or portfolio.

PDF files are simple to make, read and use. However, PDFs are difficult to modify, and extracting data from them is a difficult operation. It is a way to show structured data. Research articles have wealth of information which is in PDF format only. Mostly, process flow or algorithm or pseudo code or even formula can be represented in the form of images in research articles. It is not possible to extract text or copy the images. So a novel method is used to explore images in the research articles that can be useful to researchers. It is seen that the retrieved images are good in size as well as in good resolution. To include graphics in PDF, a device-independent Cartesian Coordinate System is utilized to represent the surface of the page. Analytic geometry, sometimes known as Cartesian geometry, is a branch of Mathematics that describes every point in two-dimensional space using two coordinates, x and y.

The x-axis is commonly defined or portrayed in Mathematics as horizontal and oriented to the right, while the y-axis is vertical and oriented upwards. The y-axis is directed downwards in Computer Graphics. Fig. 1 illustrates this. In a two-dimensional system, two coordinate spaces are maintained. The first is User Space, which is where graphics primitives are defined. The PDF page coordinate system is referred to as User Space. Device Space, on the other hand, is the coordinate system of an output device like a screen, window, or printer. Device space refers to the coordinate system where the images are drawn.

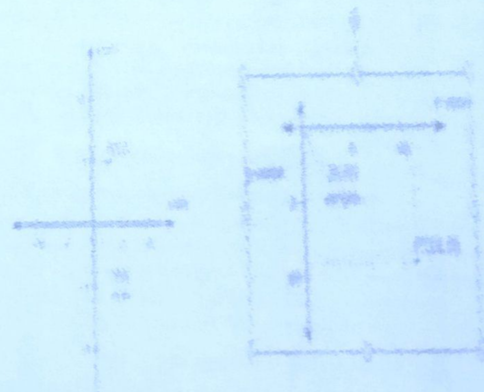


Fig. 1. Cartesian Coordinate System: a.Mathematics Perspective b.Computer Perspective

The surface of the page is considered a two-dimensional space. Two coordinates are used for getting the location and size of the image on each page of the PDF files. The internal data structure of PDF is graphics state which is widely helpful in mapping user space to device space coordinate system.

The retrieved image size is larger than the raw image in the PDF file and stored in the system folder.

II. RELATED WORKS

A vast number of papers are now preserved in Portable Document Format (PDF) files, thanks to developments in information technology and extensive use of the Internet. For several years, document structure analysis has been studied (Mao et al., 2003) have provided a description of some approaches, and there are now a variety of freely available tools that can be used to extract information from scientific papers. Several information extraction methods had been introduced. "Scholarly figures" is a term used by Bonchev et al. [1] to describe data visualizations such as plots or charts

Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022)
IEEE Xplore Part Number: CFP22BC3-DVD; ISBN: 978-1-6654-9710-1

SEACKER [2] is a method for extracting hierarchical information. The extracted data is separated into levels, with lower levels containing more specific and concrete information than higher level.

CCWRAP [3] is another approach that separates data into tuples, with each tuple including certain attributes. LAPDFText [4] methodology only extracts text from research articles and is meant to be used as a starting point for more advanced extraction methods that handle multi-modal content such as images and graphs.

The system works in three stages: (a) by means of spatial layout detection to get and recognize blocks of adjoining text, (b) by means of a rule-based system to organize text blocks into rhetorical categories, and (c) by adjoining categorized text blocks simultaneously in the exact arrange to make text from section-wise clustered blocks. CERMINE [5] is a comprehensive system that can extract not only the metadata of a document, but also the contents of the document, but also bibliographic references, and structured metadata as well as the document's structured body content, directly from a PDF file.

Piotr Adam Peszczyk et al. [6] have proposed the layout of a PDF page and understand a particular division of a page into the area called columns. Each area on a page is a stack of disjoint rectangles. This can be defined as a rectangle, let it be P representing the page. The set Q comprising columns of a page in the document is called a page division.

$$\bigcup_{i\in S} Q = P$$

$$\forall_{i,j\in S, i\neq j}\ Q_i \cap Q_j = \emptyset$$

$$\forall_{i\in S}\ Q_i \neq \emptyset$$

$$\forall_{Q\in S}\ \exists_{R\ \text{is a rectangle's page rectangle}}\ Q = \bigcup_{r\in R} r$$

Lexicographical states detected by a simple classifier do not directly translate to images. The main reason for this is that images may contain not only graphics but also portions of text. For instance, common graphical elements not belonging to a figure include logos and text separators like lines and for a figure include legend and some parts of mathematical formulas include lines and some parts of mathematical formulas graphical operations. So this graphical layout should not be considered as an image.

Mingyan Shao et al.[9] have proposed a system that can extract and analyze figures from PDF documents and classify them using machine learning and built an interpreter that iterated to a sequence of self-contained graphic objects represents the PDF content stream. This system mainly focused on vector-based diagrams that mostly images published in digital articles are raster-based. Document Image Analysis defined by George Nagy [10] is the concept of theory and practice of recovering the symbol structure of digital images scanned from paper or produced by a computer system.

Lawrence O'Gorman [11] described the Docstrum method for page layout analysis. This method can be used to segment independently oriented smaller documents like receipts, index cards, and business cards in a single image.

Structural blocks are groups of one or more text lines assembled on the basis of spatial and geometric characteristics. These structural blocks are determined by finding regions in contour line patterns [12],[13] with properties such as parallelness, perpendicular proximity, and overlap.

Kasturgo and Haralick [14] proposed a method named geometric registration for aligning the OCR (Optical Character Recognition) tags on a page produced from a file of the text. This method produces the exact character locations and tolerates for degradation through copying and scanning the printout but containing string matching techniques, it applies only to synthetically generated pages.

A Cartesian coordinate system, as explained by Manish R. Joshi et al. [15], translates a pair of numbers in a particular position on a two-dimensional plane. A data position (u,v) is a pair off of points with a space "u" beside the x-axis and a detachment "v" corresponding to the y-axis revealed simultaneously. The x-axis and y-axis distances have symbols.

Jun Zhang et al. [16] described the method to consider the relative difference of means of pixel neighborhood to be the feature of every pixel due to axis' movement is decided by the relative difference of means of pixel circle neighborhood.

III. TECHNICAL DETAILS OF PDF

Text, vector graphics, and bitmap or raster graphics make up a PDF file. Vector graphics are illustrations and designs made up of shapes and lines. Photographs and other forms of images are represented using raster or bitmap graphics. To depict the surface of a page, PDF graphics use a device-independent Cartesian coordinate system.

A. Coordinate System of PDF

PDF contents are represented in the two- dimensional coordinate system. Each coordinate can be represented as a vector which is an ordered pair(x, y) or column matrix. The origin of the PDF coordinate system (0,0) denotes the bottom-left corner of the PDF page. PDF file specifies 72 points to 1 physical inch.

One important data structure is the graphics state that is associated with a PDF file. The Graphics state holds the information on how graphics are rendered to the screen. Its values are graphics control parameters such as the Current Transformation Matrix (CTM) and Color Space. The graphical parameter Current Transformation Matrix determines the coordinate system. These parameters are used to map from the user coordinate system into the device coordinate system when extracting the images from PDF pages.

IV. METHODOLOGY

Images are stored as a separate object XObject which contains the raw binary data for the image in the PDF file. However the image is part of an x-object container, the image location is not available. First, the PDF file is parsed to create a sequence of object instances. Next text stripped pages are created. Then it is a checked object which is an image object that was extracted with the help of a Cartesian coordinate system with two parameters x and y-axis values. Here it is considered user space to find the coordinates of

images in the PDF pages and device space to fix the size of the extracted images according to the monitor or screen.

In order to find the exact size of the image, the x coordinate is increased to the right, and the y coordinate is increased downwards. The conversion between user space and device space is performed automatically during extraction. Because the PDF page coordinate system is user space, it is necessary to refer coordinate system of the screen or window which is device space. The targeted device is a screen or window. This was helpful in finding the location and size of the image object in the PDF file. Graphics state is the most important component of our work. It is associated with PDF files. The key factor of graphics state is the Current Transformation Matrix used to map from user space coordinate system that denotes PDF page coordinates to device space coordinate system which denotes targeted device which is monitor or screen.

A. Proposed Methodology

- Collect Research papers that are in PDF
- Apply parsing method to create a sequence of object instances
- Create text stripped pages
- Detect graphics contents in the text stripped pages
- If graphics contents are detected, extract images from those pages with the help of a two-dimensional Cartesian coordinate system
- Create an image file for each extracted image which may be JPG or GIF or PNG or BMP and store it in the specified computer folder path as a separate file

The above-proposed methodology is depicted in Fig. 2. The purpose of system is to extract all the images in the entire PDF file and created separate image files that are stored in the folder directory. The extracted image size is somewhat larger than the raw image. Because device space is a device-dependent coordinate system. So extracted image size varies from the raw image which is on the PDF page.

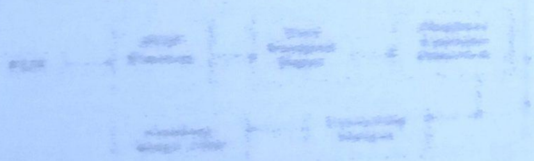


Fig. 2. Block Diagram Illustrating Proposed Approach

The resolution of the extracted image is 96dpi which has a good resolution for displaying the image on the monitor. The proposed method has extracted high-resolution images by doing the mapping process from user coordinates to device coordinates using CTM.

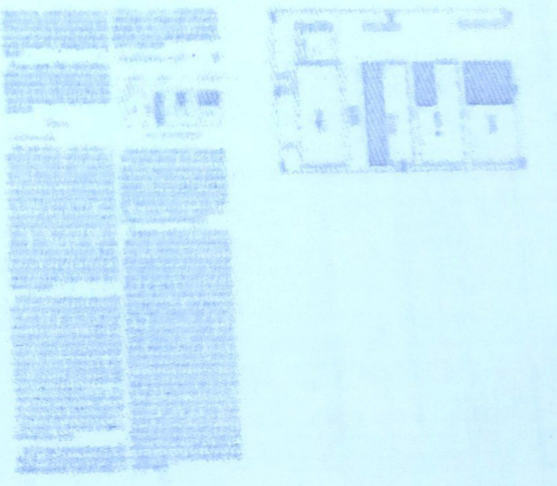


Fig. 3. Sample Paper (Alexandra Bosch et al., [?]) and Extracted Coloured Image

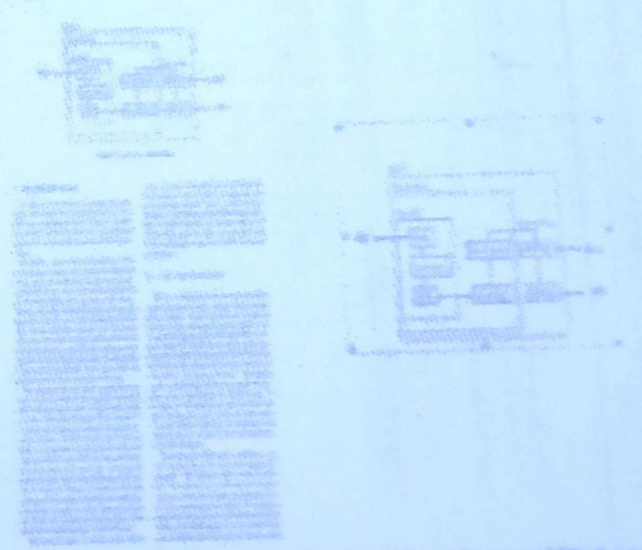


Fig. 4. Sample Paper (Simone Marinai [?]) and Extracted Black and White Image

The proposed work can extract black and white as well as coloured images from PDF pages. Extracted images are the same as original images which are in the PDF document. Fig. 3. and Fig. 4. show the PDF pages and extracted images. The proposed approach is with 300 PDF files. It has extracted 196 images from PDF files. The extraction procedure is complex due to the structural format of the PDF files. But images are more extractable than the text. It is interesting that most of the results are satisfactory from the user's point of view. These images can be used in a variety of ways.

V. RESULTS AND DISCUSSION

n = 347	Relevant	Irrelevant	Total
Retrieved	337	3	340
Not Retrieved	10	0	10
Total	347	3	350

Total Number of PDF files tested	300
Total Number of Images extracted correctly	337
Incorrect Images extracted	3
Not extracted Images	10
Precision	97.61%
Recall	97.89%
F-Score	97.31%
Accuracy	94.78%

The performance of the proposed system on 300 PDF documents is focused on Computer Science files which are published in 2009-2014 years. These documents contain 345 images.

The performance of the new system is assessed by two methods such as one is execution (response) time and another one is standard performance metrics (Precision, Recall, F-Score, and Accuracy).

For the first method, the response time is calculated on two different machines. The proposed work is executed in two different machines. For one input, the first machine with 8GB RAM took a response time of 0.35 seconds and the next was. For another input, the first machine with 8GB RAM took 42 seconds, and the next machine with 4 GB RAM took 44 seconds. The system time can differ by frequency of the clock, the frequency of the clock must be considered while calculating time. Factors affecting response time are:

a. Program b. Compiler c. Instruction Set Architecture d. CPU Design (Organization) e. Technology (VLSI) f. Data Set (PDF Documents) g. system clock. So the response time varies from one input to another input as well as different configurations of machines.

In the second method for evaluating the extraction applied factor is the standard metrics of precision, recall, F-Score and Accuracy by using the confusion matrix which is defined as below Table I. It is seen that the performance is better than the existing method by producing an acceptable accuracy value.

TABLE I. [illegible]

Confusion Matrix

	Relevant	Irrelevant	Total
Retrieved	a (Hit)	b (Junk)	a + b
Not Retrieved	c (Missed)	d (Rejected)	c + d
Total	a + c	b + d	a + b + c + d

$$Precision = \left(\frac{a}{a+b}\right) \times 100 \qquad Recall = \left(\frac{a}{a+c}\right) \times 100$$

$$F-Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \qquad Accuracy = \frac{a+d}{a+b+c+d}$$

The results are tabulated in Table II. Which are obtained on executing the test data set. This experiment is conducted with nearly 300 PDF documents which have produced Precision 97.61%, Recall 97.89%, F-Score 97.31%, and Accuracy 94.78%. It is found that the proposed system has performed better in extracting images from the PDF documents produces good quality images.

TABLE II. [illegible]

Fig. 5. shows the graph of our system performance in metrics.

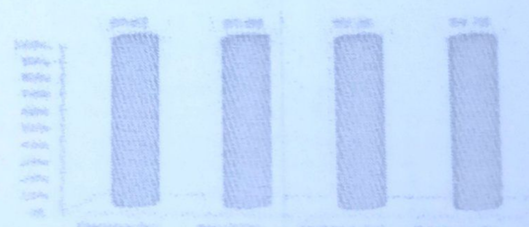


Fig. 5 [illegible]

VI. CONCLUSION

An image or a picture is worth a thousand words. A single image can convey complex and multiple ideas. The images in the scientific articles are very much helpful to the researchers. Complex concepts are better explained by an image than the text. The PDF file format is an entirely different structure and uses a different coordinate system. So extracting text and images from PDF files is a complex task. This work is a foundation for extracting all images from the PDF files. The proposed method has parsed the PDF file and created the text stripped pages. This process is helpful to

[illegible]

REFERENCES

[1] [illegible]

[2] [illegible]

[3] [illegible]

[4] [illegible]

[5] [illegible]

[6] [illegible]

[7] [illegible]

[8] [illegible]

[9] [illegible]

[10] [illegible]

[11] [illegible]

[12] [illegible]

[13] [illegible]

[14] [illegible]

[15] [illegible]

[16] [illegible]

[17] [illegible]

[18] [illegible]

[19] [illegible]

RESEARCH PUBLICATION DURING 2022 – 2023

DEPARMENT OF MICROBIOLOGY

TITLE	AUTHOR	DEPARTMENT	YEAR OF APPLICATION
Exploring Anti-Diabetic Compounds From The Ethanolic Extraction Of Strychnos Potatorum Seeds: Ligand- Based Design, Molecular Dynamics	Dr K Arunthathi	Microbiology	2023
Extraction of phytochemical constituents from the leaves of Vitex trifolia in tha inhibition of Lipoxygenase, Cyclooxygenase,Tyrosine and Xanthine oxidase	Dr K Arunthathi	Microbiology	2023
Analysis Of Phytochemicals In Root Sample Of Abutilon Indicum For Treatment Of Various Diseases By Using Molecular Docking	Dr K Arunthathi	Microbiology	2023
The culmination of multidrug - resistant efflux pumps vs. meager antibiotc arsenal era: Urgent need for an improved new generation of EPIs	DrShielaChetri	Microbiology	2023

Original Article

Imperative Linear Algebra for Data Science with R-Software

[1]R. Nagarathinam, [2]R. Punitha, [3]M. Nazreen Banu

[1]Associate Professor, Department of Mathematics, Dr.MGR Educational and Research Institute, Chennai, nagarathinam.hs@drmgrdu.ac.in

[2]Associate Professor, Department of Mathematics, Thassim Beevi Abdul Kader college for Women, Kilakarai, Tamilnadu, India, varshre@gmail.com

[3]Associate Professor, Department of Mathematics, Thassim Beevi Abdul Kader college for Women, Kilakarai, Tamilnadu, India, nazz_reen@yahoo.com

Abstract

Data science and machine learning are built on linear algebra. Machine learning and data science make extensive use of linear algebra, a branch of Mathematics. Machine learning relies heavily on linear algebra. Matrix representations are commonly used in machine learning models. Using linear algebra in data science means regularizing, reducing to dimensions, recognizing images, learning algorithms, and analyzing images. Many data science algorithms are based on linear algebra. This article will cover three uses of linear algebra in three different data science domains. We will discuss loss functions from the perspective of machine learning, and image convolution from the perspective of computer vision. Any prospective data scientist must learn R since it is a very strong language designed specifically for data analysis and data visualisation. With linear algebra, R is extremely useful. It has built-in data types like matrices and vectors.

Keywords: Linear Algebra, Data Science, Algorithms, R-Software.

Introduction

The area of mathematics known as "linear algebra" deals with linear equations and how they are represented in vector spaces and through matrices. The study of vectors and linear functions is what linear algebra is, to put it simply. Linear algebra involves the study of matrices, determinants, linear transformations, vector spaces, and subspaces and uses closed vectors that operate under addition and scalar multiplication. It enables us to do mathematical operations and comprehend geometric notions in greater dimensions. Nearly all branches of mathematics, including geometry and functional analysis, are based on linear algebra. Understanding its ideas is essential for comprehending the theory underlying data science.

Data science is a branch of study that combines subject-matter knowledge, programming abilities, and a working understanding of mathematics and statistics to draw out valuable insights from data.

Data scientists use machine learning algorithms to analyse data from a variety of sources, including text, images, videos, and audio, to create artificial intelligence systems that can carry out activities that often require human intellect. To study and analyse real-world phenomena using data, "data science" is a concept that combines statistics, data analysis, informatics, and their related methodologies. In the context of mathematics, statistics, computer science, information science, and domain knowledge, it makes use of methods and theories from a variety of domains. A data scientist is a person who writes programming code and uses statistical expertise in conjunction with it to derive insights from data.

Address for correspondence: R. Nagarathinam
Associate Professor, Department of Mathematics, Dr.MGR Educational and Research Institute, Chennai
Email: nagarathinam.hs@drmgrdu.ac.in

Access this article online

Quick Response Code:

Website: www.pnrjournal.com

DOI: 10.47750/pnr.2022.13.04.255

How to cite this article: R. Nagarathinam, R. Punitha, M. Nazreen Banu, Imperative Linear Algebra for Data Science with R-Software, J PHARM NEGATIVE RESULTS 2022;13(4): 1852-1857.

Data scientists may be able to avoid using linear algebra for a while, but not for very long. Here are several ways that linear algebra can help with computer vision and machine learning. Few people think of linear algebra when discussing data science in general or specific subfields like machine learning or computer vision. Because the modern tools we use to perform data science algorithms do a great job of disguising the underlying math that makes things function, linear algebra is sometimes overlooked. In data science, linear algebra is used.

Impact of Linear Algebra in Data Science

Area of mathematics called linear algebra is very helpful in machine learning and data science. The following data science fields make use of linear algebra:

- Regularization
- Reduction to dimension
- Image recognization
- Working with data sets
- Machine learning
- Computer vision

REGULARIZATION

Regularization is one of the biggest hurdles in machine learning, especially for beginners in data science. It is when a model is too close a fit for the available data, to the point that it does not perform well with any new or outside data. A concept called "regularization" is used to prevent the model from overfitting. This concept makes use of linear algebra as it uses the "norm". The norm can be defined simply as the magnitude of a vector. This magnitude can be calculated in various ways one popular way is using the Euclidean Distance. ie, using the distance from the origin. Regularization prevent overfitting as it adds the norm of the weight vector to the cost function. This makes sure that model does not become overly complex as aim is always to reduce the cost function, and therefore have to reduce this norm. This is much better understood by someone who knows the basics of linear algebra.

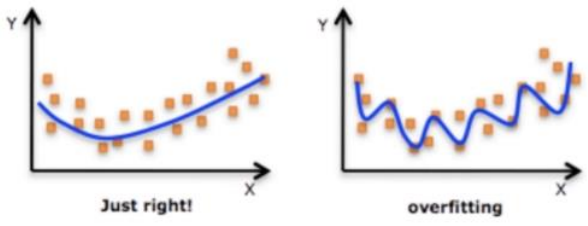


Figure 1

REDUCTION TO DIMENSION

While making Machine Learning Models, often come across data that is made up a hundreds or even thousands of variable. Our model becomes more and more complicated as these variables increase. Dimensionality reduction is the technique that reduces the number of input variables in data set. Since datasets can be easily represented as matrices, certain matrix factorization methods can be used to reduce a matrix and hence the dataset into its constituents parts. Then any operation that used to be performed on the original matrix, could be performed on the smaller matrices. Decomposition method like LU matrix decomposition and QR matrix decomposition can be easily performed using python programming.

LU Decomposition

Consider the system of equation written as a matrix equation:

$X_1+X_2-X_3=4$

$X_1-2X_2+3X_3=-6$

$2X_1+3X_2+X_3=7$

$$\begin{bmatrix}1 & 1 & 1\\ 1 & -2 & 3\\ 2 & 3 & 1\end{bmatrix}\begin{bmatrix}X1\\ X2\\ X3\end{bmatrix}=\begin{bmatrix}4\\ -6\\ 7\end{bmatrix}$$

$$AX = B$$

We can solve the system using LU Decomposition

Let A= LU and substitute into AX=B

Solve LUX=B FOR X to solve the system

Let UX=Y

LY=B and UX=Y

First solve LY=B and Y and the solve UX=Y for X

LINEAR ALGEBRA IN IMAGE RECOGNIZATION

When implementing data science models, especially in deep learning, it come across data in the form of image, however we cannot just pass an image to a model and expect it to understand it. It need to convert each image into something Mathematical or Statistical to be understood by the model. This is where linear algebra come in.

IMPACT OF LINEAR ALGEBRA IN IMAGE PROCESSING

Image processing is the manipulation of images using mathematical operations. With the introduction of computers, processing is now done using computer graphic algorithms on digital images obtained through a digitization process or directly using any digital device. Digital image processing is the use of a computer to perform image processing on digital images. Linear algebra can be used to perform computer graphics operations such as rotation, skewing, scaling, Bezier curves, reflections, dot and cross products, projections, and vector fields. Other more complex operations, such as filters, necessitate the use of linear algebra in conjunction with other mathematical tools. Linear algebra deals with matrices and all the operation to be

Performed on matrix. Any image is made of pixels, which are nothing but coloured square of varying Intensitics (for gray scale image it could be a single number with the intensity, and for coloured images, it could be the RGB value).

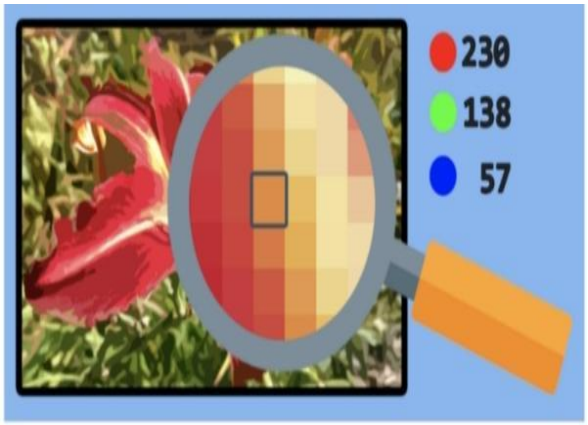


Figure 2

WORKING WITH DATA SETS

When building a machine learning model, it will most probably dealing with large data sets having multiple rows and columns. These are nothing but matrices when you spilt your dataset into training and testing data, you are performing operations on these matrices. Matrices are the key data structure in linear algebra and it deals with the various operations performed on matrix, including row and column transformations, transpose of a matrix, addition or scalar multiplication in matrices

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

This is nothing but a 5*4 matrix each record is a row and is indexed with the numbers 0,1,….4 . Each column has its name on top.

MACHINE LEARNING

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programmes can predict outcomes more accurately without having to be explicitly instructed to do so. Machine learning algorithms predict new results using historical data as input. The way in which a machine learning algorithm learns to improve its prediction accuracy is a common way to classify traditional machine learning. There are four fundamental strategies: reinforcement learning, semi-supervised learning, unsupervised learning, and supervised learning. The kind of data that scientists wish to predict determines the kind of algorithm they use.

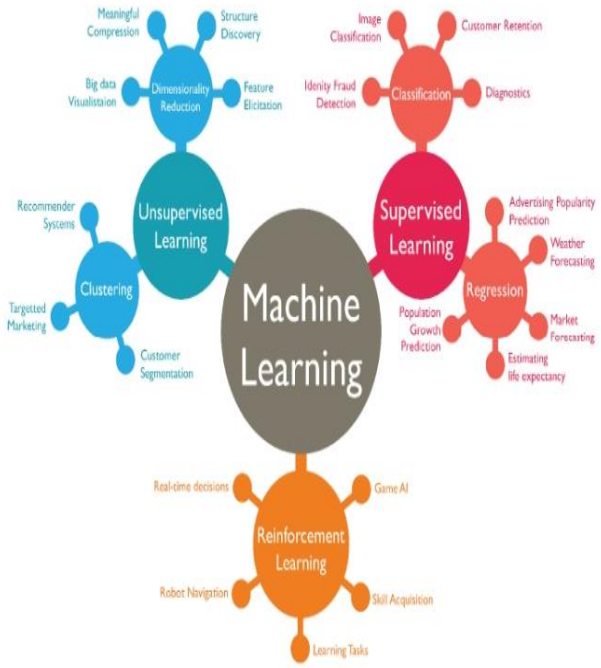


Figure.3

LINEAR ALGEBRA IN MACHINE LEARNING

Vectors, matrices, and linear transforms are the focus of the mathematics subject known as linear algebra. From the notation used to describe the operation of algorithms through the implementation of algorithms in code, it serves as a crucial basis for the field of machine learning. Despite the fact that linear algebra is essential to the subject of machine learning, the close connection is frequently ignored or described using impersonal ideas like vector spaces or certain matrix operations.

The following are significant fields of application made possible by linear algebra:

Data representation

Representation using learnt models

Eigenvectors for word embeddings

REPRESENTATION OF DATA

Data, which serves as the fuel for ML models, must first be transformed into arrays before being fed to your models. Matrix multiplication is one of the operations carried out on these arrays (dot product). Additionally, the output is returned, which can also be seen as a changed matrix or tensor of integers.

EMBEDDING WORDS

Just above it is the idea of using a lesser dimensional vector to express large-dimensional data (consider a large number of variables in your data).

EIGENVECTORS (SVD) (SVD)

Principal component analysis allows us to decrease the number of characteristics or dimensions in the data without sacrificing the quality of any of them.

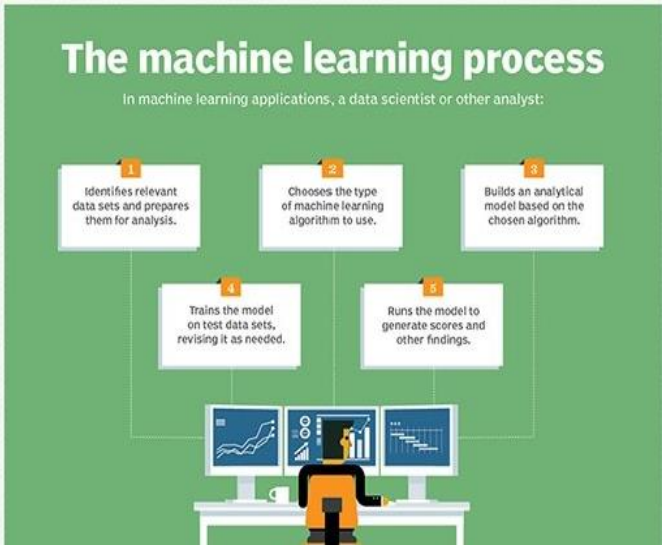


Figure 4

LINEAR ALGEBRA IN COMPUTER VISION

For machine learning, data science, and other related fields like computer vision and natural language processing, linear algebra is a potent tool. With its own implementations in the form of matrices, vectors, and tensors, linear algebra is used in computer vision. This includes basic operations, such as linear transformation, matrix operations, linear combination, and dependency of variables.

Significant contributions from linear algebra have been made to the difficult calculations used in computer vision. Large and complex matrices are necessary for the computations and operations involved in the complex matrix multiplications used in computer vision techniques. The computer vision system uses a variety of techniques to extract data from images, including compression, rotation, flip-flopping, convolution, noise reduction, object detection, facial recognition, etc.

The study of visual knowledge extraction is known as computer vision. and the fundamental idea of linear algebra is applied to the extraction. Vectors, matrices, and tensors are LA concepts.

VECTOR

A vector is a type of 1D array that is typically described with magnitude and direction.

MATRIX

It is a 2D array of numbers called a matrix. Think about an image's matrix-based pixel representation as an example. Projections, translations, rotations, scaling, and affine transformations are some of the operations of matrices.

TENSOR

It is a generalisation of matrices and vectors.

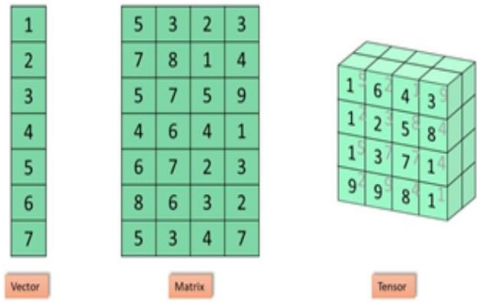


Figure. 5

VECTOR PRODUCT: MATRIX

The identity matrix is created by combining the unit vectors e1,..., em. A simpler approach to refer to random vectors a1,...an from the same vector space in a matrix is to use the letter I.

VECTORS AND MATRICE IN DATA SCIENCE

The some essential of linear algebra in the context of data science applications.

Vectors organize information that cannot be expressed as a single number and for which there exist a concept of scaling and addition. Matrices group together multiple vectors. The matrix –vector product expresses a linear combination of the column vectors of the matrix. solving a linear system

Ax=b=Ib, to find x€Rm for given b€Rm , re –expresses the linear algebra b= b1e1+…bmem ,

I=[e1e2….em] , as another linear combination b=x1a1 + x2a2+…..xnan , A=[a1a2…an] for certain problems the linear combination Ax might be more insightful, but the above transformation is information preserving, with b1x both having the same number of components

Finding the best approximation of some given b€Rm by a linear combination Ax of the column vector of A€Rm*n is known as a least squares problem and transforms the information from the m components of b into n components of x, and knowledge of the form of the column vectors. If m>n and the form of the column of A can be successfully stated, the transformation compresses information.

USAGE OF R SOFTWARE

COMMANDS TO FIND MULTIPLICATION OF MATRIX USING R

```
>
> #Multiplication of matrix
>
> A<-matrix(c(11,12,13,14,15,16,17,18,19),nrow = 3,byrow = T)
> B<-matrix(c(20,21,22,23,24,25,26,27,28),nrow = 3,byrow = T)
>
> A*B
     [,1] [,2] [,3]
[1,]  220  252  286
[2,]  322  360  400
[3,]  442  486  532
```

COMMANDS TO FIND INVERSE OF MATRIX USING R

```
> #Inverse of matrix
>
> B<-matrix(c(30,31,40,41,50,51,60,61,70),nrow = 3,byrow = T)
>
> A<-solve(B)
> A
            [,1]    [,2]        [,3]
[1,] -0.16208333 -0.1125  0.17458333
[2,] -0.07916667  0.1250 -0.04583333
[3,]  0.20791667 -0.0125 -0.09541667
>
```

COMMANDS TO FIND DETERMINANT OF A USING R

```
>
> #Determinant of A
> det(A)
[1] -0.0004166667
> #Calculating eigenvalues and eigenvectors
> A<-matrix(c(30,31,40,41,50,51,60,61,70),nrow = 3,byrow = T)
> e <- eigen(A)
> e$values
[1] 147.737576   5.317459  -3.055035
```

COMMANDS TO FIND EIGENVECTOR USING R

```
> e$vectors
           [,1]       [,2]        [,3]
[1,] -0.3948374  0.4437557 -0.74478185
[2,] -0.5497457 -0.8199420 -0.06303763
[3,] -0.7361271  0.3616296  0.66432391
> |
```

EIGENVECTORS APPLICATIONS IN DATA SCIENCE

The principal component analysis of a machine learning method makes use of the idea of an eigenvector.

Assume you have data that is very high in dimensionality and has a lot of features. It's possible that the data contains redundant features. A lot of features will also decrease efficiency and take up more disc space. In this case, PCA eliminates some of the less significant features; eigenvectors save the day. Let's go over the PCA algorithm. Let's say we wish to condense a "n"-dimensional dataset into a "k"-dimensional dataset. We'll proceed in stages.

Step 1: The data are subjected to mean normalisation and feature scaling.

Step 2: We find out the covariance matrix of our data set.

Step 3: Finding the covariance matrix's Eigenvectors is step three. It is a sophisticated statistical notion. We will discover "n" Eigenvectors matching "n" Eigenvalues because our data is in "n" dimensions.

Step 4: The fourth step entails choosing "k" Eigenvectors that correspond to the "k" biggest Eigenvalues and creating a matrix in which each Eigenvector is a column. It's time to locate the less data points now. Let's say you want to shrink a data point from the data set, "a," to dimensions of "k." Multiply the dimensions of the matrix U and transpose it.

After discussing Eigenvectors, let's move on to a more complex and valuable idea in linear algebra known as singular value decomposition, or SVD for short. To fully comprehend it, a thorough investigation of linear

SINGULAR VALUE DECOMPOSITION

Consider receiving a feature matrix A. We divide our matrix A into three constituent matrices for a specific purpose, as the name would imply. Additionally, it has been asserted on occasion that SVD is a generalisation of Eigen value decomposition. A data set's redundant characteristics are removed using SVD. Let's say you have a data set with 1000 features. There will undoubtedly be redundant features in any real data set with this many attributes.

Figure. 6

This tiger can be rendered in monochrome and seen as a matrix whose elements stand in for the pixel intensity and pertinent position. In other words, the matrix comprises data in the form of rows and columns concerning the intensity of pixels in the image. This image displays many images with various resolutions that correlate to various levels. Just assume for the time being that a higher rank indicates that there is more information about pixel intensity.

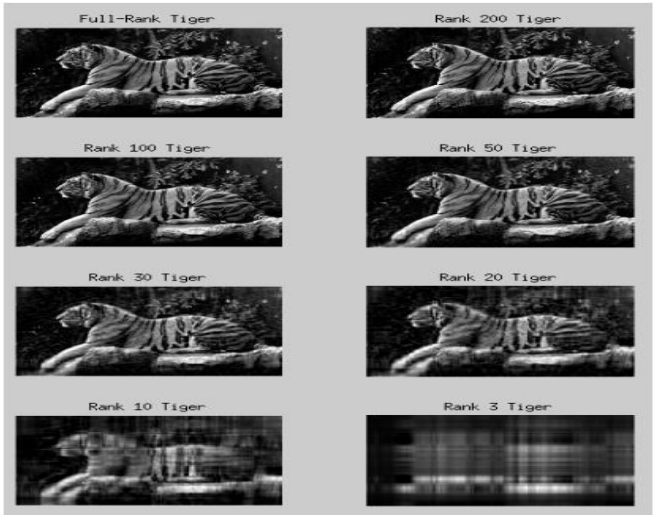


Figure. 7

It is obvious that we can achieve a reasonably good image with 20 or 30 ranks rather than 100 or 200 ranks, and in a scenario of highly redundant data, that is what we want to do. What I'm trying to say Is it true that we don't need to save every piece of data from the original dataset in order to generate a valid hypothesis? However, some of the features make it difficult to determine the optimum algorithm. For instance, multicolinearity in linear regression is caused by the presence of redundant characteristics. Additionally,

some qualities don't matter for our model. The method fits better, is more time efficient, and uses less disc space when these features are removed. The application of singular value decomposition

Conclusion

In the actual world, linear algebra is quite useful. In data science, linear algebra techniques are utilised to increase algorithm performance and produce more accurate findings. In this paper, it has gathered thatarethe uses of linear algebra in data science and provided an overview of each technique. To analyse the data sets, the data scientists can utilise linear algebra as a tool. Given the continually growing search outputs and the accessibility of the available evidence, which is a specific issue for the study sector in terms of quality improvement, machine learning algorithms are of special relevance. Regularization, dimension reduction, image identification, machine learning, and computer vision were all topics I covered. R is very useful for linear algebra.

References

[1] Sorin Mitran - Linear algebra for data science – University of North Carolina at chapel hill
[2] Herstein. L.N (2016) topic in algebra, second edition, wiley student edition
[3] Santiago M.L(2001) Modern algebra, Tata Megraw- hill publishing co ltd
[4] victor A. Bloomfield university of Minnesota Minneopolis, USA – using R for Numerical Analysis in Science and Engineering.
[5] Mike X Cohen, Practical Linear Algebra for Data Science, Released September 2022, Publisher(s): O'Reilly Media, Inc.,ISBN: 9781098120610.
[6] Kenneth Hoffman, Ray Kunze (1996) Linear Algebra,

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 17 April 2023
DOI 10.3389/fmicb.2023.1149418

OPEN ACCESS

EDITED BY
Rami M. Elshazli,
Horus University, Egypt

REVIEWED BY
Padmani Sandhu,
Institute of Microbial Technology (CSIR), India
Martijn Zwama,
Osaka University, Japan

*CORRESPONDENCE
Shiela Chetri
✉ sony240689@gmail.com

SPECIALTY SECTION
This article was submitted to Antimicrobials, Resistance and Chemotherapy, a section of the journal Frontiers in Microbiology

RECEIVED 21 January 2023
ACCEPTED 13 March 2023
PUBLISHED 17 April 2023

CITATION
Chetri S (2023) The culmination of multidrug-resistant efflux pumps vs. meager antibiotic arsenal era: Urgent need for an improved new generation of EPIs. *Front. Microbiol.* 14:1149418. doi: 10.3389/fmicb.2023.1149418

The culmination of multidrug-resistant efflux pumps vs. meager antibiotic arsenal era: Urgent need for an improved new generation of EPIs

Shiela Chetri*

Department of Microbiology, Thassim Beevi Abdul Kader College for Women, Kilakarai, Tamil Nadu, India

Efflux pumps function as an advanced defense system against antimicrobials by reducing the concentration of drugs inside the bacteria and extruding the substances outside. Various extraneous substances, including antimicrobials, toxic heavy metals, dyes, and detergents, have been removed by this protective barrier composed of diverse transporter proteins found in between the cell membrane and the periplasm within the bacterial cell. In this review, multiple efflux pump families have been analytically and widely outlined, and their potential applications have been discussed in detail. Additionally, this review also discusses a variety of biological functions of efflux pumps, including their role in the formation of biofilms, quorum sensing, their survivability, and the virulence in bacteria, and the genes/proteins associated with efflux pumps have also been explored for their potential relevance to antimicrobial resistance and antibiotic residue detection. A final discussion centers around efflux pump inhibitors, particularly those derived from plants.

KEYWORDS

efflux pump, MDR, antibiotic resistance, resistance nodulation division superfamily, efflux pump inhibitors (EPIs), *Escherichia coli*

Introduction

Bacteria are already well known for their ability to persist in extreme environments *via* their evolving numerous defense mechanisms against various substances that are lethal for their survival. Currently, antimicrobial resistance, the most complicated health concern across the world, which is primarily associated with an extended hospital stay, expensive treatment, and high mortality and morbidity rates, needs undeniable attention and multiple operative approaches involving the "One Health" concept. As the concept specifies, the interrelatedness among food, animals, and human health with the environment and the intentional collaboration between different sectors' efforts to confront the standing challenge (Taneja and Sharma, 2019), but regrettably, of antimicrobial resistance (AMR) development due to environmental influences have received less care as compared to human and animal health in spite of the significant deviations in the geographical dissemination of AMR associated with environment (O'Neill, 2015). The ultimate result of the increasing incidences of AMR is the development of multidrug-resistant phenotypes, and the multidrug-resistant (MDR) bacteria is a serious problem worldwide (Cooper and Shlaes, 2011; Blair et al., 2014; O'Neill, 2015; Venter, 2019). Moreover, unexpectedly, this problem occurs at the same time when several pharmaceutical companies stopped investing in research to find

new drugs to treat MDR due to a lack of high financial returns. In the bacterial cell membrane, the presence of transporters helps in the extrusion of noxious compounds, ultimately decreasing the concentration of external and unwanted or irrelevant substances and enhancing growth. However, the role of some transporters in the biodegradation of various deadly substances present in our environment is also well known (Ganas et al., 2007). As per the previous reports available, there are various resistance mechanisms contributing toward the development of resistant phenotypes—both extrinsic and intrinsic (Ganas et al., 2007). However, the role of intrinsic mechanisms such as the overexpression of multidrug-resistant efflux pump systems in evolving MDR phenotypes has gained partial attention (Nikaido and Pagès, 2012). The role of these efflux pumps is not only limited to functions as an efficient transporter but also limited to combat the stress from external environmental for bacterial survivability. However, it should be noted that some efflux pumps are substrate specific while others have a broad range of substrate specificity, which means that they can pump out multiple substrates including different classes of antibiotics, thus prompting multidrug resistance in bacteria (Nikaido and Pagès, 2012). Therefore, the efflux pump systems which form an important antibiotic resistance mechanism are widely considered due to their unique ability to extrude a variety of noxious compounds including dyes, detergents, different classes of antibiotics, and disinfectants (Nikaido and Pagès, 2012; Venter et al., 2015) outside the cell and thus are intensely associated with MDR expansion. Moreover, these specific features of the efflux pump systems have made them more prominent as multidrug efflux pumps (Webber and Piddock, 2003; Piddock, 2006a).

For the survivability of bacteria, these transporters located in the bacterial cell membrane play a crucial role in facilitating the reduction of external and irrelevant substance concentrations and promoting bacterial growth. However, evidence indicating the role of these transporters in the biodegradation of deadly compounds that are present in the environment has previously been reported (Ganas et al., 2007), suggesting the role of these efflux pump systems not only as a transporter but also in defending bacterial growth and survivability under various environmental stresses. Substrate specificity of these efflux pumps against various substrates, such as different antimicrobial classes, induces the pathogens to become multidrug resistant (Hernando-Amado et al., 2016). However, in some cases it has also been observed that the ejection of the substrates or organic solvents results into overexpression of transporters, which further, triggers the co-selection of the altered features of antimicrobial resistance in bacteria (Blair and Piddock, 2016). As a consequence of overexpression of the efflux pump systems, biofilm formation and quorum sensing (QS) will also be impacted by bacterial pathogenicity (Alcalde-Rico et al., 2016; Kong et al., 2017). In addition to antimicrobials, efflux pumps also have the potential to export virulence determinants, such as adhesins, toxins, and other crucial proteins, which are very critical for colonization within host cells (Piddock, 2006b). Several studies have already been reported on unfolding the development of antibiotic resistance associated with efflux pump systems along with numerous novel efflux pump transporters and simultaneous proteins. Additionally, in recent years, several other functions of an efflux pump system besides antibiotic resistance, such as virulence and bacterial self-protection against various environmental pollutants, were investigated. However, the particular mechanisms of regulation of these efflux pump systems and the active domains of the transporters are still not clearly understood. There are several inducing factors located on the inner and outer membranes of bacteria that stimulate the activity of efflux pump systems and promote their structural alterations within the fluid membrane environment. A brief overview of the role of efflux pumps in Gram-positive and Gram-negative bacteria is presented in this review article. Furthermore, the applications of efflux pump regulators to identify antimicrobial resistance and antibiotic residues, as well as the discovery of new efflux pump inhibitors derived from plants, have been discussed in this study.

Efflux pump families

Efflux pump systems are predominantly classified into six major superfamilies based on their energy source utilized to export the substrates: ATP-binding cassette (ABC) superfamily, multidrug and toxic compound extrusion (MATE) superfamily, major facilitator superfamily (MFS), resistance nodulation and cell division (RND) superfamily, small multidrug resistance (SMR) superfamily, and proteobacterial antimicrobial compound efflux (PACE) (Hassan et al., 2018) (Figure 1). One of the chief differentiating features of these efflux pumps is the source of energy they utilize; for example, the ABC superfamily draws energy *via* ATP hydrolysis (Verchère et al., 2012); however, other EP families such as RND, MFS, SMR, and MATE utilize either proton motive force (PMF) derived from H+ or electrochemical gradient of Na+ for exporting their substrates and for extruding several compounds (Kim and Hummer, 2012). Additionally, there are some other features as well for differentiating these efflux pumps based upon their composition or structure of the efflux pump transporters; for example, in an RND efflux pump system, three components, namely, an outer membrane protein (OMP) channel, an inner membrane transporter protein, and two modules that were connected *via* a membrane fusion protein (MFP) combine to form a composite known as a tripartite complex, which ultimately works together for the extrusion of noxious agents (Daury et al., 2016; Neuberger et al., 2018). However, research on novel multidrug-resistant (MDR) efflux pump systems in bacteria is still underway.

ATP-binding cassette superfamily

The MacAB-TolC efflux pump system is a tripartite complex belonging to the ABC and is the most comprehensively studied EP system among gram negative bacteria which is actively associated with the extrusion of macrolides and polypeptides virulence factors driven *via* ATPase MacB and further participates in enterotoxin heat stable enterotoxins (TII) secretion in *Escherichia coli* (Lu and Zgurskaya, 2013; Fitzpatrick et al., 2017; Jo et al., 2017). Additionally, lipopolysaccharides (LPS) and analogous glycolipids can be considered as organic substrates of the MacAB-TolC pump

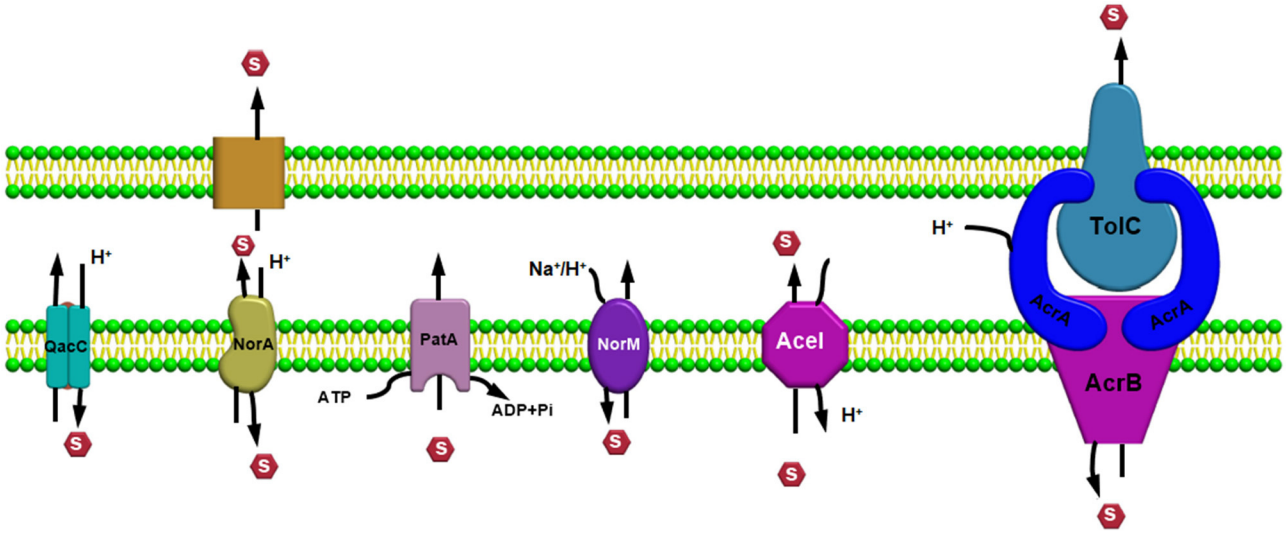


FIGURE 1
Efflux pump major super-families based on their energy source utilized to export the substrates.

systems (Lu and Zgurskaya, 2013). As the MacAB-TolC (pump system) is a tripartite complex with three components, including MacB, which is an inner membrane protein that functions as a homodimer complex and consists of two domains, namely, the (i) N-terminal, a nucleotide-binding domain that enables power generation *via* ATP hydrolysis and the (ii) C-terminal cytoplasmic tail (Lu and Zgurskaya, 2013). Other important components of the EP system are MacA, a periplasmic adaptor protein (PAP) that gets stimulated when ATPase binds specifically with the lipopolysaccharide core, and an outer membrane channel protein TolC that functions as an exit duct for substrate transport (Lu and Zgurskaya, 2013; Fitzpatrick et al., 2017). We observed from previous literature that, in *Serratia marcescens*, the absence of the MacAB (efflux pump system) resulted in increased susceptibility to polymixins and aminoglycosides, with a significant decrease in swimming motility and biofilm formation ability, and even caused impairment in defense capability to combat superoxide stress (Shirshikova et al., 2021). In addition, in their research study, Shi et al. (2019) observed that, in *Agrobacterium tumefaciens* 5A {a strain that is arsenite [As (III)] resistant}, MacAB efflux pump confers resistance against a variety of macrolide and penicillin-like antibiotics along with arsenite [As (III)].

However, Gram-positive bacteria harboring the ABC efflux pump systems, containing single component or transmembrane protein, have also been observed: for example, EfrAB transporter in *Enterococcus faecalis*, Msr protein in *Streptococcus*, LmrA in *Lactococcus lactis* (Hellmich et al., 2015), and PatA/B in *Streptococcus pneumoniae* (Garvey et al., 2011a; Iannelli et al., 2014). EfrAB, an MDR efflux pump system, is a heterodimeric protein conferring resistance toward different antibiotics such as gentamicin, chloramphenicol, and streptomycin (Lerma et al., 2014). In addition, the expression of the EfrAB pump was found to be highly induced when exposed to these antibiotics in sub-inhibitory concentration, and a decline in the expression was observed in the presence of a sub-inhibitory concentration of EDTA (3 mM) (Lerma et al., 2014). The LmrA transporter protein in *L. lactis* is a homodimer protein having one nucleotide-binding domain and six α-helix domains, which initially identifies and then proceeds toward the extrusion of macrolides and lincosamides (Hellmich et al., 2012). In *Streptococcus*, the Msr protein, comprising only two nucleotide-binding domains [without any transmembrane domain (TMD)], was found to be responsible for conferring resistance against macrolides and to collaborate with the Mef transport family (Zhang et al., 2016; Tatsuno et al., 2018). Further, in the case of the PatA/B efflux pump system, nearly the entire hydrophilic fluoroquinolone family members, such as ciprofloxacin, function as a substrate for this pump (Baylay and Piddock, 2015).

Multidrug and toxic compound extrusion superfamily

Members of the multidrug and toxic compound extrusion (MATE) family are broadly disseminated among the three domains, namely, archaea, bacteria, and eukarya (Kuroda and Tsuchiya, 2009). This class of transporters belongs to the multidrug/oligosaccharide-lipid/polysaccharide (MOP) family, which is further categorized into two subfamilies, namely, (i) NorM and DinF found in prokaryotes and (ii) the eMATE subfamily in eukaryotes based on their sequence similarity (Brown et al., 1999; Omote et al., 2006). MATE transporter proteins in bacteria are mainly involved in the extrusion process of amphiphilic cationic drugs, such as norfloxacin, outside the cell. However, in plants, these transporters are involved in physiological roles like developing resistance toward herbicides, leaf senescence, auxin synthesis, tolerance against aluminum, homeostasis, and sequestration of organic compounds (which are plant-based) inside the vacuoles (Remy and Duque, 2014). Further, the MATE transporters were found to be located in both the proximal

(convoluted and straight) tubules in the kidney in the case of mammals and in the canalicular domain inside hepatocytes, where the organic compounds are transported across this membrane domain during the final round of drug elimination process (Omote et al., 2006; Terada and Inui, 2008).

All life forms on earth need a balanced extrusion of exogenous noxious substances outside to sustain homeostasis. Therefore, the extrusion of lethal compounds such as xenobiotics requires the efficient involvement of the transporters. First, MATE family transporters were known for their ability to develop antibiotic resistance in *Vibrio parahaemolyticus* (Morita et al., 1998) but they later became well-known in other domains as well. Moreover, MATE is abundant throughout the human body, but its expression is highly concentrated in the bile canaliculi of the liver and within the brush border membrane of the kidney (Otsuka et al., 2005). Metformin, cimetidine, procainamide, and acyclovir (Masuda et al., 2006; Tanihara et al., 2007) are the cationic drugs that are recognized by the MATE transporter, and therefore, they act as influential factors in controlling plasma concentrations. However, previous literature suggests the significant role of MATE in inducing drug-associated nephrotoxicity during the last phase of pharmacodynamics of cationic drugs. An example is cisplatin, a clinically vital anticancer agent which has not been recognized by the MATE efflux transporter and has ultimately resulted in nephrotoxicity due to renal accumulation (Yokoo et al., 2007). Moreover, cimetidine, which is a histamine (H2) receptor antagonist that is a well-known MATE inhibitor functioning as a gastric acid inhibitor (Ito et al., 2012; Wagner et al., 2016), must be carefully handled and administered. Therefore, a deep understanding of the MATE transporter's working mechanisms is necessary for refining the efficiencies of cationic drugs. The MATE transporters are energetically involved in several physiological functions, which makes them attractive pharmaceutical targets. The MATE transporter protein structure is composed of 12-transmembrane helices arranged in two six-transmembrane (TM) domains, namely, the N- and C-lobes.

However, the positioning of these helices is arranged in an altered manner so as to make it different from the transporters belonging to the major facilitator superfamily (MFS) (He et al., 2010; Van Veen, 2010), supporting the previous assumptions that these two transporters should be considered as distinct superfamilies (Brown et al., 1999). Previous studies on the effective changes in the mechanisms, which are required either for the survivability of a particular microbe or for the development of a multidrug-resistant variant, have been observed in the case of a variety of transporter proteins. As a result of consecutive conformational variations, the transporter protein substitutes within two main conformations, one of which functions *via* exposing the substrate binding pocket (located centrally) to the inside and the other disclosures the pocket toward the outside of the cell. However, the structure and catalytic reactions mediated desired conformational changes are still not properly understood and one of the main reason behind this is the evidence indicating two crystal structures of the MATE transporter protein having an outward fronting conformation. Additionally, transporter proteins belonging to the MATE family were primarily well-known for mediating drugs or ions antiport *via* utilizing Na+ or H+ energy sources. The difference in the utilization of different energy sources such as H+ or Na+ gradient in different family members within the MATE superfamily has been suggested according to previous x-ray crystallographic studies concerning the NorM subfamily depending upon Na+ coupling in the C-lobe (He et al., 2010; Lu et al., 2013a), and a quite similar pattern was noticed in case of the eMATE subfamily (Miyauchi et al., 2017; Tanaka et al., 2017) but the DinF subfamily members was found to utilize H+ in the N-lobe (Lu et al., 2013b; Tanaka et al., 2013; Radchenko et al., 2015; Mousa et al., 2017; Kusakizako et al., 2019).

Major facilitated superfamily

Several secondary carrier families having similar topological features were identified before 1993, but no information was available to prove or to relate these families by a common origin. Among these families, a sugar porter's family including glucose facilitators found in mammals, two families of specific and multidrug efflux pump transporters, a metabolic uptake transporter family, and an oligosaccharide transporting family, including lactose permease present in *E. coli*, have been well studied. In addition, the knowledge and effort of bioinformatics helped us to understand the mechanisms and provided evidence to relate these families, and further, the name "major facilitator superfamily" (MFS) was coined (Marger and Saier, 1993). The major facilitator superfamily (MFS) transporters can be categorized into three main groups based on the mode of transportation of their substrates: (i) uniporters, which are responsible for the transference of a single substrate, (ii) symporters, which transport a substrate along with the simultaneous transportation of a coupling ion (e.g., H+), and (iii) antiporters, which are responsible for the simultaneous transportation of two substrates in opposite directions in such a way that the transportation of one substrate is dependent on prior discharge of the co-substrate (Forrest et al., 2011) (Figures 2A–C). In addition, it has been observed that the uniporters are able to transport their substrates without utilizing any energy source below their concentration gradient, while in the case of antiporters and symporters, the stored energy of the coupling ion concentration gradient is utilized for substrate transportation against their concentration gradient. However, irrespective of the differences in the mode of transportation of the substrates, all the members belonging to the major facilitator superfamily have the identical core structure comprising 12 TM helices, which are organized on two domains, namely, the N-domain and the C-domain (having similar structure) (Hirai et al., 2002; Yan, 2013), and are further partitioned into two inverted repeats of three helices.

One of the most well-studied MFS transporters in bacteria is the lactose permease (LacY) present in *E. coli,* which has worked as an appropriate protein model, for understanding the diverse features of the mechanisms involved in transportation associated with MFS transporters (Guan and Kaback, 2006; Smirnova et al., 2011; Kaback, 2015). However, in humans, the most studied MFS efflux pump (EP) system is the glucose transporter which plays a very crucial role in maintaining glucose homeostasis

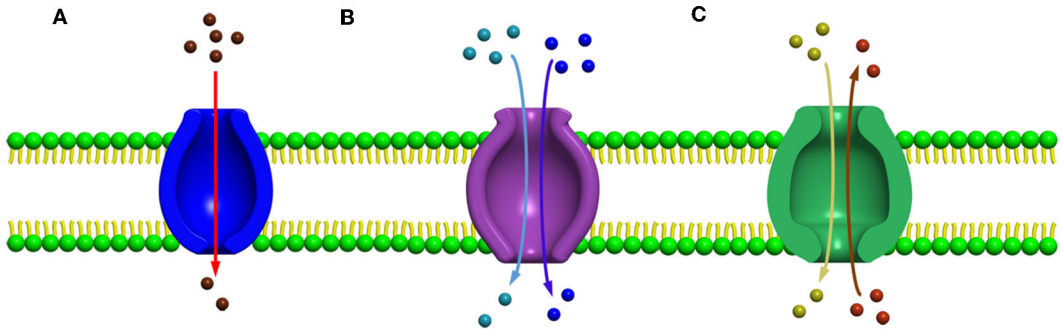


FIGURE 2
Major facilitator super family (MFS) transporters can be categorized into three main groups based upon their mode of transportation of their substrates **(A)** uniport, **(B)** symport, and **(C)** antiport.

in the human body, which is also referred to as the SLC2 protein. Moreover, the deterioration or faulty regulation of these proteins may lead to important diseases such as diabetes (Type II) (Augustin, 2010; Cura and Carruthers, 2012). Further, the roles of the Msr transporter belonging to the ABC superfamily and the Mef system of the MFS family were found to increase the macrolides extrusion co-actively, which further enhanced the resistance toward 14–15-member ring macrolides (Nunez-Samudio and Chesneau, 2013). Besides their active participation in the extrusion process, these MFS transporters are also linked with other biological pathways, such as the role of MdrT and MdrM in promoting the immune response in hosts by activating the production of interferon-β (IFN-B, which is a type I interferon response) and in maintaining the stability of the cell wall (Pasqua et al., 2021). In the case of *Staphylococcus aureus*, the presence of a constitutively expressing MFS efflux pump system Tet38 has been reported to influence the progression of host cell invasion comprising internalization, adhesion, and relocation within epithelial cells and the consequent step involved in epithelial cell infection due to *Staphylococcus aureus*, which includes the viability of bacteria and trafficking within phagolysosomes (Pasqua et al., 2021).

However, in *Acinetobacter baumannii*, loss of the MFS transporter protein AbaQ resulted in a significant decrease in bacterial virulence and motility capability (Pasqua et al., 2021), and furthermore, a substantial reduction in bacterial motility and virulence was detected when the efflux pumps encoding genes (such as RND, MATE, SMR, and ABC) are inactivated (Pérez-Varela et al., 2019). The working mechanisms of MFS transporters still require more explanation. However, an updated form of the classical rocker–switch model for understanding the changes emerging in the conformation of the MFS transporter associated with a form of substrate transportation known as the "clamp-and-switch" model has been previously proposed (Quistgaard et al., 2016), which also discussed the function of gating residues, binding of substrates, and the conformational change observed in the updated model (Quistgaard et al., 2016). In the past decades, many molecular biology, biochemical, and biophysical experiments have been conducted to understand the function of MFS transporters due to their physiological and pharmacological significance.

Resistance nodulation division superfamily

Resistance nodulation division (RND) superfamily members are the most well-known efflux pump systems because of their astonishing capability of extruding a wide range of substrates, which includes various classes of antibiotics, such as tetracyclines, chloramphenicol, different β-lactam antibiotics, and besides antibiotics, several other compounds such as dyes, detergents, metals, bacterial metabolites are also utilized as a substrate by this pump (Munita and Arias, 2016). The only feature which is common among these compounds is the amphiphilic nature of the substrates utilized by this pump (Yamaguchi et al., 2015). RND family members are part of multiprotein complexes which are arranged within the bacteria in a way that connects the outer membrane, periplasmic space, and the inner membrane to form a tripartite complex; thus, the RND transporters can facilitate the transport of compounds into these compartments by allowing them to work in conjunction with other transmembrane transporters. RND transporters are capable of capturing substrates not only from the cytoplasm but also from the inner membrane's outer leaflets or from the periplasm (Du et al., 2015). Consequently, it has been observed that these protein complexes are larger than several other transmembrane proteins of bacteria, which usually range over only a single membrane. Resistance nodulation division (RND) family members is comprised of three components, including (i) a periplasmic adaptor protein (PAP) (ii) an outer membrane protein (OMP) channel, and (iii) an MFP, which connects IMP and OMP, and are referred as a tripartite system for having three components together to form a complex (Figure 3). The main function of this tripartite complex is to identify and bind the substrates found in the periplasmic space, outer leaflet of the inner membrane, or the cytoplasm and lastly extrude or expel outside the cell through the outer membrane channel. However, other transporter protein functions as a single unit within the inner membrane for transporting substrates *via* the membrane bilayer (Nikaido, 2011; Du et al., 2018). Besides the participation of efflux pumps in the development of antibiotic resistance, other mechanisms such as alteration in drug target or inactivation or modification of antibiotics also have a vital role in AMR (Walsh, 2003; Blair et al., 2015; Colque et al., 2020). However,

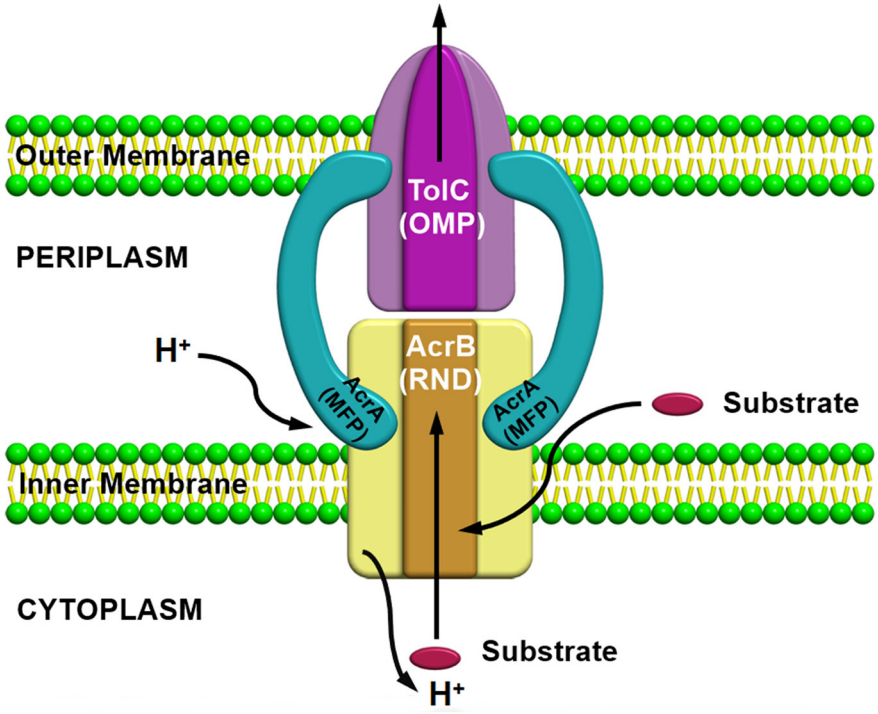


FIGURE 3
AcrAB-TolC a tripartite efflux pump system containing three components together to form a complex.

it has been noticed that comparative to the single factor that contributes to the resistance development against a particular class or group of antibiotics, multidrug resistance (MDR) can progress due to a reduction in bacterial membrane permeability (Nikaido, 2009) and *via* overexpression of the MDR efflux pump system in both Gram-positive and Gram-negative bacteria (Nikaido, 2009; Allen et al., 2010; Li et al., 2015). Besides being intrinsically expressed or present, bacteria also acquire resistance by RND efflux pumps *via* plasmids, etc. (Levy and Marshall, 2004; Allen et al., 2010), and mutation is intensely associated with the significant rise in the expression level of both intrinsic and extrinsic efflux pump systems in the clinical strains (Blair et al., 2015). This resulted in an increase in the expression level, which is one of the chief reasons for MDR development (Blair et al., 2014). In the current scenario, resistance nodulation division (RND) superfamily members are considered among the core contributors to MDR progress in clinical isolates (Nikaido, 1996; Blair et al., 2014).

Besides, the ability of these RND transporters to work as a tripartite pump in the extrusion of a wide variety of substrates such as different classes of antibiotics, detergents (Zwama and Yamaguchi, 2018), dyes, and various other noxious compounds outside the cell (Du et al., 2014), the overexpression of the efflux pump genes as a result of an increase in the transcriptional level of the operonic genes due to mutations in their regulatory proteins and genes (Nishino et al., 2008, 2009) has already been well studied and established in pathogenic bacteria within clinical settings (Oethinger et al., 1998; Nishino et al., 2008, 2009; Blair et al., 2015; Yu et al., 2020; Salehi et al., 2021). Examples of some of the well-studied RND efflux pump systems observed in Gram-negative bacteria are AcrAB-TolC in *E. coli* and *Salmonella enterica*, MexAB-OprM in *Psedomonas aeruginosa*, and the AdeABC system in *A. baumannii* (Bhakdi et al., 1988; Nikaido, 1996; Xu et al., 2020; Lv et al., 2021) are intensely associated with the extrusion process of a wide range of lethal compounds. Other examples such as the CmeABC system in *Campylobacter jejuni*, SmeABC in *S. enterica*, and the MtrCDE efflux pump system in *Neisseria gonorrhoeae*, have also been reported previously (Gilson et al., 1990; Singh et al., 2006; Wehmeier et al., 2009; Guérin et al., 2014; Wright and Tate, 2015; Lv et al., 2021). An important role of the RND transporter proteins, such as MtrD, MexB, AcrB, and AdeB, within the system, is in the substrate binding step (where the substrate may be specific) and in the drug transportation process, which is crucial for resistance development clinically (Delmar et al., 2014). An example is the mutation in AcrB gene which resulted in ciprofloxacin treatment failure (B). According to the previous literature, the transcriptional level of the regulatory proteins belonging to the TetR family such as AcrR (Blanco et al., 2016; Chetri et al., 2019a), MtrR (Lin et al., 2017), CmeR (Zwama and Nishino, 2021), TtgR (Bay and Turner, 2009), MexR (Srinivasan and Rajamohan, 2013), NalC/NalD (Bay et al., 2008), SmeT (Lytvynenko et al., 2016), and RarA belonging to the AraC family (Chetri et al., 2018b) influences the expression of these efflux pumps. The amino acid (AA) residues within the efflux pump systems are very crucial and actively participate in the substrate binding course and the substitution, or any changes in the AA residues probably can modify the substrate affinity (Banigan et al., 2015). Further, alteration in prominent AA residues was demonstrated to associate with efflux pump-facilitated

resistance toward various drugs (Padariya et al., 2018). AcrAB-TolC, MexAB-OprM, and CmeABC are among the most well-characterized multidrug resistant efflux pump systems belonging to the RND superfamily and are responsible for survivability and pathogenicity in bacteria. There are reports indicating the co-existence of different defense mechanisms including porin deficiencies and β-lactamase production with the overexpression of efflux pumps in *E. coli* (Chetri et al., 2019a,c), but rare evidence of insertions into penicillin-binding protein 3 (PBP3) of *E. coli* affecting their activity against monobactams and cephalosporins has already been reported (Alm et al., 2015). In carbapenem-resistant *E. coli* isolates, the overexpression of the AcrAB-TolC efflux pump simultaneously with the downregulation of OmpF due to regulation *via* the MarA gene was observed (Chetri et al., 2020). Further, for a comparative analysis study in carbapenem-resistant *E. coli* isolates, the β-lactamase gene (*bla*NDM-1, *bla*CTX-M-14, and *bla*CTX-M-3) were targeted and amplified (Shimizu et al., 2017; Chetri et al., 2019b). However, a study showing the capability of the AcrAB-TolC efflux pump system in influencing bacterial attachment, then penetration within the host, and further proliferation and persistence within animals has been reported (Piddock, 2006b) previously. Besides, the role of the AcrAB-TolC efflux pump in *E. coli*, AcrEF-TolC has also been reported to have a significant role in developing resistance toward different antibiotics (Chetri et al., 2018a).

Extensive usage of heavy metals (HME) in antimicrobials is not an infrequent event, for example, in disinfectants, and as a result, the RND efflux pump superfamily has developed the ability to resist these compounds for their survivability. Further, these RND superfamily is divided into two main categories: (i) hydrophobic and amphiphilic (HAE)-RND family, for example, AcrB, AcrD, AcrF, MdtC, MdtD, and YhiV, (ii) HME-RND family, for example, CusA which utilizes Cu(I) and Ag(I) as the most important substrate (Delmar et al., 2014), where CusA functions as a membrane fusion protein, CusC as an outer membrane channel protein to form a tripartite complex, the CusCBA efflux pump system (Long et al., 2012). However, the mechanical pathways required for the exportation of metal ions have previously been predicted (Long et al., 2012). Further, for survival, bacteria adapts in different ways such as efflux pump-associated biodegradation to get rid of the harmful effects of the lethal chemical compounds resulting from organic pollutants, for example, in *Pseudomonas Putida*, the occurrence of the TtgABC efflux pump system is responsible for conferring tolerance toward toluene (Blanco et al., 2016). Consequently, due to pressure exerted *via* overuse of antimicrobials, mutations in functional genes occurs, which may influence various changes in the AA residues located in the binding sites, which affect the substrate specificity of an efflux pump, and as a result, bacteria become less receptive toward antimicrobials, i.e., antimicrobial-resistant phenotype significantly altering the scenario in the environmental, clinical, and laboratory settings (Zwama and Nishino, 2021). The genetic changes occurring in an efflux pump denote an evolutionary adaptation within a microorganism against antimicrobials, and it is undoubtedly complexing the entire therapeutic option management system of pathogenic bacteria (Zwama and Nishino, 2021).

Small multidrug resistance family

Small multidrug resistance (SMR) family members utilize proton gradient to export a wide range of antibiotics, hence contributing toward MDR development. On the basis of the sequence, the SMR family can be divided into two main physiological subtypes (Kermani et al., 2020): (i) guanidinium exporter (representative of "Gdx") which is responsible for the allocation of bacterial metabolites such as guanidinium ion in exchange of two protons (Nelson et al., 2017; Kermani et al., 2018) and (ii) quaternary ammonium compound (Qac) representative subtype, which is responsible for the export of certain hydrophobic, cationic, and quaternary ammonium compounds (Nelson et al., 2017; Kermani et al., 2018). After the discovery of the first antiseptics containing quaternary ammonium around 100 years ago, Qac cluster proteins were found to be nearly associated with the propagation of multidrug resistance elements (Russell, 2002; Pal et al., 2015; Gillings, 2017; Zhu et al., 2017). Several bacteria harboring both subtypes of SMR transporter proteins were detected. Additionally, it was observed that the two subtypes do not interfere with each other's physiological role, for example, Qac protein will not participate in the transportation of guanidinium ion (Gdm+) and likewise, the guanidinium exporter will not involve in the export of any cationic or quaternary ammonium compounds but will require a substrate that is having guanidine moiety (Kermani et al., 2020). In *E. coli*, an example of the most well-studied Qac transporter protein is EmrE, and also an EmrE mutant, i.e., S64V that is having diminished conformational exchange dynamics based on extensive NMR measurements has been recently reported (Shcherbakov et al., 2021). SMR family members comprise short polypeptide chains of 100–150 amino acids and can stretch the cytoplasmic membrane as 4 TM α-helices (Bay et al., 2008). Proteins with short hydrophilic loops allow solubilization of a broad range of drugs, noxious lipophilic compounds that comprise DNA intercalating agents, lethal metabolites such as nicotine, and polyamine compounds such as spermidine (Bay et al., 2008; Bay and Turner, 2009). There are several reports indicating the active role of these family members in developing resistance toward different substrates—like in *A. baumannii*—AbeS transporter is found to be responsible for the transportation of ethidium, benzalkonium, and acriflavine, which simultaneously increases resistance toward amikacin (Lytvynenko et al., 2016; Lin et al., 2017). Similarly, in *Klebsiella pneumoniae* activation of the efflux pump system, KpnEF confers resistance against chlorhexidine, other antiseptics, benzalkonium chloride, etc. (Srinivasan and Rajamohan, 2013). In both *P. aeruginosa* and *E. coli*, the role of EmrE transporter protein in recognizing the substrates and mediating the elimination of polyaromatic compounds has already been reported (Banigan et al., 2015; Chetri et al., 2018b), but the Qac proteins are particularly associated in resistance development against different antibiotics and antiseptics (Jaglic and Cervinkova, 2012). Frequent incidences of QacA or QacB genes in *E. faecalis* and *S. aureus* have been observed, whereas in the case of *Enterobacteriaceae* and *Pseudomonas* spp., the dissemination of qacE gene was more prevalent (Jaglic and Cervinkova, 2012). SMR family members are plasmid-encoded and can often occur on integrons (the

mobile genetic elements [MGEs]), which influences the menace of horizontal spread of resistance (Bay and Turner, 2009). Incidences of QAC exposure influencing the overexpression of the efflux pump supports the horizontal transfer of integrons or MGEs harboring fluoroquinolone resistance determinants within Class I intergron (QacED1) (Anandapadamanaban et al., 2016; Buffet-Bataillon et al., 2016). The transfer of both the antibiotic-resistant and disinfectant-resistant genes between diverse species regulates the bactericidal effect of these lethal compounds.

Proteobacterial antimicrobial compound efflux family

The abundance of these efflux pump systems in proteobacteria has provided a purpose for the naming of the family as the proteobacterial antimicrobial compound efflux (PACE) family (Hassan et al., 2015a). One of the well-studied transporter proteins of the PACE family, *Acinetobacter* chlorhexidine efflux (AceI) found to be responsible for developing resistance toward different compounds and additionally, numerous homologs of AceI conferring resistance against various biocides and mediating transportation of fluorescent dyes, such as proflavine and acriflavine, has been observed which differentiates this family from other families of efflux pump (Hassan et al., 2015a). Presently, PACE family members responsible for the extrusion of around 10 substrates have been reported and the transporter proteins belonging to this family are included in the Database of Transporter Classification (Saier et al., 2016), which comes under proteobacterial chlorhexidine efflux (CHX) family (Elbourne et al., 2017). Transporter protein of the PACE superfamily has been reported for the export of biocides such as acriflavine and chlorhexidine; however, transporter AbgT having a role in the extrusion of sulphonamides was observed. Regulators play an important role in modulation the efflux pump systems for drug extrusion, and the factors influencing drug resistance such as under tetracycline exposure, TetR regulates TetB expressional level and in case of cationic antimicrobials pressure, QacR controls qacA expression (Grkovic et al., 2002). In bacterial membranes, the regulators control the systems in such a way that, when there is a requirement of efflux, and only the expression of the efflux pump will proceed; thus, avoiding lethal effect of overexpressed efflux pumps that are constitutively expressing within the bacteria and saving the cellular resources (Andersson and Levin, 1999). From a research viewpoint, the identification of the efflux pump (EP) system recognizing the substrates or novel factors associated with resistance or tolerance against drugs can be analyzed through observation of the changes occurring in their transcriptional level, and it is possible because of the constricted regulatory control of EP genes. The analysis of the transcriptional response of *A. baumannii* toward chlorhexidine (a membrane-active biocide) revealed the role of *Acinetobacter* chlorhexidine efflux protein (AceI) system in *A. baumannii* (Hassan et al., 2013, 2015b). Antibiotic-resistant strains of *A. baumannii* are becoming increasingly prevalent in hospitals because they have emerged as a major cause of Gram-negative infections.

Function and regulation mechanism of efflux pump

Efflux pump genes are present in all species of bacteria, and the efflux pump encoding genes are located either on a chromosome or a plasmid, especially in clinical cases (Piddock, 2006b; Poole, 2007). The understanding of the regulation of these efflux pumps, finding their function, their participation in developing resistance toward specific or multiple substrates, and their response under different environmental conditions, all of which were very crucial for the implementation of an accurate action plan. For a variety of membrane transporters, alternating access mechanisms have shown evolutionary persistence over the past few decades (Jardetzky, 1966). During sequential conformational modifications, a substrate-binding pocket in the middle of the transporter is exposed to the inner cellular space, while the other exposes the pocket to the outer cellular space through multiple conformational changes (Raturi et al., 2021). However, the structure and catalysis of MATE transporters underlying conformational changes are still not clearly understood. Partly, this is due to the fact that, particularly, two previously reported crystal structures of MATE transporters were with an outward-facing orientation (Raturi et al., 2021). Therefore, the MATE transporters facilitating drug/ion coupled transportation *via* Na+ or H+ coupling (Lu et al., 2013a) have already been described. The functionally diverse ABC family members utilize ATP to export their substrates and some act as ion channel modulators (Li et al., 2017). The structure of the ABC family transporter consists of two domains—one with substrate binding pocket (TMDs) and another with nucleotide-binding domain (NBDs)—which binds and initiates hydrolysis of ATP to stimulate the transport cycle (Dawson and Locher, 2006; Jin et al., 2012; Choudhury et al., 2014; Kodan et al., 2014; Johnson and Chen, 2017; Verhalen et al., 2017). An "alternating access" mechanism involved in the transportation of the substrate has been observed based on the functional and structural data available, in which the transportation triggers the switch in the conformation between inward-open, occluded, and outward-open states required for substrate transportation, and additionally, these conformational alterations are associated with the dimerization and dissociation of the nucleotide-binding domain-mediated *via* binding and hydrolysis of ATP (Khare et al., 2009; Korkhov et al., 2012; Woo et al., 2012). Consequently, in *E. coli*, the presence of ABC transporter protein MsbA utilizing the ATPase activity to transfer lipopolysaccharide precursor lipid A from the inner-cytoplasmic membrane to the outer-periplasmic membrane has been observed (Woebking et al., 2005), and as a result of its catalytic action, this enzyme is referred to as a flippase. Further, the cryogenic electron microscopy (cryo-EM) studies of the structural aspects of MsbA protein helped to understand the functional states and the visualization of the transportation route (Mi et al., 2017). However, some ABC transporters have shown that the MacB exhibits only outward-directed transport mechanism for the substrate entering through the periplasmic side, and as a result, the state of the substrate binding pocket remains in the outward direction. The transporters have the ability to interfere with their substrates from the outside facing binding pocket followed by a conformational alteration coupled with ATP hydrolysis, which

enables the reduction in the affinity toward the ligand inducing its shifting into the exterior part (Perez et al., 2015; Locher, 2016). Similar to the ABC transporter family, the mechanism of alternating access was also reported in MFS transporters as well where in the course of a transport cycle the conformational switch of the two domains within inward to outward open states has been reported (Jardetzky, 1966). According to the data obtained from lactose permease symporter (LacY), the conformational switch of the alternating access mechanism is activated *via* induced fit from the ligand binding and further, and the transport rate is governed by the electrochemical proton gradients (Kaback, 2015). The fundamental component of the model is the formation of a ternary complex when the proton and substrate bind in an ordered manner. However, various MFS transporters bind in different sequences as well, such as initiated with binding with proton, then substrate, or with a substrate first and then proton. Additionally, it has been observed that the transportation cycle of MFS protein generally involves the binding (as an initial step) followed by the release of substrate and proton, however, variations in the stoichiometric drug–proton interconversion within family members have been reported (Schaedler and Van Veen, 2010; Fluman et al., 2014). Small multidrug resistance (SMR) efflux pump system utilizes proton motive force as an energy source was previously reported *via* an experiment with *S. aureus* Smr (Sau-Smr) conducted first by Grinius and Goldberg in 1994 (Littlejohn et al., 1992; Grinius and Goldberg, 1994). Moreover, this experiment revealed that an electrochemical proton gradient was utilized as the energy source for tetraphenylphosphonium (TPP) efflux due to the reconstitution of Sau-Smr efflux within proteoliposomes. Further, under similar experimental conditions, similar results were obtained for Eco-EmrE, the SMR homolog (Yerushalmi et al., 1995, 1996). Therefore, SMR proteins are considered a proton-dependent MDR efflux pump system (Paulsen et al., 1996). Despite QACs, SMR proteins are able to transport neutral and negatively charged compounds (Jack et al., 2000; Nishino and Yamaguchi, 2001). However, previous studies revealed that the nature of the compound has a significant effect on the energy required for substrate transportation observed in Eco-EmrE studies, for example, the transportation of TPP requires movement of the charge for an active efflux besides the transportation of methyl viologen (MV) (a lipophilic divalent cation) (Yerushalmi et al., 1995; Rotem and Schuldiner, 2004). Further, as a result of several structural studies of the AcrB protein of *E. coli* by utilizing X-ray crystallography (Murakami et al., 2002; Nakashima et al., 2011; Eicher et al., 2012) and cryo-EM techniques (Du et al., 2014), the transporter protein was well-studied and became the most expansively characterized member of RND family (Zgurskaya, 2009; Zwama et al., 2019). Additionally, the interaction of AcrB with a wide range of substrates, which are chemically distinct, was analyzed to understand the poly-specificity feature of the transporter protein (Murakami et al., 2006; Nakashima et al., 2011; Eicher et al., 2012, 2014; Hung et al., 2013; Mousa et al., 2016; Sjuts et al., 2016; Zwama et al., 2018). Further, the movement of the substrates *via* the transporter is driven by the conformational changes taking place within the system, which is also revealed by the previous studies (Du et al., 2014; Blair et al., 2015). Moreover, the understanding of the function of these EP systems and the inhibition mediated *via* small molecules was simplified with the help of molecular simulation studies (Vargiu and Nikaido, 2012; Vargiu et al., 2018). The entire protein domains present were defined by the initial crystal structure developed for AcrB, which revealed that the region of transmembrane consists of 12 α-helices that anchor the inner membrane protein and supports proton movement and is accompanied by the transportation of the substrates and energy supply to drive translocation (Murakami et al., 2006; Seeger et al., 2006). Further, the entry of the substrates into the multiprotein complex is facilitated by the groove located between 8 and 9 transmembrane helices, which provide one out of two distinct access sites for the substrates, and the remaining protein residue projects into periplasm which is required for binding of the substrates and to link up with AcrA (periplasmic protein) as well as with the outer membrane channel protein TolC that is rooted in the outer membrane. The carrier domain of AcrB having the ability to span periplasm consist of four sub-domains, namely, PN1, PN2, PC1, and PC2, each of which comprises a common motif pattern, namely, two β-strand–α-helix–β-strand motif. The carrier sub-domains of all three protomers are arranged in such a way that a central pore along with a long channel extending through the membrane has been formed for the translocation of the substrate (Zwama and Yamaguchi, 2018).

Further, it has been reported that a second access site for substrate was found to be located in a cleft that formed in between PC1 and PC2 domains. The binding of AcrA and TolC is facilitated by the docking domains (DN and DC), which are located on the top of the porter domain. Additionally, tight binding of the top of AcrB with the TolC base is very essential for preventing the substrate leakage into periplasm aided by β-strands, which are placed antiparallel to generate a funnel-like construction. The crystal structure initially reported by Murakami et al. (2002) revealed the crystallized symmetrical homotrimeric structure without any bound ligand and explained the entire structural design of AcrB. According to three consequent studies published in 2006, it has been first reported that the establishment of an AcrB complex with various ligands that demonstrated the adoption of diverse conformations by each of the three protomers resulted in an asymmetrical homotrimeric structure assembly (Murakami et al., 2006; Seeger et al., 2006; Su et al., 2006), which has been indicated to be an active state. Therefore, under these circumstances, three different conformations have been well-defined as follows: (i) loose (L) that is mainly associated with the access for the substrates, (ii) tight (T) that is linked with the binding course, and (iii) open (O) state coupled with the extrusion conformation (Seeger et al., 2009; Nikaido, 2011; Eicher et al., 2012). As the substrates move over the protein, the monomer rotates between L, T, and O and returns back to L, and this "L conformer" which is accessible for substrate binding is also recognized as the resting state (Murakami et al., 2006; Seeger et al., 2006; Nakashima et al., 2011). First, the substrate enters into AcrB during the unliganded L state either from the periplasm or inner membrane envelope *via* one of the two defined entry sites, which are then released into a proximal binding pocket (PBP). Additionally, as a result of the movement of substrates *via* AcrB in the direction of the distal binding pocket (DBP), which is situated deep in the carrier domain, a conformational alteration toward the T state assisting the transfer of the substrate from the central pore of AcrB toward TolC entrance has been observed.

Subsequently, the ejection of the substrate from AcrB into TolC occurs *via* the rotating functional mechanism, which involves the protein transitional step, i.e., from state T to state O (Murakami et al., 2006; Seeger et al., 2006). Under that situation, state O which indicates an open state, a channel which is generated for substrate extrusion from AcrB to TolC and during the transportation of substrate, and several compressions and hindrances occurred inside the porter domain which facilitated the transportation of substrate in a unidirectional way toward central funnel referred as "peristaltic pump mechanism" (Seeger et al., 2006, 2008; Pos, 2009). In AcrB, the substrate entry *via* the two primary sites, primarily, *via* the proximal binding pocket (PBP) and then toward the distal binding pocket (DBP), plays an important role in defining the poly-specificity feature of the transporter.

However, in the case of hydrophobic substrates which are separated in the inner membrane's outer leaflet can enter through the TMD, containing hydrophobic grooves and are well defined as TM8 and TM9 (Murakami et al., 2006; Pos, 2009). Thereafter, tunnel 1 is formed which then elongates from the periplasmic membrane plane at the end of the TM8/TM9 groove till it links the proximal binding pocket (PBP) 2. However, comparatively, the entry of compounds confined to periplasm occurs *via* another site placed within the porter domains PC1 and PC2, ∼15 Å on top of the membrane plane, leads toward tunnel 2, which further elongates in the direction of the protein center (Eicher et al., 2012). Finally, after linking tunnels 1 and 2, it leads toward a distal binding pocket located deep within AcrB protein, where the presence of a glycine-rich switch loop that is proposed to contribute to the unidirectional movement of the substrates further facilitates the separation of distal and proximal binding pockets (Eicher et al., 2012; Zwama et al., 2018; Tam et al., 2020). As per the previous AcrB structural studies, the high molecular mass drugs, such as erythromycin and rifampicin, co-crystallize within the proximal binding pocket; however, the distal binding pockets are found to be responsible for holding both the low and high molecular mass drugs such as minocycline, doxorubicin, and neopentyl glycol derivative C7NG (Nakashima et al., 2011; Ababou and Koronakis, 2016; Sakurai et al., 2019). AcrB in *E. coli* and MexB homolog in *P. aeruginosa* and have similar structures and substrate-binding sites; therefore, both have similarities in a broad substrate specificity as well. Distal binding pocket (DBP) is a huge chamber surrounded by hydrophobic phenylalanine (Phe) residues and polar residues (Vargiu et al., 2018); further, the side chains present on these hydrophobic residues interact with the substrates through Pi–Pi stacking and van der Waals interaction, and on the contrary, polar residues interact *via* H-bonding. There are several pieces of evidence supporting the manifestation of broad substrate specificity of the pump as demonstrated by the different binding positions adopted by minocycline and doxorubicin. Moreover, according to previous structural studies, the role of a switch loop containing Gln 616–Gly 619 in regulating substrate's entrance into the distal binding pocket (Vargiu et al., 2018) and also in functioning as a barrier to separate the distal binding pocket from proximal pocket has been observed. Additionally, it has been suggested that physical contact between the substrate with the switch loop is essential for the translocation of the substrate before leaving the proximal binding pocket, and the movement instigated *via* the flexible loop facilitated the transportation of the substrates deep inside the distal binding site; however, without passing through the loop, the extrusion of the substrates *via* TolC is not possible (Nakashima et al., 2011; Eicher et al., 2012; Vargiu and Nikaido, 2012; Jamshidi et al., 2018). Immediately, after the binding of the substrate into the distal binding pocket occurs, it leads to a conformational change in the protein from state T to O and subsequently, the disintegration of tunnels 1 and 2 because of a transition of the coil to helix at N-terminal end of TM8 and extensive reorientation of carrier domains has been noticed (Zwama et al., 2017). Furthermore, for substrate transportation, a third tunnel that directs the substrates from AcrB on the way to TolC for extrusion is created in the state O (Eicher et al., 2012). Besides AcrB, the role of AcrA is also very crucial as repacking of AcrA helps to inhibit the leaking of the substrate into the periplasm *via* closing the cavities that exist inside the pump, and further, the changes arising in AcrA which result into the opening of the closed base of TolC are obviously initiated *via* the ligand binding phase due to the conformational changes occurring in AcrB. In addition, if the assemblage is exposed to the periplasm, it will restrict the opening of the TolC channel and the absence of any substrate inside a pump or extrusion of a substrate facilitates the conformational changes in AcrB from the drug transportation state to the resting state L (Wang Y. et al., 2017; Wang Z. et al., 2017). According to the aforementioned discussion regarding the homotrimeric symmetrical structure of AcrB which crystallizes immediately and forms substrate or inhibitor bound protein co-complexes because of their flexibility and multiple binding site mechanisms of AcrB, the movement for transportation of the substrates driven *via* multiple structural changes has been observed (Sakurai et al., 2019). A well-organized protein crystal required for the investigation of x-ray diffraction patterns can be difficult to grow when either of these processes is involved. However, currently, a number of compound-bounded AcrB structural studies have been carried out, which facilitated our understanding of the molecular foundation of inhibition (Murakami et al., 2006; Sennhauser et al., 2009; Nakashima et al., 2011, 2013; Eicher et al., 2012; Sjuts et al., 2016; Wang Y. et al., 2017; Wang Z. et al., 2017). Distal binding pockets (DBP) contain two characteristics crucial to inhibitor binding: (i) channels that translocate hydrophilic substrates and bind multiple drugs and (ii) a hydrophobic catcher that is located proximate to the distal binding pocket that divides this channel, which gives more attention to AcrB targeted small molecule inhibitors designing.

Efflux pumps influencing biofilm foundation and quorum sensing

Besides the contribution of efflux pumps in the development of antibiotic resistance, some other mechanisms such as biofilm formation, which is clearly a community of microbes affixed to a surface, also contribute to the progress of bacterial resistance and tolerance; additionally, it has been suggested that the efflux pump system has the capability to influence the function of biofilm either directly or indirectly (Hall and Mah, 2017; Alav et al., 2018). In *A. baumannii*, biofilm formation is inhibited due to tigecycline exposure in sub-inhibitory concentration *via* the downregulation of the adeG gene encoding for efflux pumps (Chen et al., 2017). Quorum sensing (QS) is a network that is generated

within cells to encourage mutual communication at the cellular level, and the utilization of signal transduction for participation in activities occurring inside bacteria has already been observed. The association between biofilm and QS improves bacterial viability in response to different variations in environmental signals. As an example, the ejection of acylated homoserine lactone (AHL) *via* the efflux pump system MexAB-OprM in *P. aeruginosa* is maintained by the QS contribution and overexpression of the efflux pump, which ultimately resulted in the release of QS signals (Piddock, 2006b). According to the available literature, it has been assumed that the bacterial QS mechanism is restricted if inhibitors hinder the efflux pump activity (Gupta et al., 2017). Similarly, in *S. aureus*, the inhibition of the NorA efflux pump activity due to the presence of an appropriate concentration of efflux pump inhibitor has also been described in a previous study (Sabatini et al., 2017). Additionally, the role of efflux pump systems in facilitating the pathogens in adapting to the harsh environment present within a mature biofilm, which involves response against envelope stress in maintaining the membrane rigidity and in conferring resistance toward high osmotic stress (Robin et al., 2022), such as the essential role played by the MacAB-TolC efflux pump in biofilm formation (Robin et al., 2022), has been reported. Pmt and AbaF are the two well-known MFS transporter proteins found in *A. baumannii*, which play an important role in the formation of biofilms such as Pmt in extracellular DNA extrusion and active participation of AbaF in emancipating biofilm materials (Pasqua et al., 2021). There is evidence indicating variations in bacterial membrane influencing the functioning of biofilm formation mediated *via* different MDR efflux pump systems belonging to the RND family (Yoon et al., 2016), as shown in *A. baumannii*, where the inhibition of adeAB gene expression or deletion resulting in reduced expression or hindrance of biofilm and QS systems have been described earlier (Richmond et al., 2016; Dou et al., 2017). A previous study proposed the incidence of a positive association between mRNA levels of efflux pump genes and biofilm formation, which can be influenced by the MIC levels of antibiotics such as polymyxin B or colistin (Sato et al., 2018). For example, MexGHI, an efflux pump system recognized in *P. aerugiosa* responsible for phenazine transportation, plays a crucial role in biofilm morphogenesis (Sakhtah et al., 2016). Similarly, the QS system also has a similar correlation with the efflux pump system (Liang et al., 2016). Additionally, a study evidenced the mutation in mexI, an efflux pump gene of *P. aeruginosa,* mediating the loss and reduced expression of quorum sensing molecules or virulence factors (Wolloscheck et al., 2018). The role of ATP-binding cassette (ABC) type efflux pump systems in conferring resistance toward antifungal agents that are used against fungi, mostly *Candida* species, is well known and their involvement in QS molecules secretion and influential capability toward biofilm formation is found to be similar to that of RND family members (Cannon and Holmes, 2015).

Role of efflux pump inhibitors toward antimicrobial resistance

Efflux pumps largely contribute toward multidrug resistance, both intrinsic and extrinsic in bacteria; therefore, the active participation of efflux pump inhibitors (EPIs) have a potential impact in preventing antimicrobial resistance globally. A significant challenge in the discovery of EPI has been its highly hydrophobic nature and the substrate-binding poly-specific site for the target. However, the primary studies on the x-ray crystal structures of different compounds such as pyranopyridine (Sjuts et al., 2016) and pyridopyrimidine (Nakashima et al., 2013) (Table 1) revealed that the EPIs bound to different efflux pumps of RND family of *P. aeruginosa* and *E. coli* have helped to facilitate the rapid advancement in our understanding of different mechanisms involved and the binding interactions of these EPIs. Moreover, a strong binding interface between the Phe-abundant hydrophobic trap and the inhibitor D13-9001 participates in inhibiting the conformational change from state L to T and further hinders the functional rotation mechanism. Further, it has been suggested that the entry of the substrates into the binding cleft situated between the domains PC1 and PC2 is blocked by the hydrophilic moiety existing on the inhibitor D13-9001. Phe 178 is an imperative amino acid that exists in the hydrophobic trap, which plays a vital role in the binding of inhibitors with the aromatic substrates *via* utilizing π-π stacking (Nakashima et al., 2013; Sjuts et al., 2016), and it has been reported that, by applying a structure-guided design for an effective inhibitors development, the hydrophobic trap discovery has a major role (Aron and Opperman, 2018). As these efflux pumps are considered excellent targets for antimicrobial combination treatment, facilitating the usage of synthetic (Table 1) or plant-based (Table 2) efflux pump inhibitors (EPIs) to support antimicrobial therapeutic options toward bacterial infections (Opperman and Nguyen, 2015; Alibert et al., 2017). There are several effective approaches to restrict or avoid the activity of an active efflux pump system found in bacteria such as decreasing the antibiotic binding affinity toward transporters *via* chemical structural modification of the drug, escalating the OM permeability to increase the drug concentration within cell, knocking out or impeding the efflux pump associated genes, and disrupting the supply of ATP or *via* designing compounds which have the capability to compete with the antimicrobial agents for active sites for their action to obstruct the efflux pump activity (Jamshidi et al., 2016). Additionally, approaches such as computational analysis or artificial abstraction from plants were utilized for the detection of numerous inhibitors active against both Gram-positive and Gram-negative bacteria. Researchers working on screening different natural EP inhibitor compounds *via* utilizing high-throughput methods have observed a natural compound referred to as *Terminalia chebula*, which is isolated from an Indian medicinal plant, having the ability to abolish the binding of ligands with Na+, which resulted in a change in the conformation of the transporter protein NorM found in *N. gonorrhea* to closed state (Kesherwani et al., 2017). In *E. coli*, *P. aeruginosa*, and *A. baumannii*, one of the most widely studied efflux pump inhibitors of AcrB, phenylalanyl-arginine-β-naphthylamide (PAβN), is recognized for its ability to bind with the hydrophobic pocket located within AcrB, further inhibiting the drug extrusion activity (Lomovskaya et al., 2001; Ribera et al., 2002; Kinana et al., 2016). However, few other EPIs (MBX inhibitors) having the ability to bind strongly to AcrB have also been reported (Sjuts et al., 2016). Additionally, this inhibitor can be combined with a secondary metabolite, i.e., carolacton produced by mycobacteria, which can be utilized as a

TABLE 1 Examples of synthetic efflux pump inhibitors (EPIs) having inhibitory action against Gram-negative bacteria.

S. No	Efflux pump inhibitor (EPI) types	Efflux pump substrate (s)	Effective against bacteria	References
1	Peptidomimetics	Chloramphenicol, fluoroquinolones, macrolides, carbapenem, tetracyclines	*E. coli, P. aeruginosa, K. pneumoniae, S. enterica, Campylobacter* spp., *E. aerogenes, A. baumanii*	Renau et al., 1999; Malléa et al., 2002; Mamelli et al., 2003; Lomovskaya and Bostian, 2006; Cortez-Cordova and Kumar, 2011; Vera-Leiva et al., 2018
2	Quinoline derivatives	Chloramphenicol, tetracycline, norfloxacin	*E. aerogenes, K. pneumoniae*	Chevalier et al., 2001; Gallo et al., 2003; Malléa et al., 2003; Ghisalberti et al., 2006; Mahamoud et al., 2006
3	Trimethoprim	Chloramphenicol, tetracycline, ciprofloxacin, erythromycin	*Enterobacteriaceae, P. aeruginosa*	Piddock et al., 2010
4	Aminoglycoside analogs	Tetracycline, gentamicin	*H. influenza*	Van Bambeke and Lee, 2006
5	Microbial EPIs	Levofloxacin	*P. aeruginosa*	Lee et al., 2001
6	Arylpiperazines	Chloramphenicol, tetracyclines, fluoroquinolones, macrolides, linezolid	*A. baumanii, P. aeruginosa, C. jejuni Enterobacteriaceae*	Bohnert and Kern, 2005; Kern et al., 2006; Pannek et al., 2006; Schumacher et al., 2006; Kurinčič et al., 2012; Sonnet et al., 2012
7	Indole derivatives	Chloramphenicol, erythromycin, ciprofloxacin, tetracycline	*E. coli*	Zeng et al., 2010
8	sRNA and antisense oligonucleotides	Ciprofloxacin, erythromycin	*E. coli, C. jejuni*	Van Bambeke and Lee, 2006; Mu et al., 2013
9	Arylpiperidines	Linezolid	*E. coli*	Thorarensen et al., 2001
10	Substituted polyamines	Unknown	*H. influenza*	Mahmood et al., 2016
11	Hydantoins	Chloramphenicol, nalidixic acid, sparfloxacin, doxycycline, erythromycin	*E. coli, E. aerogenes*	Handzlik et al., 2011; Otreebska-Machaj et al., 2016
12	Epinephrine	Chloramphenicol, tetracycline, ciprofloxacin, erythromycin	Enterobacteriaceae, *P. aeruginosa*	Piddock et al., 2010
13	Antibiotic analogs, tetracycline analogs	Tetracyclines	*E. coli*	Nelson et al., 1993, 1994
14	Quinazoline derivatives	Chloramphenicol, nalidixic acid, sparfloxacin	*E. aerogenes, P. aeruginosa*	Chevalier et al., 2010
15	Pyranopyridines	Fluoroquinolones, piperacillin	Enterobacteriaceae	Opperman et al., 2014; Vargiu et al., 2014; Aron and Opperman, 2016; Sjuts et al., 2016
16	Phenothiazines	Chloramphenicol, tetracyclines, nalidixic acid, levofloxacin, triclosan, erythromycin, aminoglycosides	*E. coli, S. enterica, B. pseudomallei*	Chan et al., 2007; Bailey et al., 2008; Martins et al., 2008
17	Serum compounds	Minocycline, ciprofloxacin	*A. baumanii, P. aeruginosa*	Blanchard et al., 2014
18	Antibiotics globomycin	Chloramphenicol, norfloxacin	*E. aerogenes*	Malléa et al., 2002
19	Fluoroquinolone analogs	Fluoroquinolones, macrolides	*E. coli, P. aeruginosa*	Van Bambeke and Lee, 2006
20	Pyridopyrimidines	Fluoroquinolones, B-lactams	*P. aeruginosa*	Yoshida et al., 2006, 2007
21	Naphthamides	Chloramphenicol, tetraphenylphosphonium, erythromycin	*E. coli*	Wang Y. et al., 2017; Wang Z. et al., 2017; Wang et al., 2018

potential therapeutic option against pathogens over-expressing the AcrAB-TolC efflux pump system (Donner et al., 2017). Another example of EPIs is tannic acid, which inhibits the multidrug efflux pump systems, such as Tet and Msr (Tintino et al., 2017), and is well studied in *S. aureus*. Tannic acid has also been reported to reduce the MIC levels of different antibiotics such as erythromycin and tetracycline significantly (Tintino et al., 2017). However, in clinical settings, the usage of an antibiotic along with an efflux pump inhibitor (EPI) as combination therapy is a potential challenge that depends upon the internal permeability characteristics of the bacterial outer membrane. Further, in a study done by Yang et al. (2017) the use of combination therapy against multidrug-resistant *P. aeruginosa* comprising tobramycin and EPI (NMP, paroxetine, or DBP) encourages the binding of EPI with tetracycline; additionally, the response of the combination of tobramycin plus EPI along with rifampicin, fluoroquinolone, and fosfomycin was also analyzed, which revealed that they exert a robust effect cooperatively upon the bacteria, which resulted into reduction in the MIC80 level against these antibiotics (Yang et al., 2018). Another example of combination therapy containing levofloxacin and EPI (trimethoprim and sertraline) used against *P. aeruginosa* harboring the MexAB-OprM efflux

TABLE 2 List of plant-derived efflux pump inhibitors (EPIs) having inhibitory action toward Gram-negative bacteria.

Plant	Compound extracted	Effective against	References
Ipomoea muricata	Lysergol	*E. coli*	Maurya et al., 2013
Glycine max	Daidzein	*E. coli*	Dwivedi et al., 2015
Baccharoides adoensis, Callistemon citrinus	Ethanolic extract	*Pseudomonas aeruginosa*	Chitemerere and Mukanganyama, 2014
Larrea tridentata	Nordihydroguaretic acid	*E. coli*	Ohene-Agyei et al., 2014
Lithospermum erythrorhizon	Shikonin		
Plumbagin	*Plumbago indica*		
Digitalis lanata	Lanatoside C	*E. coli, Pseudomonas aeruginosa*	Aparna et al., 2014
Acer saccharum	Phenolic-rich maple syrup extracts (PRMSE)	*E. coli, Proteus mirabilis, Pseudomonas aeruginosa*	Maisuria et al., 2015
Ammannia spp.	4-Hydroxy-alpha-tetralone + semisynthetic derivatives	*E. coli*	Dwivedi et al., 2014
Eucalyptus tereticornis	Ursolic acid	*E. coli*	Dwivedi et al., 2015
Berveris bulgaris	Berberine, palmatine	*Pseudomonas aeruginosa*	Aghayan et al., 2017
Holarrhena antidysenterica	Conessine		Siriyong et al., 2017

pump system showed a progressive advantage over monotherapy (using levofloxacin alone) (Adamson et al., 2015). Similar research conducted by Prasch and Bucar (2015) has also validated that a reduction in the antibiotic dose was observed when the combination therapy comprising EPI inhibitor and antibiotics was co-administered. Besides the aforementioned EPIs, several plant-based EPIs, have also been investigated, and above 20 different EPIs derived from plants have been reported based on their extraction mechanisms (Garvey et al., 2011b; Shiu et al., 2013). Additionally, vegetable-derived compounds, such as artesunate, berberine (Table 2), and curcumin, have also been reported to have the ability to inhibit EP activity in *E. coli* and *P. aeruginosa*, besides their anti-inflammatory and antibacterial activities (Li et al., 2011; Negi et al., 2014; Laudadio et al., 2019). Different food items such as vegetables, including *Momoedica balsamina*, spices, such as cumin and pepper, and oils extracted from aromatic plants are known for being an exceptional source of EPIs (Karumathil et al., 2018; Li et al., 2019; Tariq et al., 2019; Solnier et al., 2020; Muniz et al., 2021); further, the application of flavonoids such as flavonolignans in combating MDR *via* inhibiting efflux pump systems have been described previously (Chambers et al., 2019).

However, incidences of another resistance mechanism associated with membrane-embedded MDR efflux pumps which can extrude a wide range of noxious compounds such as antibiotics, dyes, and biocides have been reported (Mitchell et al., 2019; Jesin et al., 2020), which can be a promising novel approach for tackling the increasing risk of antibiotic-resistant bacteria. Based on previous studies, all the bacterial species possess MDR efflux pumps, conferring intrinsic resistance against antibiotics (Bay and Turner, 2009), and six different superfamilies of efflux pumps have already been reported, but mostly, the small multidrug resistance efflux pump family was studied extensively (Sun et al., 2014). In hospitals, the commonly used disinfectants, the quaternary ammonium compounds (QACs), toxic compounds such as ethidium bromide, and many other compounds such as biocides, cetyltrimethylammonium bromide (CTAB), and benzalkonium chloride (BZK) are utilized as a substrate for acquiring energy by the SMRs family members (Bay et al., 2008). A study by Kadry et al. (2017) revealed the existence of a strong correlation between the gene encoding SMR protein with the enhanced level of resistance toward compounds such as benzalkonium chloride in clinical isolates of *P. aeruginosa*.

Conclusions

In this review study, we have discussed the important role played by efflux pump family members in developing resistance against single or multiple antibiotics as well as against various compounds other than the antimicrobial agents including heavy metals, preservatives, and toxins. Apart from different classes of antibiotics, efflux pumps also provide resistance against disinfectants and virulence factors and even develop cross-resistance against other different noxious compounds, which further leads to the development of a multidrug-resistant phenotype. The working mechanism along with the regulation of an efflux pump system is an intricate process that actively participates in different biological processes in bacteria such as biofilm formation, quorum sensing, and adhesion of bacteria to host, and their progressive multiplication have been highlighted in this review. This review further highlights the importance of the discovery of novel EPIs, which should be plant-based, safe, and non-toxic, and many efflux pump inhibitors that are under clinical trial as well. Efflux pump inhibitors derived from plants are of great importance and are studied widely presently. However, the variations in function and regulatory mechanisms of the efflux pump systems along with the progress

and utilization of a variety of efflux pump inhibitors still need further exploration. According to the progressive research work, it has been suggested that TetR family members can be studied further for the discovery of antibiotic residues associated with multidrug resistance. In addition to the role of the aforementioned active regulatory proteins, some other proteins with an ability to identify single or multiple substrates and specific substrate-binding domains that can be utilized to identify antibiotic residues involved in substrate binding have been reported earlier. Therefore, other than the TetR regulatory protein, the detection of residues of antibiotics and transcriptional regulatory protein requires the utilization of efflux pump concomitant genes. Additionally, the proteins with a high substrate binding specificity toward diverse antimicrobials with more appropriate features are very demanding. Moreover, this review discussed the association of the expressional levels of efflux pump systems with various other defense mechanisms such as biofilm formation and quorum sensing, which provides novel insights for expanding the extent of fundamental research. In this article, we provided an overview of efflux genes and significant transporters indicating an area which requires persistent exploration to provide supervision for clinical practices.

Author contributions

SC was involved in the literature review, writing of the original draft, figure illustrations, conceptualization, preparing the draft of the manuscript, reviewing, editing, revising the manuscript, has contributed to the article, and approved the submitted version.

Acknowledgments

SC would like to acknowledge the support of management, Dr. S. Sumaya (Principal, Thassim Beevi College, Kilakarai, Tamil Nadu, India) and Dr. M. S. Irfan Ahmed (Director—Research and Industry–Institute Relations) for providing a stimulating environment.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ababou, A., and Koronakis, V. (2016). Structures of gate loop variants of the AcrB drug efflux pump bound by erythromycin substrate. *PLoS ONE* 11, e159154. doi: 10.1371/journal.pone.0159154

Adamson, D. H., Krikstopaityte, V., and Coote, P. J. (2015). Enhanced efficacy of putative efflux pump inhibitor/antibiotic combination treatments versus MDR strains of *Pseudomonas aeruginosa* in a Galleria mellonella *in vivo* infection model. *J. Antimicrob. Chemother.* 70, 2271–2278. doi: 10.1093/jac/dkv111

Aghayan, S. S., Mogadam, H. K., Fazli, M., Darban-Sarokhalil, D., Khoramrooz, S. S., Jabalameli, F., et al. (2017). The effects of berberine and palmatine on efflux pumps inhibition with different gene patterns in *Pseudomonas aeruginosa* isolated from burn infections. *Avicenna J. Med. Biotechnol.* 9, 2.

Alav, I., Sutton, J. M., and Rahman, K. M. (2018). Role of bacterial efflux pumps in biofilm formation. *J. Antimicrob. Chemother.* 73, 2003–2020. doi: 10.1093/jac/dky042

Alcalde-Rico, M., Hernando-Amado, S., Blanco, P., and Martínez, J. L. (2016). Multidrug efflux pumps at the crossroad between antibiotic resistance and bacterial virulence. *Front. Microbiol.* 7, 1483. doi: 10.3389/fmicb.2016.01483

Alibert, S., N'gompaza Diarra, J., Hernandez, J., Stutzmann, A., Fouad, M., Boyer, G., et al. (2017). Multidrug efflux pumps and their role in antibiotic and antiseptic resistance: a pharmacodynamic perspective. *Expert Opin. Drug Metab. Toxicol.* 13, 301–309 doi: 10.1080/17425255.2017.1251581

Allen, H. K., Donato, J., Wang, H. H., Cloud-Hansen, K. A., Davies, J., and Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* 8, 251–259. doi: 10.1038/nrmicro2312

Alm, R. A., Johnstone, M. R., and Lahiri, S. D. (2015). Characterization of *Escherichia coli* NDM isolates with decreased susceptibility to aztreonam/avibactam: role of a novel insertion in PBP3. *J. Antimicrob. Chemother.* 70, 1420–1428. doi: 10.1093/jac/dku568

Anandapadamanaban, M., Pilstål, R., Andresen, C., Trewhella, J., Moche, M., Wallner, B., et al. (2016). Mutation-induced population shift in the MexR conformational ensemble disengages DNA binding: a novel mechanism for MarR family derepression. *Structure* 24, 1311–1321. doi: 10.1016/j.str.2016.06.008

Andersson, D. I., and Levin, B. R. (1999). The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* 2, 489–493. doi: 10.1016/s1369-5274(99)00005-3

Aparna, V., Dineshkumar, K., Mohanalakshmi, N., Velmurugan, D., and Hopper, W. (2014). Identification of natural compound inhibitors for multidrug efflux pumps of *Escherichia coli* and *Pseudomonas aeruginosa* using *in silico* high-throughput virtual screening and *in vitro* validation. *PLoS ONE* 9, e101840. doi: 10.1371/journal.pone.0101840

Aron, Z., and Opperman, T. J. (2016). Optimization of a novel series of pyranopyridine RND efflux pump inhibitors. *Curr. Opin. Microbiol.* 33, 1–6. doi: 10.1016/j.mib.2016.05.007

Aron, Z., and Opperman, T. J. (2018). The hydrophobic trap—the Achilles heel of RND efflux pumps. *Res. Microbiol.* 169, 393–400. doi: 10.1016/j.resmic.2017.11.001

Augustin, R. (2010). The protein family of glucose transport facilitators: it's not only about glucose after all. *IUBMB Life* 62, 315–333. doi: 10.1002/iub.315

Bailey, A. M., Paulsen, I. T., and Piddock, L. J. (2008). RamA confers multidrug resistance in *Salmonella enterica via* increased expression of acrB, which is inhibited by chlorpromazine. *Antimicrob. Agents Chemother.* 52, 3604–3611. doi: 10.1128/AAC.00661-08

Banigan, J. R., Gayen, A., Cho, M. K., and Traaseth, N. J. (2015). A structured loop modulates coupling between the substrate-binding and dimerization domains in the multidrug resistance transporter EmrE. *J. Biol. Chem.* 290, 805–814. doi: 10.1074/jbc.M114.601963

Bay, D. C., Rommens, K. L., and Turner, R. J. (2008). Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochim. Biophys. Acta. Biomembr.* 1778, 1814–1838. doi: 10.1016/j.bbamem.2007.08.015

Bay, D. C., and Turner, R. J. (2009). Diversity and evolution of the small multidrug resistance protein family. *BMC Evol. Biol.* 9, 1–27. doi: 10.1186/1471-2148-9-140

Baylay, A. J., and Piddock, L. J. (2015). Clinically relevant fluoroquinolone resistance due to constitutive overexpression of the PatAB ABC transporter in *Streptococcus pneumoniae* is conferred by disruption of a transcriptional attenuator. *J. Antimicrob. Chemother.* 70, 670–679. doi: 10.1093/jac/dku449

Bhakdi, S., Mackman, N., Menestrina, G., Gray, L., Hugo, F., Seeger, W., et al. (1988). The hemolysin of *Escherichia coli*. *Eur. J. Epidemiol.* 4, 135–143. doi: 10.1007/BF00144740

Blair, J., Webber, M. A., Baylay, A. J., Ogbolu, D. O., and Piddock, L. J. (2015). Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* 13, 42–51. doi: 10.1038/nrmicro3380

Blair, J. M., and Piddock, L. J. (2016). How to measure export *via* bacterial multidrug resistance efflux pumps. *MBio* 7, e00840-16. doi: 10.1128/mBio.00840-16

Blair, J. M., Richmond, G. E., and Piddock, L. J. (2014). Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future Microbiol.* 9, 1165–1177. doi: 10.2217/fmb.14.66

Blanchard, C., Barnett, P., Perlmutter, J., and Dunman, P. M. (2014). Identification of *Acinetobacter baumannii* serum-associated antibiotic efflux pump inhibitors. *Antimicrob. Agents Chemother.* 58, 6360–6370. doi: 10.1128/AAC.03535-14

Blanco, P., Hernando-Amado, S., Reales-Calderon, J. A., Corona, F., Lira, F., Alcalde-Rico, M., et al. (2016). Bacterial multidrug efflux pumps: much more than antibiotic resistance determinants. *Microorganisms* 4, 14. doi: 10.3390/microorganisms4010014

Bohnert, J. A., and Kern, W. V. (2005). Selected arylpiperazines are capable of reversing multidrug resistance in *Escherichia coli* overexpressing RND efflux pumps. *Antimicrob. Agents Chemother.* 49, 849–852. doi: 10.1128/AAC.49.2.849-852.2005

Brown, M. H., Paulsen, I. T., and Skurray, R. A. (1999). The multidrug efflux protein NorM is a prototype of a new family of transporters. *Mol. Microbiol.* 31, 394–395. doi: 10.1046/j.1365-2958.1999.01162.x

Buffet-Bataillon, S., Tattevin, P., Maillard, J. Y., Bonnaure-Mallet, M., and Jolivet-Gougeon, A. (2016). Efflux pump induction by quaternary ammonium compounds and fluoroquinolone resistance in bacteria. *Future Microbiol.* 11, 81–92. doi: 10.2217/fmb.15.131

Cannon, R. D., and Holmes, A. R. (2015). Learning the ABC of oral fungal drug resistance. *Mol. Oral Microbiol.* 30, 425–437. doi: 10.1111/omi.12109

Chambers, C. S., Viktorová, J., Rehorová, K., Biedermann, D., Turková, L., Macek, T., et al. (2019). Defying multidrug resistance! Modulation of related transporters by flavonoids and flavonolignans. *J. Agric. Food Chem.* 68, 1763–1779. doi: 10.1021/acs.jafc.9b00694

Chan, Y. Y., Ong, Y. M., and Chua, K. L. (2007). Synergistic interaction between phenothiazines and antimicrobial agents against *Burkholderia pseudomallei*. *Antimicrob. Agents Chemother.* 51, 623–630. doi: 10.1128/AAC.01033-06

Chen, H., Cao, J., Zhou, C., Liu, H., Zhang, X., and Zhou, T. (2017). Biofilm formation restrained by subinhibitory concentrations of tigecyclin in *Acinetobacter baumannii* is associated with downregulation of efflux pumps. *Chemotherapy* 62, 128–133. doi: 10.1159/000450537

Chetri, S., Bhowmik, D., Dhar, D., Chakravarty, A., and Bhattacharjee, A. (2019b). Effect of concentration gradient carbapenem exposure on expression of blaNDM-1 and acrA in carbapenem resistant *Escherichia coli*. *Infect. Genet. Evol.* 73, 332–336. doi: 10.1016/j.meegid.2019.05.024

Chetri, S., Bhowmik, D., Paul, D., Pandey, P., Chanda, D. D., Chakravarty, A., et al. (2019a). AcrAB-TolC efflux pump system plays a role in carbapenem non-susceptibility in *Escherichia coli*. *BMC. microbial.* 19, 1–7. doi: 10.1186/s12866-019-1589-1

Chetri, S., Das, B. J., Bhowmik, D., Chanda, D. D., Chakravarty, A., and Bhattacharjee, A. (2020). Transcriptional response of mar, sox and rob regulon against concentration gradient carbapenem stress within *Escherichia coli* isolated from hospital acquired infection. *BMC Res. Notes.* 13, 1–7. doi: 10.1186/s13104-020-04999-2

Chetri, S., Dolley, A., Bhowmik, D., Chanda, D. D., Chakravarty, A., and Bhattacharjee, A. (2018a). Transcriptional response of AcrEF-TolC against fluoroquinolone and carbapenem in *Escherichia coli* of clinical origin. *Indian J. Med. Microbiol.* 36, 537–540. doi: 10.4103/ijmm.IJMM_18_308

Chetri, S., Singha, K., Bhowmik, D., Chanda, D. D., Chakravarty, A., and Bhattacharjee, A. (2018b). Sub-inhibitory concentration of ertapenem induces overexpression of regulator of antibiotic resistance A in *Escherichia coli*. *Indian J. Med. Microbiol.* 36, 569–572. doi: 10.4103/ijmm.IJMM_18_436

Chetri, S., Singha, M., Bhowmik, D., Nath, K., Chanda, D. D., Chakravarty, A., et al. (2019c). Transcriptional response of OmpC and OmpF in *Escherichia coli* against differential gradient of carbapenem stress. *BMC Res. Notes.* 12, 1–6. doi: 10.1186/s13104-019-4177-4

Chevalier, J., Atifi, S., Eyraud, A., Mahamoud, A., Barbe, J., and Pagès, J. M. (2001). New pyridoquinoline derivatives as potential inhibitors of the fluoroquinolone efflux pump in resistant enterobacter a erogenes strains. *J. Med. Chem.* 44, 4023–4026. doi: 10.1021/jm010911z

Chevalier, J., Mahamoud, A., Baitiche, M., Adam, E., Viveiros, M., Smarandache, A., et al. (2010). Quinazoline derivatives are efficient chemosensitizers of antibiotic activity in Enterobacter aerogenes, Klebsiella pneumoniae and Pseudomonas aeruginosa resistant strains. *Int. J. Antimicrob. Agents.* 36, 164–168. doi: 10.1016/j.ijantimicag.2010.03.027

Chitemerere, T. A., and Mukanganyama, S. (2014). Evaluation of cell membrane integrity as a potential antimicrobial target for plant products. *BMC Complement. Altern. Med.* 14, 1–8. doi: 10.1186/1472-6882-14-278

Choudhury, H. G., Tong, Z., Mathavan, I., Li, Y., Iwata, S., Zirah, S., et al. (2014). Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state. *Proc. Natl. Acad. Sci. USA* 111, 9145–9150. doi: 10.1073/pnas.1320506111

Colque, C. A., Albarracín Orio, A. G., Feliziani, S., Marvig, R. L., Tobares, A. R., Johansen, H. K., et al. (2020). Hypermutator *Pseudomonas aeruginosa* exploits multiple genetic pathways to develop multidrug resistance during long-term infections in the airways of cystic fibrosis patients. *Antimicrob. Agents Chemother.* 64, e02142-19. doi: 10.1128/AAC.02142-19

Cooper, M. A., and Shlaes, D. (2011). Fix the antibiotics pipeline. *Nature* 472, 32–32. doi: 10.1038/472032a

Cortez-Cordova, J., and Kumar, A. (2011). Activity of the efflux pump inhibitor phenylalanine-arginine β-naphthylamide against the AdeFGH pump of *Acinetobacter baumannii*. *Int. J. Antimicrob. Agents.* 37, 420–424. doi: 10.1016/j.ijantimicag.2011.01.006

Cura, A. J., and Carruthers, A. (2012). The role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism and homeostasis. *Compr. Physiol.* 2, 863. doi: 10.1002/cphy.c110024

Daury, L., Orange, F., Taveau, J. C., Verchere, A., Monlezun, L., Gounou, C., et al. (2016). Tripartite assembly of RND multidrug efflux pumps. *Nat. Commun.* 7, 1–8. doi: 10.1038/ncomms10731

Dawson, R. J., and Locher, K. P. (2006). Structure of a bacterial multidrug ABC transporter. *Nature*. 443, 180–185. doi: 10.1038/nature05155

Delmar, J. A., Su, C. C., and Yu, E. W. (2014). Bacterial multidrug efflux transporters. *Annu. Rev. Biophys.* 43, 93–117. doi: 10.1146/annurev-biophys-051013-022855

Donner, J., Reck, M., Bunk, B., Jarek, M., App, C. B., Meier-Kolthoff, J. P., et al. (2017). The biofilm inhibitor carolacton enters gram-negative cells: studies using a TolC-deficient strain of *Escherichia coli*. *Msphere* 2, e00375-17. doi: 10.1128/mSphereDirect.00375-17

Dou, Y., Song, F., Guo, F., Zhou, Z., Zhu, C., Xiang, J., et al. (2017). *Acinetobacter baumannii* quorum-sensing signalling molecule induces the expression of drug-resistance genes. *Mol. Med. Rep.* 15, 4061–4068. doi: 10.3892/mmr.2017.6528

Du, D., van Veen, H. W., and Luisi, B. F. (2015). Assembly and operation of bacterial tripartite multidrug efflux pumps. *Trends Microbiol.* 23, 311–319. doi: 10.1016/j.tim.2015.01.010

Du, D., Wang, Z., James, N. R., Voss, J. E., Klimont, E., Ohene-Agyei, T., et al. (2014). Structure of the AcrAB–TolC multidrug efflux pump. *Nature* 509, 512–515. doi: 10.1038/nature13205

Du, D., Wang-Kan, X., Neuberger, A., Van Veen, H. W., Pos, K. M., Piddock, L. J., et al. (2018). Multidrug efflux pumps: structure, function and regulation. *Nat. Rev. Microbiol.* 16, 523–539. doi: 10.1038/s41579-018-0048-6

Dwivedi, G. R., Maurya, A., Yadav, D. K., Khan, F., Darokar, M. P., and Srivastava, S. K. (2015). Drug resistance reversal potential of ursolic acid derivatives against nalidixic acid-and multidrug-resistant *Escherichia coli*. *Chem. Biol. Drug Des.* 86, 272–283. doi: 10.1111/cbdd.12491

Dwivedi, G. R., Upadhyay, H. C., Yadav, D. K., Singh, V., Srivastava, S. K., Khan, F., et al. (2014). 4-Hydroxy-α-tetralone and its derivative as drug resistance reversal agents in multi drug resistant *Escherichia coli*. *Chem. Biol. Drug Des.* 83, 482–492. doi: 10.1111/cbdd.12263

Eicher, T., Cha, H. J., Seeger, M. A., Brandstätter, L., El-Delik, J., Bohnert, J. A., et al. (2012). Transport of drugs by the multidrug transporter AcrB involves an access and a deep binding pocket that are separated by a switch-loop. *Proc. Natl. Acad. Sci. USA* 109, 5687–5692. doi: 10.1073/pnas.1114944109

Eicher, T., Seeger, M. A., Anselmi, C., Zhou, W., Brandstätter, L., Verrey, F., et al. (2014). Coupling of remote alternating-access transport mechanisms for protons and substrates in the multidrug efflux pump AcrB. *Elife* 3, e03145. doi: 10.7554/eLife.03145

Elbourne, L. D., Tetu, S. G., Hassan, K. A., and Paulsen, I. T. (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic. Acids. Res.* 45, D320–D324. doi: 10.1093/nar/gkw1068

Fitzpatrick, A. W., Llabrés, S., Neuberger, A., Blaza, J. N., Bai, X. C., Okada, U., et al. (2017). Structure of the MacAB–TolC ABC-type tripartite multidrug efflux pump. *Nat. Microbiol.* 2, 1–8. doi: 10.1038/nmicrobiol.2017.70

Fluman, N., Adler, J., Rotenberg, S. A., Brown, M. H., and Bibi, E. (2014). Export of a single drug molecule in two transport cycles by a multidrug efflux pump. *Nat. Commun.* 5, 4615. doi: 10.1038/ncomms5615

Forrest, L. R., Krämer, R., and Ziegler, C. (2011). The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta. Bioenerg.* 1807, 167–188. doi: 10.1016/j.bbabio.2010.10.014

Gallo, S., Chevalier, J., Mahamoud, A., Eyraud, A., Pagès, J. M., and Barbe, J. (2003). 4-alkoxy and 4-thioalkoxyquinoline derivatives as chemosensitizers for the

chloramphenicol-resistant clinical Enterobacter aerogenes 27 strain. *Int. J. Antimicrob. Agents.* 22, 270–273. doi: 10.1016/S0924-8579(03)00215-2

Ganas, P., Mihasan, M., Igloi, G. L., and Brandsch, R. (2007). A two-component small multidrug resistance pump functions as a metabolic valve during nicotine catabolism by Arthrobacter nicotinovorans. *Microbiology* 153, 1546–1555. doi: 10.1099/mic.0.2006/004234-0

Garvey, M. I., Baylay, A. J., Wong, R. L., and Piddock, L. J. (2011a). Overexpression of patA and patB, which encode ABC transporters, is associated with fluoroquinolone resistance in clinical isolates of *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 55, 190–196. doi: 10.1128/AAC.00672-10

Garvey, M. I., Rahman, M. M., Gibbons, S., and Piddock, L. J. (2011b). Medicinal plant extracts with efflux inhibitory activity against Gram-negative bacteria. *Int. J. Antimicrob. Agents.* 37, 145–151.doi: 10.1016/j.ijantimicag.2010.10.027

Ghisalberti, D., Mahamoud, A., Chevalier, J., Baitiche, M., Martino, M., Pagès, J. M., et al. (2006). Chloroquinolines block antibiotic efflux pumps in antibiotic-resistant *Enterobacter aerogenes* isolates. *Int. J. Antimicrob. Agents.* 27, 565–569. doi: 10.1016/j.ijantimicag.2006.03.010

Gillings, M. R. (2017). Class 1 integrons as invasive species. *Curr. Opin. Microbiol.* 38, 10–15. doi: 10.1016/j.mib.2017.03.002

Gilson, L., Mahanty, H. K., and Kolter, R. (1990). Genetic analysis of an MDR-like export system: the secretion of colicin V. *EMBO J.* 9, 3875–3884. doi: 10.1002/j.1460-2075.1990.tb07606.x

Grinius, L. L., and Goldberg, E. B. (1994). Bacterial multidrug resistance is due to a single membrane protein which functions as a drug pump. *J. Biol. Chem.* 269, 29998–30004. doi: 10.1016/S0021-9258(18)43980-4

Grkovic, S., Brown, M. H., and Skurray, R. A. (2002). Regulation of bacterial drug export systems. *Microbiol. Mol. Biol. Rev.* 66, 671–701. doi: 10.1128/MMBR.66.4.671-701.2002

Guan, L., and Kaback, H. R. (2006). Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* 35, 67. doi: 10.1146/annurev.biophys.35.040405.102005

Guérin, F., Galimand, M., Tuambilangana, F., Courvalin, P., and Cattoir, V. (2014). Overexpression of the novel MATE fluoroquinolone efflux pump FepA in Listeria monocytogenes is driven by inactivation of its local repressor FepR. *PLoS ONE* 9, e106340. doi: 10.1371/journal.pone.0106340

Gupta, D., Singh, A., and Khan, A. U. (2017). Nanoparticles as efflux pump and biofilm inhibitor to rejuvenate bactericidal effect of conventional antibiotics. *Nanoscale Res. Lett.* 12, 1–6. doi: 10.1186/s11671-017-2222-6

Hall, C. W., and Mah, T. F. (2017). Molecular mechanisms of biofilm-based antibiotic resistance and tolerance in pathogenic bacteria. *FEMS Microbiol. Rev.* 41, 276–301. doi: 10.1093/femsre/fux010

Handzlik, J., Szymańska, E., Chevalier, J., Otrebska, E., Kieć-Kononowicz, K., Pagès, J. M., et al. (2011). Amine–alkyl derivatives of hydantoin: new tool to combat resistant bacteria. *Eur. J. Med. Chem.* 46, 5807–5816. doi: 10.1016/j.ejmech.2011.09.032

Hassan, K. A., Elbourne, L. D., Li, L., Gamage, H. K., Liu, Q., Jackson, S. M., et al. (2015b). An ace up their sleeve: a transcriptomic approach exposes the AceI efflux protein of *Acinetobacter baumannii* and reveals the drug efflux potential hidden in many microbial pathogens. *Front. Microbiol.* 6, 333. doi: 10.3389/fmicb.2015.00333

Hassan, K. A., Jackson, S. M., Penesyan, A., Patching, S. G., Tetu, S. G., Eijkelkamp, B. A., et al. (2013). Transcriptomic and biochemical analyses identify a family of chlorhexidine efflux proteins. *Proc. Natl. Acad. Sci. USA* 110, 20254–20259. doi: 10.1073/pnas.1317052110

Hassan, K. A., Liu, Q., Elbourne, L. D., Ahmad, I., Sharples, D., Naidu, V., et al. (2018). Pacing across the membrane: the novel PACE family of efflux pumps is widespread in Gram-negative pathogens. *Res. Microbiol.* 169, 450–454. doi: 10.1016/j.resmic.2018.01.001

Hassan, K. A., Liu, Q., Henderson, P. J., and Paulsen, I. T. (2015a). Homologs of the *Acinetobacter baumannii* AceI transporter represent a new family of bacterial multidrug efflux systems. *MBio* 6, e01982–14. doi: 10.1128/mBio.01982-14

He, X., Szewczyk, P., Karyakin, A., Evin, M., Hong, W. X., Zhang, Q., et al. (2010). Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature* 467, 991–994. doi: 10.1038/nature09408

Hellmich, U. A., Lyubenova, S., Kaltenborn, E., Doshi, R., van Veen, H. W., Prisner, T. F., et al. (2012). Probing the ATP hydrolysis cycle of the ABC multidrug transporter LmrA by pulsed EPR spectroscopy. *J. Am. Chem. Soc.* 134, 5857–5862. doi: 10.1021/ja211007t

Hellmich, U. A., Mönkemeyer, L., Velamakanni, S., van Veen, H. W., and Glaubitz, C. (2015). Effects of nucleotide binding to LmrA: a combined MAS-NMR and solution NMR study. *Biochim. Biophys. Acta- Biomembr.* 1848, 3158–3165. doi: 10.1016/j.bbamem.2015.10.003

Hernando-Amado, S., Blanco, P., Alcalde-Rico, M., Corona, F., Reales-Calderón, J. A., Sánchez, M. B., et al. (2016). Multidrug efflux pumps as main players in intrinsic and acquired resistance to antimicrobials. *Drug Resist. Updat.* 28, 13–27. doi: 10.1016/j.drup.2016.06.007

Hirai, T., Heymann, J. A., Shi, D., Sarker, R., Maloney, P. C., and Subramaniam, S. (2002). Three-dimensional structure of a bacterial oxalate transporter. *Nat. Struct. Biol.* 9, 597–600. doi: 10.1038/nsb821

Hung, L. W., Kim, H. B., Murakami, S., Gupta, G., Kim, C. Y., and Terwilliger, T. C. (2013). Crystal structure of AcrB complexed with linezolid at 3.5 Å resolution. *J. Struct. Funct. Genomics.* 14, 71–75. doi: 10.1007/s10969-013-9154-x

Iannelli, F., Santagati, M., Santoro, F., Oggioni, M. R., Stefani, S., and Pozzi, G. (2014). Nucleotide sequence of conjugative prophage *Φ*1207. 3 (formerly Tn 1207.3) carrying the mef (A)/msr (D) genes for efflux resistance to macrolides in Streptococcus pyogenes. *Front. Microbiol.* 5, 687. doi: 10.3389/fmicb.2014.00687

Ito, S., Kusuhara, H., Yokochi, M., Toyoshima, J., Inoue, K., Yuasa, H., et al. (2012). Competitive inhibition of the luminal efflux by multidrug and toxin extrusions, but not basolateral uptake by organic cation transporter 2, is the likely mechanism underlying the pharmacokinetic drug-drug interactions caused by cimetidine in the kidney. *J. Pharmacol. Exp. Ther.* 340, 393–403. doi: 10.1124/jpet.111.184986

Jack, D. L., Storms, M. L., Tchieu, J. H., Paulsen, I. T., and Saier Jr, M. H. (2000). A broad-specificity multidrug efflux pump requiring a pair of homologous SMR-type proteins. *J. Bacteriol.* 182, 2311–2313. doi: 10.1128/JB.182.8.2311-2313.2000

Jaglic, Z., and Cervinkova, D. (2012). Genetic basis of resistance to quaternary ammonium compounds–the qac genes and their role: a review. *Vet. Med.* 57, 275–281. doi: 10.17221/6013-VETMED

Jamshidi, S., Sutton, J. M., and Rahman, K. M. (2016). An overview of bacterial efflux pumps and computational approaches to study efflux pump inhibitors. *Future Med. Chem.* 8, 195–210. doi: 10.4155/fmc.15.173

Jamshidi, S., Sutton, J. M., and Rahman, K. M. (2018). Mapping the dynamic functions and structural features of acrb efflux pump transporter using accelerated molecular dynamics simulations. *Sci. Rep.* 8, 1–13. doi: 10.1038/s41598-018-28531-6

Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature* 211, 969–970. doi: 10.1038/211969a0

Jesin, J. A., Stone, T. A., Mitchell, C. J., Reading, E., and Deber, C. M. (2020). Peptide-based approach to inhibition of the multidrug resistance efflux pump AcrB. *Biochemistry* 59, 3973–3981. doi: 10.1021/acs.biochem.0c00417

Jin, M. S., Oldham, M. L., Zhang, Q., and Chen, J. (2012). Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. *Nature* 490, 566–569. doi: 10.1038/nature11448

Jo, I., Hong, S., Lee, M., Song, S., Kim, J. S., Mitra, A. K., et al. (2017). Stoichiometry and mechanistic implications of the MacAB-TolC tripartite efflux pump. *Biochem. Biophys. Res. Commun.* 494, 668–673. doi: 10.1016/j.bbrc.2017.10.102

Johnson, Z. L., and Chen, J. (2017). Structural basis of substrate recognition by the multidrug resistance protein MRP1. *Cell.* 168, 1075–1085. doi: 10.1016/j.cell.2017.01.041

Kaback, H. R. (2015). A chemiosmotic mechanism of symport. *Proc. Natl. Acad. Sci. USA* 112, 1259–1264. doi: 10.1073/pnas.1419325112

Kadry, A. A., Serry, F. M., El-Ganiny, A. M., and El-Baz, A. M. (2017). Integron occurrence is linked to reduced biocide susceptibility in multidrug resistant *Pseudomonas aeruginosa*. *Br. J. Biomed. Sci.* 74, 78–84. doi: 10.1080/09674845.2017.1278884

Karumathil, D. P., Nair, M. S., Gaffney, J., Kollanoor-Johny, A., and Venkitanarayanan, K. (2018). Trans-cinnamaldehyde and eugenol increase *Acinetobacter baumannii* sensitivity to beta-lactam antibiotics. *Front. Microbiol.* 9, 1011. doi: 10.3389/fmicb.2018.01011

Kermani, A. A., Macdonald, C. B., Burata, O. E., Ben Koff, B., Koide, A., Denbaum, E., et al. (2020). The structural basis of promiscuity in small multidrug resistance transporters. *Nat. Commun.* 11, 1–9. doi: 10.1038/s41467-020-19820-8

Kermani, A. A., Macdonald, C. B., Gundepudi, R., and Stockbridge, R. B. (2018). Guanidinium export is the primal function of SMR family transporters. *Proc. Natl. Acad. Sci. USA* 115, 3060–3065. doi: 10.1073/pnas.1719187115

Kern, W. V., Steinke, P., Schumacher, A., Schuster, S., Baum, H. V., and Bohnert, J. A. (2006). Effect of 1-(1-naphthylmethyl)-piperazine, a novel putative efflux pump inhibitor, on antimicrobial drug susceptibility in clinical isolates of *Escherichia coli*. *J. Antimicrob. Chemother.* 57, 339–343. doi: 10.1093/jac/dki445

Kesherwani, M., Michael Gromiha, M., Fukui, K., and Velmurugan, D. (2017). Identification of novel natural inhibitor for NorM—a multidrug and toxic compound extrusion transporter—an insilico molecular modeling and simulation studies. *J. Biomol. Struct. Dyn.* 35, 58–77. doi: 10.1080/07391102.2015.1132391

Khare, D., Oldham, M. L., Orelle, C., Davidson, A. L., and Chen, J. (2009). Alternating access in maltose transporter mediated by rigid-body rotations. *Mol. Cell.* 33, 528–536. doi: 10.1016/j.molcel.2009.01.035

Kim, Y. C., and Hummer, G. (2012). Proton-pumping mechanism of cytochrome c oxidase: a kinetic master-equation approach. *Biochim. Biophys. Acta. Bioenerg.* 1817, 526–536. doi: 10.1016/j.bbabio.2011.09.004

Kinana, A. D., Vargiu, A. V., May, T., and Nikaido, H. (2016). Aminoacyl β-naphthylamides as substrates and modulators of AcrB multidrug efflux pump. *Proc. Natl. Acad. Sci. USA* 113, 1405–1410. doi: 10.1073/pnas.1525143113

Kodan, A., Yamaguchi, T., Nakatsu, T., Sakiyama, K., Hipolito, C. J., Fujioka, A., et al. (2014). Structural basis for gating mechanisms of a eukaryotic P-glycoprotein homolog. *Proc. Natl. Acad. Sci. USA* 111, 4049–4054. doi: 10.1073/pnas.1321562111

Kong, E. F., Tsui, C., Kucharíková, S., Van Dijck, P., and Jabra-Rizk, M. A. (2017). Modulation of *Staphylococcus aureus* response to antimicrobials by the *Candida albicans* quorum sensing molecule farnesol. *Antimicrob. Agents. Chemother.* 61, e01573-17. doi: 10.1128/AAC.01573-17

Korkhov, V. M., Mireku, S. A., and Locher, K. P. (2012). Structure of AMP-PNP-bound vitamin B12 transporter BtuCD–F. *Nature*. 490, 367–372. doi: 10.1038/nature11442

Kurinčič, M., Klančnik, A., and Smole Možina, S. (2012). Effects of efflux pump inhibitors on erythromycin, ciprofloxacin, and tetracycline resistance in *Campylobacter* spp. isolates. *Microb. Drug Resist.* 18, 492–501. doi: 10.1089/mdr.2012.0017

Kuroda, T., and Tsuchiya, T. (2009). Multidrug efflux transporters in the MATE family. *Biochim. Biophys. Acta. Proteins Proteom*. 1794, 763–768. doi: 10.1016/j.bbapap.2008.11.012

Kusakizako, T., Claxton, D. P., Tanaka, Y., Maturana, A. D., Kuroda, T., Ishitani, R., et al. (2019). Structural basis of H+-dependent conformational change in a bacterial MATE transporter. *Structure* 27, 293–301. doi: 10.1016/j.str.2018.10.004

Laudadio, E., Cedraro, N., Mangiaterra, G., Citterio, B., Mobbili, G., Minnelli, C., et al. (2019). Natural alkaloid berberine activity against *Pseudomonas aeruginosa* MexXY-mediated aminoglycoside resistance: *in silico* and *in vitro* studies. *J. Nat. Prod.* 82, 1935–1944. doi: 10.1021/acs.jnatprod.9b00317

Lee, M. D., Galazzo, J. L., Staley, A. L., Lee, J. C., Warren, M. S., Fuernkranz, H., et al. (2001). Microbial fermentation-derived inhibitors of efflux-pump-mediated drug resistance. *IL Farmaco* 56, 81–85. doi: 10.1016/S0014-827X(01)01002-3

Lerma, L. L., Benomar, N., Valenzuela, A. S., Muñoz, M. D. C. C., Gálvez, A., and Abriouel, H. (2014). Role of EfrAB efflux pump in biocide tolerance and antibiotic resistance of *Enterococcus faecalis* and Enterococcus faecium isolated from traditional fermented foods and the effect of EDTA as EfrAB inhibitor. *Food Microbiol.* 44, 249–257. doi: 10.1016/j.fm.2014.06.009

Levy, S. B., and Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.* 10, S122–S129. doi: 10.1038/nm1145

Li, B., Yao, Q., Pan, X. C., Wang, N., Zhang, R., Li, J., et al. (2011). Artesunate enhances the antibacterial effect of β-lactam antibiotics against *Escherichia coli* by increasing antibiotic accumulation *via* inhibition of the multidrug efflux pump system AcrAB-TolC. *J. Antimicrob. Chemother.* 66, 769–777. doi: 10.1093/jac/dkr017

Li, J., Liu, D., Tian, X., Koseki, S., Chen, S., Ye, X., et al. (2019). Novel antibacterial modalities against methicillin resistant *Staphylococcus aureus* derived from plants. *Crit. Rev. Food Sci. Nutr.* 59, S153–S161. doi: 10.1080/10408398.2018.1541865

Li, N., Wu, J. X., Ding, D., Cheng, J., Gao, N., and Chen, L. (2017). Structure of a pancreatic ATP-sensitive potassium channel. *Cell*. 168, 101–110. doi: 10.1016/j.cell.2016.12.028

Li, X. Z., Plésiat, P., and Nikaido, H. (2015). The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin. Microbiol. Rev.* 28, 337–418. doi: 10.1128/CMR.00117-14

Liang, Z. B., Chen, Y. M., Chen, Y. F., Cheng, Y. Y., and Zhang, L. H. (2016). RND efflux pump and its interrelationship with quorum sensing system. *FEMS Microbiol. Lett.* 38, 894–901. doi: 10.16288/j.yczz.16-139

Lin, M. F., Lin, Y. Y., Tu, C. C., and Lan, C. Y. (2017). Distribution of different efflux pump genes in clinical isolates of multidrug-resistant *Acinetobacter baumannii* and their correlation with antimicrobial resistance. *J. Microbiol. Immunol. Infect.* 50, 224–231. doi: 10.1016/j.jmii.2015.04.004

Littlejohn, T. G., Paulsen, I. T., Gillespie, M. T., Tennent, J. M., Midgley, M., Jones, I. G., et al. (1992). Substrate specificity and energetics of antiseptic and disinfectant resistance in *Staphylococcus aureus*. *FEMS Microbiol. Lett*. 95, 259–265. doi: 10.1111/j.1574-6968.1992.tb05376.x

Locher, K. P. (2016). Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat. Struct. Mol. Biol*. 23, 487–493. doi: 10.1038/nsmb.3216

Lomovskaya, O., and Bostian, K. A. (2006). Practical applications and feasibility of efflux pump inhibitors in the clinic—a vision for applied use. *Biochem. Pharmacol.* 71, 910–918. doi: 10.1016/j.bcp.2005.12.008

Lomovskaya, O., Warren, M. S., Lee, A., Galazzo, J., Fronko, R., Lee, M. A. Y., et al. (2001). Identification and characterization of inhibitors of multidrug resistance efflux pumps in *Pseudomonas aeruginosa*: novel agents for combination therapy. *Antimicrob. Agents Chemother*. 45, 105–116. doi: 10.1128/AAC.45.1.105-116.2001

Long, F., Su, C. C., Lei, H. T., Bolla, J. R., Do, S. V., and Yu, E. W. (2012). Structure and mechanism of the tripartite CusCBA heavy-metal efflux complex. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 1047–1058. doi: 10.1098/rstb.2011.0203

Lu, M., Radchenko, M., Symersky, J., Nie, R., and Guo, Y. (2013b). Structural insights into H+-coupled multidrug extrusion by a MATE transporter. *Nat. Struct. Mol. Biol.* 20, 1310–1317. doi: 10.1038/nsmb.2687

Lu, M., Symersky, J., Radchenko, M., Koide, A., Guo, Y., Nie, R., et al. (2013a). Structures of a Na+-coupled, substrate-bound MATE multidrug transporter. *Proc. Natl. Acad. Sci. USA* 110, 2099–2104. doi: 10.1073/pnas.1219901110

Lu, S., and Zgurskaya, H. I. (2013). MacA, a periplasmic membrane fusion protein of the macrolide transporter MacAB-TolC, binds lipopolysaccharide core specifically and with high affinity. *J. Bacteriol*. 195, 4865–4872. doi: 10.1128/JB.00756-13

Lv, Z., Zhao, M., Wang, W., Wang, Q., Huang, M., Li, C., et al. (2021). Changing Gly311 to acidic amino acid in the MATE family protein DTX6 leads to enhanced resistance of *Arabidopsis* to the dihydropyridine herbicides. *Mol. Plant.* 14, 2115–2125. doi: 10.1016/j.molp.2021.09.002

Lytvynenko, I., Brill, S., Oswald, C., and Pos, K. M. (2016). Molecular basis of polyspecificity of the small multidrug resistance efflux pump AbeS from *Acinetobacter baumannii*. *J. Mol. Biol.* 428, 644–657. doi: 10.1016/j.jmb.2015.12.006

Mahamoud, A., Chevalier, J., Davin-Regli, A., and Barbe, J. (2006). Quinoline derivatives as promising inhibitors of antibiotic efflux pump in multidrug resistant Enterobacter aerogenes isolates. *Curr. Drug Targets*. 7, 843–847. doi: 10.2174/138945006777709557

Mahmood, Y. H., Jamshidi, S., Mark Sutton, J. M., and Rahman, K. (2016). Current advances in developing inhibitors of bacterial multidrug efflux pumps. *Curr. Med. Chem.* 23, 1062–1081. doi: 10.2174/0929867323666160304150522

Maisuria, V. B., Hosseinidoust, Z., and Tufenkji, N. (2015). Polyphenolic extract from maple syrup potentiates antibiotic susceptibility and reduces biofilm formation of pathogenic bacteria. *Appl. Environ. Microbiol.* 81, 3782–3792. doi: 10.1128/AEM.00239-15

Malléa, M., Chevalier, J., Eyraud, A., and Pagès, J. M. (2002). Inhibitors of antibiotic efflux pump in resistant Enterobacter aerogenes strains. *Biochem. Biophys. Res. Commun.* 293, 1370–1373. doi: 10.1016/S0006-291X(02)00404-7

Malléa, M., Mahamoud, A., Chevalier, J., Alibert-Franco, S., Brouant, P., Barbe, J. et al. (2003). Alkylaminoquinolines inhibit the bacterial antibiotic efflux pump in multidrug-resistant clinical isolates. *Biochem. J.* 376, 801–805. doi: 10.1042/BJ20030963

Mamelli, L., Amoros, J. P., Pagès, J. M., and Bolla, J. M. (2003). A phenylalanine–arginine β-naphthylamide sensitive multidrug efflux pump involved in intrinsic and acquired resistance of Campylobacter to macrolides. *Int. J. Antimicrob. Agents.* 22, 237–241. doi: 10.1016/S0924-8579(03)00199-7

Marger, M. D., and Saier Jr, M. H. (1993). A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem. Sci.* 18, 13–20. doi: 10.1016/0968-0004(93)90081-w

Martins, M., Dastidar, S. G., Fanning, S., Kristiansen, J. E., Molnar, J., Pages, J. M., et al. (2008). Potential role of non-antibiotics (helper compounds) in the treatment of multidrug-resistant Gram-negative infections: mechanisms for their direct and indirect activities. *Int. J. Antimicrob. Agents.* 31, 198–208. doi: 10.1016/j.ijantimicag.2007.10.025

Masuda, S., Terada, T., Yonezawa, A., Tanihara, Y., Kishimoto, K., Katsura, T., et al. (2006). Identification and functional characterization of a new human kidney–specific H+/organic cation antiporter, kidney-specific multidrug and toxin extrusion 2. *J. Am. Soc. Nephrol.* 17, 2127–2135. doi: 10.1681/ASN.2006030205

Maurya, A., Dwivedi, G. R., Darokar, M. P., and Srivastava, S. K. (2013). Antibacterial and synergy of clavine alkaloid lysergol and its derivatives against nalidixic acid-resistant *Escherichia coli*. *Chem. Biol. Drug Des.* 81, 484–490. doi: 10.1111/cbdd.12103

Mi, W., Li, Y., Yoon, S. H., Ernst, R. K., Walz, T., and Liao, M. (2017). Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature*. 549, 233–237. doi. 10.1038/nature23649

Mitchell, C. J., Stone, T. A., and Deber, C. M. (2019). Peptide-based efflux pump inhibitors of the small multidrug resistance protein from Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* 63, e00730–e00719. doi: 10.1128/AAC.00730-19

Miyauchi, H., Moriyama, S., Kusakizako, T., Kumazaki, K., Nakane, T., Yamashita, K., et al. (2017). Structural basis for xenobiotic extrusion by eukaryotic MATE transporter. *Nat. Commun.* 8, 1–11. doi: 10.1038/s41467-017-01541-0

Morita, Y., Kodama, K., Shiota, S., Mine, T., Kataoka, A., and Mizushima, T. (1998). NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrob. Agents Chemother.* 42, 1778–1782. doi: 10.1128/aac.42.7.1778

Mousa, J. J., Newsome, R. C., Yang, Y., Jobin, C., and Bruner, S. D. (2017). ClbM is a versatile, cation-promiscuous MATE transporter found in the colibactin biosynthetic gene cluster. *Biochem. Biophys. Res. Commun*. 482, 1233–1239. doi: 10.1016/j.bbrc.2016.12.018

Mousa, J. J., Yang, Y., Tomkovich, S., Shima, A., Newsome, R. C., Tripathi, P., et al. (2016). MATE transport of the *E. coli*-derived genotoxin colibactin. *Nat. Microbiol.* 1, 1–7. doi: 10.1038/nmicrobiol.2015.9

Mu, Y., Shen, Z., Jeon, B., Dai, L., and Zhang, Q. (2013). Synergistic effects of anti-CmeA and anti-CmeB peptide nucleic acids on sensitizing *Campylobacter jejuni* to antibiotics. *Antimicrob. Agents Chemother.* 57, 4575–4577. doi: 10.1128/AAC.00605-13

Munita, J. M., and Arias, C. A. (2016). Mechanisms of antibiotic resistance. *Microbiol. Spectr.* 4, 4–2. doi: 10.1128/microbiolspec.VMBF-0016-2015

Muniz, D. F., dos Santos Barbosa, C. R., de Menezes, I. R. A., de Sousa, E. O., Pereira, R. L. S., Júnior, J. T. C., et al. (2021). *In vitro* and *in silico* inhibitory effects of synthetic and natural eugenol derivatives against the NorA efflux pump in *Staphylococcus aureus*. *Food. Chem.* 337, 127776. doi: 10.1016/j.foodchem.2020.127776

Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T., and Yamaguchi, A. (2006). Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* 443, 173–179. doi: 10.1038/nature05076

Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* 419, 587–593. doi: 10.1038/nature01050

Nakashima, R., Sakurai, K., Yamasaki, S., Hayashi, K., Nagata, C., Hoshino, K., et al. (2013). Structural basis for the inhibition of bacterial multidrug exporters. *Nature* 500, 102–106. doi: 10.1038/nature12300

Nakashima, R., Sakurai, K., Yamasaki, S., Nishino, K., and Yamaguchi, A. (2011). Structures of the multidrug exporter AcrB reveal a proximal multisite drug-binding pocket. *Nature* 480, 565–569. doi: 10.1038/nature10641

Negi, N., Prakash, P., Gupta, M. L., and Mohapatra, T. M. (2014). Possible role of curcumin as an efflux pump inhibitor in multi drug resistant clinical isolates of *Pseudomonas aeruginosa*. *J. Clin. Diagnostic. Res.* 8, DC04. doi: 10.7860/JCDR/2014/8329.4965

Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., and Breaker, R. R. (2017). Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol. Cell* 65, 220–230. doi: 10.1016/j.molcel.2016.11.019

Nelson, M. L., Park, B. H., Andrews, J. S., Georgian, V. A., Thomas, R. C., and Levy, S. B. (1993). Inhibition of the tetracycline efflux antiport protein by 13-thio-substituted 5-hydroxy-6-deoxytetracyclines. *J. Med. Chem.* 36, 370–377. doi: 10.1021/jm00055a008

Nelson, M. L., Park, B. H., and Levy, S. B. (1994). Molecular requirements for the inhibition of the tetracycline antiport protein and the effect of potent inhibitors on the growth of tetracycline-resistant bacteria. *J. Med. Chem.* 37, 1355–1361. doi: 10.1021/jm00035a016

Neuberger, A., Du, D., and Luisi, B. F. (2018). Structure and mechanism of bacterial tripartite efflux pumps. *Res. Microbiol.* 169, 401–413. doi: 10.1016/j.resmic.2018.05.003

Nikaido, H. (1996). Multidrug efflux pumps of gram-negative bacteria. *J. Bacteriol.* 178, 5853–5859. doi: 10.1128/jb.178.20.5853-5859.1996

Nikaido, H. (2009). Multidrug resistance in bacteria *Annu. Rev. Biochem.* 78, 119–146. doi: 10.1146/annurev.biochem.78.082907.145923

Nikaido, H. (2011). Structure and mechanism of RND-type multidrug efflux pumps. *Adv. Enzymol. Relat. Areas Mol. Biol.* 77, 1. doi: 10.1002/9780470920541.ch1

Nikaido, H., and Pagès, J. M. (2012). Broad-specificity efflux pumps and their role in multidrug resistance of Gram-negative bacteria. *FEMS Microbiol. Rev.* 36, 340–363. doi: 10.1111/j.1574-6976.2011.00290.x

Nishino, K., Nikaido, E., and Yamaguchi, A. (2009). Regulation and physiological function of multidrug efflux pumps in *Escherichia coli* and Salmonella. *Biochim. Biophys. Acta Proteins. Proteom.* 1794, 834–843. doi: 10.1016/j.bbapap.2009.02.002

Nishino, K., Senda, Y., and Yamaguchi, A. (2008). The AraC-family regulator GadX enhances multidrug resistance in *Escherichia coli* by activating expression of mdtEF multidrug efflux genes. *J. Infect. Chemother.* 14, 23–29. doi: 10.1007/s10156-007-0575-y

Nishino, K., and Yamaguchi, A. (2001). Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J. Bacteriol.* 183, 5803–5812. doi: 10.1128/JB.183.20.5803-5812.2001

Nunez-Samudio, V., and Chesneau, O. (2013). Functional interplay between the ATP binding cassette Msr (D) protein and the membrane facilitator superfamily Mef (E) transporter for macrolide resistance in *Escherichia coli*. *Res. Microbiol.* 164, 226–235. doi: 10.1016/j.resmic.2012.12.003

Oethinger, M., Podglajen, I., Kern, W. V., and Levy, S. B. (1998). Overexpression of the marA or soxS regulatory gene in clinical topoisomerase mutants of *Escherichia coli*. *Antimicrob. Agents. Chemother.* 42, 2089–2094. doi: 10.1128/aac.42.8.2089

Ohene-Agyei, T., Mowla, R., Rahman, T., and Venter, H. (2014). Phytochemicals increase the antibacterial activity of antibiotics by acting on a drug efflux pump. *Microbiologyopen* 3, 885–896. doi: 10.1002/mbo3.212

Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M., and Moriyama, Y. (2006). The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends Pharmacol. Sci.* 27, 587–593. doi: 10.1016/j.tips.2006.09.001

O'Neill, J. (2015). *Antimicrobials in Agriculture and the Environment: Reducing Unnecessary Use and Waste. The Review on Antimicrobial Resistance*. 1–41.

Opperman, T. J., Kwasny, S. M., Kim, H. S., Nguyen, S. T., Houseweart, C., D'Souza, S., et al. (2014). Characterization of a novel pyranopyridine inhibitor of the AcrAB efflux pump of *Escherichia coli*. *Antimicrob. Agents Chemother.* 58, 722–733. doi: 10.1128/AAC.01866-13

Opperman, T. J., and Nguyen, S. T. (2015). Recent advances toward a molecular mechanism of efflux pump inhibition. *Front. Microbiol.* 6, 421. doi: 10.3389/fmicb.2015.00421

Otreebska-Machaj, E., Chevalier, J., Handzlik, J., Szymańska, E., Schabikowski, J., Boyer, G., et al. (2016). Efflux pump blockers in Gram-negative bacteria: the new generation of hydantoin based-modulators to improve antibiotic activity. *Front. Microbiol.* 7, 622. doi: 10.3389/fmicb.2016.00622

Otsuka, M., Matsumoto, T., Morimoto, R., Arioka, S., Omote, H., and Moriyama, Y. (2005). A human transporter protein that mediates the final excretion step for toxic organic cations. *Proc. Natl. Acad. Sci. USA* 102, 17923–17928. doi: 10.1073/pnas.0506483102

Padariya, M., Kalathiya, U., and Baginski, M. (2018). Structural and dynamic insights on the EmrE protein with TPP+ and related substrates through molecular dynamics simulations. *Chem. Phys. Lipids.* 212, 1–11. doi: 10.1016/j.chemphyslip.2017.12.004

Pal, C., Bengtsson-Palme, J., Kristiansson, E., and Larsson, D. G. (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genom.* 16, 1–14. doi: 10.1186/s12864-015-2153-5

Pannek, S., Higgins, P. G., Steinke, P., Jonas, D., Akova, M., Bohnert, J. A., et al. (2006). Multidrug efflux inhibition in *Acinetobacter baumannii*: comparison between 1-(1-naphthylmethyl)-piperazine and phenyl-arginine-β-naphthylamide. *J. Antimicrob. Chemother.* 57, 970–974. doi: 10.1093/jac/dkl081

Pasqua, M., Bonaccorsi di Patti, M. C., Fanelli, G., Utsumi, R., Eguchi, Y., Trirocco, R., et al. (2021). Host-Bacterial pathogen communication: the Wily role of the multidrug efflux pumps of the MFS family. *Front. Mol. Biosci.* 8, 723274. doi: 10.3389/fmolb.2021.723274

Paulsen, I. T., Brown, M. H., and Skurray, R. A. (1996). Proton-dependent multidrug efflux systems. *Microbiol. Rev.* 60, 575–608. doi: 10.1128/mr.60.4.575-608.1996

Perez, C., Gerber, S., Boilevin, J., Bucher, M., Darbre, T., Aebi, M., et al. (2015). Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature.* 524, 433–438. doi: 10.1038/nature14953

Pérez-Varela, M., Corral, J., Aranda, J., and Barbé, J. (2019). Roles of efflux pumps from different superfamilies in the surface-associated motility and virulence of *Acinetobacter baumannii* ATCC 17978. *Antimicrob. Agents Chemother.* 63, e02190–e02118. doi: 10.1128/AAC.02190-18

Piddock, L. J. (2006a). Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. *Clin. Microbiol. Rev.* 19, 382–402. doi: 10.1128/CMR.19.2.382-402.2006

Piddock, L. J. (2006b). Multidrug-resistance efflux pumps? Not just for resistance. *Nat. Rev. Microbiol.* 4, 629–636. doi: 10.1038/nrmicro1464

Piddock, L. J., Garvey, M. I., Rahman, M. M., and Gibbons, S. (2010). Natural and synthetic compounds such as trimethoprim behave as inhibitors of efflux in Gram-negative bacteria. *J. Antimicrob. Chemother.* 65, 1215–1223. doi: 10.1093/jac/dkq079

Poole, K. (2007). Efflux pumps as antimicrobial resistance mechanisms. *Ann. Med.* 39, 162–176. doi: 10.1080/07853890701195262

Pos, K. M. (2009). Drug transport mechanism of the AcrB efflux pump. *Biochim. Biophys. Acta Proteins. Proteom.* 1794, 782–793. doi: 10.1016/j.bbapap.2008.12.015

Prasch, S., and Bucar, F. (2015). Plant derived inhibitors of bacterial efflux pumps: an update. *Phytochem. Rev.* 14, 961–974. doi: 10.1007/s11101-015-9436-y

Quistgaard, E. M., Löw, C., Guettou, F., and Nordlund, P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Rev. Mol. Cell. Biol.* 17, 123–132. doi: 10.1038/nrm.2015.25

Radchenko, M., Symersky, J., Nie, R., and Lu, M. (2015). Structural basis for the blockade of MATE multidrug efflux pumps. *Nat. Commun.* 6, 1–11. doi: 10.1038/ncomms8995

Raturi, S., Nair, A. V., Shinoda, K., Singh, H., Bai, B., Murakami, S., et al. (2021). Engineered MATE multidrug transporters reveal two functionally distinct ion-coupling pathways in NorM from Vibrio cholerae. *Commun. Biol.* 4, 558. doi: 10.1038/s42003-021-02081-6

Remy, E., and Duque, P. (2014). Beyond cellular detoxification: a plethora of physiological roles for MDR transporter homologs in plants. *Front. Physiol.* 5, 201. doi: 10.3389/fphys.2014.00201

Renau, T. E., Léger, R., Flamme, E. M., Sangalang, J., She, M. W., Yen, R., et al. (1999). Inhibitors of efflux pumps in *Pseudomonas aeruginosa* potentiate the activity of the fluoroquinolone antibacterial levofloxacin. *J. Med. Chem.* 42, 4928–4931. doi: 10.1021/jm9904598

Ribera, A., Ruiz, J., Jiminez de Anta, M. T., and Vila, J. (2002). Effect of an efflux pump inhibitor on the MIC of nalidixic acid for *Acinetobacter baumannii* and *Stenotrophomonas maltophilia* clinical isolates. *J. Antimicrob. Chemother.* 49, 697–698. doi: 10.1093/jac/49.4.697

Richmond, G. E., Evans, L. P., Anderson, M. J., Wand, M. E., Bonney, L. C., Ivens, A., et al. (2016). The Acinetobacter baumannii two-component system AdeRS

regulates genes required for multidrug efflux, biofilm formation, and virulence in a strain-specific manner. *MBio* 7, e00430–e00416. doi: 10.1128/mBio.00430-16

Robin, B., Nicol, M., Le, H., Tahrioui, A., Schaumann, A., Vuillemenot, J. B., et al. (2022). MacAB-TolC contributes to the development of *Acinetobacter baumannii* biofilm at the solid–liquid interface. *Front. Microbiol.* 12, 3878. doi: 10.3389/fmicb.2021.785161

Rotem, D., and Schuldiner, S. (2004). EmrE, a multidrug transporter from *Escherichia coli,* transports monovalent and divalent substrates with the same stoichiometry. *J. Biol. Chem.* 279, 48787–48793. doi: 10.1074/jbc.M408187200

Russell, A. D. (2002). Introduction of biocides into clinical practice and the impact on antibiotic-resistant bacteria. *J. Appl. Microbiol.* 92, 121S—135S. doi: 10.1046/j.1365-2672.92.5s1.12.x

Sabatini, S., Piccioni, M., Felicetti, T., De Marco, S., Manfroni, G., Pagiotti, R., et al. (2017). Investigation on the effect of known potent *S. aureus* NorA efflux pump inhibitors on the staphylococcal biofilm formation. *RSC. Adv.* 7, 37007–37014. doi: 10.1039/C7RA03859C

Saier Jr, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C., and Moreno-Hagelsieb, G. (2016). The transporter classification database (TCDB): recent advances. *Nucleic Acids Res.* 44, D372–D379. doi: 10.1093/nar/gkv1103

Sakhtah, H., Koyama, L., Zhang, Y., Morales, D. K., Fields, B. L., Price-Whelan, A., et al. (2016). The *Pseudomonas aeruginosa* efflux pump MexGHI-OpmD transports a natural phenazine that controls gene expression and biofilm development. *Proc. Natl. Acad. Sci. USA* 113, E3538–E3547. doi: 10.1073/pnas.1600424113

Sakurai, K., Yamasaki, S., Nakao, K., Nishino, K., Yamaguchi, A., and Nakashima, R. (2019). Crystal structures of multidrug efflux pump MexB bound with high-molecular-mass compounds. *Sci. Rep.* 9, 1–9. doi: 10.1038/s41598-019-40232-2

Salehi, B., Ghalavand, Z., Yadegar, A., and Eslami, G. (2021). Characteristics and diversity of mutations in regulatory genes of resistance-nodulation-cell division efflux pumps in association with drug-resistant clinical isolates of *Acinetobacter baumannii*. *Antimicrob. Resist. Infect. Control.* 10, 1–12. doi: 10.1186/s13756-021-00924-9

Sato, Y., Unno, Y., Ubagai, T., and Ono, Y. (2018). Sub-minimum inhibitory concentrations of colistin and polymyxin B promote *Acinetobacter baumannii* biofilm formation. *PLoS ONE* 13, e0194556. doi: 10.1371/journal.pone.0194556

Schaedler, T. A., and Van Veen, H. W. (2010). A flexible cation binding site in the multidrug major facilitator superfamily transporter LmrP is associated with variable proton coupling. *FASEB J.* 24, 3653–3661. doi: 10.1096/fj.10-156927

Schumacher, A., Steinke, P., Bohnert, J. A., Akova, M., Jonas, D., and Kern, W. V. (2006). Effect of 1-(1-naphthylmethyl)-piperazine, a novel putative efflux pump inhibitor, on antimicrobial drug susceptibility in clinical isolates of Enterobacteriaceae other than *Escherichia coli*. *J. Antimicrob. Chemother.* 57, 344–348. doi: 10.1093/jac/dki446

Seeger, M. A., Diederichs, K., Eicher, T., Brandstatter, L., Schiefner, A., Verrey, F., et al. (2008). The AcrB efflux pump: conformational cycling and peristalsis lead to multidrug resistance. *Curr. Drug. Targets* 9, 729–749. doi: 10.2174/138945008785747789

Seeger, M. A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., and Pos, K. M. (2006). Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* 313, 1295–1298. doi: 10.1126/science.1131542

Seeger, M. A., von Ballmoos, C., Verrey, F., and Pos, K. M. (2009). Crucial role of Asp408 in the proton translocation pathway of multidrug transporter AcrB: evidence from site-directed mutagenesis and carbodiimide labeling. *Biochemistry* 48, 5801–5812. doi: 10.1021/bi900446j

Sennhauser, G., Bukowska, M. A., Briand, C., and Grütter, M. G. (2009). Crystal structure of the multidrug exporter MexB from *Pseudomonas aeruginosa*. *J. Mol. Biol.* 389, 134–145. doi: 10.1016/j.jmb.2009.04.001

Shcherbakov, A. A., Hisao, G., Mandala, V. S., Thomas, N. E., Soltani, M., Salter, E. A., et al. (2021). Structure and dynamics of the drug-bound bacterial transporter EmrE in lipid bilayers. *Nat. Commun.* 12, 1–13. doi: 10.1038/s41467-020-20468-7

Shi, K., Cao, M., Li, C., Huang, J., Zheng, S., and Wang, G. (2019). Efflux proteins MacAB confer resistance to arsenite and penicillin/macrolide-type antibiotics in *Agrobacterium tumefaciens* 5A. *World J. Microbiol. Biotechnol.* 35, 1–10. doi: 10.1007/s11274-019-2689-7

Shimizu, T., Harada, K., Tsuyuki, Y., Kimura, Y., Miyamoto, T., Hatoya, S., et al. (2017). *In vitro* efficacy of 16 antimicrobial drugs against a large collection of β-lactamase-producing isolates of extraintestinal pathogenic *Escherichia coli* from dogs and cats *J. Med. Microbiol.* 66, 1085–1091. doi: 10.1099/jmm.0.000535

Shirshikova, T. V., Sierra-Bakhshi, C. G., Kamaletdinova, L. K., Matrosova, L. E., Khabipova, N. N., Evtugyn, V. G., et al. (2021). The ABC-type efflux pump MacAB is involved in protection of *Serratia marcescens* against aminoglycoside antibiotics, polymyxins, and oxidative stress. *Msphere* 6, e00033–e00021. doi: 10.1128/mSphere.00033-21

Shiu, W. K., Malkinson, J. P., Rahman, M. M., Curry, J., Stapleton, P., Gunaratnam, M., et al. (2013). A new plant-derived antibacterial is an inhibitor of efflux pumps in *Staphylococcus aureus*. *Int. J. Antimicrob. Agents.* 42, 513–518. doi: 10.1016/j.ijantimicag.2013.08.007

Singh, A. K., Haldar, R., Mandal, D., and Kundu, M. (2006). Analysis of the topology of Vibrio cholerae NorM and identification of amino acid residues involved in norfloxacin resistance. *Antimicrob. Agents. Chemother.* 50, 3717–3723. doi: 10.1128/AAC.00460-06

Siriyong, T., Srimanote, P., Chusri, S., Yingyongnarongkul, B. E., Suaisom, C., Tipmanee, V., et al. (2017). Conessine as a novel inhibitor of multidrug efflux pump systems in *Pseudomonas aeruginosa*. *BMC Complement Altern. Med.* 17, 1–7. doi: 10.1186/s12906-017-1913-y

Sjuts, H., Vargiu, A. V., Kwasny, S. M., Nguyen, S. T., Kim, H. S., Ding, X., et al. (2016). Molecular basis for inhibition of AcrB multidrug efflux pump by novel and powerful pyranopyridine derivatives. *Proc. Natl. Acad. Sci. USA*. 113, 3509–3514. doi: 10.1073/pnas.1602472113

Smirnova, I., Kasho, V., and Kaback, H. R. (2011). Lactose permease and the alternating access mechanism. *Biochemistry* 50, 9684–9693. doi: 10.1021/bi2014294

Solnier, J., Martin, L., Bhakta, S., and Bucar, F. (2020). Flavonoids as novel efflux pump inhibitors and antimicrobials against both environmental and pathogenic intracellular mycobacterial species. *Molecules* 25, 734. doi: 10.3390/molecules25030734

Sonnet, P., Izard, D., and Mullié, C. (2012). Prevalence of efflux-mediated ciprofloxacin and levofloxacin resistance in recent clinical isolates of *Pseudomonas aeruginosa* and its reversal by the efflux pump inhibitors 1-(1-naphthylmethyl)-piperazine and phenylalanine-arginine-β-naphthylamide. *Int. J. Antimicrob. Agents.* 39, 77–80. doi: 10.1016/j.ijantimicag.2011.08.005

Srinivasan, V. B., and Rajamohan, G. (2013). KpnEF, a new member of the *Klebsiella pneumoniae* cell envelope stress response regulon, is an SMR-type efflux pump involved in broad-spectrum antimicrobial resistance. *Antimicrob. Agents. Chemother.* 57, 4449–4462. doi: 10.1128/AAC.02284-12

Su, C. C., Li, M., Gu, R., Takatsuka, Y., McDermott, G., Nikaido, H., et al. (2006). Conformation of the AcrB multidrug efflux pump in mutants of the putative proton relay pathway. *J. Bacteriol.* 188, 7290–7296. doi: 10.1128/JB.00684-06

Sun, J., Deng, Z., and Yan, A. (2014). Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochem. Biophys. Res. Commun.* 453, 254–267. doi: 10.1016/j.bbrc.2014.05.090

Tam, H. K., Malviya, V. N., Foong, W. E., Herrmann, A., Malloci, G., Ruggerone, P., et al. (2020). Binding and transport of carboxylated drugs by the multidrug transporter AcrB. *J. Mol. Biol.* 432, 861–877. doi: 10.1016/j.jmb.2019.12.025

Tanaka, Y., Hipolito, C. J., Maturana, A. D., Ito, K., Kuroda, T., Higuchi, T., et al. (2013). Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature* 496, 247–251. doi: 10.1038/nature12014

Tanaka, Y., Iwaki, S., and Tsukazaki, T. (2017). Crystal structure of a plant multidrug and toxic compound extrusion family protein. *Structure* 25, 1455–1460. doi: 10.1016/j.str.2017.07.009

Taneja, N., and Sharma, M. (2019). Antimicrobial resistance in the environment: the Indian scenario. *Indian. J. Med. Res.* 149, 119–128. doi: 10.4103/ijmr.IJMR_331_18

Tanihara, Y., Masuda, S., Sato, T., Katsura, T., Ogawa, O., and Inui, K. I. (2007). Substrate specificity of MATE1 and MATE2-K, human multidrug and toxin extrusions/H+-organic cation antiporters. *Biochem. Pharmacol.* 74, 359–371. doi: 10.1016/j.bcp.2007.04.010

Tariq, A., Sana, M., Shaheen, A., Ismat, F., Mahboob, S., Rauf, W., et al. (2019). Restraining the multidrug efflux transporter STY4874 of Salmonella Typhi by reserpine and plant extracts. *Lett. Appl. Microbiol.* 69, 161–167. doi: 10.1111/lam.13196

Tatsuno, I., Isaka, M., Masuno, K., Hata, N., Matsumoto, M., and Hasegawa, T. (2018). Functional predominance of msr (D), which is more effective as mef (A)-associated than mef (E)-associated, over mef (A)/mef (E) in macrolide resistance in Streptococcus pyogenes. *Microb. Drug. Resist.* 24, 1089–1097. doi: 10.1089/mdr.2017.0277

Terada, T., and Inui, K. I. (2008). Physiological and pharmacokinetic roles of H+/organic cation antiporters (MATE/SLC47A). *Biochem. Pharmacol.* 75, 1689–1696. doi: 10.1016/j.bcp.2007.12.008

Thorarensen, A., Presley-Bodnar, A. L., Marotti, K. R., Boyle, T. P., Heckaman, C. L., Bohanon, M. J., et al. (2001). 3-Arylpiperidines as potentiators of existing antibacterial agents. *Bioorganic Med. Chem. Lett.* 11, 1903–1906. doi: 10.1016/S0960-894X(01)00330-4

Tintino, S. R., Morais-Tintino, C. D., Campina, F. F., Costa, M. D. S., Menezes, I. R., de Matos, Y. M. L., et al. (2017). Tannic acid affects the phenotype of *Staphylococcus aureus* resistant to tetracycline and erythromycin by inhibition of efflux pumps. *Bioorg. Chem.* 74, 197–200. doi: 10.1016/j.bioorg.2017.08.004

Van Bambeke, F., and Lee, V. J. (2006). Inhibitors of bacterial efflux pumps as adjuvants in antibiotic treatments and diagnostic tools for detection of resistance by efflux. *Recent Pat. Antiinfect. Drug Discov.* 1, 157–175. doi: 10.2174/157489106777452692

Van Veen, H. W. (2010). Last of the multidrug transporters. *Nature* 467, 926–927. doi: 10.1038/467926a

Vargiu, A. V., and Nikaido, H. (2012). Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* 109, 20637–20642. doi: 10.1073/pnas.1218348109

Vargiu, A. V., Ramaswamy, V. K., Malloci, G., Malvacio, I., Atzori, A., and Ruggerone, P. (2018). Computer simulations of the activity of RND efflux pumps. *Res. Microbiol.* 169, 384–392. doi: 10.1016/j.resmic.2017.12.001

Vargiu, A. V., Ruggerone, P., Opperman, T. J., Nguyen, S. T., and Nikaido, H. (2014). Molecular mechanism of MBX2319 inhibition of *Escherichia coli* AcrB multidrug efflux pump and comparison with other inhibitors. *Antimicrob. Agents Chemother.* 58, 6224–6234. doi: 10.1128/AAC.03283-14

Venter, H. (2019). Reversing resistance to counter antimicrobial resistance in the World Health Organisation's critical priority of most dangerous pathogens. *Biosci. Rep.* 39:BSR20180474. doi: 10.1042/BSR20180474

Venter, H., Mowla, R., Ohene-Agyei, T., and Ma, S. (2015). RND-type drug efflux pumps from Gram-negative bacteria: molecular mechanism and inhibition. *Front. Microbiol.* 6, 377. doi: 10.3389/fmicb.2015.00377

Vera-Leiva, A., Carrasco-Anabalón, S., Lima, C. A., Villagra, N., Domínguez, M., Bello-Toledo, H., et al. (2018). The efflux pump inhibitor phenylalanine-arginine β-naphthylamide (PAβN) increases resistance to carbapenems in Chilean clinical isolates of KPC-producing *Klebsiella pneumoniae*. *J. Glob. Antimicrob. Resist.* 12, 73–76. doi: 10.1016/j.jgar.2017.12.003

Verchère, A., Broutin, I., and Picard, M. (2012). Photo-induced proton gradients for the *in vitro* investigation of bacterial efflux pumps. *Sci. Rep.* 2, 1–4. doi: 10.1038/srep00306

Verhalen, B., Dastvan, R., Thangapandian, S., Peskova, Y., Koteiche, H. A., Nakamoto, R. K., et al. (2017). Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature.* 543, 738–741. doi: 10.1038/nature21414

Wagner, D. J., Hu, T., and Wang, J. (2016). Polyspecific organic cation transporters and their impact on drug intracellular levels and pharmacodynamics. *Pharmacol. Res.* 111, 237–246. doi: 10.1016/j.phrs.2016.06.002

Walsh, C. (2003). Where will new antibiotics come from? *Nat. Rev. Microbiol.* 1, 65–70. doi: 10.1038/nrmicro727

Wang, Y., Mowla, R., Guo, L., Ogunniyi, A. D., Rahman, T., Lopes, M. A. D. B., et al. (2017). Evaluation of a series of 2-napthamide derivatives as inhibitors of the drug efflux pump AcrB for the reversal of antimicrobial resistance. *Bioorg. Med. Chem. Lett.* 27, 733–739. doi: 10.1016/j.bmcl.2017.01.042

Wang, Y., Mowla, R., Ji, S., Guo, L., Lopes, M. A. D. B., Jin, C., et al. (2018). Design, synthesis and biological activity evaluation of novel 4-subtituted 2-naphthamide derivatives as AcrB inhibitors. *Eur. J. Med. Chem.* 143, 699–709. doi: 10.1016/j.ejmech.2017.11.102

Wang, Z., Fan, G., Hryc, C. F., Blaza, J. N., Serysheva, I. I., Schmid, M. F., et al. (2017). An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *Elife* 6, e24905. doi: 10.7554/eLife.24905

Webber, M. A., and Piddock, L. J. V. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *J. Antimicrob. Chemother.* 51, 9–11. doi: 10.1093/jac/dkg050

Wehmeier, C., Schuster, S., Fähnrich, E., Kern, W. V., and Bohnert, J. A. (2009). Site-directed mutagenesis reveals amino acid residues in the *Escherichia coli* RND efflux pump AcrB that confer macrolide resistance. *Antimicrob. Agents. Chemother.* 53, 329–330. doi: 10.1128/AAC.00921-08

Woebking, B., Reuter, G., Shilling, R. A., Velamakanni, S., Shahi, S., Venter, H., et al. (2005). Drug-lipid A interactions on the *Escherichia coli* ABC transporter MsbA. *J. Bacteriol.* 187, 6363–6369. doi: 10.1128/JB.187.18.6363-6369.2005

Wolloscheck, D., Krishnamoorthy, G., Nguyen, J., and Zgurskaya, H. I. (2018). Kinetic control of quorum sensing in *Pseudomonas aeruginosa* by multidrug efflux pumps. *ACS Infect. Dis.* 4, 185–195. doi: 10.1021/acsinfecdis.7b00160

Woo, J. S., Zeltina, A., Goetz, B. A., and Locher, K. P. (2012). X-ray structure of the Yersinia pestis heme transporter HmuUV. *Nat. Struct. Mol. Biol.* 19, 1310–1315. doi: 10.1038/nsmb.2417

Wright, D. J., and Tate, C. G. (2015). Isolation and characterisation of transport-defective substrate-binding mutants of the tetracycline antiporter TetA (B). *Biochim. Biophys. Acta. Biomembr.* 1848, 2261–2270. doi: 10.1016/j.bbamem.2015.06.026

Xu, S., Chen, G., Liu, Z., Xu, D., Wu, Z., Li, Z., et al. (2020). Site-directed mutagenesis reveals crucial residues in *Escherichia coli* resistance-nodulation-division efflux pump OqxB. *Microb. Drug. Resist.* 26, 550–560. doi: 10.1089/mdr.2019.0165

Yamaguchi, A., Nakashima, R., and Sakurai, K. (2015). Structural basis of RND-type multidrug exporters. *Front. Microbiol.* 6, 327. doi: 10.3389/fmicb.2015.0032

Yan, N. (2013). Structural advances for the major facilitator superfamily (MFS) transporters. *Trends. Biochem. Sci.* 38, 151–159. doi: 10.1016/j.tibs.2013.01.003

Yang, X., Domalaon, R., Lyu, Y., Zhanel, G. G., and Schweizer, F. (2018). Tobramycin-linked efflux pump inhibitor conjugates synergize fluoroquinolones, rifampicin and fosfomycin against multidrug-resistant *Pseudomonas aeruginosa*. *J. Clin. Med.* 7, 158. doi: 10.3390/jcm7070158

Yang, X., Goswami, S., Gorityala, B. K., Domalaon, R., Lyu, Y., Kumar, A., et al. (2017). A tobramycin vector enhances synergy and efficacy of efflux pump inhibitors against multidrug-resistant Gram-negative bacteria. *J. Med. Chem.* 60, 3913–3932. doi: 10.1021/acs.jmedchem.7b00156

Yerushalmi, H., Lebendiker, M., and Schuldiner, S. (1995). EmrE, an *Escherichia coli* 12-kDa multidrug transporter, exchanges toxic cations and H+ and is soluble in organic solvents. *J. Biol. Chem.* 270, 6856–6863. doi: 10.1074/jbc.270.12.6856

Yerushalmi, H., Lebendiker, M., and Schuldiner, S. (1996). Negative dominance studies demonstrate the oligomeric structure of EmrE, a multidrug antiporter from *Escherichia coli*. *J. Biol. Chem.* 271, 31044–31048. doi: 10.1074/jbc.271.49.31044

Yokoo, S., Yonezawa, A., Masuda, S., Fukatsu, A., Katsura, T., and Inui, K. I. (2007). Differential contribution of organic cation transporters, OCT2 and MATE1, in platinum agent-induced nephrotoxicity. *Biochem. Pharmacol.* 74, 477–487. doi: 10.1016/j.bcp.2007.03.004

Yoon, E. J., Balloy, V., Fiette, L., Chignard, M., Courvalin, P., and Grillot-Courvalin, C. (2016). Contribution of the Ade resistance-nodulation-cell division-type efflux pumps to fitness and pathogenesis of *Acinetobacter baumannii*. *MBio* 7, e00697–e00616. doi: 10.1128/mBio.00697-16

Yoshida, K. I., Nakayama, K., Kuru, N., Kobayashi, S., Ohtsuka, M., Takemura, M., et al. (2006). MexAB-OprM specific efflux pump inhibitors in *Pseudomonas aeruginosa*. Part 5: Carbon-substituted analogues at the C-2 position. *Bioorg. Med. Chem. Lett.* 14, 1993–2004. doi: 10.1016/j.bmc.2005.10.043

Yoshida, K. I., Nakayama, K., Ohtsuka, M., Kuru, N., Yokomizo, Y., Sakamoto, A., et al. (2007). MexAB-OprM specific efflux pump inhibitors in *Pseudomonas aeruginosa*. Part 7: highly soluble and in vivo active quaternary ammonium analogue D13-9001, a potential preclinical candidate. *Bioorg. Med. Chem. Lett.* 15, 7087–7097. doi: 10.1016/j.bmc.2007.07.039

Yu, K., Zhang, Y., Xu, W., Zhang, X., Xu, Y., Sun, Y., et al. (2020). Hyper-expression of the efflux pump gene adeB was found in *Acinetobacter baumannii* with decreased triclosan susceptibility. *J. Glob. Antimicrob. Resist.* 22, 367–373. doi: 10.1016/j.jgar.2020.02.027

Zeng, B., Wang, H., Zou, L., Zhang, A., Yang, X., and Guan, Z. (2010). Evaluation and target validation of indole derivatives as inhibitors of the AcrAB-TolC efflux pump. *Biosci. Biotechnol. Biochem.* 74, 2237–2241. doi: 10.1271/bbb.100433

Zgurskaya, H. I. (2009). Covalently linked AcrB giant offers a new powerful tool for mechanistic analysis of multidrug efflux in bacteria. *J. Bacteriol.* 191, 1727–1728. doi: 10.1128/JB.01718-08

Zhang, Y., Tatsuno, I., Okada, R., Hata, N., Matsumoto, M., Isaka, M., et al. (2016). Predominant role of msr (D) over mef (A) in macrolide resistance in *Streptococcus pyogenes*. *Microbiology* 162, 46–52. doi: 10.1099/mic.0.000206

Zhu, Y. G., Zhao, Y. I., Li, B., Huang, C. L., Zhang, S. Y., Yu, S., et al. (2017). Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat. Microbiol.* 2, 1–7. doi: 10.1038/nmicrobiol.2016.270

Zwama, M., Hayashi, K., Sakurai, K., Nakashima, R., Kitagawa, K., Nishino, K., et al. (2017). Hoisting-loop in bacterial multidrug exporter AcrB is a highly flexible hinge that enables the large motion of the subdomains. *Front. Microbiol.* 8, 2095. doi: 10.3389/fmicb.2017.02095

Zwama, M., and Nishino, K. (2021). Ever-adapting RND efflux pumps in gram-negative multidrug-resistant pathogens: a race against time. *Antibiotics* 10, 774. doi: 10.3390/antibiotics10070774

Zwama, M., and Yamaguchi, A. (2018). Molecular mechanisms of AcrB-mediated multidrug export. *Res. Microbiol.* 169, 372–383. doi: 10.1016/j.resmic.2018.05.005

Zwama, M., Yamaguchi, A., and Nishino, K. (2019). Phylogenetic and functional characterisation of the *Haemophilus influenzae* multidrug efflux pump AcrB. *Commun. Biol.* 2, 1–11. doi: 10.1038/s42003-019-0564-6

Zwama, M., Yamasaki, S., Nakashima, R., Sakurai, K., Nishino, K., and Yamaguchi, A. (2018). Multiple entry pathways within the efflux transporter AcrB contribute to multidrug recognition. *Nat. Commun.* 9, 124. doi: 10.1038/s41467-017-02493-1

EXPLORING ANTI-DIABETIC COMPOUNDS FROM THE ETHANOLIC EXTRACTION OF STRYCHNOS POTATORUM SEEDS: LIGAND- BASED DESIGN, MOLECULAR DYNAMICS

Arunthathi, K ., Srividya, K., Abhinav, J[1] and P. Dhasarathan[1]
Dept. of Microbiology, TassimBeevi Abdul Kader College for Women, Kilakarai
[1]Dept. of Biotechnology, Prathyusha Engineering College, Tiruvallur – 602025
Corresponding Author: pdhasarathan@gmail.com

ABSTRACT

In this classical lifestyle, people are mostly affected by various diseases. The 2:3 ratios of men and women are affected by diabetics because of food habits. Currently, it is necessary to identify the particular compounds for diabetics. Type 2 diabetes mellitus is a multi factorial, multi systemic chronic disorder characterized by insufficient insulin secretion. S. potatorum seeds have long been utilized in traditional medicine to treat a variety of non-communicable disorders, including diabetes. S. potatorum is a widely utilized plant used in ayurvedic medicine. The seeds of S. potatorum are utilized in the treatment of diabetics and it contains numerous bio-active compounds such as alkaloids, which have valuable outcomes for the treatment of patients with various sicknesses in India. In this present study, an effort will be made to confirm the presence of anti-diabetic compound in the ethanolic extract of S. potatorum by using TLC and GC-MS. Molecular docking will be done with the activated protein kinase (6YYE & 6Y6C) & GC-MS results of the S. potatorum compounds. PyRx software will be used for docking studies. The anti-oxidant and the anti-bacterial activity of the ethanol extract also done for identifying the anti-diabetic compounds.

Keywords: Diabetics, *Strychnos potatorum*, molecular docking.

1. INTRODUCTION

In the next ten years, diabetes will become a major threat. Diabetes mellitus affects at least 30 million people around the world at any given time. Diabetes is one of the most common diseases, afflicting millions of individuals worldwide. The problem is expected to worsen in the coming year as a result of food addiction and the prevalence of sedentary lifestyles in India. While diabetes is not the most lethal disease, it is a main cause of adult blindness, kidney failure, neuropathy, heart attacks, and strokes. Long-term headaches, abnormal insulin and glucagon secretions, massive disruptions in carbohydrate, protein, and lipid metabolism, thickening of capillary basement membrane via the main frame to Microangiopathy and macroangiopathy, and thickening of capillary basement membrane via the main frame to Microangiopathy and macroangiopathy, and thickening of capillary basement membrane via the main frame to Microangiopathy and macroangiopathy, and thickening of capillary The eye, kidney failure, and the circulatory system will all be affected by abnormal secretions and insulin and glucagon actions [1]

There are two forms of diabetes: type 1 and type 2. Diabetes insipidus is a condition in which the urine has a low specific gravity and contains a large quantity of ADH deficiency. A tumour in the posterior pituitary is generally to blame, however it can also be idiopathic. Diabetes mellitus is a metabolic disorder marked by a high blood glucose level. It's not a matter of not having enough insulin or having insulin that isn't working properly. There are two forms of diabetes. One is insulin-dependent diabetes mellitus, whereas the other is insulin-independent diabetes mellitus.

At this time, there is no drug that can completely cure diabetes. Diabetes can be managed by following a healthy diet and exercising on a regular basis. The use of hypoglycemic tablets relieves the headaches associated with diabetes, which is a critical and life-saving situation. These, on the other hand, aren't therapeutic and have a slew of negative side effects. As a result, in recent years, searching for anti-diabetic energy components in plants has become more essential [2].

Strychonos potatorum Linn (Loganiaceae), also known as katakam in Ayurveda and Tettankottai in Tamil. It's a little tree that grows in India's southern and central regions, as well as Sri Lanka and Burma. Emetic, diaphoretic, and alexiteric, with inflammation, anaemia, jaundice, and biliousness as therapeutic options, the ripe fruit is emetic, diaphoretic, and alexiteric [3]. *S. potatorum* seed is used as a stomachic demulcent emetic and for acute diarrhoea, diabetes, gonorrhoea, and eye problems such as conjunctivitis, lachrymation, or copious watery eyes. Since the Rigveda (5600 BC), where over 67 medical works were documented, medicinal plants have played a significant role in Indian culture [4]. Traditional medications are used by 80 percent of the world's 4 billion people due to high costs, a lack of viable therapies, and personal preferences [5]. India is one among the world's 12 biodiversity hotspots, with more than half of advanced shops having medicinal potential. Traditional societies are thought to use over 7500 species, with about factories having good therapeutic potential [6].

The present study aims at identifying the pharmacological compounds of *Strychnos potatorum*, which is effective against diabetics through the phytochemical screening, TLC (thin layer chromatography) & Molecular Docking using the GC-MS result by PyRx software. Antimicrobial activity and antioxidant activity will be checked by using bacterial culture.

2. METHODOLOGY

The seeds of *S.potatorum* seed were collected from local market, Ramanathapuram, Tamilnadu, India and the transported seeds are crushed with mortar and pestle the powder is stored in the laboratory for the further research. Fresh crushed seeds of *S. potatorum*10 g was taken and dissolved in 100 ml of ethanol. The solution was stirred in the magnetic stirrer for 30 minutes. Then the extract was filtered in the Whatman no:1filterpaper and then stored in the room temperature for the further research.

Thin layer chromatography techniques performed to purify substance present in *S. potatorum* seed extracts partially. Antibacterial activity of *S. Potatorum* was tested against pathogenic bacteria such as *Escherichia coli*. The antibacterial activity of ethanol extract utilising *S. Potatorum* seed extract was investigated on nutritional agar using the disc diffusion method. Test pathogens were spread on nutritional agar for bacteria test plates using sterile swabs. Before being placed on the agar plate's surface, the disc was dipped in the sample. To see if there was a zone of inhibition, the test plates were incubated for 24 hours.

Total Antioxidant Capacity assay: According to the procedure, the total antioxidant capacity of the sample extract was determined using the phosphomolybdenum method. A sample of 5 to 10 grammes was obtained, and extraction was carried out with the appropriate solvent.0.3 mL of extract was combined with 3 mL of reagent solution (0.6 M sulfuric acid, 28 mM sodium phosphate and 4 mM ammonium molybdate). In tubes, the reaction solution was incubated at 95°C for 90 minutes. The absorbance of the solution was measured at 695 nm using a UV-VIS spectrophotometer against a blank after cooling to room temperature. In place of extract, 0.3 mL of solvent was utilised as a control.In gram equivalents of ascorbic acid, the overall antioxidant activity is evaluated. The calibration curve has finally reached its conclusion. Ascorbic acid (1000, 500, 250, 125, 62.5, and 31.25 g/mL) was mixed with solvent to create the calibration curve.

GC-MS analysis: Qualitative and quantitative analysis of phytochemicals can be done using Gas Chromatography Mass Spectroscopy (GCMS). GCMS can be applied to solid, liquid and gaseous samples. First the samples are converted into gaseous state then analysis is carried out on the basis of mass to charge ratio [7].

The ethanolic extract of S. potatorum seeds was GC-MS analysed using an Agilent GC 7890 with a triple axis 5975 MS detector. Agilient HP-5MS (30m x 250m x 0.25m) was the capillary, which was made up of 5% phenyl methyl silox. The initial oven temperature was 55 C for 0 minutes, then 10/C min up to 200 C for 0 minutes, then 5C/min up to 260 C for 5 minutes of hold time. The injector had a 10 litre capacity. The source temperature was 250°C (maximum 300°C), the quad temperature was 150°C (maximum 200°C), and the solvent delay period was 3 minutes. The NIST search library was used to identify compounds based on their RT values and mass spectra.The obtained compounds were searched for detailed pharmacological activities.

Docking studies of *s. Potatorum* compounds with 6yye &6y6c protein: Totally 66 bioactive compound which are belongs to *S.potatorum*, obtained by theliterature study. Those

compounds were revealed by the GC-MS analysis. The crystal of 6YYE &6Y6Cprotein structure was retrieved from Protein Data Bank RCSBPDB (http://www.rcsb.org/pdb/home/home.do)Then the PDB format of protein was subjectedfor docking studies.

Molecular docking: Sixty six ligands were used against the protein 6YYE &6Y6C. These compounds have been studied. Protein and ligands were converted into special file format PDBQT. Ligand preparation included the following steps (i) addition of hydrogen atoms, (ii)neutralization of the charge groups and (iii) removal of any miscellaneous structures from the ligand. Prepared and optimized structures of ligands were used for docking simulation. Grid values for protein the grid point set at 80 x 60 x 95. Before starting the docking study, Vina wizard software was used to perform molecular docking in PyRx virtual screening tool. Discovery studio visualizer I was used to examine the docking poses of the complexes [8].

Lipinski's rule: Lipinski's rule of 5 was developed by Christopher A. Lipinski in 1997, this rule wasalso called as Pfizer's rule of five or simply the rule of five (R05). This rule was developed toset "draggability" guidelines. In the drug discovery setting, the rule of 5 predicts that poorphysiochemical and structural properties within certain ranges [9].

3. RESULTS AND DISCUSSION

Phytochemicals in medicinal plants have been reported to be the active principles responsible for the pharmacological potentials of plants [10]. Preliminary phytochemical analysis of ethanol extracts of *S. Potatorum*indicated the presence of certain metabolites in a different manner (Table 1). Reports were revealed the identification of major chemical constituents such as alkaloids, protein, coumarin, carbocyclic acid, fixed oils, fats, carbohydrate, gums and mucilages and rest of the metabolites were absent such as Phenolics, flavonoids, quinones, anthraquinone, phlobatannin, glycosides, amino acids, and cardiac glycosides.

Table 1: Phytochemical analysis of *S.potatorum*

S.No	Phytochemical test	Results
1	Alkaloids	+
2	Phenols	-
3	Coumarins	+
4	Terpenoids	-
5	Quinones	-
6	Anthraquinones	-
7	Tannins	-
8	Phlobatannins	-
9	Carbohydrates	+
10	Glycosides	-
11	Cardiac glycosides	-
12	Proteins	+
13	Aminoacids	-
14	Steroids	-

15	Phytosteroids	-
16	Saponins	-
17	Acids	-
18	Carbocyclic acids	+
19	Fixed oils	+
20	Fats	-

The presence of alkaloids, protein, coumarin, carbocyclic acid, fixed oils, fats,carbohydrate,gums and mucilages. Compare to my results, Flavonoids, carbohydrates, and titerpenes were discovered in chloroform extract, carbohydrate in ethyl acetate extract, glycosides, flavonoids, saponin, carbohydrate, proteins, amino acids, and diterpenes in ethanol extract [11].

The ethanolic extract of the *S.Potatorum* seeds was used to done the TLC. The result of the TLC place is given table 2.

Table 2: TLC analysis of *S.potatorum*

Sample	**RF VALUE**
S.Potatorum seed ethanol extract	0.1538
	0.0473
	0.3846
	0.4615
	0.5384

Tannin and flavonoids are present. compare to my results, The approach was used to separate the lipids on silica gel plates. The lipid-containing portions of the gel were agitated with 1 to 9 ml dichromate reagent, 2.5 g K2Cr2O7 in 1 litre 36 N H2SO4, and the solutions were boiled in boiling water for 45 minutes. The absorbance was measured at 350 m after chilling and centrifuging 0.5 ml supernatant solution with 20 ml water. It is possible to estimate as little as 15 g of lipid. The presence of phenolic components was detected in the test sample based on the Rf value [12].

The ethanol extracts of *Strychnos potatorum* showed antibacterial activity against pathogenic gram positive, gram negative, and acid-fast bacteria. *S. potatorum* seed ethanol extract showed maximum antibacterial activity against *Escherichia Coli.* Compare with current results [13] examined the antimicrobial activity in the alkaloid fractions of seeds of *Strychnos potatorum* which exhibited considerable activity against the tested bacteria and fungi. Inhibition of gram-positive bacterium *Staphylococcus aureus* and gram-negative bacteria *Proteus*.

By using a UV-VIS Spectrophotometer, the total anti-oxidant activity of *S. potatorum* seed ethanol extract was found to be 275 g/TAE mL-1 extract. In comparison to my findings, total antioxidant activity of SPP and SPE was determined using FTC and TBA procedures and compared to vitamins E and C.The drugs and standards in the FTC approach showed low absorption values as compared to the control. SPP exhibited the least increase in absorbance,

followed by vitamin E, vitamin C, and SPE. The TBA technique was used to determine the total peroxide levels produced by the oxidation of linoleic acid. The higher the absorption levels, the poorer the antioxidant activity. The absorbance values of SPP, vitamin E, vitamin C, and SPE were all greater than the control. The results were the same as those obtained using the FTC method [14].

GC-MS analysis was performed on an ethanolic extract of S. Potatorum seeds. The GC-MS analysis of an ethanol extract of S. potatorum seed extract revealed the following peaks (graph 1), indicating the presence of different phytoconstituents. The mass spectrum GC-MS interpretation was done with the help of a database (NIST) (Sathyaprabha,G et al.2010). Sixty six bio active compounds in the GC-MS results.Compare to my results,

Only a few of the bio compounds present have known therapeutic properties. Ethyl acetate, 3-Trifluoroacetoxypentadecane, Dodecane, 1,2-dibromo-, 4-Trifluoroacetoxypentadecane, Trichloroacetic acid, pentadecyl ester, Trichloroacetic acid hexadecyl ester, E14 Hexadecenal, E-11,13-Tetradecadien-1-ol There are a few additional molecules that are present in little amounts but could play a large or minor role in medicine. More research is being done to assess these compounds and correlate their activities to Katakakhadiradi Kashayam in order to better understand the medicine's mechanism of action as well as other pharmacological aspects.

The molecular docking was done using the proteins (6Y6C & 6YYE)and the ligands of GC-MS results by the PyRx software.

Table 3: Binding affinity analysis of 6Y6C Protein

PROTEIN	PUBCHEM ID	PROBABLE COMPOUND	BINDING AFFINITY
6Y6C	536772	Imidazole, 2-fluoro-1-triacetylribofuranosyl	-6.6
	574859	Dispiro[1,3-dioxolane-2,2'-bicyclo[2.2.1]heptane-3',2"-(1",3"-dioxolane)], 4',7',7'-trimethyl	-6.2
	590903	Phosphine, (myrtenoyl)diphenyl	-7
	45382147	2-Pyridinecarbonitrile, 3-nitro	-6.1
	102070384	Bicyclo[2.2.1]heptan-2-ol, 1,3,3-trimethyl	-6.4
	66831472	Thiophene, 2,5-dihydro-	-7.7

The docking studies against 6Y6C revealed in GC-MS analysis of bioactiveCompounds such as Imidazole, 2-fluoro-1-triacetyl ribofuranosyl, Dispiro [1,3-dioxolane-2, 2'-bicyclo[2.2.1] heptane-3',2"- (1",3"-dioxolane)], 4',7',7'-trimethyl, Phosphine, (myrtenoyl) diphenyl,2-Pyridine carbonitrile, 3-nitro, Bicyclo [2.2.1] heptan-2-ol, 1,3,3-trimethyl, Thiophene, 2,5-dihydro-, showed the docking binding energy of -6.6,-6.2,-7,-6.1,-6.4,-7.7 Kcal/mol respectively. Thiophene, 2, 5-dihydro-have good anti-diabetic activity against 6Y6C protein. This study identified anti-diabetic compounds from the ethanol extract of *S.Potatorum* seeds against 6Y6C protein, which is considered as a new research task in the technology of current docking studies.

Table 3: Binding affinity analysis of 6YYE protein:

PROTEIN	PUBCHEM ID	PROBABLE COMPOUNDS	BINDING AFFINITY
6YYE	28813	1,2-Benzenedicarboxylic acid, butyl 2-methylpropyl ester	-6.2
	536772	Imidazole, 2-fluoro-1-triacetylribofuranosyl	-6.5
	538744	2,3;5,6-Diacetone-4-O-methylmannitol	-6.7
	574859	Dispiro[1,3-dioxolane-2,2'-bicyclo[2.2.1]heptane-3',2"-(1",3"-dioxolane)], 4',7',7'-trimethyl	-7.4
	590903	Phosphine, (myrtenoyl)diphenyl	-8.3
	638072	2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)-	-6.5
	23571816	Bicyclo[2.2.1]hept-2-ene, 2-methyl	-7.2
	45276392	1,4-Hexadiene, 4-methyl	-6.1
	45382147	2-Pyridinecarbonitrile, 3-nitro	-6.6
	66831472	Thiophene, 2,5-dihydro-	-8.4
	91720425	Phthalic acid, cycloheptyl isohexyl ester	-6.8
	102070384	Bicyclo[2.2.1]heptan-2-ol, 1,3,3-trimethyl	-8.2

The docking studies against 6YYE revealed in GC-MS analysis of bioactiveCompounds such as1,2-Benzenedicarboxylic acid, butyl 2-methylpropyl ester,Imidazole, 2-fluoro-1-triacetylribofuranosyl,2,3;5,6-Diacetone-4-O-methyl mannitol, Dispiro [1,3-dioxolane-2, 2'-bicyclo[2.2.1] heptane-3',2"-(1",3"-dioxolane)], 4',7',7'-trimethyl, Phosphine, (myrtenoyl)diphenyl,2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)-, Bicyclo [2.2.1]hept-2-ene, 2-methyl,1,4-Hexadiene, 4-methyl,2-Pyridine carbonitrile, 3-nitro,Thiophene, 2,5-dihydro-,Phthalic acid, cycloheptyl isohexyl ester, Bicyclo[2.2.1]heptan-2-ol, 1,3,3-trimethyl,showed the dockingbinding energy of -6.2,-6.5,-6.7,-7.4,-8.3,-6.5,-7.2,-6.1,-6.6,-8.4,-6.8,-8.2 Kcal/mol respectively. Thiophene, 2,5-dihydro-have goodanti-diabetic activity against 6YYE protein. This study identified anti-diabetic compounds from the ethanol extract of *S.Potatorum* seeds against 6YYE protein, which is considered as a new research task in the technology of current docking studies.

LIPINSKI RULE:

S.NO	COMPOUNDS	Molecular Weight (g/mol)	Log p<5	H-bond Donor <5	H-bond Acceptor <10	No. of vio-lation
1.	Imidazole, 2-fluoro-1-triacetylribofuranosyl	344.29	0.8	0	9	0
2.	Dispiro[1,3-dioxolane-2,2'-bicyclo[2.2.1]heptane-3',2"-(1",3"-dioxolane)], 4',7',7'-trimethyl	254.32	1.5	0	4	0

3.	Phosphine, (myrtenoyl)diphenyl	334.4	5.2	0	1	0
4.	2-Pyridinecarbonitrile, 3-nitro	217.10	1.6	0	7	0
5.	Bicyclo[2.2.1]heptan-2-ol, 1,3,3-trimethyl	330.5	4.6	1	3	0
6.	Thiophene, 2,5-dihydro-	86.16	1.1	0	1	0
7.	1,2-Benzenedicarboxylic acid, butyl 2-methylpropyl ester	278.34	4.8	0	4	0
8.	2,3;5,6-Diacetone-4-O-methylmannitol	276.33	0	1	6	0
9.	2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)-	410.7	11.6	0	0	1
10.	Bicyclo[2.2.1]hept-2-ene, 2-methyl	272.23	5.3	0	6	1
11.	1,4-Hexadiene, 4-methyl	186.25	3.2	0	1	0
12.	2-Pyridinecarbonitrile, 3-nitro	217.10	1.6	0	7	0
13.	Phthalic acid, cycloheptyl isohexyl ester	346.5	6.6	0	4	0

From the GC-MS results received 13 best compounds from the *S.potatorum* seed ethanolic extract based on their binding affinity. For these 13 compounds have checked the following rules.

- ✓ Molecular mass < 500.
- ✓ Calculated octanol/water partition coefficient (CLogP) < 5.
- ✓ Number of hydrogen bond donors < 5.
- ✓ Number of hydrogen bond acceptors < 10(Walters *et al*.,2012).

In this study, identified anti-diabetic compounds from the ethanol extract of *S.Potatorum* seeds against 6Y6C & 6YYE proteins by docking studies. Present study checked the compounds that obey to the Lipinski rule which is considered as a new research task in the technology of current docking studies.

4. SUMMARY AND CONCLUSION:

Phytochemical study of *S. potatorum* ethanol extracts revealed the existence of bioactive substances.The TLC (thin layer chromatography) results revealed the presence of several bioactive substances. Medicinal plants are used to find and screen phytochemical ingredients and play an important role in illness prevention, such as antibacterial, anti-diabetic, and antioxidant properties. Preliminary phytochemical examination of *S. potatorum* seed ethanol extract reveals primary and secondary metabolites that are of commercial relevance to both research institutes and pharmaceutical corporations for the development of new medications

to cure a variety of ailments. Phytochemicals with varied activity are found in the therapeutic seeds chosen. We concentrated on anti-diabetic activity in this study.

The ethanol extract of *S.potatorum* seeds revealed good bactericidal activities against *E.coli* and it has the 275 µg/TAE mL^{-1} extract total anti-oxidant capacity. The docking studies against 6Y6C revealed in GC-MS analysis of bioactive Compounds such as Imidazole, 2-fluoro-1-triacetylribofuranosyl,Dispiro[1,3-dioxolane-2,2'-bicyclo[2.2.1] heptane-3',2"-(1",3"-dioxolane)], 4',7',7'-trimethyl,Phosphine, (myrtenoyl)diphenyl,2-Pyridinecarbonitrile, 3-nitro,Bicyclo[2.2.1]heptan-2-ol, 1,3,3-trimethyl,Thiophene, 2,5-dihydro-,showed the dockingbinding energy of -6.6,-6.2,-7,-6.1,-6.4,-7.7 Kcal/mol respectively. Thiophene, 2,5-dihydro-have goodanti-diabetic activity against 6Y6C protein.

In this study, identified anti-diabetic compounds of the ethanolic extraction of *S. potatorum* seeds. So, it has been considered as a novel research work in the current technology of docking studies. Now a day's drug designing plays a major role in the competitiveworld. In future studies, myresearch work is most useful and essential for the drug designing.

REFERENCES:

[1] Dhasarathan P and Theriappan P. Evaluation of anti-diabetic activity of *Strychonous potatorum*in alloxan induced diabetic rats. *Journal of Medicine and Medical Sciences*, 2(2): (2011) 670-674.

[2] Kritikar KR and Basu BD. In: (eds.) Mhaskar, K.S., Blatter, E and Cains, J.F. Illustrated Indian Medicinal plants, Sri satyaguru's publications, New Delhi. (2001) 2271.

[3] Swati Biswas, Murugesan T, Sanghamsitra sinha, Kuntal Maiti, Jiaur Rahaman Gayan, Pal M and Saha BP. Antidiarrhoeal activity of *Strychnos potatorum* Seed extract in rats. *Fitoterapia*. 73: (2002) 43-47.

[4] Pushpahasni Kannan, Sharmila Raja S and Dhasarathan P.Efficiency of in-vitro antibacterial activity of *Syzygium cumini* phenolic extract from leaves, 8 (6): (2015) 211-214.

[5] Arunthathi K, Valivittan K, and Dhasarathan P. 'Characterization of bioactive compound isolated from *strychnos potatorum*using bacterial pathogens'. International Journal of Advanced Research in science and Engineering.Volume No.07 (2018) pp:39-61.

[6] Miura T, Itoh Y, Iwamoto N, Kato M and Ishida T. Suppressive activity of the fruit of *Momordica charantia* with exercise on blood glucose in type 2 diabetes mice. *Biological Pharmaceutical Bulletin*. 27: (2004) 248-250

[7] Leahy Jack L. "Pathogenesis of type 2 diabetes mellitus." *Archives of medical research* 36.3 (2005): 197-209.

[8] Manjula M, Indira V and Dhasarathan P. "Phytochemical characterization and anti bacterial activity screening of Achyranthes aspera", Asian J. Microbiol. Biotechnol. Environ. Sci, Vol. 11(2009) pp.365-368.

[9] Mishra SB, Verma A and Vijayakumar M. Preclinical valuation of anti-hyperglycemic and antioxidant action of Nirmali (Strychnos potatorum) seeds in streptozotocin-nicotinamide-induced diabetic Wistar rats: A histopathological investigation. *Biomarkers and Genomic medicine*, *5*(4): (2013) 157-163.

[10] Morris GM and Lim-Wilby M. Molecular docking. In *Molecular modeling of proteins* (2008) pp. 365-382. Humana Press.

[11] Nagaraja K, Rao KM, Hemalatha D, Zo S, Han SS and Rao KK. Strychnos Potatorum L. Seed Polysaccharide-Based Stimuli-Responsive Hydrogels and Their Silver Nanocomposites for the Controlled Release of Chemotherapeutics and Antimicrobial Applications. (2022) *ACS omega*.

[12] Poole Colin F. "Thin-layer chromatography: challenges and opportunities." *Journal of Chromatography A* 1000.1-2: (2003) pp 963-984.

[13] Uttu AJ, Sallau MS, Iyun OR and Ibrahim H. Antimicrobial Efficacy of Selected Strychnos Species: A Mini Review. *Journal of Chemical Reviews, 4*(1): (2022) pp 1-5.

[14] Zheng Yan, Sylvia H. Ley and Frank B. Hu. "Global aetiology and epidemiology of type 2 diabetes mellitus and its complications." *Nature reviews endocrinology* 14.2 (2018): 88-98.

A-5959
[1-5]

RESEARCH ARTICLE Indian Journal of Agricultural Research,

Extraction of Phytochemical Constituents from the Leaves of *Vitex trifolia* in the Inhibition of Lipoxygenase, Cyclooxygenase, Tyrosinase and Xanthine Oxidase

M. Thenmozhi[1], K. Arunthathi[2], J. Abhinav[1], K.V. Sharika[1], P. Dhasarathan[3]

10.18805/IJARe.A-5959

ABSTRACT

Background: Herbal medicines are often used to stimulate the immune system in an attempt to prevent disease, as well as to induce specific cures. The use of phyto medicines is becoming more scientifically based, with increasing emphasis placed on proven product safety and efficacy.

Methods: In the present study, the leaves of *Vitex trifolia* were extracted with distilled water, 1M Na_2CO_3 and 70% methanol. The extracted contents were screened to identify the presence of various phytochemicals. The contents of Na_2CO_3 and methanol were filtered, centrifuged and the supernatant was collected. The supernatant of Na_2CO_3 extract was collected and the alkaloid layer was separated in a separating funnel using alcohol. The contents were analysed for thin layer chromatography and phytochemical constituents.

Result: Thin layer chromatography was performed at360 nm the Rf value were 0.953, 0.230, 0.184, 0.153, 0.123 and in 240 nm Rf values 0.953, 0.338, 0.184, 0.153, 0.123 were obtained. The methanol extract was condensed at 40°C and dissolved in chloroform. Rf at 360 nm of 0.714, 0.396, 0.158 and at 240 nm with Rf values 0.714, 0.492, 0.349. The activity of extracts on Xanthine oxidase, tyrosinase and cyclooxygenase were estimated and the inhibition percentage of all the three was calculated as 59.34, 61.23, 71.02 and 51 respectively. The values obtained with flavanoid extracts for the same set of enzymes were 70.22, 81.48, 94.8 and 74.36 respectively. In this study, inhibition properties of flavanoid extracts were greater than that of alkaloid extracts.

Key words: Enzymes, Inhibition, Phytochemicals, Rf values, *Vitex trifolia*.

INTRODUCTION

Enzymes are the sparks that start the essential chemical reactions in our bodies need to live. They are necessary for digesting food, for stimulating the brain, for providing cellular energy and for repairing all tissues, organs and cells. The use of plant-based medications has become extremely popular in the United States and Europe, with the botanical industry in the US earning $1.5 billion per annum and the European market nearly three times as much (Gupta *et al*., 2006). Pure compounds are used when the activity is strong and specific and has a small therapeutic index. Natural products, such as plant extract, either as pure compounds or as standardized extracts, provide unlimited opportunities for new drug discoveries because of the unmatched chemical diversity they can provide (Cos *et al*., 2006). According to the World Health Organization (WHO), more than 80% of the world's population relies on traditional medicine for their primary healthcare needs. This has captured the interest of many researchers to explore local medicinal plants for valuable medicinal traits. Several studies indicate that medicinal plants contain compounds like peptides, unsaturated long chain fatty acids, aldehydes, alkaloids, essential oils, phenols and water or ethanol soluble compounds (Zerihun, 2022). These compounds are significant in therapeutic applications against human and animal pathogens, including bacteria, fungi and viruses (Pavrez *et al*., 2005).

[1]Department of Biotechnology, Prathyusha Engineering College, Tiruvallur-602 025, Tamil Nadu, India.
[2]Department of Microbiology, Tassim Beevi Abdul Kader College for Women, Kilakarai-623 517, Tamil Nadu, India.
[3]Department of Biotechnology, Madha Engineering College, Chennai-600 069, Tamil Nadu, India.

Corresponding Author: P. Dhasarathan, Department of Biotechnology, Madha Engineering College, Chennai-600 069, Tamil Nadu, India. Email: pdhasarathan@gmail.com

How to cite this article: Thenmozhi, M., Arunthathi, K., Abhinav, J., Sharika, K. and Dhasarathan, P. (2023). Extraction of Phytochemical Constituents from the Leaves of *Vitex trifolia* in the Inhibition of Lipoxygenase, Cyclooxygenase, Tyrosinase and Xanthine Oxidase. Indian Journal of Agricultural Research. doi: 10.18805/IJARe.A-5959.

Submitted: 30-12-2021 **Accepted:** 20-02-2023 **Online:** 14-03-2023

Alkaloids primarily act as feeding deterrents and toxins to insects and other herbivores, in many cases by directly interacting with molecular targets within the nervous system (Wink, 2000). Flavonoids are a secondary metabolite of plants and the most abundant polyphenols in the human diet (Burda and Oleszek 2001). Cyclo oxygenase (COX) are lipid metabolising enzymes that catalyse the oxygenation of polyunsaturated fatty acids (PUFA), preferably

arachidonicacid (AA), to form the prostanoids, which are potent cell-signalling molecules associated with the initiation, maintenance and resolution of inflammatory processes (Charlier and Michaux, 2003). Lipoxygenases (LOXs) is involved in terminal differentiation of keratinocytes and has a major role in skin barrier function. The subsequent dopaquinone is converted to dopa and dopachrome through auto-oxidation (Roh *et al*., 2004). The plant *Vitex trifolia*Linn (Verbaneaceae) is stout aromatic shrubor a small tree, found from the foot of Himalayas southwards throughout greater part of India, Western Ghat and in Andamans (Agrawal, 1997). The leaves are simple or 3-foliolate. *V. negundo* closely resembles. The flowers are numerous and borne in terminal, oblong panicles 5 to 10 centimeters in length. The fruit is rounded and 4to 5 millimeters in diameter (Padmalatha *et al*., 2009). Hence, the further studies were decided to be carried over on the coastal plant *Vitex trifolia* and also to check how well it can inhibit the above mentioned inflammatory enzymes.

MATERIALS AND METHODS

Collection of plant material

Vitex trifolia leaves were collected early in the morning from Medicinal Plant Garden at SriSairam Siddha Medical College and Research Centre, West Tambaram, Chennai600 044, a recognized institution of Government of Tamil Nadu and the Department of AYUSH, Government of India. The sample was collected during the winter session (November-December, 2020) and experiments were carried out in Department of Biotechnology, Prathyusha Engineering College, Tiruvallur, India.

Phytochemical analysis of plant material

The collected plant samples were shadow dried and powdered with help of mixer grinders. The powder samples were extracted with aqueous and methanol solvents using Soxlet apparatus. Freshly prepared extracts of aqueous and methanol were divided into different test tubes. Phyto constituents such as alkaloids, terpenoids, glycosides, tannins, saponin, flavonoids were analysed in test extracts by methods of Harborne and Williams (2000).

Partial characterizations of thin layer chromatography (TLC) in *Vitex trifolia* anthocyanin extract

The Methanol extract of *Vitex trifolia* were loaded on to pre coated TLC(60 F2 54) and it was developed using solvent system in the ratio of 1:0.5:0.1(Hexane, Chloroform and Methanol) visible and the non visible spot given and it is fluorescent with UV light at 360 nm.

Cyclooxygenase inhibition assay

The cyclooxygenase inhibition assay was performed according to amodified method of Lin and Kao (2001). Leuco-2,7-dichlorofluorescein diacetate (5 mg) was hydrolysed at room temperature in 1 MNaOH (50 µL) for 10 min, then 1 M HCl (30 µL) was added to neutralise excess NaOH before the resulting 1-DCF was diluted in 0.1 M Tris-buffer, pH 8.Cyclooxygenase enzyme (COX-1 or COX-2) was diluted in 0.1 M Tris-buffer, pH 8, so that a known aliquot gave an absorbance change of 0.05/min in the testreaction. Test samples were pre-incubated with enzyme at room temperature for 5 min in the presence of hematin. Premixed phenol, 1-DCF and arachidonic acid were added to the enzyme mixture to begin the reaction and to give a final reaction mixture of arachidonic acid (50 µM), phenol (500 µM), 1-DCF (20 ìM) and hematin (1 µM) in 1 mL final volume of 0.1 M Tris-buffer, pH 8. The reaction was recorded. Lyprinol was investigated at multiple concentrations and exhibited 50% inhibition of the enzyme reaction at approximately 1 µg/mL (final concentration), therefore the other lipid samples were tested at this final concentration for comparison.

Lipoxygenase inhibition assay

A spectrophotometric assay for determination of LOX activity was used as reported (Khan *et al*., 2006) with slight modification. Inhibition experiments were run by measuring the loss of soybean 15-LOX activity (5 µg) with 0.2 µM linoleic acid (Sigma) as the substrate prepared in solubilized state in 0.2M borate buffer (pH 9.0). Inhibition studies in presence of various concentrations of extracts (5, 10, 15, 20 µg/mL) and reference compound *viz*., quercetin was recorded at 234 nm using UV-Vis spectrophotometer. The inhibitory effect of the extracts was also expressed as percentage of enzyme activity inhibition. IC_{50} indicating the concentration required to inhibit 50 % LOX activity was also calculated (Patel *et al*., 2022).

Tyrosinase inhibition assay

Tyrosinase inhibition activity is measured by the dopachrome method (Agarwal, 1997).

Enzymatic Assay of Xanthine oxidase

Xanthine combines with hydrogen peroxide and enzyme xanthine oxidase in the presence of oxygen to form uric acid, at pH 7.5, 25°C. The reaction is investigated in the lab conditions using a continuous spectrophotometric rate determination method, absorbance of 290 nm, light path of 1 cm (Cotran *et al*., 1994).

RESULTS AND DISCUSSION

The leaves of *Vitex trifolia* were collected, dried and powdered and was subjected to various phytochemical screening, extraction of crude phytochemicals and estimation of its response to various enzymes like lipoxygenase, cyclooxygenase, xanthine oxidase and tyrosinase. The preliminary phytochemical screening with the various qualitativechemical tests of various leaf extracts revealed the presence of carbohydrates, flavonoids, protein and amino acids, tannins, phytosterols and saponins phytoconstituents. The phytochemical screening of the *Vitex trifolia* studied presently showed the presence of alkaloids, flavonoids and Glycosides (Table 1).

The alkaloid and flavonoid extract of *Vitex trifolia* were loaded on Precoated TLC plates (60 F2 54 Merck) and developed with a solvent system of hexane, ethyl acetate and acetic acid in the ratio of 10:5:0.5. The developed plate was viewed under UV 240 nm and 360 nm (Table 2). The concentration of flavonoids in various concentration extracts of the *Vitex trifolia* was determined using spectrophotometric method with aluminum chloride. The content of flavonoids was expressed in terms of quercetin equivalent. The concentration of flavonoids in plant extract from *Vitex trifolia* ranged from 20.72 to 59.26 mg/g. Methanolic extract contains the flavonoid concentration. The inhibitory effects of COX-mediated TMPD oxidation activity were examined using purified COX as enzyme sources COX-2 activity was strongly inhibited by used in this study alkaloid and flavonoid extract from *vitex trifolia* inhibited COX-2 the inhibition potency quite different based on the results obtained in Graph 1. The following experiments focused on the inhibition potential of CBDA for COX-2 activity. Whereas the standard drug Celecoxib inhibited the COX2 enzyme with an IC50 of 52 nM. The results are shown in Graph 1.

The anti-inflammatory activity of the alkaloid and flavonoid extract of *vitex trifolia* was evaluated by measuring the inhibition of LOX using linoleicacid as substrate. The results were reported in Graph 2 flavonid extract showed inhibition percentage above 70.22% at 20 µg/mL. The standard n-propylgallate showed 72% inhibition at 20 µg/mL. The alkaloid and flavonoid extract of *Vitex trifolia* were tested for the effects on the oxidation of DOPA by mushroom tyrosinase. With increasing the concentrations of the alkaloid and flavonoid extract of *Vitex trifolia*, the diphenolase activity of mushroom tyrosinase markedly decreased concentration dependently. The values of IC50, the inhibitor concentration leading to 50% activity lost, of *Vitex trifolia* were estimated to be 5, 10, 15 and 20 µg/ml, respectively (Graph 3). The four different concentrations of *Vitex trifolia* alkaloid and flavonoid extract were tested for the inhibition activity of xanthine oxidase. *Vitex trifolia* extracts in different concentrations inhibited the xanthine oxidase activity. The maximum inhibition was found at 20 µg/ml concentration (Graph 4).

In many inflammatory disorders there is excessive activation of phagocytes, production of 0_2-, OH radicals as well as non free radicals species (H_20_2), which can harm severely tissues either by powerful direct oxidizing action or indirect with hydrogen peroxide and -OH radical formed from O_2- which initiates lipid peroxidation resulting in membrane destruction. Tissue damage then provokes inflammatory response by production of mediators and chemotactic factors. The reactive oxygen species are also known to activate matrix metello proteinase damage seen in various arthritic tissues (Cotran *et al*., 1994).

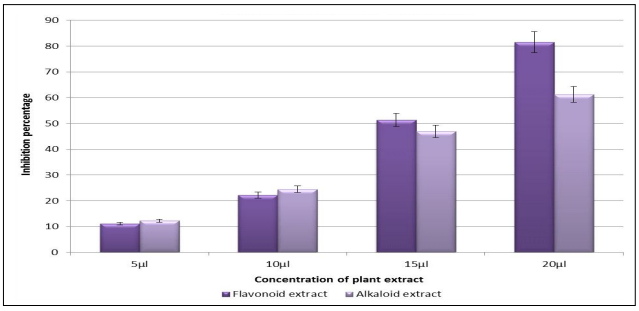
Phytochemicals from medicinal plants showing anti-inflammatory activities have the potential of filling this need because of structures are different from those of the more studied and their those of the more action may too very likely differ (Fabricant and Fanworth, 2001 and Prachayasittikul *et al*., 2008). Some studies have demonstrated that flavonoid possessanti-inflammatory activities by inhibition of cyclooxygenase-2 (COX-2) expression in lipopolysaccharide (LPS)-activated RAW 264 cells or

Table 1: Phytochemical screenings of aqueous methanol leaf extract *Vitex trifolia*.

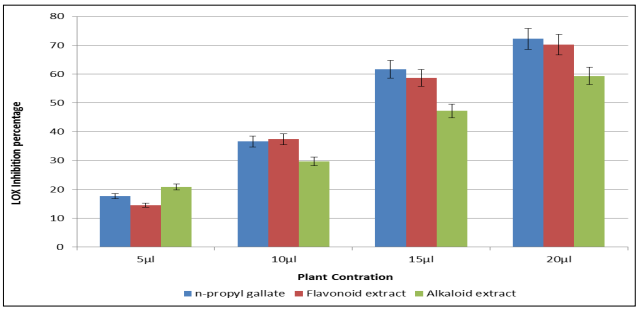
Phytochemicals	Result
Alkaloids	Positive
Flavonoids	Positive
Glycosides	Positive
Tannin	Negative
Saponin	Negative
Terpenoids	Negative
Polyphenols	Negative

Table 2: Partial characterization of different phytochemical extract from the *Vitex trifolia*leaves by TLC.

Extracts	Rf Value at UV 240 nm	Rf Value at UV 360 nm	Rf Value at Visible nm
Alkaloid	0.953	0.963	-
	0.338	0.230	-
	0.184	0.184	-
	0.153	0.153	-
	0.123	0.123	-
Flavonoid	0.714	0.714	-
	0.492	0.396	-
	0.349	0.158	-



Graph 1: Cycloxygenase inhibiton activity of flavonoid and alkaloid extract from *Vitex trifolla.*

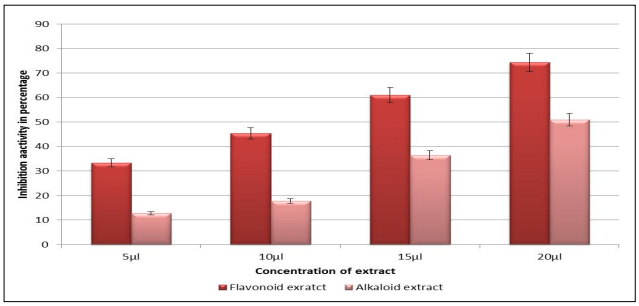


Graph 2: Lipoxygenase inhibiton activity of alkaloid and flavonoid extract of *Vitex trifolla.*

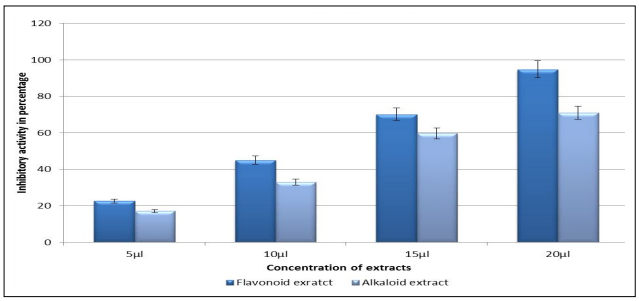
inhibiting inducible nitric oxide (iNOS) protein and mRNA expression in LPS-activated murine J774 macrophages (Hou *et al*., 2005) and such activities appear to be structure dependent. COX-2 seems to be involved in many inflammatory processes. Some antioxidants inhibit the expression of COX-2 by interfering with the signalling mechanisms that regulate the COX-2 gene (Hou *et al*., 2005). From the result, it is clear that caragennan induced paw oedema for the administered dose (500 mg/kg, p.o.) is comparable with reference standard indomethacine, which is a cycloxygenase inhibitor. But anti-inflammatory activity against caragennan-induced paw oedema is also shown by lipoxygenase inhibitor. Hence inhibition of caragennan-induced paw oedema by crude extract may be due to inhibitory activity of lipoxygenase enzymes (Gupta *et al*., 2006). Therefore, from the present study, we can conclude that alkaloid and flavonoid extracts of *Vitex trifolia* for the dose of 20 µl/ml shows anti-inflammatory activity in the early stage as well as in the late stage (up to 180 min) and after that, the effect becomes similar to that of negative control.

Lipoxygenases (LOXs) (LOX; EC 1.13.11.12) are a family of non-hemeiron-containing dioxygenases catalyzing the biosynthesis of leukotrienes. Leukotrienes function as initiators of inflammation and their inhibition is considered to be partly responsible for the anti-inflammatory activity. In the present study alkaloid and flavonoid extracts of *Vitex trifolia* showed good anti-LOX activity with an IC50 value of 29.87 µg/ml (Graph 2). LOX inhibition was used to evaluate anti-inflammatory activity of a few medicinal plants used in Limousin country. *Filipendula ulmaria* (Meadow sweet) recorded LOX inhibition with IC50 of 60 µg/ml and *Urtica dioica* (Nettle) methanolic extract inhibited LOX with IC50 of 348 µg/ml (Trouillas *et al*., 2003). In another study, eight methanolic extract out of 18 undomesticated plants of South Africa showed significant inhibition of 5-lipoxygenase (5-LOX) activity. *Bidenspilosa* extract exhibited IC_{50} of 21.8 µg/ml and *Emexaus trails* extract recorded IC_{50} of 81.4 µg/ml for LOX inhibition (Wang *et al*., 2006). LOXs are sensitive to antioxidants as antioxidants are involved in inhibition of lipid hydroperoxide formation due to scavenging of lipidoxy orlipidperoxy-radicals. This could lead to less availability of lipid hydro peroxide substrate required for LOX catalysis (Rackova *et al*., 2007 and Gentallan *et al*., 2019).

The effects of alkaloid and flavonoid extract of *Vitex trifolia*on the monophenolase and the diphenolase activities of tyrosinase were studied. The results showed that alkaloid and flavonoid extracts of *Vitex trifolia* could lengthen the lag phase of the monophenolase and decrease the steady-state rate of both monophenolase and diphenolase activity. The inhibition was displayed as reversible and the inhibition type was found to be mixed type. The inhibition constants for alkaloid and flavonoid extracts of *Vitex trifolia* binding with the free enzyme (E), KI, were obtained to be 0.507 and 1.543 mM, respectively, indicating that alkaloid and flavonoid extracts of *Vitex trifolia* was three times as potent for inhibiting the free enzyme. However, the inhibition constants for these two inhibitors binding with the enzyme substrate



Graph 3: The inhibitions of flavonoid alkaloid extract of *Vitex trifolla on* the diphenolase activity of mushroom tyrosinase for the catalysis of DOPA.



Graph 4: *In vitro* xanthine inhibitory activity of flavoniod and alkaliod extract from *Vitex trifolla.*

complex (ES), KIS, were obtained to be 1.354 and 1.321 mM, respectively. The values are almost the same, indicating alkaloid and flavonoid extracts of *Vitex trifolia* inhibited ES complex as potently as 4-cyanobenzoicacid. In addition, anthocynin extracts of *Eclipta alba* inhibited the free enzyme more potently than it did the ES complex. However, the inhibitory effect of alkaloid and flavonoid extracts of *Vitex trifolia*was the contrary.

CONCLUSION

Alkaloid and flavonoid extracts of *Vitex trifolia* hold great promise for natural treatments of arthritis that are safe and effective and can be provided as dietary supplements, added to multiple vitamins and incorporated into food products to create functional foods. In addition, the novel bioactives identified in the alkaloid and flavonoid extracts of *Vitex trifolia* extracts, when fully characterized, could prove to be promising new drug leads for Cycloxygenase, Lipoxygenases, Tyrosinase and Xanthine oxidase inhibition as well as triple inflammatory enzyme inhibitors for treatment of a range of inflammatory diseases that are safe and efficacious.

Conflict of interest: None.

REFERENCES

Agrawal, V.S. (1997). Drug plants of India. Kalyani Publisher, Cuttak; 2: 721.

Burda, S. and Oleszek, W. (2001). Antioxidant and antiradical activities of flavonoids. Journal Agriculture food chemcial. 49: 2774-2779.

Charlier, C. and Michaux, C. (2003). Dual inhibition of cyclooxygenase-2 (COX-2) and 5 lipoxygenase (5-LOX) as a new strategy to provide safer non-steroidalanti-inflammatory drugs. European Journal of Medicinal chemistry. 38: 645-659.

Cos, P., Vlietinck, A.J., Berghe, D.V. and Maes. (2006). L. Anti-infective potentialof natural products: How to develop a stronger *in vitro* proof-of-concept. Journal of Ethnopharmacology. 106: 290-302.

Cotran, R.S., Kumar, V. and Robbins, S.L. (1994). Pathologic basis of disease. Philadelphia. Saunders WB Company.

Fabricant, D.S. and Fansworth, N.R. (2001). The value of plants used in traditional medicine for drug discovery. Environment Health Perspective. 109: 69-75.

Gentallan, R.P., Altoveros, N.C., Borromeo, T.H. and Macabecha, C.G.A. (2019). Antimicrobial and phytochemical properties of atunarace mosa raf. Kernel Extract. Indian Journal of Agricultural Research. 53: 733-736.

Gupta, M., Mazumdar, U.K., Gomathi, P. and Selven, V.T. (2006). Anti-inflamatory evaluation of leaves of Plumeriaacuminata. BMC Complementary and Alternative Medicine. 6: 36. https://doi.org/10.1186/1472-6882-6-36.

Harborne, J.B. and Williams, C.A. (2000). Advances in flavonoid research since1992. Phytochemistry. 55: 481-504.

Hou, D., Yanagita, T., Uto, T., Masuzaki, S. and Fujii, M. (2005). Anthocyanidins inhibit cyclooxygenase-2 expression in LPS-evoked macrophages: Structure-activity relationship and molecular mechanisms involved. Biochemical Pharmacology. 70: 417-425.

Khan, K.M., Maharvi, G.M., Khan, M.T.H., Shaikh, A.J., Perveen, S., Begum, S. and Choudhary, M.I. (2006). Tetraketones: A new class of tyrosinase inhibitors. Bioorganic Medical Chemistry. 14: 344-351.

Lin, C.C. and Kao, C.H. (2001). Cell wall peroxidase activity, hydrogen peroxide level and NaCl-inhibited root growth of rice seedlings. Plant and Soil. 230: 135-143.

Padmalatha, K., Jayaram, N.L., Raju, M.N., Prasad, V. and Rajesh Arora. (2009). Ethnopharmacological and biotechnological significance of vitex. Bioremediation, Biodiversity and Bioavailability. 3(1): 6-14.

Patel, H.B., Patel,U.D., Modi, C.M., Ladumor, V.C., Makwana, C.N. and Rao, S.S. (2022). Pharmacokinetics of marbofloxacin following oral administration in piperine, quercetin alone and both in combination pretreated broiler chickens. Indian Journal of Animal Research. 56: 1269-1273.

Pavrez, M., Mahboob, H.K., Zahuul, I. and Shek, M.H. (2005). Antimicrobialactivities of the petroleum ether, methanol and acetone extracts of kaempferiagalangal. rhizome. Journal of Life Earth Science. 1: 25-29.

Prachayasittikul, S., Buraparuangsang, P., Worachartcheewan, A., Isarankura- Na-Ayudhya, C., Ruchirawat, S. and Prachayasittikul, V. (2008). Antimicrobial andantioxidant activity of bioreactive constituents from Hydnophytum formicarum Jack. Molecules. 13: 904-921.

Rackova, L., Oblozinsky, M., Kostalova, D., Kettmann, V. and Bezakova, L. (2007). Free radical scavenging activity and lipoxygenase inhibition of Mahoniaaquifolium extract and isoquinoline alkaloids. Journal of Inflammation. 4: 15-21.

Roh, J.S., Han, J.Y., Kim, J. and Hwang, J.K. (2004). Inhibitory effects of active compounds isolated from safflower (*Curthamustin ctorius* L.) seeds for melanogenesis. Biology Pharma Bulletin. 27: 1976-1978.

Trouillas, P., Calliste, C.A., Allais, D.P., Simon, A., Marfak, A., Delage, C. and Duroux, J.L. (2003). Antioxidant, anti-inflammatory and anti-proliferative properties of sixteen water plant extracts used in Limous in countryside as herbal teas. Food Chemistry. 80: 399-407.

Wang, K.H., Lin, R.D., Hsum, F.L., Huang, Y.H., Chan, H.C. and Lee, M.H. (2006). Cosmetic applications of selected traditional Chinese herbal medicine. Journal of Ethnopharmacology. 106: 353-359.

Zerihun, J. (2022). Production, challenges, potential and prospects of oromo potato (*Plectranthus edulis*) and anchote (*Coccinia abyssinica*) in Ethiopia: A review. Agricultural Reviews. 43: 348-354..

ANALYSIS OF PHYTOCHEMICALS IN ROOT SAMPLE OF *ABUTILON INDICUM* FOR TREATMENT OF VARIOUS DISEASES BY USING MOLECULAR DOCKING

Arunthathi, K., Jummana kasina, K and P.Dhasarathan*

Dept. of Microbiology, Thassim Beevi Abdul Kader College for Women, Kilakarai.
*Dept. of Biotechnology, Madha Engineering College, Kundrathur, Chennai-69, India.

ABSTRACT

Abutilon indicum is a medicinal plant that belongs to the Malvaceae family, which is allotted for duration of a wide variety of tropical and subtropical areas and has been used for more than a few issues in regular medicine. Its leaves, root, and bark has enormous medicinal value, and are used as an aphrodisiac, anti-diabetic, and diuretic. The leaf extract of *Abutilon indicum* is used to treat piles. *Abutilon indicium* root extract was tested to treat different diseases. The phytochemical screening was also carried out to find various phyto constituents present in the *Abutilon indicum*. The ethanolic extract was subjected to gas chromatography-mass spectrometry (GC-MS) analysis for the identification of compounds present in the root sample. The identified compounds will be docked with different diseases (Tuberculosis, Leprosy, and Gonnorrhoeae).

Keywords: *Abutilon indicum*, phytochemical screening, GC-MS, Tuberculosis, Leprosy, Gonorrhoea.

INTRODUCTION

Abutilon indicum is a South Asian native plant that is also known as "Thuthi" or "Kanghi" in Hindi. Natural medications are effective without causing negative effects. Indian mallow is a medicinal herbal plant and is highly used in most traditional medicine practices like Ayurveda, Siddha, Unani, and Folk. The root extract of the plant has given for ulcer and hematuria treatments. It is good medicine for cough and fever. The root infusion is even natural eyewash to use. Traditionally, Root and bark are used as aphrodisiac, anti-diabetic, nervine tonic, and diuretic. Seeds are used in urinary disorders (Chen *et al*., 2008).

In 2020, 10 million people were off worldwide (TB), 5.6 million men, 3.3 million women, and 1.1 million children. TB is the 13th leading cause of death and the second-leading infectious killer after COVID-19. It is a contagious disease caused by *Mycobacterium tuberculosis*. It usually attacks the lungs, although it can also spread to other parts of the body, and is transmitted through coughing and sneezing. The simplicity with which TB infection spreads, for instance, by inhalation of droplets nuclei 2–5 mm in diameter containing as few as 1–3 bacilli has helped to sustain this scourge at current levels (Arunthathi *et al*., 2023).

In global level 1, 27, 558 new leprosy cases were detected during 2020. It includes 8,629 children below 15 years. Leprosy affects humanity and it causes two to three million people to disable permanently. *Mycobacterium leprae* is an etiologic agent, identified by the Norwegian doctor Gerhard Hansen, causes a chronic infectious disease called leprosy or Hansen's disease. For ages, it is considered an incurable disease. Nerve damage, disfiguring skin sores, and progressive debilitation are some of the characteristics of this disease. New gonococcal infections occur 1.6 million of United States during 2018 and more than half occur among young people aged 15-24. Gonnorrhoeae is the second most commonly reported bacterial sexually transmitted infection in the United States (Arunthathi *et al*., 2023). From this background the present study planned to find various phyto constituents present in the *Abutilon indicum* for treatment of various diseases.

Materials and Methods

Plant material: The fresh roots were collected from Devipattinam, Ramanathapuram (district), Tamilnadu, India

Preparation of extract: Leaves were separated from the *A. indicum* plant and then washed with water and then dried under shade at room temperature for 7 days. The dried leaves were grounded into a fine powder using the blender mixture. The extraction was carried out by using a solvent of ethanol.

Preparation of ethanolic extract: 10g of dried finely powdered root sample was taken in 200ml of the ethanolic extract by dissolving the overnight. After, the extracts were magnetic stirred over 1-2 hours and then filtered using No.1 Whatman filter paper and stored in an airtight container for further analysis

Phytochemical screening: Qualitative analysis of extract was carried out to determine the presence of various bioactive compounds using the standard qualitative procedure.

Determination and screening of bioactive compounds by using GCMS method: Qualitative and quantitative analysis of phytochemicals can be done using Gas Chromatography Mass Spectroscopy (GCMS). GC-MS can be applied to solid, liquid and gaseous samples. First the samples are converted into gaseous state then analysis is carried out on the basis of mass to charge ratio (Sahira and Catherine, 2015). The GC-MS analysis of extract of *A. indicum* root was carried out using Agilent GC 7890 with triple axis 5975 MS detector. The capillary column was Agilient HP-5MS (30 m x 250 µm x 0.25 µm) composed of 5% phenyl methyl silox. The initial oven temperature was 55 C for 0 min which was raised at rate of 10 C/min upto 200C for 0 min and then at rate of 5 C/min upto 260C for the hold time of 5 min. The injector volume was 10 µl. The Helium gas used as a carrier with constant flow rate of 1 ml/min with split ratio of 10:1. The MS operating conditions were; source temperature 250C (max-300C), quad temperature 150C (max-200C), solvent delay time of 3 min. Compounds were identified in terms of Rt values and mass spectra with those obtained from the NIST search library. The obtained compounds were searched for detailed pharmacological activities.

Studies of *Abutilon indicum* roots with three main proteins: Bioactive compounds from root substances which belong to *A. indicum* were obtained. Those compounds were revealed by the GCMS analysis. The crystal of three main protein structures was retrieved from Protein Data Bank RCSB PDB (http://www.rcsb.org/pdb/home/home.do)Then the PDB format of protein was subjected for docking studies.

Lipinski's rule: Lipinski's rule of 5 was developed by Christopher A. Lipinski in 1997, this rule was also called as Pfizer's rule of five or simply the rule of five (R05). This rule was developed to set "draggability" guidelines. In the drug discovery setting, the rule of 5 predicts that poor physiochemical and structural properties within certain ranges (Devvret *et al*., 2017). **Preparation of target protein structure: LSR2** is the target protein against the phyto compounds identified from *A. indicum*. The protein LSR2 is responsible for Tuberculosis. The three-dimensional crystal structure of LSR2 (PDB ID: 4EIP). **3AFQ** is the target protein against the phyto compounds identified from *A. indicum*. The protein 3AFQ is responsible for Leprosy. The three-dimensional crystal structure of 3AFQ (PDB ID: 3AFQ). **6JTI** is the target protein against the phytocompounds identified from *A. indicum*. The protein 6JTI is responsible for Gonnorrhoeae. The three-dimensional crystal structure of 6JTI (PDB ID: 6JTI).

Molecular docking: Among the 66 ligands were used against the protein **4EIP, 3AFQ, 6JTI**. These compounds have been studied. Protein and ligands were converted into special file format PDBQT. Ligand preparation included the following steps (i) addition of hydrogen atoms, (ii) neutralization of the charge groups and (iii) removal of any miscellaneous structures from the ligand. Prepared and optimized structures of ligands were used for docking simulation. Grid values for protein the grid point set at 80 x 60 x 95. Before starting the docking study, Vina wizard software was used to perform molecular docking in PyRx virtual screening tool. Discovery studio visualizer I was used to examine the docking poses of the complexes (Kailasam, 2015).

Result and Discussion

Phytochemical analysis of ethanolic extracts of *Abutilon Indicum* root indicated the presence of bioactive compounds. Reports were revealed the identification of major chemical constituents such as Alkaloids, Coumarins, Saponins, Carboxylic acid, Fixed oils, Flavonoids are present (Table 1-2 and Graph 1)

Abutilon indicum roots were collected from Devipattinam, and an ethanolic extract was prepared. Phytochemical screening was done on ethanolic extract. Docking studies were done by using compounds obtained from ethanolic extract (GC-MS analysis) with proteins **4EIP, 3AFQ, and 6JTI** (Table 3-5 and Figure 1-3)**.** Phytochemical analysis of Ethanol extracts of *A. indicum* (roots) indicated the presence of bioactive compounds. The results obtained from the phytochemical analysis had shown similarities with the literature. Molecular docking has become an increasingly important tool for drug discovery. The completion of the human genome project has resulted in an increasing number of new therapeutic targets for drug discovery. From the molecular docking studies, we conclude that identified novel bioactive ligands of *Abutilon indicum* (root) extract were checked for their three different diseases (Tuberculosis, Leprosy, Gonorrhoeae) using PyRx-Python prescription (version 0.8) virtual screening tool (Nwude and Ebong, 1980 and Nag *et al*., 2013).

The docking studies against **4E1P** revealed in GC-MS analysis of 4 bioactive showed the docking binding energy of (-6.8,-7.2,-6.1,-6.2) Kcal/mol respectively have good Tuberculosis activity against 4E1P. The docking studies against **3AFQ** revealed in GC-MS analysis of 17

bioactive Compounds showed the docking binding energy of (-7.5,-6.1,-6.3,-7.1,-7.1,-8.9,-8.1,-6.3,-8.5,-7.4,-7,-7.3,-7.1,-7,-6.3,-6.8,-6.8) Kcal/mol respectively have good Leprosy activity against 3AFQ (Olangeli *et al*., 2009 and Yadav and Agarwala, 2011). The docking studies against **6JTI** revealed in GC-MS analysis of 17 bioactive Compounds showed the docking binding energy of (-6.6,-6.4,-8.4,-8.9,-8.6,-7,-6.4,-6.6,-9,-7.8,-8.3,-9.3,-8.1,-6.3,-8.1,-7.6,-7.4) Kcal/mol respectively have good Gonnorrhoeae activity against 6JTI . Linker rule analysis performed and confirmed the findings (Table 6)

CONCLUSION

The properties of the root of *A. indicum* were studied *in silico* level. In the future, in vitro, and in vivo studies can be done to develop or transform *A. indicum* into nanogel and nano tablet as a painkiller and drug to cure targeted diseases (tuberculosis, gonorrhoea, leprosy).

Table 1- phytochemical screening of *Abutilon indicum* in root

S.NO	TESTS		REAGENT	ETHANOL EXTRACT
1.	Alkaloids	a) Mayer's test	Mayer's reagent	-
		b) Wagner's test	Wagner's reagent	+
2.	Phenols	Ferric chloride test	FeCl3	-
3.	Coumarins		NaOH	+
4.	Terpenoids	Salkowski's test	$CHCl_3$, H_2SO_4	_
5.	Quinones	Alkali test	10% NaOH	_
6.	Anthraquinones	Bontrager's test	HCL	_
7.	Tannins	Gelatin test	NaCl	_
8.	Phlobatannins		NH_4OH	_
9.	Carbohydrates	a) Molish's test	Alpha naphthol	+
		b) Benedict's test	Benedict's reagent	−
		c)Fehling's test	Fehling's reagent	
10.	Glycosides	Legal's test	$CHCl_3$, NH_40H	_
11.	Cardiac glycosides	Keller-killani test	CH_3COOH, $FeCl_3$, H_2SO_4	−
12.	Proteins	a) Millon's test	Millon's reagent	−
		b) Biuret test	$CuSO_4$, C_2H_5OH, KOH	
13.	Aminoacids	Ninhydrin test	Ninhydrin ragent	_
14.	Steroids	Ferric chloride-Acetic acid test	$CHCl_3$,$FeCl_3$, CH_3COOH, H_2SO_4	−
15.	Phytosteroids	Foam test	$CHCl_3$, H_2SO_4	_
16.	Saponins		H_2O	+
17.	Acids		$NaHCO_3$	_
18.	Carboxyclic acids	Phenolphthalein test	$C_{20}H_{14}O_4$, NaOH	+

19.	Fixed oils	Oil detection test	Whatman filter paper	+
20.	Fats	Saponification test	KOH, $C_{20}H_{14}O_4$	_
21.	Gums and Mucilages		C_2H_5OH	_
22.	Flavonoids	a) Aqueous NaOH	NaOH	+
		b) Sulphuric acid	H_2SO_4	-

Graph 1: Gas Chromatography-Mass spectrometry chromatogram of ethanol extract of *Abtilon indicum* (Root)

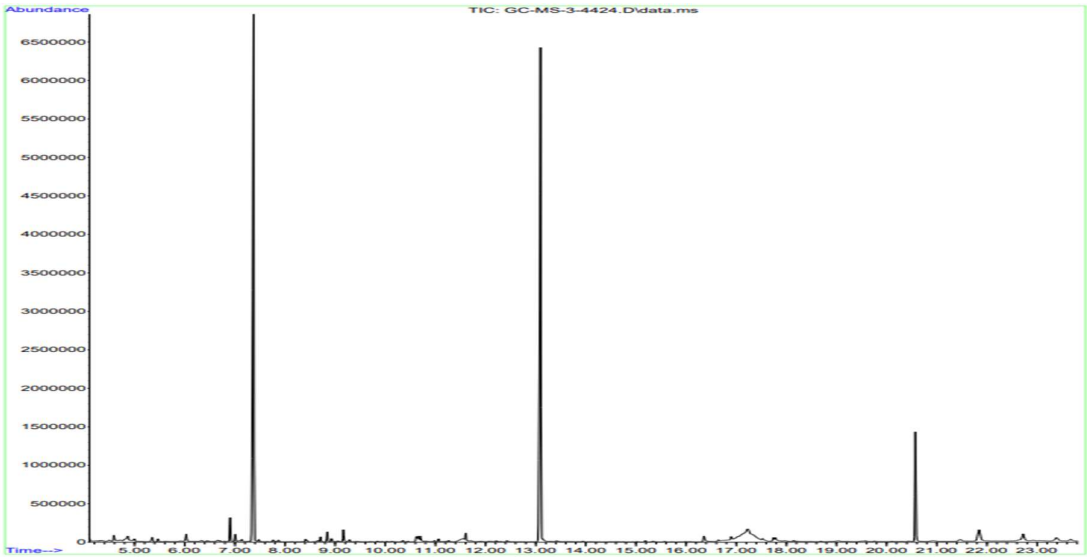


Table 2 : Phyto-components identified from the ethanolic extract of Root of *Abutilon indicum*

S.No	Rt	Pubchem Id	Compounds	Molecular Weight	Molecular Formula	Peak Area%
1	7.009	24624	Propane, 1,1,3-triethoxy-	176.25	$C_9H_{20}O_3$	0.34
2	7.009	41027	Ethanethioamide, N,N-dimethyl-	255.4	$C_{16}H_{17}NS$	0.34
3	7.375	5369371	cis-4-Hepten-1-al diethyl acetal	186.29	$C_{11}H_{22}O_2$	36.97
4	7.375	77223	Pentane, 1,1-diethoxy-	160.25	$C_9H_{20}O_2$	36.97
5	7.375	77225	Butane, 1,1-diethoxy-	146.23	$C_8H_{18}O_2$	36.97
6	13.097	6781	Diethyl Phthalate	222.24	$C_{12}H_{14}O_4$	34.53
7	20.574	20393	1,2-Benzenedicarboxylic acid, mono(2-ethylhexyl) ester	278.34	$C_{16}H_{22}O_4$	5.80
8	20.574	6423885	Phthalic acid, 2-ethylhexyl isohexyl ester	362.5	$C_{22}H_{34}O_4$	5.80
9	20.929	610065	1,2,4-Triazol-3-amine, 5-(1,3,5-trimethyl-4-pyrazolyl)amino	207.24	$C_8H_{13}N_7$	0.23

Table 3 - Bioactive compounds of *A. indicum* root against 4EIP Protein with best Binding affinity.

PROTEIN	PUBCHEM ID	COMPOUNDS	BINDING AFFINITY
4EIP	73145	beta-Amyrin	-7.2
	73170	alpha-Amyrin	-6.8
	457801	gamma-Sitosteol	-6.2
	5280794	Stigmasterol	-6.1

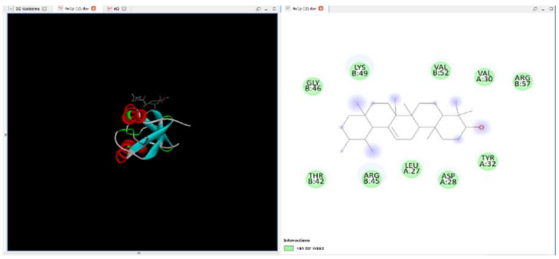


Figure 1:2D&3D interaction of protein 4EIP with beta-Amyrin

Table 4 - Bioactive compounds of *A. indicum* root against 3AFQ Protein with best Binding affinity.

PROTEIN	PUBCHEM ID	COMPOUNDS	BINDING AFFINITY
3AFQ	73170	alpha-Amyrin	-8.9
	610065	1,2,4-Triazol-3-amine, 5-(1,3,5-trimethyl-4-pyrazolyl)amino	-8.5
	73145	beta-Amyrin	-8.1
	22215476	Methyl3,4-ethylidene-alpha-D-galactopyranoside	-7.5
	974047	[1,2,4]Triazolo[1,5-a]pyrimidine-6-carboxylic acid, 7-amino-, ethyl ester	-7.4
	5280794	Stigmasterol	-7.3
	457801	gamma-Sitosterol	-7.1
	573406	2H-Pyran-2,4(3H)-dione, dihydro-6-(4-fluorophenyl)-3,3-dimethyl-5-spirocyclohexane-	-7.1
	712442	1H-Pyrimidine-2,4-dione, 1-(2,2-diethoxyethyl)-5-methyl-	-7.1
	173183	Campesterol	-7
	222284	beta-Sitosterol	-7
	610182	Benzo[h]quinoline, 2,4-dimethyl-	-6.8
	258555	3-Methyl-2-phenylindole	-6.8
	573508	Cyclohexane, [2-[(2-ethylhexyl)oxy]ethyl]-	-6.3
3AFQ	610161	2-Ethylacridine	-6.3
	526607	1,5-Anhydro-3-O-acetyl-2,4,6-tri-O-methyl-D-galactitol	-6.1

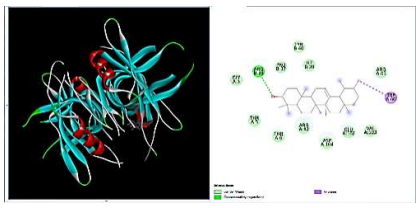


Figure 2:2D&3D interaction of protein 3AFQ with alpha-Amyrin

Table 5 - Bioactive compounds of *A. indicum* root against 6JTI Protein with best Binding affinity.

PROTEIN	PUBCHEM ID	COMPOUNDS	BINDING AFFINITY

6JTI	457801	gamma-Sitosterol	-9.3
	91703800	Ergost-5-en-3-ol, (3.beta.)-	-9
	73170	alpha-Amyrin	-8.9
	73145	beta-Amyrin	-8.6
	573406	2H-Pyran-2,4(3H)-dione, dihydro-6-(4-fluorophenyl)-3,3-dimethyl-5-spirocyclohexane-	-8.4
	5280794	Stigmasterol	-8.3
	222284	beta-Sitosterol	-8.1
	610161	2-Ethylacridine	-8.1
	173183	Campesterol	-7.8
	610182	Benzo[h]quinoline, 2,4-dimethyl-	-7.6
	258555	3-Methyl-2-phenylindole	-7.4
	75524	Pyrene, hexadecahydro-	-7
	41027	Ethanethioamide, N,N-dimethyl-	-6.6
	580058	Cyclohexane,1,1'-(oxydi-2,1-ethanediyl)bis-	-6.4
	6431151	Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1.alpha.,2.beta.,4.beta.)]-	-6.4
	612605	2(1H)Naphthalenone, 3,5,6,7,8,8a-hexahydro-4,8a-dimethyl-6-(1-methylethenyl)-	-6.3

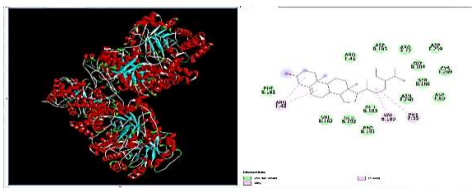


Figure 3: 2D&3D interaction of protein 6JTI with gamma-Sitosterol

Table 6: Lipinski rule of *A. Indicum* bioactive compounds analysed using admetSAR

S.No	Compound	Molecular weight	Log p<5	H-Bond donor< 5	H-Bond acceptor<10	No. of violation
1	Methyl 3,4-ethylidene-alpha-D-galactopyranoside	220.22	-1.2	2	6	0
2	1,5-Anhydro-3-O-acetyl-2,4,6-tri-O-methyl-D-galactitol	248.27	-0.4	0	6	0
3	Ethanethioamide, N,N-dimethyl-	255.4	3.6	0	1	0
4	Cyclohexane, 1,1'-(oxydi-2,1-ethanediyl)bis-	266.5	6.8	0	1	1
5	Cyclohexane, [2-[(2-ethylhexyl)oxy]ethyl]-	240.42	6.3	0	1	1
6	2H-Pyran-2,4(3H)-dione, dihydro-6-(4-fluorophenyl)-3,3-dimethyl-5-spirocyclohexane-	304.4	4.3	0	4	0
7	1H-Pyrimidine-2,4-dione, 1-(2,2-diethoxyethyl)-5-methyl-	242.27	0.1	1	4	0
8	alpha-Amyrin	426.7	9	1	1	1
9	beta-Amyrin	426.7	9.2	1	1	1
10	Pyrene, hexadecahydro-	218.38	6.3	0	0	1
11	Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1.alpha.,2.beta.,4.beta.)]-	204.35	6.1	0	0	1
12	1,2,4-Triazol-3-amine, 5-(1,3,5-trimethyl-4-pyrazolyl)amino	207.24	0.6	3	5	0
13	[1,2,4]Triazolo[1,5-a]pyrimidine-6-carboxylic acid, 7-amino-, ethyl ester	207.19	0.5	1	6	0
14	Ergost-5-en-3-ol, (3.beta.)-	470.8	8.8	1	1	1
15	Campesterol	400.7	8.8	1	1	1
16	Stigmasterol	412.7	8.6	1	1	1
17	gamma-Sitosterol	414.7	9.3	1	1	1
19	beta-Sitosterol	414.7	9.3	1	1	1
20	2(1H)Naphthalenone, 3,5,6,7,8,8a-hexahydro-4,8a-dimethyl-6-(1-methylethenyl)-	218.33	3.2	0	1	0
21	2-Ethylacridine	207.27	4.2	0	1	0
22	Benzo[h]quinoline, 2,4-dimethyl-	207.27	4.2	0	1	0
23	3-Methyl-2-phenylindol	207.27	4.1	1	0	0

REFERENCE

Arunthathi, K ., Srividya, K., Abhinav, J and P. Dhasarathan. 2023. Exploring anti-diabetic compounds from the ethanolic extraction of *Strychnos potatorum* seeds: ligand- based design, molecular dynamics. YMER, 22 (1): 435-444.

Chen JM, Ren H, Shaw JE, Wang YJ and M. Li. 2008. Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. Nucleic Acids Res 36: 2123–2135.

Devvret, Kumud Pant, Ashish Thapliyal, and Neema Tufchi. 2017. In Silico Docking analysis of Mycobacterium tuberculosis potential targets AftB and EmbA with selected phytochemicals." *International Journal of Pharmacy Research & Technology,* 7(2): 15-22.

Kailasam K.V. 2015. Abutilon indicum L (Malvaceae) - medicinal potential review. *Pharmacognosy Journal,* 7(6): 330-332.

Nag, Sudipa, Anirban Paul, and Rituparna Dutta. 2013. Phytochemical analysis of methanolic extracts of leaves of some medicinal plants. *IJSRP* 3(4): 1-5.

Nwude, N and O. Ebong. 1980. Medicinal plants used in the treatment of leprosy in Africa." *Leprosy review,* 51(1):11-18.

Olangeli R, Haq A, Arcus VL, Summers E and RS. Magliozzo. 2009. The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. Proc Natl Acad Sci USA, 106: 4414–4418.

Sahira, Banu J and L Cathrine. 2015. General techniques involved in phytochemical analysis. *International Journal of Advanced Research in Chemical Science,* 2 (4): 25-32.

Yadav RNS and M. Agarwala. 2011. Phytochemical analysis of some medicinal plant. Journal of phytotology, 3(12):10-14.