ANALYSIS OF *MYCOBACTERIUM FORTUITUM* PROTEOME USING MACHINE LEARNING TECHNIQUES

Thesis submitted in partial fulfillment of the requirement for the

Degree of Masters of Science

IN

BIOTECHNOLOGY

BY

SHAN GHAI (225111004)

Under the supervision of

Dr. Rahul Shrivastava and Prof. Shruti Jain

2024

Department of Biotechnology and Bioinformatics

Jaypee University of Information and Technology,Waknaghat,

Solan – 173234, Himachal Pradesh

DECLARATION

I hereby declare that the project work entitled "**Analysis of *Mycobacterium fortuitum* Proteome Using Machine Learning Techniques.**" has been solely submitted to the Department of Biotechnology and Bioinformatics, **Jaypee University of Information Technology, Waknaghat (Solan)** is a record of an original work done by me under the supervision of **Dr. Rahul Shrivastava and Prof. Shruti Jain.**

Signature:

Name: **Shan Ghai (225111004)**
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology, Waknaghat
Solan.
Date: 20.05.2024

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, P.O. – WAKNAGHAT,
TEHSIL – KANDAGHAT, DISTRICT – SOLAN (H.P.)
PIN – 173234 (INDIA) Phone Number- +91-1792-257999
(Established by H.P. State Legislature vide Act No. 14 of 2002)

CERTIFICATE

This is to certify that the work reported in the M.Sc. project report entitled **"Analysis of *Mycobacterium fortuitum* Proteome Using Machine Learning Techniques"** which is being submitted by **Shan Ghai (225111004)** in fulfillment for the award of **Masters of Science in Biotechnology and Bioinformatics** by the Jaypee University of Information Technology, is the record of candidate's own work carried out by him under our supervision. This work is original and has not been submitted partially or fully anywhere else for any other degree or diploma.

Dr. Rahul Shrivastava
Associate Professor
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology
Waknaghat, Distt-Solan, H.P. - 173234
Date:20.05.20.24

Prof. Shruti Jain
Associate Dean (Innovation)
Jaypee University of Information Technology
Waknaghat, Distt-Solan, H.P. - 173234
Date: 20.05.20.24

ACKNOWLEDGEMENT

I would like to express my profound gratitude to my guide Dr. Rahul Shrivastava and Dr. Shruti Jainfor their guidance, support and constant encouragement throughout the course of this project work. They were more than just my project guides; at times a mentor to rescue me out of my doubts. They always helped me to work hard and also taught me how to implement different ideas to deal with the problem. Moreover, they taught me to not give up and many other valuable lessons.

Furthermore, I would like to acknowledge Vice-Chancellor Prof. (Dr.) Rajendra Kumar Sharma, Prof. (Dr.) Ashok Kumar Gupta, Dean of academics & research for providing me with an opportunity to be a part of the institute and to complete my **Master's Degree**.

I also want to mention the **HoD** of Biotechnology and Bioinformatics **Prof. Sudhir Kumar** has been a source of immense motivation and inspiration both for my academic and personal life. He was never, and I know will never be, more than just a phone call away. He has helped me in almost every aspect I have asked him for.

In addition, I would like to thank all the faculty members of the BT/BI Department of JUIT, who have helped me whenever I needed and also would like to thank all the lab engineers.

I would also like to appreciate the part that my classmates (Prajjwal, Bishal, Akshita, Shashank) have played in shaping this project work. They have been my constant support and cheered me up at hard times. They helped me whenever I had any doubts. Thanks a lot!

Shan Ghai
M.Sc. Biotechnology
JUIT, Solan

Table of Content

List of Tables

List of Figures

Abstract

Background: *Mycobacterium fortuitum*is an ubiquitous, opportunistic pathogen responsible for a causing number of nosocomial infections, particularly affecting immunocompromised patients. Its ability to survive under diverse environments and resist common antibiotics presents a significant challenge in healthcare settings. Understanding the molecular mechanisms underlying *M. fortuitum's* pathogenesis is crucial for developing effective diagnostic tools and therapeutic strategies.

Objective: This thesis investigates the proteome of *M. fortuitum*, aiming to identify key proteins involved in its biofilm formation and antibiotic resistance. By employing a machine learning-based approach, we sought to extract valuable insights from the complex proteomic data and discover potential targets for future interventions.

Methods: We performed a comprehensive proteomic analysis of *M. fortuitum*, isolating and identifying its global proteome. Subsequently, we implemented various machine learning algorithms to analyze the vast dataset and identify informative patterns within the protein profiles.

Results: Utilizing Machine learning, we successfully identified a set of proteins potentially associated with *M. fortuitum* biofilm formation. These proteins may represent as potential drug targets or biomarkers for diagnosis. Among the algorithms used, k Nearest Neighbor(kNN) emerged as the most effective tool for our specific data, demonstrating superior performance in protein classification and feature selection.

Conclusion: The findings presented in this thesis contribute significantly to our understanding of the molecular basis of *M. fortuitum* survival and biofilm formation. By leveraging machine learning tools and proteomics.

Chapter 1

Introduction

Introduction

Mycobacterium fortuitum, a ubiquitous bacterium, has become a growing concern in healthcare settings due to its opportunistic nature [1]. This pathogen is responsible for an escalating number of nosocomial infections, particularly affecting immunocompromised individuals such as those undergoing medical procedures or with weakened immune systems [2]. The ability of *M. fortuitum*to thrive in diverse environments, including water and medical devices, coupled with its resistance to common antibiotics, makes it a challenging pathogen to manage [3, 4]. Traditional methods for diagnosing and treating *M. fortuitum*infections are often time-consuming and limited in efficacy [5].

Understanding the underlying molecular mechanisms of *M. fortuitum's* pathogenesis is crucial for developing effective strategies to combat this growing threat. Proteomics, a powerful technique that analyzes an organism's entire protein complement, offers a unique window into its biological processes [6]. By investigating the proteome of *M. fortuitum*, we can identify key proteins involved in its survival, virulence, and antibiotic resistance mechanisms [7].

This thesis delves into the application of proteomics in conjunction with machine learning to unravel the intricacies of *M. fortuitum*'s biology. We employed various machine learning algorithms to analyze the complex proteomic data, allowing us to identify informative patterns and extract valuable insights. Among the algorithms tested, the Support Vector Machine (SVM) demonstrated superior performance in our analysis [8].

This thesis presents the comprehensive findings of our investigation. We detail the methodology employed for the proteomic analysis of *M. fortuitum*and the subsequent application of different machine learning algorithms. With a particular focus on the effectiveness of SVM, we discuss the identified proteins and their potential roles in the context of *M. fortuitum*'s pathogenesis. The insights gained from this study can

significantly contribute to our understanding of this emerging pathogen and pave the way for the development of targeted interventions, potentially leading to improved diagnostic tools and more effective therapeutic strategies against *M. fortuitum* infection.

Chapter 2

Review of literature

Review of literature has been divided into two sections – first section is regarding *M.fortuitum* and second sections deals with the domain of machine learning.

2.1 *Mycobacterium fortuitum*

M. fortuitum, a rapidly growing non tuberculous mycobacterium of the Runyon group IV, is one of the currently most studied bacteria, mainly due to the exceptional features and the difficulty in treating drug resistance, which it can create [9]. The bacterium's discovery from frogs and the fact that it is widespread in soil and water indicates the scope of the threat associated with its ubiquity, both in the indoor and outdoor environmental. Moreover, the changing classification of mycobacteria, especially rapidly growing Mycobacteria, elucidates the difficulty in understanding and detecting mycobacterial infections. This complexity is additionally associated with the pathogen's ambiguity when it comes to the affected population, as *M. fortuitum* can infect immune-compromised subjects, which creates the heterogeneity of the clinical picture, mostly depending on the route of entry [9].

As we delve into exploring *M.fortuitum*, this section aims to provide a comprehensive review of its characteristics, pathogenicity, transmission modes, and the challenges posed bydrug resistance in treating infections caused by this rapid growing mycobacterium causing clinical infections[9][10]. With advances in molecular techniques and microbiologic methods leading tothe identification of new NTM species, including *M. fortuitum*, understanding the dynamics of drug resistance becomes pivotal in developing effective treatment strategies. The subsequent sections will cover the epidemiology, clinical manifestations, and both current and emerging therapeutic responses to combat these infections[10].

1. Characteristics of *Mycobacterium fortuitum*

Physical Characteristics: The physical form of the bacterium is Gram-positive, nonmotile, and an acid-fast rod. They are between 1-3 µm long and 0.2-0.4 µm wide. Some bacteria form colonies, which are smooth and hemispheric and range in color from off-white to cream. The colonies also range in texture but may appear waxy, multilobate, and rosette-clustered in some conditions [11].

Growth Conditions: *M. fortuitum* has been found to grow amazingly fast in the laboratory test. It forms visible colonies on Löwensteini-Jensen media in 2-4 days. Moreover, it also has a capacity to be grown on MacConkey agar, which is very rare among many mycobacteria [11]. This occurrence is assisted by its capacity to take up blue dye specifically from the Löwenstein-Jensen media dyed with Malachite green [11].

Genetic and Taxonomic Considerations: The taxonomy of *M. fortuitum* is regularly changing and is now known worldwide with at least 50 different strains. The reason for the complex nature of *Mycobacterium tuberculosis* lies in the fact that there is an ongoing discovery and taxonomy of the Mycobacterium family of organisms. A new taxonomic species (TNTM28), which is genetically similar with *M. fortuitum*complex has been recently established. This bacterial strain is so pathogenic that the species likely has a genome containing a high G + C content among the genes responsible for virulence, including those involved in within macrophage replication and persistence [12].

Environmental Ubiquity and Pathogenicity: The*M. Fortuitum* are not only common environmental bacteria but also an opportunistic pathogen. It has been isolated from different sources such as soil, dust, rivers, lakes and also tap water. The opportunistic nature of this microbe is emphasized by its association with *M. fortuitum*-PD and various skin and soft tissue infections [13][14]. The widespread nature of the bacterium and its competence to exhibit growth in diversified environments makes it

more of a concern for public health, especially in the context of its high drug-resistance [15][16].

2. Epidemiology

Global Incidence and Distribution: The*M.fortuitum*is a non tuberculous pathogen (NTM), commonly reported worldwide without specific endemic foci. It has been shown that NTM-pulmonary disease (NTM-PD) patients are more prone to be affected by this type of pathogens especially those who were already diagnosed with other pre-existing diseases like pulmonary tuberculosis and lung cancers [14]. *M.fortuitum*, especially, has been detected in different parts of the world through monitoring of isolates from humans, animals, and the environment as well and has become a common species [11][15].

Clusters and Prevalence Patterns: Recently, certain clusters of *M. fortuitum*have been shown to be connected epidemiologically. For example, cluster I covers isolates that were got from India, Mozambique, and Cambodia between 2008 and 2012. Cluster number two consists of isolates mainly from South Africa during 2011 to 2012. Cluster three, being the most diverse, includes isolates from nearly a century from different sources such as humans, animals, and the environment (1923 to 2020) [15]. Apparently, this grouping points to geographical disparities, and maybe with differences in the strains' pathogenic ability.

Epidemiological Challenges: Study of NTM infection, in which *M. fortuitum*is included, is still in its infancy. Moreover, an added difficulty arises because of the fact that these infections are globally spread having no recognizable endemic areas, which makes the epidemiological tracking and management difficult [11]. Furthermore, it is highly possible that the toll of these diseases is grossly under calculated as they exhibit non-specific symptoms and the requirement of advanced laboratory techniques for accurate diagnosis [15].

Demographic Insights : A survey at a regional tuberculosis clinical center in Southern China, that NTM organisms presented with the highest prevalence were *Mycobacterium avium* complex (44.5%), *Mycobacterium abscessus*complex (40%) Also, the study sheds light on a relatively rare cause of pulmonary infection and the dominant causes are *M.tuberculosis* (15.7%), *Mycobacterium kansasii* (10. 0%) , and *M. fortuitum*(2.0%)[8]. The observation of this research as well, that rapidly-growing mycobacterium (RGM) diseases like those caused by *M. fortuitum*were higher in migrants compared to the resident population, implied that migration patterns and disease prevalence may also have connection [16].

Gender and Age Group Diversity: According to the demography of NTM diseases, the male-female ratio found in cases of RGM infections is noticeably lower compared to those of the SGM diseases. Furthermore, the rate of acquiring full form (SGM) diseases elevates with age, which implies the elders are more vulnerable in these infections. Bronchiectasis was a prevalent comorbidity in those patients with pulmonary RGM diseases than in those of those with SGM diseases [15].

Medical tourism and risk of infection: In NTM medical tourism has been found to be the risk that springs from the postoperative wound infections. *M. abscessus*was found to be the most common isolate, followed by *M. fortuitum*, hence a stringent surveillance by the healthcare settings to contain infections by international patients need to be emphasized [1].

3. Pathogenicity of *M.fortuitum*, Infection Types and Clinical Manifestations

i. **Skin and Soft Tissue Infections:** *Mycobacterium securum* often appears during a kind of tissue infection and can arise from surgery or a direct contact with the bacterium from the environment [11].

ii. **Respiratory Infections:** People who had beforelung diseases may be at risk of acquiring transient or chronic lunginfections, which prove the bacterium's capability to seize on lung structural deficiencies [12].

iii. **Disseminated Disease:** *M. fortuitum* might cause systemic illnesses such as endocarditis, meningitis and osteomyelitis on the severe cases with serious health risks [13]. Notable instances of the infection Infections occur particularly after heart and thoracic surgeries during which *M. fortuitum*is usually a causative agent, which delays wound healing and necessitates long courses of antimicrobial therapy.

The bacterium has been proved to be a source of postoperative infections following breast augmentation, hence a hint that it can also infect surgical implants and prosthetic devices [13].

The Treatment Issues

M. fortuitum infections are highly resistant to multiple antibiotics, with treatment therefore being a challenge [17]. The development of biofilm by *M. fortuitum*is responsible for its resistance to common disinfection procedures and makes it more difficult to get rid of the microbes from the clinical settings. An *erm* gene is a genetic factor that complicates macrolide resistance and narrow margins when using this class of drugs in treatment protocols [18].

4. Transmission and Risk factors

It can easily get inside the body through open traumatic or surgical wounds and through exposure to contaminated water causing local infections. By infections around inserted devices or at injection sites, the bacterium demonstrates its opportunistic nature. Outbreaks are attributable to contact with polluted areas like whirlpools in nail salons and tattoo inks [17].

Prognosis and Treatment

Lung or non-removable implantations agents pose significant challenges, usually resisting total eradication of the infection even with appropriate treatment [17].

Postsurgical debridement and an antibiotic approach targeted to specific cases, which underlines the issue of accuracy in diagnosis and antibiotic sensitivity test [17].

Modes of Transmission

i. **Direct Contact and Inhalation:** Transmission to humans primarily occurs through direct contact with contaminated sources or by inhaling aerosolized particles that contain the bacteria.
ii. **Contaminated Medical Equipment**: Devices such as catheters and respiratory equipment can serve as sources of transmission, especially in healthcare settings.
iii. **EnvironmentalExposure:** Individuals may be exposed to the bacterium through contaminated water used in medical procedures or personal care, such as in whirlpools at nail salons or through tattoo ink [17].

Uncommon and Indirect Transmission Routes

Person-to-person transmission of *M. fortuitum*is considered rare, highlighting its spread dominantly environmental transmission pathway [17].

Medical tourists are particularly at risk, as they may develop chronic NTM infections following procedures involving surgical sites, breast prostheses, and injection sites, often due to exposure to contaminated medical tools or environments [17].

Contamination Sources

Tap water is a frequent vector for *M.fortuitum*, as the bacteria are resistant to standard disinfecting procedures, allowing them to persist even in treated water supplies [18].

5. Clinical Manifestations of *M. fortuitum*Infections

Overview of Infection Types and Symptoms

*M. fortuitum*is known for causing a wide spectrum of clinical manifestations, affecting both immune competent and immune-compromised individuals. The diversity in

symptoms and the severity of infections are influenced by the site of the infection and the patient's immune status.

A. Skin and Soft Tissue Infections:

i. Patients with breast implant-related infections typically present with breast pain or tenderness, with symptoms manifesting after an average incubation period of 9 months [19].

ii. Tattoo-associated infections often lead to a nonpruritic papular eruption within 1-2 weeks post-procedure.

iii. Cutaneous disease may progress to nonhealing ulcers or more severe forms, including ulcerative skin lesions and subcutaneous nodules, sometimes leading to draining fistulas [16][19].

B. **Pulmonary Infections:**

i. Common symptoms include sputum production (68.6%), hemoptysis (51.4%), and cough (45.7%). Patients may also experience complications related to gastro-esophageal disease (22.9%) [13].

ii. In severe cases, infections can lead to chronic pulmonary conditions, particularly in individuals with pre-existing structural lung defects or gastroesophageal abnormalities [19].

iii. Specific cases have reported migratory infiltrates and refractory pneumonia in patients with chronic aspiration, notably post-gastrectomy [16].

C. Ocular and Cardiac Infections:

i. Eye infections may manifest as keratitis or corneal ulcers.

ii. Cardiac involvement can lead to endocarditis, often identified by a valvular murmur.

D. **Other Infections:**

i. Abdominal infections might present with diffuse tenderness indicative of peritonitis.[20]

ii. Disseminated disease can affect various organs, leading to severe health complications like meningitis and osteomyelitis.

5.2 Diagnostic Challenges and Treatment Outcomes

i. The American Thoracic Society presents treatment guidelines for *M. fortuitum*-PD; however, information on the best antibiotic protocols and prognosis for *M. fortuitum*-PD patients are yet limited [4]. A review of *M. fortuitum*-PD therapeutic outcomes showed a microbiological cure rate of 81%[4].

ii. Much of the mortality is attributed to localized infections although the extensive pulmonary and disseminated disease cases, especially in patients whose immune system is compromised, may be very high.

iii. In most cases of infections, the prognosis is usually good if an appropriate antibiotic therapy and surgery are applied. However, some kinds of chronic infections, especially at some sites, can be extremely hard to eradicate [16].

iv. The fact that these diseases can present with varying symptoms highlights the need for strong vigilance in diagnostic testing and customized treatment approaches to improve management of *M. fortuitum*infections.

6. Emerging Therapies and Research

i. NITD-916:

A Prospective Inhibitor NITD-916, aimed at the enoyl-ACP reductase InhA in *M. fortuitum*, has shown highly attractive features both in the laboratory and in a zebrafish model of infection. The compound possess low minimal inhibitory concentration (MIC) values against clinical strains of *M. fortuitum*and it showed an appreciable

antimicrobial effect in macrophages [21]. Furthermore, NITD-916 treatment led to the significant suppression of mortality, bacterial burden and abscesses in the zebrafish larvae infected with *M. fortuitum*. While resistance from particular mutation (G96S in InhAMFO) has also been recorded, thus the necessity to do more work to ascertain its clinical application [22].

Progress in Nanotechnology

The route to addressing antibiotic resistance of *M. fortuitum*by the use of carboxyl-functionalized multi-walled carbon nanotubes (MWCNT-COOH) has been disclosed by recent research. According to the research the MWCNT-COOH has capability to penetrate bacterial cell walls and increase the ability of common antibiotics like kanamycin and streptomycin. Combination of these antibiotics at a concentration of 28 µg/mL can be useful when applied simultaneously with5µg/mL of MWCNT-COOH has completely killed the bacteria. This implies a combinative synergistic effect which can be exploited to inhibit drug resistance in clinical settings [23].

Clinical Outcomes Due to Current Antibiotics

The efficacy of various antibiotics including *M. fortuitum*pulmonary disease was studied in this study which had 35 patients. Prior to the treatment, the isolates of all typeswere sensitive to amikacin, and almost all of them demonstratedsusceptibility to imipenem and moxifloxacin. Even after that, more than half of the patients did not worsen without antibiotic therapy. In the present study the patients requiring antibiotic treatment showed a good recovery with the right antibiotics and it suggests that the existing drugs may be suitable treatment for this infection [17].

Potential of Bacteriophage Therapy

Emerging research into bacteriophage therapy presents a novel approach to tackle infections caused by rapidly growing mycobacteria (RGM). This method exploits bacteriophages, viruses that infect and lyse bacteria, offering a targeted mechanism to combat bacterial infections resistant to conventional antibiotics [18].

Gepotidacin: A Novel Antibiotic Candidate

The development of gepotidacin, a new topoisomerase inhibitor, represents a significant breakthrough in antibiotic therapy. Its unique mechanism of action allows it to bypass common bacterial resistance pathways, offering a promising treatment alternative for infections caused by *M. fortuitum*[24].

7. CurrentTreatmentStrategies

Antibiotic Regimens and Susceptibility

*M. fortuitum*infections often require a combination of oral and intravenous antibiotics due to the bacterium's resistance profile. Commonly used antibiotics include macrolides, quinolones, tetracyclines, sulfonamides, and carbapenems [20]. For severe infections, a regimen of IV treatment combined with oral antibiotics is necessary, typically extending for several months to ensure efficacy [23]. In vitro susceptibility testing is crucial to select appropriate antibiotics, as most clinical isolates show sensitivity to amikacin, imipenem, or moxifloxacin [14][21].

The susceptibility of *M. fortuitum*to various antibiotics is crucial for effective treatment. Generally, isolates show favorable responses to clarithromycin, amikacin, and imipenem [24]. Specifically, *M. fortuitum*strains demonstrate varying degrees of sensitivity to several key antibiotics:

Amikacin: Intermediate to highly sensitive

Ciprofloxacin: Highly susceptible

Doxycycline: Intermediate susceptible

Clofazimine, TMP-SMX (Trimethoprim-Sulfamethoxazole), Linezolid: Susceptible However, resistance patterns also exist, particularly against all

antituberculosis agents, varying responses to macrolides, and resistance to imipenem [19]. A study of 86 isolates revealed resistance to clarithromycin and tobramycin but susceptibility to tetracyclines and quinolones [18].

Multi drug Therapy and Treatment Duration

i. **Initial Intensive Treatment:** An intensive treatment period often starts with an injection of amikacin or imipenem, recommended for several weeks to manage severe infections effectively [24].
ii. **Long-term Antibiotic Use:** Following the intensive phase, patients may need to continue with dual antibiotic therapy, including drugs like ciprofloxacin, ofloxacin, and doxycycline, to prevent the development of resistance [18].
iii. **Duration of Therapy**: Pulmonary infections require treatment until sputum results are negative for 12 months, often starting with a combination of amikacin and cefoxitin, followed by trimethoprim-sulfamethoxazole plus doxycycline or levofloxacin for 6-12 months [21].

Surgical Interventions and Specialist Consultations

Surgical excision may be necessary for pulmonary lesions if the response to antibiotic therapy is inadequate or if the organism shows significant resistance. This approach is also considered for severe cutaneous, ocular, or bone lesions. Consultations with specialists such as infectious disease experts, pulmonologists, and surgeons are recommended to guide both diagnosis and treatment [21].

Monitoring and Management

Long-term management of *M. fortuitum*infections includes regular follow-up care to monitor for adverse effects, periodic assessments of renal function and hearing, and monthly sputum cultures for patients with pulmonary disease. It is crucial for patients to adhere strictly to the prescribed antibiotic regimen to avoid the dismal outcomes associated with mono-therapy [18] [21]. Patients undergoing treatment need to be

reassured about the non-contagious nature of the infection to differentiate it from diseases like tuberculosis [24].

Special Considerations for Treatment Administration

Patients requiring long-term intravenous antibiotic therapy who cannot receive home-based care may need placement in an extended-care facility. This ensures proper administration of the treatment and helps manage the infection effectively [25].

2.2 Machine Learning

Machine Learning and Deep Learning: A Transformative Landscape

Machine learning (ML) and deep learning (DL) have revolutionized numerous fields, from computer vision and natural language processing to healthcare and finance. This review article delves into the core concepts of both, exploring their capabilities, applications, and the ever-evolving landscape they create.

Machine Learning: Foundations of Intelligent Systems

ML encompasses algorithms that empower computers to learn from data without explicit programming. This learning process allows them to identify patterns, make predictions, and improve their performance over time. There are three main learning paradigms within ML:

i. **Supervised Learning:** Involves training a model on labeled data, where each data point has a corresponding output value. The model learns the relationship between inputs and outputs, enabling it to predict future outputs for unseen data [26]. Examples include linear regression for continuous predictions and support vector machines for classification tasks.

ii. **Unsupervised Learning:** Deals with unlabeled data, where the model seeks to uncover hidden structures within the data itself. Common techniques include k-

means clustering, which groups data points based on similarities, and dimensionality reduction methods, which reduce the complexity of high-dimensional data [26].

iii. **Reinforcement Learning:** Involves an agent interacting with an environment, receiving rewards for desired actions. The agent learns through trial and error to maximize its reward, making it suitable for complex decision-making problems [27].

Applications of Machine Learning

Machine learning's versatility allows it to be applied across a spectrum of fields, driving decision-making and innovation. In healthcare, it supports earlier detection of diseases and personalized patient care. The financial industry benefits from machine learning in fraud detection and algorithmic trading, while in agriculture, it aids in predicting crop yields and monitoring soil health. Social media platforms use machine learning for targeted advertising and content recommendation, enhancing user engagement Additionally, in the realm of natural language processing, it enables machines to understand and generate human language, facilitating real-time translation and voice-activated assistants. Machine learning's ability to process and analyze vast datasets is transforming industries by making operations more efficient, enhancing customer experiences, and opening new avenues for research and development [27][28][29].

Machine Learning Architecture

ML algorithms operate on a more fundamental level, focusing on extracting meaningful patterns from data. Here's a general workflow inspired by [26]:

i. **Data Preprocessing:** Raw data is cleaned, transformed, and formatted to be suitable for the chosen learning algorithm [26].

ii. **Model Selection:** The appropriate ML algorithm is selected based on the task (classification, regression, clustering, etc.) [26].

iii. **Model Training:** The model learns from the training data by adjusting its internal parameters to minimize errors in predicting outputs for labeled data points [26].

iv. **Model Evaluation:** The trained model is tested on unseen data (testing data) to assess its performance and generalization capabilities [26].

v. **Model Tuning (Optional):** Hyperparameters of the model (e.g., learning rate, number of trees in a random forest) are adjusted to further improve performance [26].

Common ML Architectures:

i. **Linear Regression:** Uses a linear equation to model the relationship between a dependent and one or more independent variables. You can find a detailed explanation in textbooks like [27].

ii. **Support Vector Machines (SVMs):** Finds a hyperplane in high-dimensional space that best separates data points belonging to different classes. For a deeper understanding, refer to.

iii. **Decision Trees:** Tree-like structures where each internal node represents a test on a feature, and each leaf node holds a class label or a prediction value.

iv. **k-Means Clustering:** Groups data points into a predefined number of clusters based on their similarity. 'MacQueen' introduced this clustering algorithm.[28]

Deep Learning: Unveiling the Power of Neural Networks

DL is a subfield of ML inspired by the structure and function of the human brain. It utilizes artificial neural networks (ANNs) with multiple layers of interconnected nodes, mimicking the biological neural network. These layers extract progressively complex features from the data, allowing DL models to learn intricate patterns and achieve superior performance on various tasks.

Popular DL architectures include:

i. **Convolutional Neural Networks (CNNs):** Excel at tasks involving spatial data, such as image recognition and video analysis, by automatically learning spatial features from the data [29].
ii. **Recurrent Neural Networks (RNNs):** Designed for sequential data, like text and speech, by incorporating a memory mechanism that allows them to process information across time steps [30].

A World Transformed by Machine Learning and Deep Learning

The impact of ML and DL is pervasive, permeating diverse applications:

i. **Computer Vision:** Object detection, image classification, facial recognition, and self-driving cars.
ii. **Natural Language Processing:** Machine translation, sentiment analysis, text summarization, and chatbots.
iii. **Healthcare:** Medical diagnosis, drug discovery, personalized medicine, and analysis of medical images.
iv. **Finance:** Fraud detection, credit risk assessment, algorithmic trading, and market forecasting.

Challenges and Future Directions

Despite their remarkable success, ML and DL face challenges:

i. **Explainability:** Understanding the internal logic behind complex models, crucial for ensuring trust and ethical considerations.
ii. **Data Bias:** Models trained on biased data can perpetuate those biases, necessitating responsible data collection and curation practices.
iii. **Computational Cost:** Training large DL models can require significant computational resources, prompting research into efficient training algorithms and hardware solutions.

The future of ML and DL is brimming with innovation:

i. **Explainable AI (XAI):** Development of techniques to provide transparency and interpretability into model decisions.
ii. **Lifelong Learning:** Continuously learning models that adapt to new data and changing environments.
iii. **Federated Learning:** Collaborative learning without compromising data privacy, enabling data-rich training across decentralized devices.[331],[32]

Deep Learning Architectures

Deep learning (DL) is a subfield of machine learning (ML) that utilizes artificial neural networks (ANNs) with multiple interconnected layers to achieve superior performance on various tasks. These architectures are inspired by the structure and function of the human brain, allowing them to learn complex representations from data. This article explores the core principles of deep learning architectures and some popular architectures used in various applications.

Core Principles:

i. **Artificial Neural Networks (ANNs):** Deep learning models are built upon ANNs, which are networks of interconnected artificial neurons. Each neuron receives input from other neurons, performs a weighted sum and applies an activation function to generate its output. The connections between neurons are represented by weights, which are adjusted during the learning process to improve the network's performance.
ii. **Multi-Layered Architecture:** Deep learning architectures consist of multiple layers of neurons, stacked on top of each other. Each layer learns a specific level of abstraction from the input data. The first layers typically extract low-

level features, while higher layers learn more complex and abstract representations.

iii. **Learning Process:** Deep learning models learn through a process called backpropagation. The network is trained on a dataset with labeled examples. During training, the model's predictions are compared to the actual labels. Errors are propagated backward through the network, allowing adjustments to the weights and biases of each neuron to minimize the errors in future predictions.[33]

Popular Deep Learning Architectures:

i. **Convolutional Neural Networks (CNNs):** Specialized for processing grid-like data, particularly images, by efficiently extracting spatial features. CNNs utilize convolutional layers with learnable filters that convolve with the input data to identify patterns and features. Pooling layers are often used to downsample the data and reduce computational complexity.

ii. **Recurrent Neural Networks (RNNs):** Designed to handle sequential data like text and speech. RNNs incorporate internal loops that allow them to process information across time steps and capture temporal dependencies within the data. However, they can suffer from the vanishing gradient problem, making it difficult to learn long-term dependencies [34][35][36].

iii. **Long Short-Term Memory (LSTM):** A special type of RNN architecture that overcomes the vanishing gradient problem. LSTMs incorporate memory cells that can store information for longer durations, enabling them to learn long-term dependencies in sequential data [37].

Chapter 3

Materials and Methodology

3.1 Bacterial Strains

List of all the bacterial strains used in this study are listed in following table.

Table 1: List ofmicrobesusedinstudy.

***E. coli*DHα**	Institute of Microbial Technology (IMTECH), Chandigarh, India
M. fortuitumATCC 6841	CentralDrugResearchInstitute(CDRI),Lucknow,India

3.2 Media and other Chemicals

Table2:List of chemical

CrystalViolet	**LobaChemie**
GramsIodine	**LobaChemie**
Safranin	**LobaChemie**
Basicfuschin	**Merck**
MethyleneBlue	**Fisherscientific**

Table 3: List of media prepared

Luria Broth with Glycerol and Tween 80 (LBGT) media	**Culturing**
Nutrient Agar with Tween 80 (NAT)	**Isolation**

Table 4: ListofInstruments

4ºC storage	**BLUE-STAR**
Incubatorshaker	**Labnet**
Google-Colab	**Google**
PyLab	**Open source**

Google-Colab is a free cloud platform from Google that lets you write and run Python code in your web browser. This eliminates the need to install software and provides access to powerful computing resources for data analysis and machine learning tasks.

PyLabis a commonopen sourceplatform for plotting and numerical computing in Python. It combines functions from Matplotlib and NumPy into a single namespace, which can clutter your code and lead to naming conflicts**.**

3.4 Methods

Different streaking methods:

1. Simple streaking

i. The date and the microorganism's name were written on the petri-plate.
ii. To sterilize the inoculation loop, it was heated to till red hot using a Bunsen burner.
iii. To cool the loop, it was placed on a sterile nutritional agar plate's corner.
iv. Using the ring and little fingers, the cotton plug was taken out of the bacterial culture while one hand was still holding the loop.
v. The flask was placed in front of the burner for five to ten seconds in order to heat sterilize its mouth.
vi. A colony was picked up by the loop and it was introduced into the broth culture.
vii. Next, the test tube's mouth was once more sealed with a cotton plug.
viii. The loop was gently pulled in a zig-zag fashion on the nutritional agar plate.
ix. The loop was heated till red hot once again for sterilisation.
x. The plates were incubated at 37°C for in the incubator.

2. Quadrant Streaking

i. After being held in a blue flame and allowed to cool, the loop was sterilized.
ii. The agar surface was streaked with a loop containing a bacterial culture.
iii. The loop was dragged back and forth across the agar surface.

iv. Another sterilization of the loop was performed.

v. After turning the plate 90 degrees, new streaks were created starting at the end of the preceding streak.

vi. Identical procedure was carried out two more times.

vii. The plateswere incubated for 24 hours at 37°C.

3.5 Methodology for using various Machine Learning approaches

Several machine learning techniques, such as SVMs [35], KNNs [36], and ANNs [45][46], were used in this work to evaluate the intricate and dynamic proteomic data of *M. fortuitum*. These methods have shown to be successful in analysing biological sequencing data and obtaining insightful information. The suggested technique used in the article is displayed in Fig. 1**.**

The fundamental procedures used in this investigation are described in Fig.1. The global proteome study of *M. fortuitum*in both planktonic and biofilm phases yields the proteomic data. This data set is further subdivided into abundant proteins and overexpressed and underexpressed proteins during biofilm formation. (Abundant proteins are those that were present in all three triplets that were linked to biofilm.) These overexpressed and underexpressed proteins serve as model training sets, whereas an abundance of proteins serves as the testing data set. To get the proteomic data ready for analysis, the authors used strict feature engineering and data processing techniques. Then, by examining the protein profiles of each duplicate, the distinctions between the planktonic and biofilm phases were ascertained. Upon eliminating potentially irrelevant proteins from the initial set of 11,123 proteins identified through proteomic analysis, researchers were left with 2,333 "master proteins" for further investigation. In this work, proteins associated with the production of *M. fortuitum*biofilms were identified using proteomic analysis. The data acquired will be helpful in developing strategies to disrupt or prevent biofilms, which might result in more effective treatment options for illnesses caused by these bacteria. 283 proteins were utilized to test the concept as they were shared by all three biofilm triplets. Using the distinct training and test sets [46][47] for various machine-learning approaches, the models were tuned and trained to

discover the underlying patterns and connections within the data.Machine learning is an essential component of the rapidly expanding discipline of data science.

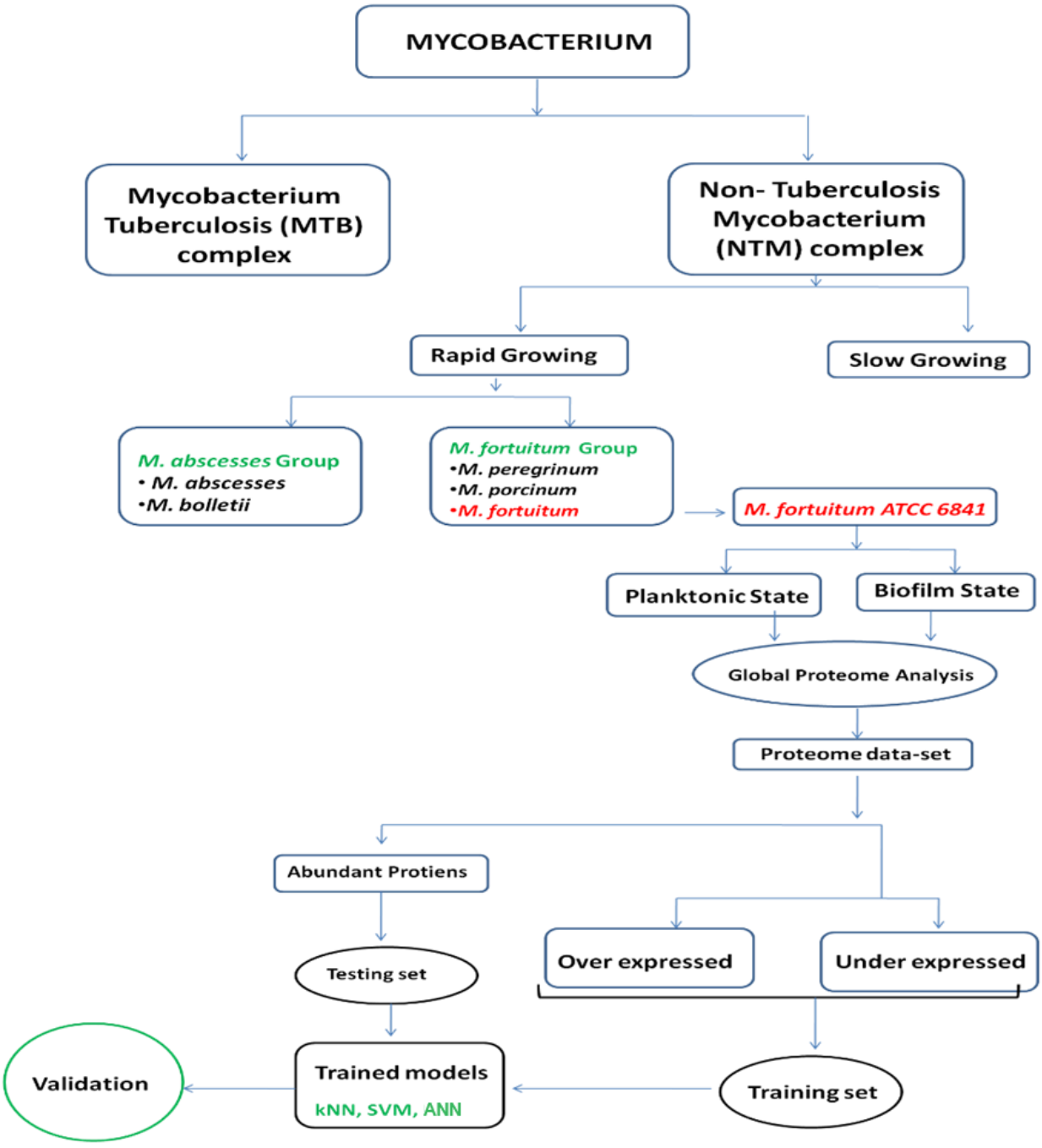


Fig-1. Methodology for using various ML approaches

The model optimisation, error function, and decision procedure are its three main processes. Machine learning algorithms are often used to provide a prediction or categorization. To estimate a pattern in the data, the algorithm makes use of certain input data that may or may not be labelled. Machine learning incorporates imitation of human behaviours and functions accordingly, all without the need for human intervention. Among the most crucial elements of machine learning are reinforcement learning, unsupervised learning, and supervised learning, among others [40, 41]. Machine learning algorithms come in several varieties. In this research, the prediction model was designed using ANN, SVM, andkNN [48].

A supervised algorithm for classification is called kNN. The approach is instance-based, non-parametric, and saves all cases that are accessible. New instances are categorized based on the majority vote of their k closest neighbours. Simple to construct and frequently used for classification tasks, kNN's performance can be influenced by the number of neighbours employed in the prediction as well as the choice of distance measure.AlthoughkNN is quick and easy to use, the number of neighbours utilized in the prediction and the distance measure selected might affect how accurate the model is. SVM is a linear classifier that determines the ideal feature space border between the classes. It performs well on short datasets and high-dimensional datasets with a distinct margin of separation and is frequently applied to binary classification issues. Compared to artificial neural networks, SVM may be quicker and simpler to train, but it might not be as effective on challenging issues. ANNs are a powerful technique for machine learning that draws inspiration from biological brains. Artificial neurons, which connect together, processing units that communicate similarly to brain cells, make up these networks. By altering the connections between these neurons in response to training data, ANNs may learn to identify patterns, forecast results, and perform complicated tasks without the need for explicit programming. For problems like financial forecasting, speech recognition, and photo identification, this makes them ideal. Moreover, parallel processing facilitates ANNs by allowing them to efficiently handle large datasets. They also have some fault tolerance because of their dispersed architecture, thus the loss of a single neuron won't have a major impact on system performance. The original features are converted using the principal component analysis (PCA) approach into a new set of linearly uncorrelated features.

Depending on how much variety in the data they collect, these components are arranged in order. PCA generates new features by combining the original features, as opposed to choosing a subset of the original features. Various parameters may be derived from true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) to analyse the classification performance of various classifiers. Eqs. (1), (2), and (3), respectively, reflect the parameters accuracy (ACC), sensitivity (SEN), and specificity (SP) [40].

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$SEN = \frac{TP}{TP+FN} \quad (2)$$

$$SEP = \frac{TN}{TN+FP} \quad (3)$$

3.6 Proposed methodology for using NN as a prediction model

Algorithms in machine learning are sets of algorithms that can learn from data and improve with each new set of data. These are used to tackle issues when more conventional coding methods and procedures prove to be ineffective or inefficient. The foundation of machine learning algorithms is the notion that machines can learn from data and improve with further data. This implies that machine learning algorithms have the ability to automatically analyse data and modify their behavior [38, 39]. The most popular machine learning algorithms are reinforcement learning, supervised learning, and unsupervised learning. Figure-2 illustrates the proposed method that was used in this study.

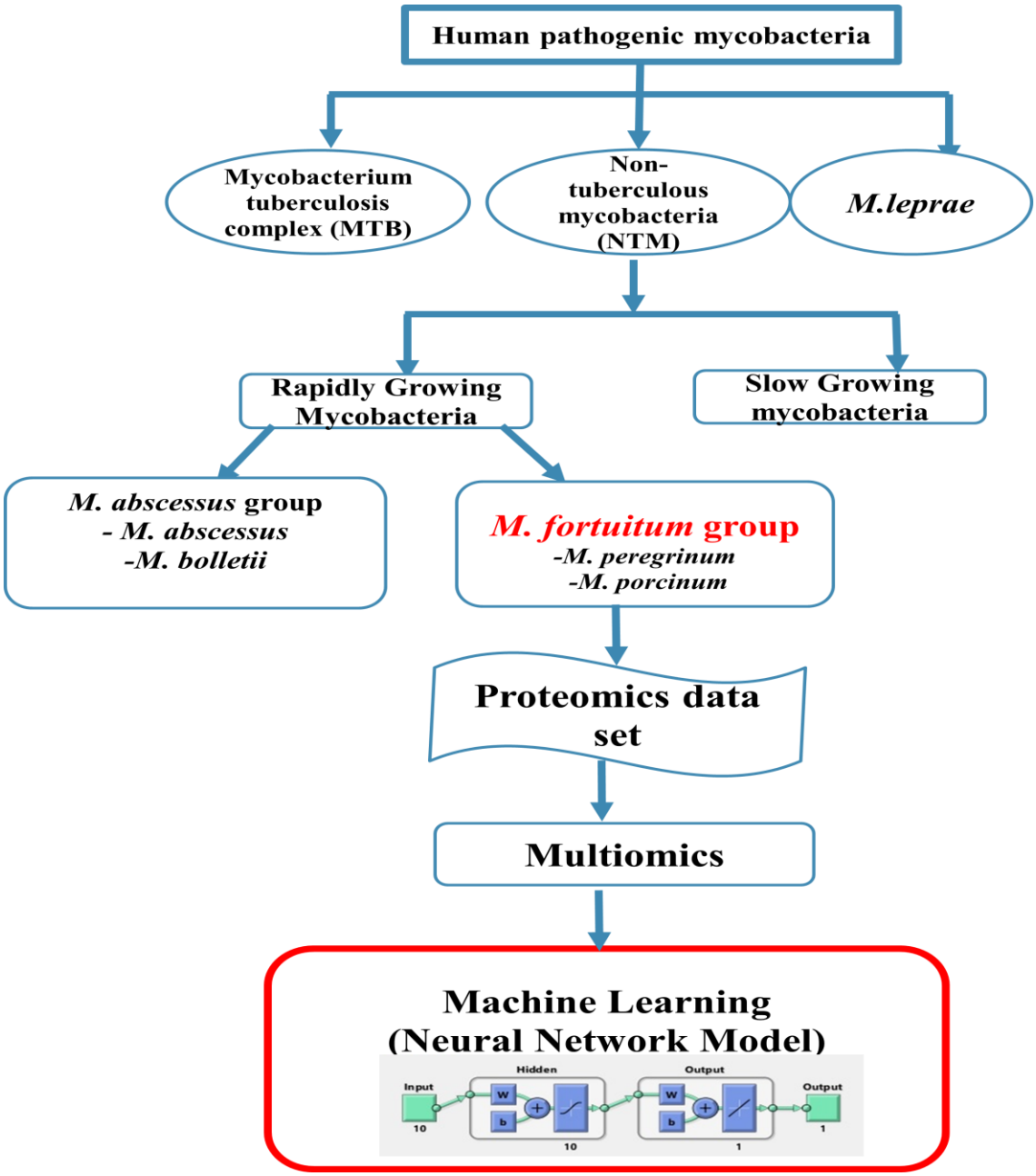


Fig. 2 – Proposed methodology of using NN as a prediction model

Human pathogenic mycobacteria are classified into three types: *M. tuberculosis*complex, NTMand*M. leprae*,. Two further subgroups of NTM are identified: mycobacteria with rapid and slow growth rates.Next, two families of the quickly proliferating mycobacteria are separated: *M. abscessus* and *M. fortuitum*. *M. fortuitum*serves as this study primary focus. Neural network (NN) analysis is used to a proteomic dataset of *M. fortuitum*as a machine

learning technique. Triplet data (P1, P2, and P3) from up- and down-regulated proteins were used to train the model. NN are models for parallel processing, as opposed to the single processor that contemporary computers employ to gather and show data. Faster processing and solution computation are also made possible by the parallel computing technique.Neural networks do not employ a set procedure to produce a certain outcome; instead, they operate inside a dynamic computational framework. These networks are based on real neurons and the architecture of the brain, with each neuron having many distinct inputs and a single output. Coefficients of determination and mean square error (MSE) values were evaluated for each of the three sets.

3.7Proposed methodology for using Support Vector Machine(SVM) as a prediction model

Machine learning (ML) is a crucial part of artificial intelligence. Machine Learning (ML) is the capacity of a computer or system to acquire new knowledge without requiring a particular programme [40][41]. An ML algorithm is essentially a procedure, or set of procedures, that helps a model adjust to data in order to achieve a goal. An ML approach usually explains how the data is transformed from input to output and how the model decides the appropriate mapping from input to output. The learning algorithm changes the parameters while the model specifies the mapping function and keeps the parameters in an effort to help the model achieve its objective [42]. The three main types of machine learning algorithms are supervised, unsupervised, and reinforcement learning. In unsupervised machine learning, where clustering techniques such as KMEAN, Hierarchical, DBSCAN, etc. are used, the data set is unlabeled (no feature or class/category labels to be predicted). The data collection is divided by these methods into clusters or similar categories. In supervised machine learning, a labelled data set with a reliably labelled feature is employed. The two categories of algorithms are regression and classification. To forecast non-integer numerical characteristics, regression techniques are applied. Predicting a label or class (e.g., cancer, non malignant in medical tests, etc.) is done through classification. The main methods for classification include Bayesian algorithms, Artificial Neural Networks (ANN), Logistic Regression, kNN, SVM, Decision Trees, etc. [43][44]. **Fig. 3** illustrates the SVM algorithm for up- and down-regulated protein prediction.

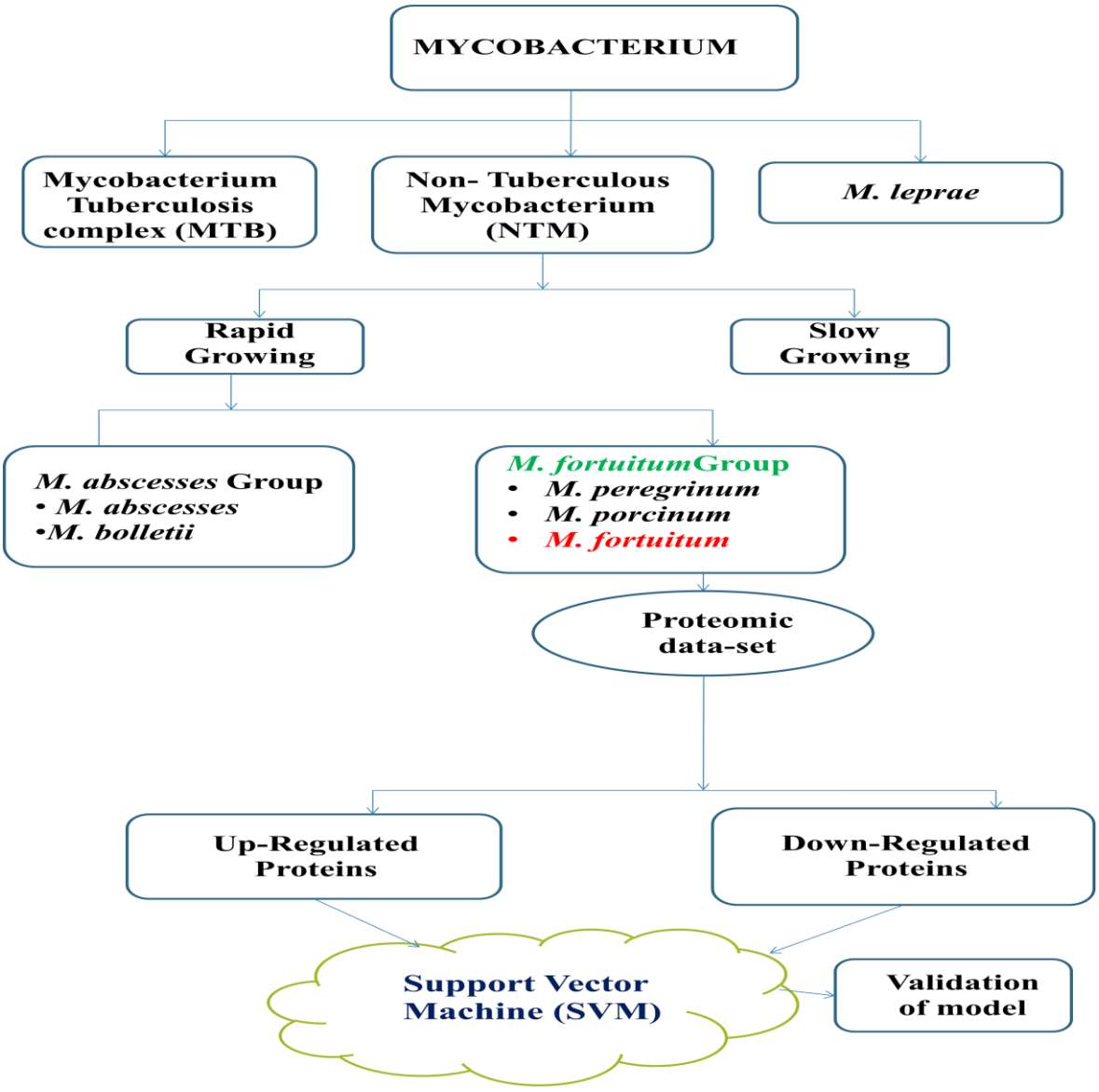


Fig. 3 – Proposed methodology of using SVM as a prediction model

The proteomic data set of *M. fortuitum*isanalysed using SVM as a classification tool. Triplet data (P1, P2, and P3) for up- and down-regulated proteins are used to train the model.Later, kNNwas used to verify the model. 2181 samples total from a 70-30 sample ratio made up the dataset. Software for the simulations is MATLAB 2019a.

The SVM approach looks for the best decision boundary or line to split n-dimensional space into classes. Hyperplanes are the best choice boundaries that exist. To help create the hyperplane, SVM chooses the extreme vectors and points. These extreme circumstances are referred recognised as support vectors, which is why the technique is called an SVM. To partition a set of points of two kinds into two groups, SVM constructs a (N-1) dimensional hyperplane in an N-dimensional space. SVM classifies data sets by utilising the concept of hyper-planes. A two-dimensional line with two categories or characteristics is called a hyper-plane, and it splits a dataset into two classes. SVM performs badly when there is greater noise in the data set, or when the target classes overlap. If there are more characteristics per data point than training data samples, the SVM will perform worse. The selection of the kernel or similarity function (Linear, Polynomial, Logistic/Sigmoid, and Gaussian/RBF kernels), the choice of cost parameter (C), and the choice of Gamma (G) (if the Gaussian kernel was used) were some of the parameters that were utilised to optimise the SVM algorithms.**Table5**displays the impact of gamma and C values, both big and small, on variance and bias.

Table 5: Effect of gamma and C value

	Large G	Small G	Large C	Small C
Bias	High	Low	Low	High
Variance	Low	High	High	Low

An SVM model's performance is determined by two hyper-parameters: gamma and cost. Bias and Variance in any machine learning model should be properly balanced. Our sole task is to optimise the C value for the linear kernel. Nonetheless, if the RBF kernel is to be utilised, simultaneous optimisation of the gamma and C parameters is necessary. The influence of C is negligible at large gamma values. Should gamma be weak, then C affects the model in the same manner that it does the linear model.

Chapter 4

Results and Discussion

4.1 Results of Quadrant streaking.

Followingare the results of quadrant streaking done for obtaining isolated colonies of *E. coli DHα*

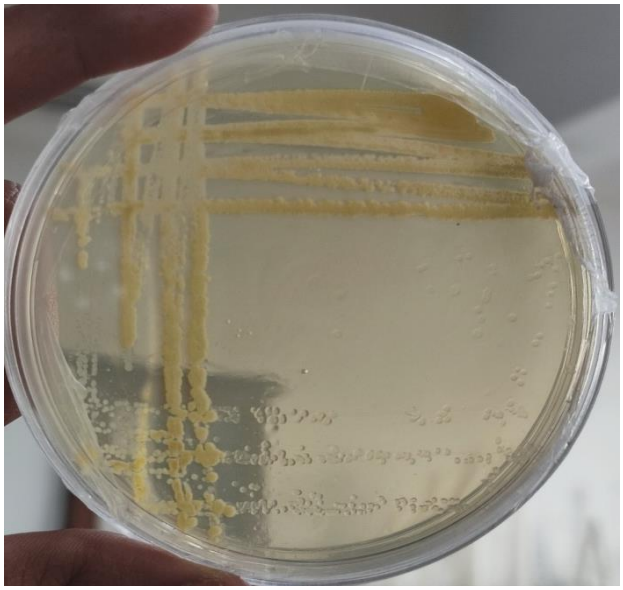


Fig. 4 Quadrant streaking of *E. coli DHα*

4.2 Results of Ziehl-NeelsenStaining

Following image shows stained *M. fortuitum*ATCC 6841 under microscope (60X)

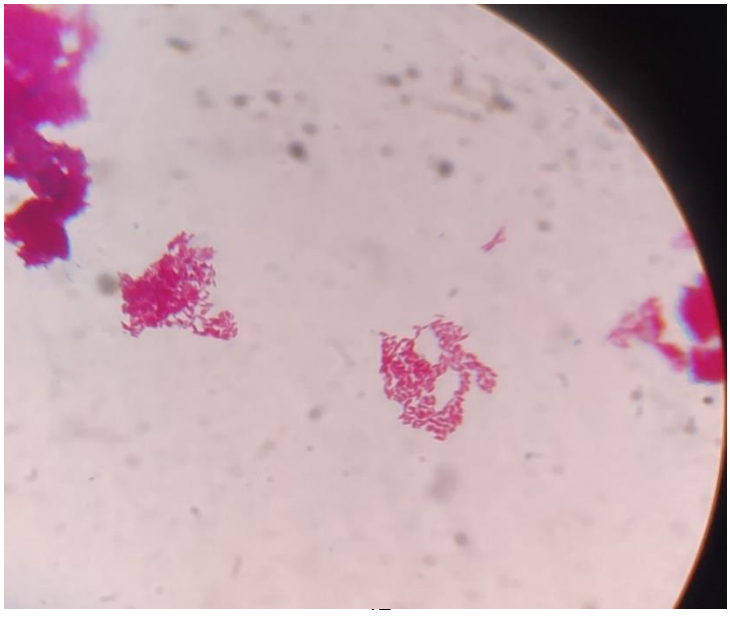
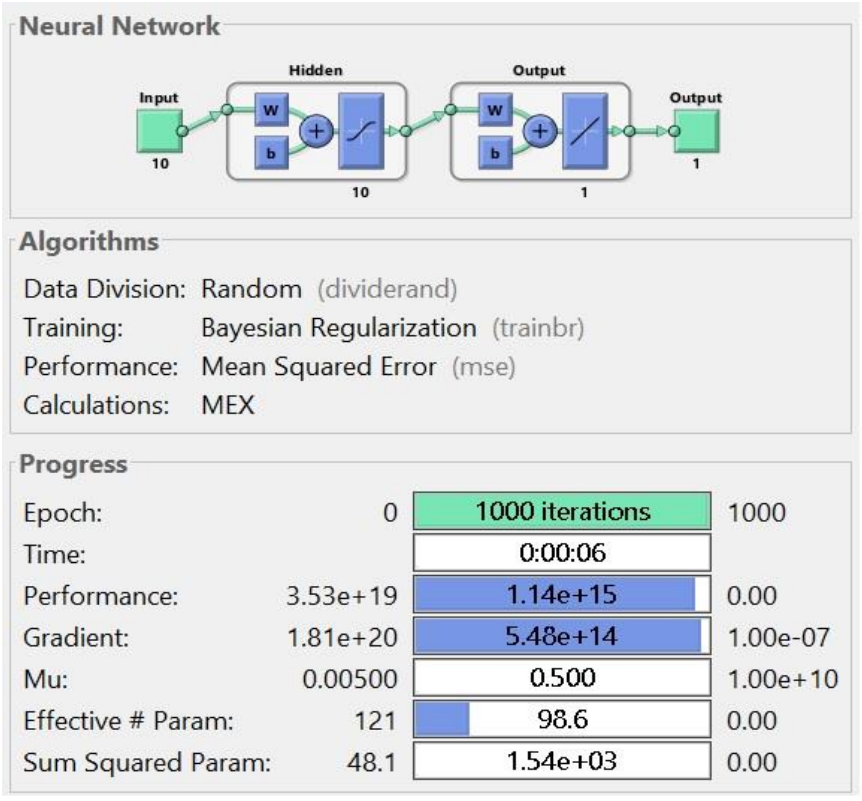


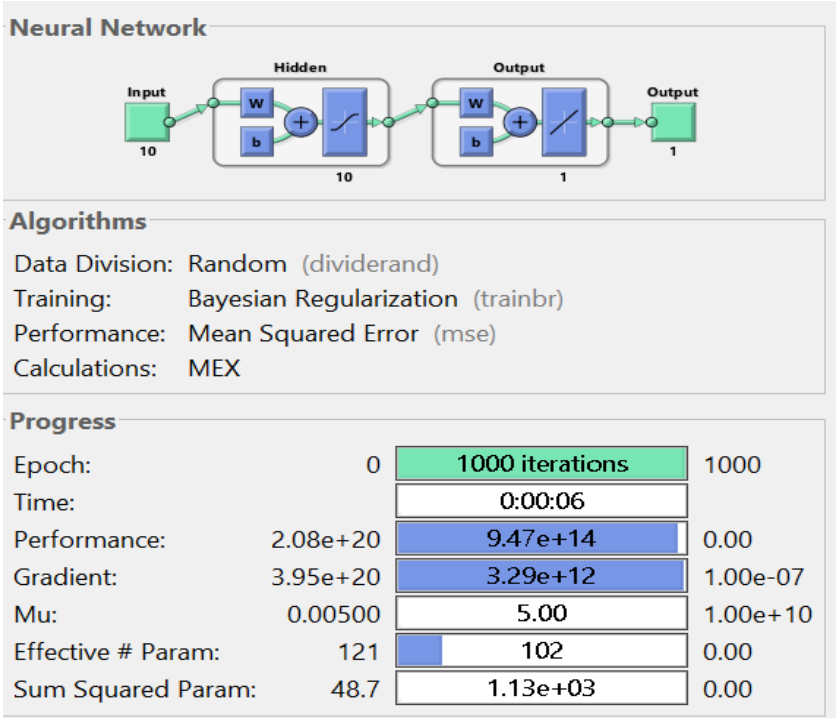
Fig. 5 Ziehl-Neelsen Staining

4.3Results for Neural Network (NN)

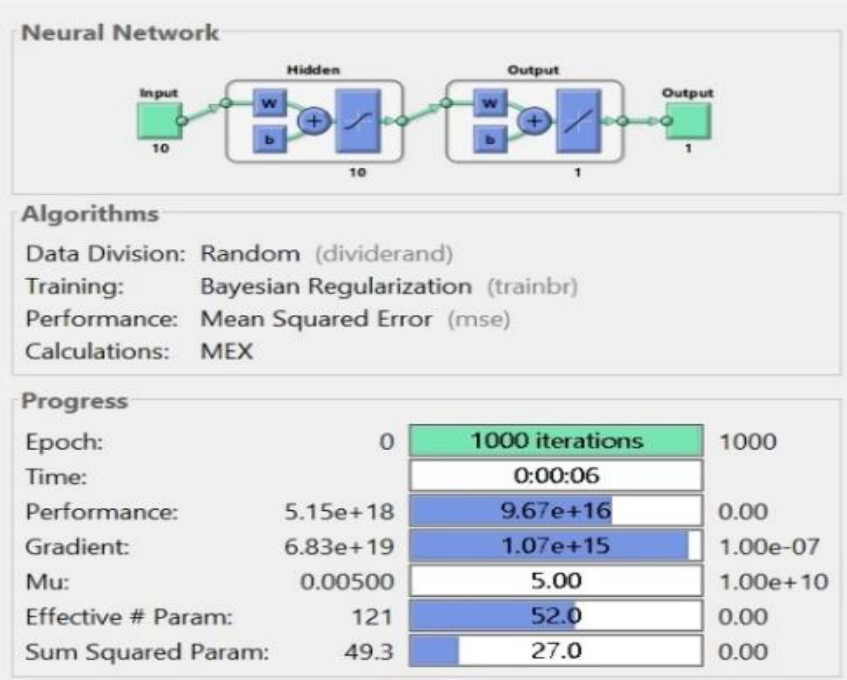
Machine learning is an essential part of artificial intelligence. computer learning (ML) is the ability of a system or computer to learn new things without the assistance of a particular programme [42, 44]. A machine learning approach typically explains how the data is converted from input to output and how the model finds the optimal mapping between input and output. The authors of this work employed NN approaches for the analysis of three distinct data sets (P1, P2, and P3) that were particular to planktonic states. Neural networks (NNs) are made up of networked nodes that function as neurons and may mimic learning and decision-making. Neural networks are capable of pattern recognition, experience-based learning, and decision-making based on patterns.



a) For P1



(b) For P2



c) For P3

Fig 6: NN model of *M. fortuitum*planktonic state proteome (a) P1, (b) P2, and (c) P3

To do this, the input and output data-based weights of the links connecting the nodes are modified. The neural network can learn from and adapt to new data by adjusting the weights in a way that minimizes the error between the input and output. In order to predict the biomarkers, the network employed the logistic sigmoid transfer function and variants of the back propagation learning technique known as Bayesian Regularization (BR). 2181 samples were included in the dataset. 70% of the data were used for training, 15% for testing, and 15% for validation over the model's 1000 training cycles. The BR method often takes longer, but for difficult, limited, or noisy datasets, it yields high generalization. Adaptive weight reduction (regularization) states that training ends. The Intel(R) Core (TM) i7-10700 CPU @ 2.90GHz 2.90 GHz with 16 GB RAM is utilised with MATLAB 2019a software for biomarker prediction. Fig. 6 displays the evaluated values of the three parameters for the NN planktonic model: gradient, mu, and MSE. Figure 6 displays the data for P1, P2, and P3 values.

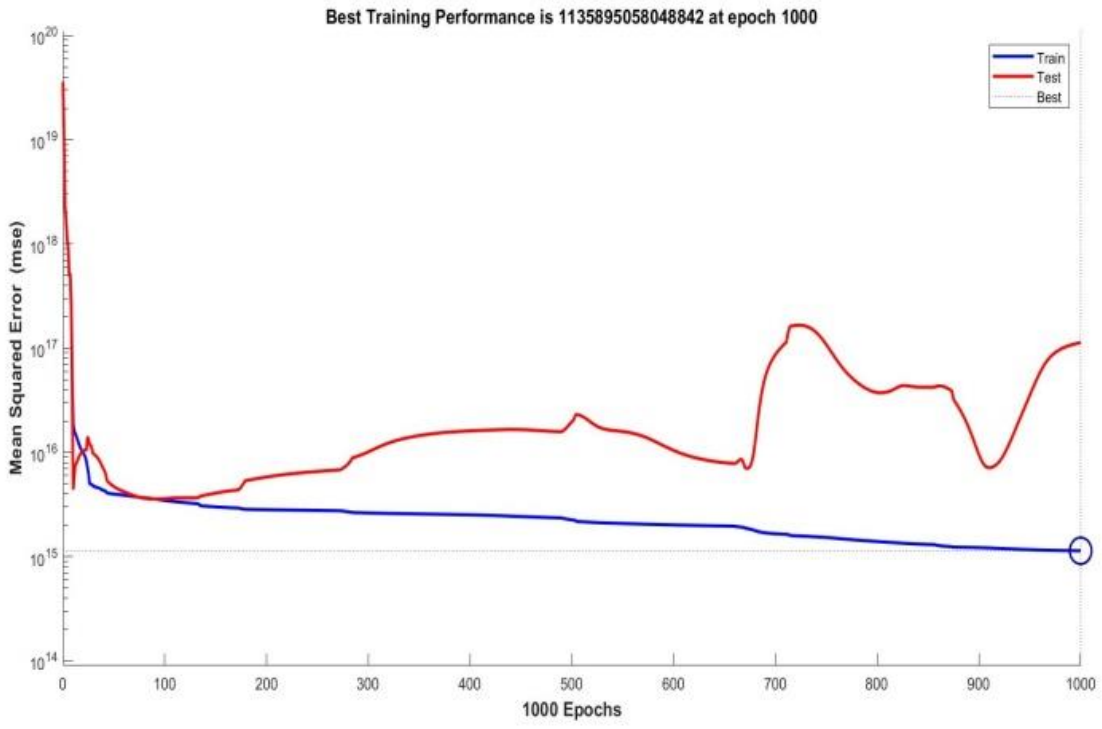
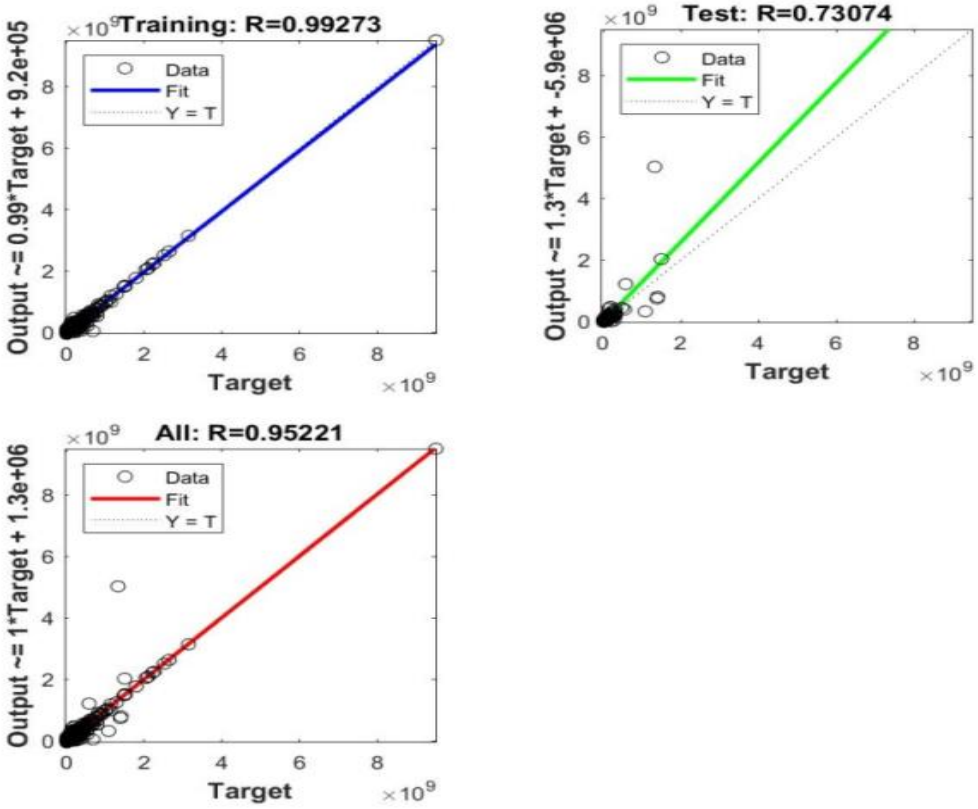


Fig 7: Network Performance Graph

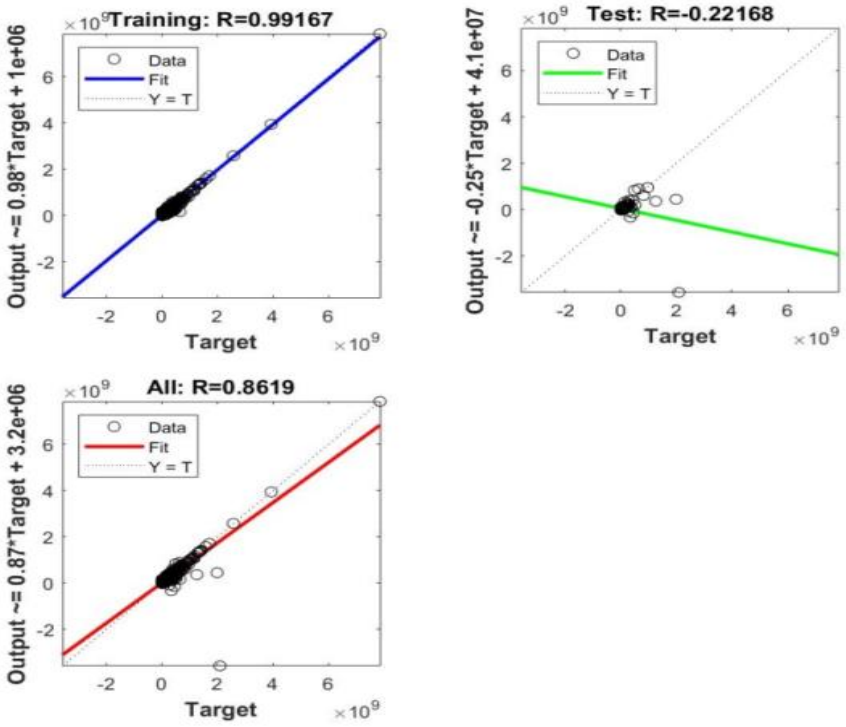
Fig. 7 displays the gradient, mu values, and NN model for our model. The data split process, a commonly used method for evaluating the prediction performance of the machine learning

model, was employed to evaluate the neural network's performance. The performance of the neural network is shown in Fig. 6 as the mean squared error (MSE) v/s iterations on the training and testing sets between the targets and network outputs. It undergoes supervised training since the inputs and results are previously known. During training, mistakes are communicated back across the neurons, changing the weights of the neurons. This training procedure keeps on till it runs out. In order to avoid the network from being overfit, whenever its performance reaches the optimal fit, it will stop training and continue for another 1000 epochs. The neural network's performance parameter's best-fit point was investigated. Epoch 1000 is the point of maximum fit.

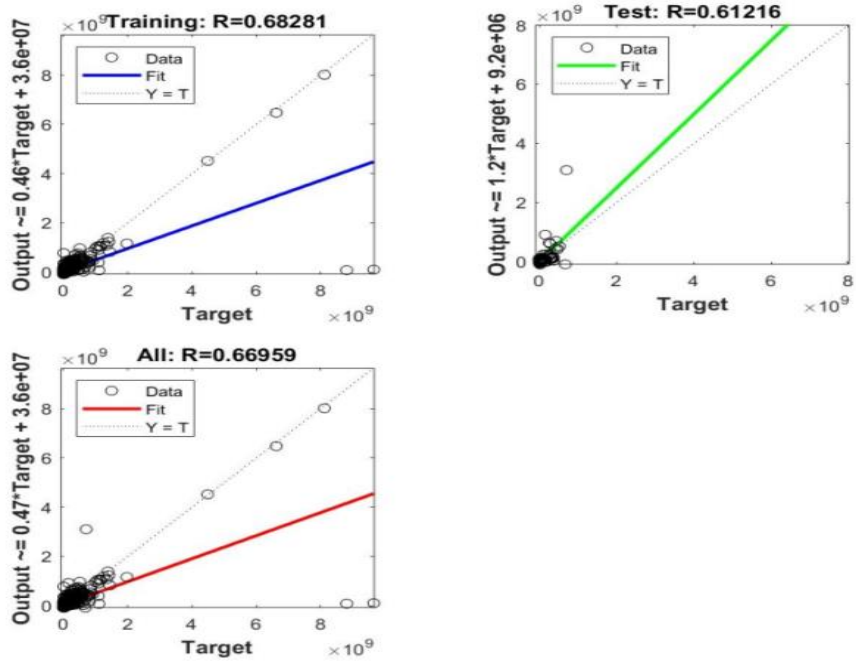
The degree to which the input and output data fit into the network is then ascertained using a regression plot. About 15% of the data sets are used to verify the network, 15% are used to test the network, and only 70% of the data sets are usually used for training throughout training. These divisions will all occur automatically. Figure 8 displays the plots, which are made up of three graphs that stand for training, testing, and merging all of these.



(A) For P1



(B) For P2



(C) For P3

Fig 8: Regression graph produced by a neural network (a) P1, (b) P2, and (c) P3.

Figure 8 shows that all of the data sets are accurately fitted to the line. It demonstrated the accuracy of our neural network's structure. It may also be used to project the outcomes for various sets of input data. The graphs show how the network's outputs relate to the actual objectives for the training and testing sets, hence verifying the network's performance. With an MSE of less than 0.2147e-10 excessively, a coefficient of determination of 0.9916 for training, -0.2216 for testing, and 0.8619 for all the sets taking into account the P1 dataset, as displayed in the charts, the network showed sufficient generalization ability and respectable performance. The coefficient of determination values for the three proteome sets are shown in following Table 6.

Table 6: Coefficient of determination values for the three models

	Data points	P1	P2	P3
Training	1527	0.9916	0.9927	0.6828
Testing and Validation	327 + 327	-0.2216	0.7307	0.612
Overall		0.8619	0.9522	0.6695

For the training dataset, the coefficient of determination was 0.9927; for the testing dataset, it was 0.7307; and for the P2 dataset as a whole, it was 0.9522. Similarly, using the P3 dataset, values of 0.6695 are recorded for the total, 0.612 for testing, and 0.6828 for training. Following Table 7 displays the R-values after accounting for the other two models.

Table 7: R-values

Tested on → / Training set ↓	P1	P2	P3
P1	-	0.846	0.551
P2	0.556	-	0.558
P3	0.884	0.847	-

4.4Results for Support Vector Machine (SVM)

Predictions were made using data on the 70:30 ratio. **Figure 9** displays the accuracy of cross-validation for three distinct planktonic datasets. The results presented in demonstrate that the P1 dataset produced the best accuracy of 58.89%. It shows the accuracy when using a 70:30 ratio for training and testing for three different planktonic datasets as well as the full dataset. The accuracy rates for P1 are 53.61%, P2 are 54.07%, and P3 are 53%.

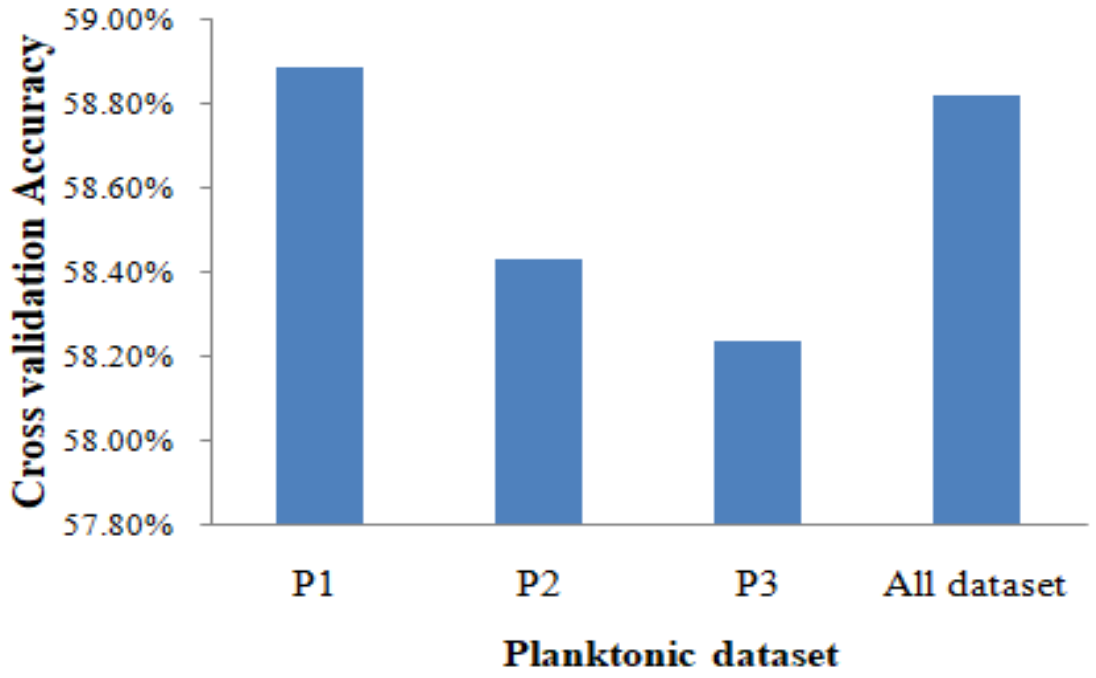


Fig. 9:Crossvalidation of accuracy obtained by Machine Learning models

The P1 dataset produced the maximum accuracy of 58.89%, as shown in Figure 9. It was discovered that the accuracy extended to other datasets. Then, three datasets related to plankton and a combination of three sets in one were evaluated for the confusion matrix (CM). Table 8 uses SVM to tabulate the CM. The table shows the values for true negatives (TN), false negatives (FN), false positives (FP), and true positives (TP).

Table8:Confusion matrices for different datasets

	Confusion matrix	
	TP	FP
	FN	TN
P1 dataset	198	106
	196	151
P2 dataset	231	94
	205	121
P3 dataset	216	112
	195	128
All datasets in one	210	105
	190	146

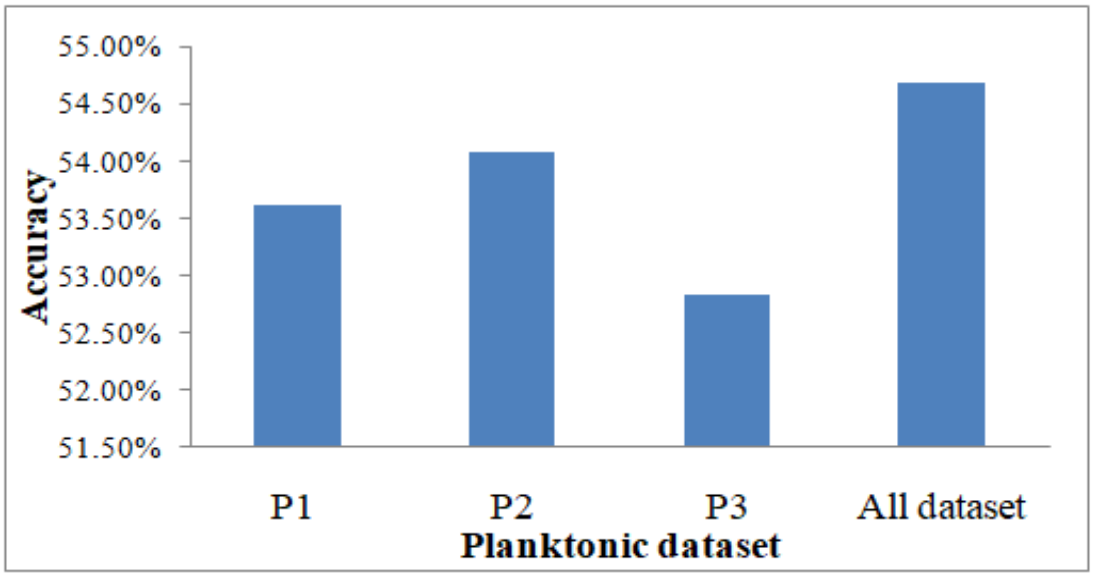


Fig. 10: Evaluation of Accuracy using SVM

CM, commonly referred to as an error matrix, provides insights into the performance of classification tasks by comparing properly classified examples with wrongly classified ones. This method works for both binary and multiclass classification problems. When the 70:30 ratio is utilised for training and testing, Figure 10 displays the accuracy for both the full dataset and three other datasets.

P1 achieves an accuracy of 53.61%, P2 achieves 54.07%, and P3 achieves 53%. Table tabulates the optimal gamma (G), CV, and cost parameter (C) values.

Table 9: Optimized parameters

	best C	**best G**	**best cv**
P1 dataset	1024	0.0156	58.8989
P2 dataset	16384	0.0078	58.4314
P3 dataset	16384	0.0078	58.23534
All datasets in one	8192	0.0039	58.8889

The C parameter applies a penalty to every misclassified data item. If C is small enough to have a minimal penalty for misclassified points, a large margin boundary choice is made at the cost of more misclassifications. Because a high penalty results in a decision boundary with a tighter margin, SVM aims to minimize the number of instances that are wrongly categorized when C is big. The RBF Gamma parameter affects the impact distance of one training point. A larger concentration of clustered points results from a broad similarity radius, which is denoted by low gamma values.

To be included in the same category, points with high gamma values must be relatively close to one another. Models with extremely high gamma values appear to be overfitting. It's also important to keep in mind that SVM needs consistent input data in order for the functions to be of the same size and compliant. In SVM, higher gamma values result in more biassed yet accurate results, and vice versa. In contrast, a large value of the C indicates low bias but poor accuracy, and vice versa.

4.6 Results for comparing outputs of various Machine Learning approaches.

In this study, we have used different experimentations which are tabulated in Tables. Tables A to Table C list the various experiments that were conducted for this article.

Table A: An explanation of the experiments done to classify *M. fortuitum's*proteomedata-set
Experiment: To attain the classification performance for two-class classification using kNN, ANN, and SVM classifier.

Table B: An explanation of the experiments done to classify *M. fortuitum's* proteome data-setusing pre-processing
Experiment: To attain the classification performance of pre-processed data for two-class classification using kNN, ANN, and SVM classifier.

Table C: An explanation of the experimentsdone to classify *M. fortuitum*proteomedata-set using pre-processing and optimization
Experiment: To attain the classification performance of pre-processed and optimized data for two-class classification using kNN, ANN, and SVM classifier.

Experiment A: To attain theclassification performance for two-class classification using kNN, ANN, and SVM classifier.

The dataset used in this experiment was taken straight from the lab, without any optimisation or pre-processing. The outcomes of SVM, kNN, and ANN without pre-processing or optimisation are displayed in Table.

Table 10: Accuracy values of different machine learning models

Techniques	***Accuracy***
kNN without pre-processing and without optimization	49.62%
SVM without pre-processing and without optimization	53.75%
ANN without pre-processing and without optimization	48.4%

The accuracy values that were achieved without the use of pre-processing or optimisation are tabulated in Table. The experiment revealed that the greatest accuracy achieved using SVM was 53.75%. Because models do not incorporate pertinent data from the training set, accuracy levels are lower. Because it doesn't require a drawn-out re-training process and has a non-linear decision boundary, the kNN classifier is both simple and effective. The authors used pre-processing techniques to increase accuracy.

Experiment B: To attain the classification performance of pre-processed data for two-class classification using kNN, ANN, and SVM classifier.

In this experimentation, authors pre-processed the dataset which was obtained from the lab. Table 11 shows the results of SVM, kNN, and ANN with pre-processing.

Table 11: Accuracy values of different machine learning models with pre-processing

Techniques	***Accuracy***
kNN with pre-processing and without optimization	81.75%
SVM with pre-processing and without optimization	81.21%
ANN with pre-processing and without optimization	75%

The accuracy levels that were discovered following the data's pre-processing are tabulated in Table 11. Authors used over- and under-expressed proteins during biofilm development to train the model for pre-processing, which increased accuracy levels—exactly what the author required for biomarker prediction. With kNN, the highest accuracy of 81.75% was reached without any optimizations. Using kNN over SVM resulted in an accuracy gain of 0.06%, whereas using kNNover ANN produced an accuracy improvement of 8.2%.Conceptually, kNN is more straightforward and intuitive. It doesn't need knowledge of intricate ideas like hyperplanes, which are essential to SVM.

Experiment 3: To attain theclassification performance of pre-processed and optimized data for two-class classification using kNN, ANN, and SVM classifier.
In this experimentation, authors pre-processed the dataset and optimized the values using PCA.

Table 12 shows the results of SVM, kNN, and ANN with pre-processing and using optimization techniques.

Table 12: Accuracy values of different machine learning models with pre-processing and optimization

Techniques	***Accuracy***
kNN with pre-processing and optimization	82.98%
SVM with pre-processing and with optimization	82.75%
ANN with pre-processing and optimization	78%

The highest accuracy of 82.98% was achieved by utilisingkNN with optimization. A 0.02% increase in accuracy was obtained when kNN was used instead of SVM.

4.7 Comparison with other techniques*:*

Authors compared their results with other machine learning algorithms like Random Forest, Naïve Bayes, and Logistic regression. The results are shown in **Fig 11.**

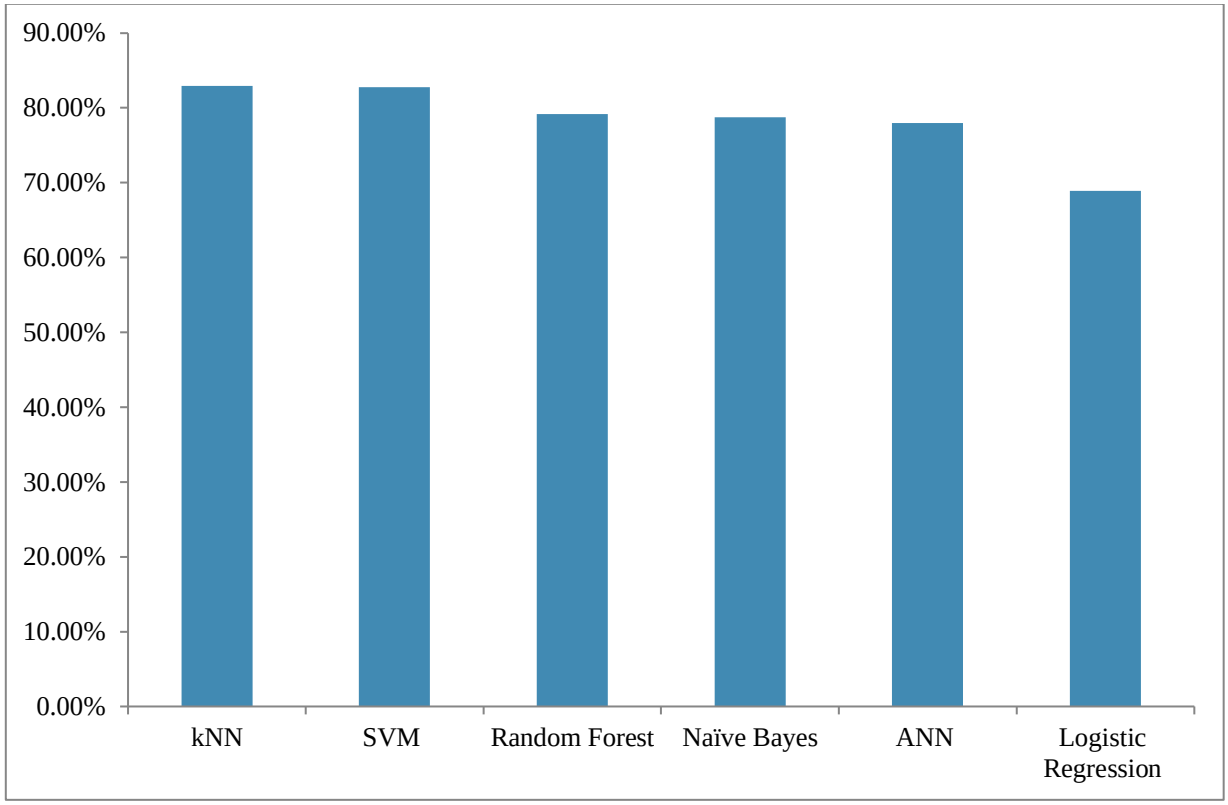


Fig 11- Comparativegraph of different machine learning tools

As seen in Fig 11,kNN yields a maximum accuracy of 82.98%.Because of their inherent randomness, random forests are also capable of arbitrary complexity and are resistant to overfitting. However, because they struggle to handle large numbers of features, they are not ideal for picture data. It has been demonstrated that Naive Bayes is a very computationally efficient algorithm that can handle big datasets rapidly. It is very accurate in both binary and multi-class prediction tasks.

Chapter 5

Conclusion and future prospective

Conclusion

The findings of this study are based on the machine learning prediction on *M. fortuitum*proteomic data.The analysis was conducted by employing different machine learning algorithms, including k Nearest Neighbor, Support Vector Machine, and Artificial Neural Network. Upon implementation of the machine learning algorithms on global proteome data from *M. fortuitum*, the research revealed that k- Nearest Neighbor the best accuracy of 82.98% after pre-processing and optimization. In addition, the research compared results classification accuracy with other previous algorithm activities and concluded that k-Nearest Neighbor model recording better results than Random Forest, Naïve Bayes, Artificial Neural Network, and Logistic Regression. The findings indicate that k -Nearest Neighbor produced the highest accuracy in predicting proteins associated with biofilm production, and therefore it may be used for prediction of proteins as diagnostic or therapeutic markersagainst *M. fortuitum*infection.

Future Directions:

Building upon this study, future investigations could develop deeper understanding into the identified proteins, unlocking their potential to revolutionize our approach to *M. fortuitum and* related infections.

i. **Functional validation:** Experimentally validating the roles of identified proteins in *M. fortuitum*'s pathogenesis is paramount. Techniques such as gene knockouts or overexpression can be employed to elucidate their specific contributions to biofilm formation, antibiotic resistance, or other crucial cellular processes. This functional validation will not only strengthen our understanding of the pathogen's biology but also provide a solid foundation for targeted drug development.

ii. **Drug target development:** The identified proteins present exciting opportunities for the development of novel therapeutics against *M. fortuitum*. Their structures can be used for structure-based drug design, a computational approach to designing drugs that specifically target these proteins. Alternatively, high-throughput screening can be employed to rapidly evaluate vast libraries of existing compounds for their ability to inhibit the function of these proteins. By pursuing these avenues, researchers can

identify promising drug candidates that can be further developed and tested for efficacy in combating *M. fortuitum*infections.

iii. **Biomarker development:** The identified proteins also hold immense potential as diagnostic biomarkers for *M. fortuitum*infections. By developing assays that can detect the presence or abundance of these proteins in clinical samples, such as blood or sputum, clinicians can diagnose infections earlier and more accurately. This early diagnosis can lead to prompt initiation of appropriate treatment, potentially improving patient outcomes and reducing the risk of complications associated with *M. fortuitum*infections. Additionally, monitoring the levels of these biomarkers during treatment can provide valuable insights into the efficacy of the therapy, allowing for adjustments to be made as needed.

By pursuing these future directions, we can significantly advance our knowledge *ofM. fortuitum*and translate these findings into tangible improvements in patient care. Unveiling the functional roles of the identified proteins will not only provide deeper biological insights but also pave the way for the development of targeted therapies. Furthermore, the establishment of these proteins as diagnostic biomarkers holds a way of earlier and more accurate diagnosis, ultimately leading to improved clinical management of *M. fortuitum*infections.

Chapter 6

References

REFERENCES

1. X. Wang et al., "Mycobacterium fortuituminfections: Clinical features, diagnosis, and treatment," Clinical Microbiology Reviews, vol. 24, no. 2, pp. 898-928, 2011.(DOI: 10.1128/CMR.00005-2010)
2. M. J. Benenson, "Immunocompromised Host Defenses in Sepsis," Clinics in Chest Medicine, vol. 21, no. 1, pp. 281-290, 2000.
3. M. B. Coyle, "Mycobacteria: Acid-Fast Bacteria," in Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, G. L. Mandell, J. E. Bennett, and R. Dolin, Eds. Philadelphia, PA: Churchill Livingstone, 2005, pp. 2468-2509.
4. M. Dhar et al., "Nosocomial infections due to Mycobacterium fortuitumcomplex: A review and proposal for diagnostic schema," International Journal of Mycobacteriology, vol. 3, no. 1, pp. 25-33, 2014.(DOI: 10.4103/0220-1868.127911)
5. C. Rodloff et al., "Mycobacterium fortuitumcomplex infections: Diagnosis and treatment challenges," Therapeutic Advances in Infectious Diseases, vol. 2, no. 6, pp. 221-239, 2013.(DOI: 10.1177/2040275813480202)
6. M. F. Wilkins et al., "Progress with proteome projects: Why all proteins are important," Trends in Biotechnology, vol. 19, no. 6, pp. 385-390, 2001.(DOI: 10.1016/S0167-7799(01)01803-7)
7. Pandey et al., "Mass spectrometry-based proteomics to understand and combat infectious diseases," Expert Review of Proteomics, vol. 5, no. 3, pp. 327-338, 2008.(DOI: 10.1080/14789450802038440)
8. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 27:1-27:27, 2011.(DOI: 10.1145/1999399.1999442)
9. W. Hoefsloot*et al.*, “The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study,” *Eur Respir J*, vol. 42, no. 6, pp. 1604–1613, Dec.2013, doi:.
10. J. Umrao*et al.*, “Prevalence and species spectrum of both pulmonary and extrapulmonary nontuberculous mycobacteria isolates at a tertiary care center,” *International Journal of Mycobacteriology*, vol. 5, no. 3, pp. 288–293, Sep. 2016, doi: 10.1016/j.ijmyco.2016.06.008.

11. K. Maurya *et al.*, "Prevalence of Nontuberculous Mycobacteria among Extrapulmonary Tuberculosis Cases in Tertiary Care Centers in Northern India," *BioMed Research International*, vol. 2015, pp. 1–5, 2015, doi: 10.1155/2015/465403.
12. S. Macente, C. Helbel, S. F. R. Souza, V. L. D. Siqueira, R. A. F. Padua, and R. F. Cardoso, "Disseminated folliculitis by Mycobacterium fortuitumin an immunocompetent woman*," *An. Bras. Dermatol.*, vol. 88, no. 1, pp. 102–104, Feb. 2013, doi: 10.1590/S0365-05962013000100014.
13. P. Chetchotisakd*et al.*, "Disseminated Infection Due to Rapidly Growing Mycobacteria in Immunocompetent Hosts Presenting with Chronic Lymphadenopathy: A Previously Unrecognized Clinical Entity," *Clinical Infectious Diseases*, vol. 30, no. 1, pp. 29–34, Jan. 2000, doi: 10.1086/313589.
14. K. Shrivastava *et al.*, "An Overview of Pulmonary Infections due to Rapidly Growing Mycobacteria in South Asia and Impressions from a Subtropical Region," *International Journal of Mycobacteriology*, vol. 9, no. 1, 2020.
15. Poonam, R.M. Yennamalli, G.S. Bisht, *et al.* Ribosomal maturation factor (RimP) is essential for survival of nontuberculous mycobacteria *M. fortuitum*under in vitro acidic stress conditions. 3 Biotech 9, 127 (2019).https://doi.org/10.1007/s13205-019-1659-y.
16. W. L. Hand, "Mycobacterium fortuitum—A Human Pathogen," *Ann Intern Med*, vol. 73, no. 6, p. 971, Dec. 1970, doi: 10.7326/0003-4819-73-6-971.
17. Sharma, S. Bansal, N. Kumari, J. Vashistt, and R. Shrivastava, "Comparative proteomic investigation unravels the pathobiology of Mycobacterium fortuitumbiofilm," *Appl MicrobiolBiotechnol*, vol. 107, no. 19, pp. 6029–6046, Oct. 2023, doi: 10.1007/s00253-023-12705-y.
18. E. Bardouniotis, H. Ceri, and M. E. Olson, "Biofilm Formation and Biocide Susceptibility Testing of Mycobacterium fortuitumand Mycobacterium marinum," *Current Microbiology*, vol. 46, no. 1, pp. 28–32, Jan. 2003, doi: 10.1007/s00284-002-3796-4.
19. P. Katoch, K. Gupta, R. M. Yennamalli, J. Vashistt, G. S. Bisht, and R. Shrivastava, "Random insertion transposon mutagenesis of *M. fortuitum*identified mutant defective in biofilm formation," Biochemical and Biophysical Research Communications, vol. 521, no. 4, pp. 991–996, Jan. 2020, doi: 10.1016/j.bbrc.2019.11.021.

20. L. Hall-Stoodley, C. W. Keevil, and H. M. Lappin-Scott, "Mycobacterium fortuitumand Mycobacterium chelonae biofilm formation under high and low nutrient conditions," *Journal of Applied Microbiology*, vol. 85, no. S1, pp. 60S-69S, Dec. 1998, doi: 10.1111/j.1365-2672.1998.tb05284.x.
21. Brown-Elliott and R. J. Wallace, "Clinical and Taxonomic Status of Pathogenic Nonpigmented or Late-Pigmenting Rapidly Growing Mycobacteria," *Clin Microbiol Rev*, vol. 15, no. 4, pp. 716–746, Oct. 2002, doi: 10.1128/CMR.15.4.716-746.2002.
22. G. El Helou, G. M. Viola, R. Hachem, X. Y. Han, and I. I. Raad, "Rapidly growing mycobacterial bloodstream infections," *The Lancet Infectious Diseases*, vol. 13, no. 2, pp. 166–174, Feb. 2013, doi: 10.1016/S1473-3099(12)70316-X.
23. Brown-Elliott and R. J. Wallace, "Clinical and Taxonomic Status of Pathogenic Nonpigmented or Late-Pigmenting Rapidly Growing Mycobacteria," *Clin Microbiol Rev*, vol. 15, no. 4, pp. 716–746, Oct. 2002, doi: 10.1128/CMR.15.4.716-746.2002.
24. M. A. De Groote and G. Huitt, "Infections Due to Rapidly Growing Mycobacteria," *Clinical Infectious Diseases*, vol. 42, no. 12, pp. 1756–1763, Jun. 2006, doi: 10.1086/504381.
25. Kumar, K. Shrivastava, A. Singh, V. Chauhan, and M. Varma-Basil, "Skin and soft-tissue infections due to rapidly growing mycobacteria: An overview," *Int J Mycobacteriol*, vol. 10, no. 3, p. 293, 2021, doi: 10.4103/ijmy.ijmy_110_21.
26. Ortíz-Pérez, N. Martín-de-Hijas, N. Alonso-Rodríguez, D. Molina-Manso, R. Fernández-Roblas, and J. Esteban, "Importance of antibiotic penetration in the antimicrobial resistance of biofilm formed by non-pigmented rapidly growing mycobacteria against amikacin, ciprofloxacin and clarithromycin," *EnfermedadesInfecciosas y MicrobiologíaClínica*, vol. 29, no. 2, pp. 79–84, Feb. 2011, doi: 10.1016/j.eimc.2010.08.016.
27. T. T. Aung *et al.*, "Biofilms of Pathogenic Nontuberculous Mycobacteria Targeted by New Therapeutic Approaches," *Antimicrob Agents Chemother*, vol. 60, no. 1, pp. 24–35, Jan. 2016, doi: 10.1128/AAC.01509-15.
28. Sharma, J. Vashistt, and R. Shrivastava, "Response surface modeling integrated microtiter plate assay for *Mycobacterium fortuitum*biofilm quantification", *Biofouling*, vol 37, no. 8, pp.830-843, doi: 10.1080/08927014.2021.1974846

29. J. O. Falkinham, “Epidemiology of infection by nontuberculous mycobacteria,” *CLIN. MICROBIOL. REV.*, vol. 9, 1996.
30. J. R. Honda, R. Virdi, and E. D. Chan, “Global Environmental Nontuberculous Mycobacteria and Their Contemporaneous Man-Made and Natural Niches,” *Front. Microbiol.*, vol. 9, p. 2029, Aug. 2018, doi: 10.3389/fmicb.2018.02029.
31. N. Prashar, M. Sood, and S. Jain, “Novel Cardiac Arrhythmia Processing using Machine Learning Techniques,” Int. J. Image Grap., vol. 20, no. 03, p. 2050023, Jul. 2020, doi: 10.1142/S0219467820500230.
32. J. Dogra, S. Jain, M. Sood, Glioma Classification of MR brain tumor employing Machine Learning, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(8), 2676-2682, 2019.
33. A.O. Salau, S. Jain, M. Sood, “Paradigms in Computational Intelligence and Data Sciences”, 2022, CRC, Taylor & Francis Group, U.K.ISBN:9781032123134
34. M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015
35. L. Hall-Stoodley, C. W. Keevil, and H. M. Lappin-Scott, “Mycobacterium fortuitumand Mycobacterium chelonae biofilm formation under high and low nutrient conditions,” *J. Appl. Microbiol.*, vol. 85, no. S1, pp. 60S-69S, 1998.
36. E. Gonzalez-Diaz, R. Morfin-Otero, H. R. Perez-Gomez, S. Esparza-Ahumada, and E. Rodriguez-Noriega, “Rapidly Growing Mycobacterial Infections of the Skin and Soft Tissues Caused by *M. fortuitum*and*M.*chelonae,” *Curr. Trop. Med. Rep.*, vol. 5, no. 3, pp. 162–169, 2018.
37. N. Zamora, J. Esteban, T. J. Kinnari, A. Celdrán, J. J. Granizo, and C. Zafra, “In-vitro evaluation of the adhesion to polypropylene sutures of non-pigmented, rapidly growing mycobacteria,” *Clin. Microbiol. Infect.*, vol. 13, no. 9, pp. 902–907, 2007.
38. R.K. Taylor, V.L. Miller, D.B. Furlong, J.J. Mekalanos, Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 2833e2837, https://doi.org/ 10.1073/pnas.84.9.2833.
39. R. Parti, R. Shrivastava, S. Srivastava, A. Subramanian, R. Roy, B.S. Srivastava, R. Srivastava, A transposon insertion mutant of Mycobacterium fortuitumattenuated in virulence and persistence in a murine infection model that is complemented by

Rv3291c of Mycobacterium tuberculosis, Microb. Pathog. 45 (2008) 370e376, https://doi.org/10.1016/j.micpath.2008.08.008.

40. T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, 1967
41. Graves, A. Mohamed, and G. Hinton, "Generating sequences with recurrent neural networks," arXiv preprint arXiv:1308.0850, 2013.
42. A.O. Salau, S. Jain, Adaptive Diagnostic Machine Learning Technique for Classification of Cell Decisions for AKT Protein, Informatics in Medicine Unlocked, 7 January 2021, 100511
43. N. Prashar, M. Sood, and S. Jain, "Novel Cardiac Arrhythmia Processing using Machine Learning Techniques," *Int. J. Image Grap.*, vol. 20, no. 03, p. 2050023, Jul. 2020, doi: 10.1142/S0219467820500230.
44. J. Dogra, S. Jain, M. Sood, Glioma Classification of MR brain tumor employing Machine Learning, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(8), 2676-2682, 2019.
45. S. Jain, A Computational Model for Detection of Lung Diseases Due to Forkhead Transcription Factors. In: Marriwala, N., Tripathi, C.C., Jain, S., Mathapathi, S. (eds) Emergent Converging Technologies and Biomedical Systems . Lecture Notes in Electrical Engineering, vol 841. Springer, Singapore, 2022.https://doi.org/10.1007/978-981-16-8774-7_7
46. T. Paul, R. Raj, P. Garg and S. Jain, "Real Time Monitoring of Water Quality for Rural Areas: A Machine Learning and Internet of Things Approach," *2023 4th International Conference on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, 2023, pp. 1-6, doi: 10.1109/ICIEM59379.2023.10165824.
47. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," in 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016, pp. 265–283.
48. A.O. Salau, S. Jain, M. Sood, "Paradigms in Computational Intelligence and Data Sciences", 2022, CRC, Taylor & Francis Group, U.K.ISBN:9781032123134

Chapter 6

Publication

LIST OF PUBLICATIONS

JOURNALS:

Shan Ghai, Rahul Shrivastava, Shruti Jain, Analysis of Proteins Involved in *Mycobacterium fortuitum*Biofilm Formation Employing Machine Learning Techniques for Prediction of Potential Drug Targets, Current Cancer Drug Targets. **(Communicated)**

CONFERENCE PUBLICATIONS:

1) **S. Ghai**, P. Jagwan, R. Shrivastava and S. Jain, "Analysis of differentially expressed *M. fortuitum*proteins for biomarker prediction using Support Vector Machine," 2023 Seventh International Conference on Image Information Processing (ICIIP), Solan, India, 2023, pp. 212-217, doi: 10.1109/ICIIP61524.2023.10537781.

2) PrajjwalJagwan, **ShanGhai**, R. Shrivastava and S. Jain, "Prediction of Protein Biomarkers for *Mycobacterium fortuitum*using Machine Learning Technique," 2023 9th International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2023, pp. 416-421, doi: 10.1109/ICSC60394.2023.10440976.

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 6th June 2024

Type of Document (Tick): PhD Thesis | M.Tech/M.Sc. Dissertation ✓ | B.Tech./B.Sc./BBA/Other

Name: Shan Ghai Department: BT-BI Enrolment No 225111004

Contact No. 6230008897 E-mail. Ghaishan25@gmail.com

Name of the Supervisor: Dr. Rahul Shrivastava and Prof. Shruti Jain.

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): ANALYSIS OF MYCOBACTERIUM PROTEOME USING MACHINE LEARNING TECHNIQUES

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages = 59
- Total No. of Preliminary pages = 7
- Total No. of pages accommodate bibliography/references = 66

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found Similarity Index at ....10..... (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) 6/6/24 (Dr. Shruti)

10/06/2024

Signature of HOD

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Abstract & Chapters Details	
06.06.24	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String	07%	Word Counts	9453
Report Generated on			Character Counts	55158
06.06.24		Submission ID	Page counts	52
		2396806691	File Size	1.3M

Checked by
Name & Signature SURESH K CHAUHAN

S Chauhan 06.06.24

LIBRARIAN
LEARNING RESOURCE CENTRE
Jaypee University of Information Technology
Waknaghat, Distt. Solan (Himachal Pradesh)
Pin Code - 173234

Librarian

Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at plagcheck.juit@gmail.com