Enhancing Tuberculosis Diagnosis and Treatment Outcomes: A Stacked Loopy Decision Tree Approach Empowered by Moth Search Algorithm Optimization

Dr. Huma Khan[1], Dr. Mithun D'Souza[2], Dr. K. Suresh Babu[3], Janjhyam Venkata Naga Ramesh[4], Dr.K.R.Praneeth[5], Pinapati Lakshmana Rao[6]

Associate Professor CSE, Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India[1]
Assistant Professor, Department of Computer Science, St. Joseph's University, Bangalore, India[2]
Professor, Department of Biochemistry, Symbiosis Medical College for Women, Symbiosis International (Deemed University) Pune, India[3]
Adjunct Professor, Department of CSE, Graphic Era Hill University, Dehradun,248002, India[4(a)].
Adjunct Professor, Department of CSE, Graphic Era Deemed To Be University, Dehradun,248002, Uttarakhand, India[4(b)]
Assistant Professor, School of Management, The Apollo University, The Apollo Knowledge City Campus, Chittoor, Andhra Pradesh, India[5]
Assistant professor, Department of Computer Science and Engineering, Research Scholar, Koneru Lakshmaiah Education Foundation (KLEF),Vaddeswaram, Guntur Dist, Andhra Pradesh, India[6].

Abstract*—Chest X-ray imaging is the main tool for detecting tuberculosis (TB), providing essential information about pulmonary abnormalities that may indicate the presence of the disease. Still, manual interpretation is a common component of older diagnostic methods, and it may be laborious and subjective. The development of sophisticated machine learning methods offers a potential way to improve TB detection through the automation of chest X-ray image interpretation. This takes a look at goals to increase a sturdy framework for TB diagnosis the usage of Stacked Loopy Decision Trees (SLDT) optimized with the Moth Search Algorithm (MSA). The objective is to improve upon current techniques with the aid of integrating sophisticated feature extraction and ensemble mastering strategies. The novelty lies in the integration of SLDT, a hierarchical ensemble model able to shooting complex styles in chest X-ray images, with MSA for optimized parameter tuning and function selection. This technique addresses the complexity of TB analysis by enhancing each interpretability and overall performance metrics. The proposed framework employs the Gray-Level Co-prevalence Matrix (GLCM) for texture characteristic extraction, accompanied with the aid of SLDT ensemble studying optimized through MSA. This methodology objectives to discern TB-particular styles from chest X-ray pictures with excessive accuracy and efficiency. Evaluation of a comprehensive dataset demonstrates advanced performance metrics including accuracy, sensitivity, specificity, and vicinity underneath the ROC curve (AUC-ROC) compared to traditional gadget gaining knowledge of procedures. The outcomes demonstrate how well the SLDT-MSA framework performs in diagnosing TB, with 99% accuracy. The observation indicates that the SLDT-MSA framework offers practitioners a trustworthy and easily understandable solutions, marking a significant advancement in TB diagnosis.

Keywords*—Tuberculosis (TB); chest x-ray; stacked loopy decision trees (SLDT); moth search algorithm (MSA); medical imaging

I. Introduction

Tuberculosis is a persistent infectious disease caused by Mycobacterium tuberculosis (MTB), a slow-growing microorganism capable of surviving in both extracellular and intracellular environments [1]. The bacterium can enter a latent phase within the host's body, remaining dormant until conditions favor its reactivation into an active, contagious state, particularly in individuals with compromised immune systems [2]. According to the World Health Organization (WHO) report in 2019, approximately 10.0 million people worldwide were infected with TB, leading to 1.4 million deaths annually, highlighting TB as a significant global health challenge [3]. The disease disproportionately affects developing regions, where limited access to specialized medical professionals and TB should be diagnostic tools exacerbate the burden of TB [4]. While TB is curable, delayed detection can sigthis study did nificantly impact health outcomes [5]. Regions such as Africa and Southeast Asia bear the highest TB burden due to socio-economic factors [6]. The gold standard for TB detection remains culture testing, complemented by methods like chest radiography, sputum smear microscopy, and nucleic acid amplification. Advanced diagnostic techniques including chest computed tomography and molecular tests offer further improvements in detecting and managing TB cases [7]. Existing diagnostic and treatment methods for tuberculosis (TB) face significant challenges that impede effective disease management across diverse global settings. The cornerstone of conventional TB diagnosis revolves around methods like sputum smear microscopy and culture, which, while established, are fraught with limitations [8]. These techniques are labor-intensive, requiring skilled personnel and prolonged processing times that delay diagnosis and treatment initiation. Furthermore, they often fail to detect TB in its early stages or in cases with extrapulmonary manifestations, leading to missed diagnoses and subsequent transmission risks [9].

The inconsistent sensitivity of conventional TB testing techniques is one of its main disadvantages. For example, in situations with low bacterial load, sputum smear microscopy may not identify TB bacilli, leading to false-negative findings that postpone necessary treatment. To make matters worse, waiting weeks for findings from culture-based techniques might prolong the time it takes to diagnose and treat patients. Though these regions have the greatest rate of tuberculosis, these issues are most severe in resource-constrained settings where access to diagnostic facilities, skilled people, and dependable laboratory infrastructure is restricted [10]. The growing prevalence of drug-resistant tuberculosis strains exacerbates these problems by complicating treatment results and calling for more specialised and efficient diagnostic techniques. Drug-resistant TB strains, such as those that are extensively drug-resistant (XDR-TB) and multidrug-resistant (MDR-TB), need specific treatment plans that depend on a quick and precise diagnosis [11]. In addition to endangering the course of treatment for individual patients, delaying the identification of drug-resistant tuberculosis (TB) promotes resistance amplification in communities and continuous transmission. Given these obstacles, it is vitally necessary to create and implement more precise, quick, and easily accessible TB testing techniques on a worldwide scale. These advancements ought to focus on increasing specificity, speeding up diagnosis, and improving sensitivity in order to distinguish tuberculosis from other respiratory disorders. Through the surmounting of these obstacles, novel diagnostic technologies have the potential to significantly enhance tuberculosis (TB) management tactics, enable prompt treatment start, reduce the rate of transmission, and ultimately enhance health outcomes for TB-affected individuals and communities around the globe. In order to overcome the shortcomings of current TB diagnosis and treatment approaches, this project will create a novel strategy that combines optimisation methods with the Moth Search Algorithm (MSA) and Stacked Loopy Decision Trees (SLDT). Through the use of cutting-edge machine learning techniques, this innovative approach seeks to improve the efficacy, scalability, and accuracy of tuberculosis diagnosis and treatment results using chest X-ray images. The study aims to accomplish multiple goals: firstly, to create a strong SLDT model that can identify complex patterns in medical images that suggest tuberculosis-related abnormalities; secondly, to use MSA to optimise model parameters and feature selection to improve diagnostic accuracy and reliability; thirdly, to compare the performance of the SLDT-MSA framework to established diagnostic techniques using large datasets, with a focus on metrics like sensitivity, specificity, and area under the ROC curve (AUC-ROC); and lastly, to offer insights into the proposed approach's potential impact on TB management strategies. Through the integration of SLDT and MSA, this research seeks to develop a more efficient and user-friendly diagnostic tool that can expedite the diagnosis of tuberculosis (TB), enable the beginning of treatment promptly, and ultimately enhance patient outcomes across a variety of global healthcare settings. The principal findings of the research are given below are as follows:

- The study integrates Stacked Loopy Decision Trees (SLDT) with the Moth Search Algorithm (MSA), offering a novel approach to tuberculosis (TB) diagnosis. This integration enhances the model's ability to capture intricate patterns in chest X-ray images indicative of TB-related abnormalities.
- The research expedites the diagnostic process, cutting down on processing times and increasing the effectiveness of tuberculosis detection by utilizing MSA for feature selection and parameter optimization inside the SLDT setup.
- The proposed approach holds significant clinical relevance by providing clinicians with interpretable diagnostic outputs, aiding in informed decision-making and facilitating prompt patient management strategies.
- Achieved a high diagnostic accuracy of 99% in distinguishing TB-positive cases from normal conditions, surpassing traditional methods. This improvement is crucial for early detection and timely treatment initiation, especially in resource-limited settings.
- Offers a scalable solution applicable across diverse healthcare settings, potentially improving access to accurate TB diagnosis globally. This scalability is critical for addressing the disparities in TB healthcare delivery and outcomes. Advances in TB diagnosis through innovative machine learning methodologies contribute to optimized TB management strategies, aiming to reduce transmission rates, mitigate drug resistance, and enhance overall public health outcomes.

The rest of the paper is organized as follows: Section II discusses related work. Section III discusses the problem statement Section IV explains the proposed methodology. Section V reports and compares the experimental results. Section VI concludes the paper and mentions future work.

II. Literature Review

Bacteria lead to the development of TB, which is a life-threatening lung disease and one of the top 10 causes of mortality. Detecting tuberculosis at an early stage and confirming the diagnosis is crucial, as failing to do so can lead to serious illness. This project involved developing a technique for precise tuberculosis detection from chest X-rays through the enhancement of images and advanced technological analysis. We used many public databases to make a new database with 3500 TB infected and 3500 normal chest X-ray pictures for our research. Nine separate deep convolutional neural networks were employed to leverage their pre-existing training, before being evaluated for their ability to distinguish between TB and non-TB normal instances. This study carried out three different. At the beginning, two separate U-net models were used for segmentation of the X-ray images. Second, it classified X-ray images. It categorized and organized lung images into different sections. Using X-ray images, the most effective model, ChexNet, is capable of detecting tuberculosis with great precision. It is responsive and demonstrates a high F1-score and specificity. However, the classification accuracy improved when utilizing lung images that were segmented into sections compared to using the entire X-ray images. The segmented lung images exhibited improved accuracy, precision, sensitivity, F1-score, and specificity with DenseNet201.It also used a way to

show that CNN mostly learns from certain parts of the lung, which made it better at finding problems. The new approach is highly effective and can aid doctors in promptly diagnosing tuberculosis with the help of computers. Despite this, the study's limitations include its use of a limited dataset comprising only 7000 images, which may compromise the accuracy of the models in diverse real-life contexts [12].

Modern health systems greatly rely on computer science for their functioning. The utilization of computers in medical practice facilitates the teamwork of doctors in diagnosing illnesses, ultimately improving the care provided to patients. They also provide support to researchers and decision-makers in the healthcare field. So, any new ideas that make it easier to diagnose health issues while still being safe and effective are really important for making healthcare better. Early detection can lead to the identification of numerous illnesses. In this research, we used different methods to study tuberculosis (TB). Our recommendation is to create an improved machine learning algorithm that identifies the optimal texture characteristics in images related to TB and configures the classifier settings. We want to make our measurements more accurate and use fewer characteristics. The challenge lies in attempting to handle numerous tasks at once and ensuring they all function at their highest capacity. The most beneficial traits are selected using a genetic algorithm (GA) and then input into a support vector machine (SVM) for classification. The new technique we implemented for the ImageCLEF 2020 data yielded better results compared to the other methods we used. According to the test results, the modified SVM classifier outperforms the standard ones. The study has some limitations. Utilizing just one set of data from ImageCLEF 2020 may not capture all the diverse presentations of TB. The choice of features by the genetic algorithm could potentially neglect important factors in diverse clinical settings [13].

The most recent research conducted by the World Health Organization (WHO) in 2018 revealed that tuberculosis leads to the deaths of 5 million people annually, with approximately 10 million people falling ill from the disease. Furthermore, over 4,000 people die from TB every day. If the sickness had been identified sooner, many of the deaths could have been prevented. Recent books and articles have talked about using deep learning to help doctors diagnose illnesses by looking at medical images. Although deep learning has shown potential in many areas, there are not many thorough studies to diagnose tuberculosis. In order to improve its performance, deep learning requires a substantial amount of good training examples. TB chest x-ray pictures are usually not very clear because the contrast is not very strong. This research focuses on the impact of improving visual representations on the problem-solving abilities of a computer program. The program for enhancing images enhanced the appearance of the images by highlighting their distinctive characteristics. Three different methods were tested to enhance the visual appeal of images: Unsharp Masking, High-Frequency Emphasis Filtering, and Contrast Limited Adaptive Histogram Equalization. The better pictures were given to ResNet and EfficientNet models to learn from. In a collection of TB pictures, we got 89. 92% accuracy in classifying them and 94. 8% in AUC scores. The Shenzhen dataset is the source of all the findings and is available to anyone. However, this study has its restrictions. A more advanced GPU and increased memory are necessary to utilize the CNN network for sending the original image at its full resolution. Moreover, the training would be extended due to the process [14].

The progression from dormant tuberculosis to active tuberculosis presents a serious problem. Although skin tests and blood tests are effective for detecting a tuberculosis (TB) infection, they cannot differentiate between latent TB infection and active TB. Diagnosis of LTBI presents difficulties as there are no accurate tests available and differentiating it from active TB is complex. Testing for tuberculosis using sputum culture takes a long time and cannot tell the difference between active tuberculosis and latent tuberculosis infection. This article discusses the way TB bacteria grows and the body's defence mechanism in latent TB infection. This involves both the innate and acquired immune responses of the body, the strategies used by TB bacteria to evade the immune system, and the impact of genetic factors on this mechanism. Given our current understanding, we elucidate the present circumstances and challenges in detecting LTBI. We also explore the potential use of machine learning (ML) in the diagnosis of LTBI, as well as the advantages and disadvantages of employing ML in this context. The study explores the ways in which machine learning could be utilized to enhance LTBI detection in the future. Although ML has benefits like better accuracy and efficiency in diagnosis, it also has some problems that need to be fixed. The constraints involve a requirement for extensive data, complexity in comprehension, dependence on particular techniques and technology, apprehensions about data security, and ambiguity in choosing features [15].

Medical professionals on the front lines must swiftly establish whether a patient showing symptoms has tested positive for COVID-19 or not. In areas with scarce resources and lacking biotechnology tests, this task becomes even more challenging. Tuberculosis remains a significant health concern in numerous impoverished nations. The primary indications include high body temperature, an unproductive cough, and fatigue, which bear resemblance to COVID-19. To aid in the detection of COVID-19, researcher propose the use of specialized technology for analysing chest X-rays, which are commonly found in hospitals. Following this, it have the option to employ computer programs to categorize and identify the X-rays, without the requirement of expensive equipment. A set of five various chest X-ray images was assembled by us. Included in this assortment are the same amount of cases for COVID-19, viral pneumonia, bacterial pneumonia, TB, and healthy individuals. The performance of different computer program combinations in extracting useful information from a dataset was evaluated. We tested out 14 advanced pre-made networks alongside conventional machine learning tools to identify the most optimal pairing. The best pipeline for classifying five different groups of items was a combination of ResNet-50 and a subspace discriminant classifier. It had the highest accuracy in detecting the classes. Additionally, the pipeline was able to accurately classify COVID-19, TB, and healthy cases in simpler problems with three categories, as well as COVID-19 and healthy images in problems with only two categories. The pipeline was really fast. It only took 0. 19 seconds to extract DF from each X-ray image and 2 minutes to train a traditional

classifier with over 2000 images on a regular computer. The findings show that our method could be helpful in finding COVID-19, especially in places with few resources. It uses X-rays that are easy to get and doesn't need a lot of computer power [16].

Tuberculosis (TB), a leading cause of death globally, necessitates accurate and early detection to prevent life-threatening outcomes. Using deep learning and sophisticated machine learning techniques to enhance tuberculosis detection from chest X-ray pictures has been the topic of several investigations. Through the use of deep learning models, data augmentation, and picture preprocessing, Rahman et al. (2020) showed that diagnosis accuracy may be greatly improved by segmentation and classification approaches, albeit the generalizability of their model may be constrained by the use of a particular dataset. A genetic approach for feature selection in conjunction with a support vector machine (SVM) classifier was presented by Hrizi et al. (2022). This method achieved good accuracy, although it may be constrained by bias in feature selection and dataset specificity. In order to improve the performance of deep learning models on low-contrast TB chest X-rays, Munadi et al. (2020) assessed image enhancement strategies. They were able to achieve a significant level of accuracy, but encountered difficulties with computing demands. In their evaluation of machine learning applications for latent tuberculosis infection (LTBI), Li et al. (2023) noted improvements in diagnostic efficiency as well as drawbacks including data needs and interpretability problems. Al-Timemy et al. (2021) highlighted the promise of accessible imaging technology in healthcare diagnostics by developing a computationally efficient pipeline for diagnosing COVID-19 and TB using deep features from chest X-rays. The pipeline demonstrated great accuracy even in resource-limited situations. When taken as a whole, this research shows the potential and difficulties of combining deep learning and machine learning to improve tuberculosis detection and emphasise the continuous need for a variety of high-quality data sources and processing power.

III. Research Gap

In the realm of tuberculosis (TB) diagnosis using chest X-ray imaging and advanced machine learning techniques, several critical research gaps hinder progress towards more accurate and efficient diagnostic methods. One major gap is the underexplored integration of ensemble learning, such as Stacked Loopy Decision Trees (SLDT), with sophisticated texture analysis methods tailored for TB detection. While SLDT frameworks excel in capturing intricate patterns in medical images, their synergy with advanced texture analysis techniques, like those derived from Gray-Level Co-occurrence Matrix (GLCM), remains insufficiently explored. This gap limits the development of models capable of effectively extracting subtle yet crucial texture features indicative of TB-related abnormalities in chest X-rays. Moreover, optimizing SLDTs for TB diagnosis presents another challenge. While metaheuristic algorithms like the Moth Search Algorithm (MSA) show promise in optimizing model parameters and feature selection, their application in fine-tuning SLDTs for medical image analysis is still emerging. Current studies often use simpler optimization techniques or focus on single-model architectures, neglecting the complexity of ensemble frameworks necessary for accurate TB diagnosis [17]. The scarcity of diverse and well-annotated datasets is another critical gap. Existing datasets often lack comprehensive representation across demographic groups, TB manifestations, and imaging conditions. This limitation hampers the development of machine learning models that can reliably generalize across different clinical scenarios and patient populations, essential for ensuring robust diagnostic accuracy and clinical applicability. Comparative studies benchmarking SLDT-MSA frameworks against deep learning architectures in TB diagnosis are sparse. Understanding the trade-offs between these methodologies, including diagnostic performance, computational efficiency, and scalability, is crucial for selecting optimal approaches based on specific clinical needs and resource constraints. Addressing these research gaps is essential for advancing TB diagnosis using machine learning, facilitating the development of more effective, interpretable, and clinically relevant diagnostic tools that enhance patient outcomes and healthcare delivery worldwide.

IV. Research Framework

The research framework for this examine on tuberculosis (TB) prognosis integrates superior device gaining knowledge of strategies with scientific imaging analysis to beautify diagnostic accuracy and efficiency. Central to the framework is the utilization of Stacked Loopy Decision Trees (SLDT) as an ensemble gaining knowledge of technique, optimized via the Moth Search Algorithm (MSA). SLDT permits the hierarchical extraction of complex patterns from chest X-ray images, which might be crucial for identifying TB-related abnormalities. The Moth Search Algorithm complements this by first-rate-tuning SLDT parameters and deciding on top of the line features derived from strategies which include Gray-Level Co-prevalence Matrix (GLCM) evaluation. This combined method ambitions to enhance diagnostic precision by means of shooting diffused texture versions indicative of TB, thereby improving sensitivity and decreasing fake-terrible effects. The framework also emphasizes version interpretability, presenting clinicians with transparent insights into the decision-making process for TB prognosis. By integrating those advanced methodologies, the studies framework seeks to pioneer a greater powerful and scalable diagnostic device for TB, probably transforming medical practice and public health techniques in TB management. Fig. 1 shows the workflow of the proposed approach.

A. Data Collection

The dataset for tuberculosis (TB) diagnosis consists of chest X-ray images collected from multiple sources, including a publicly accessible portion and additional images obtainable through the NIAID TB portal. The dataset is a collaborative effort involving researchers from Qatar University and the University of Dhaka, Bangladesh, along with collaborators from Malaysia, supported by medical professionals from Hamad Medical Corporation and Bangladesh. For research reasons, the 700 TB-positive images in this collection are freely available. Additionally, after assuming certain terms and conditions, researchers can download 2,800 TB photos from the NIAID TB site. 3,500 normal chest X-ray images are additionally contained in the collection and are openly available for comparison [18].

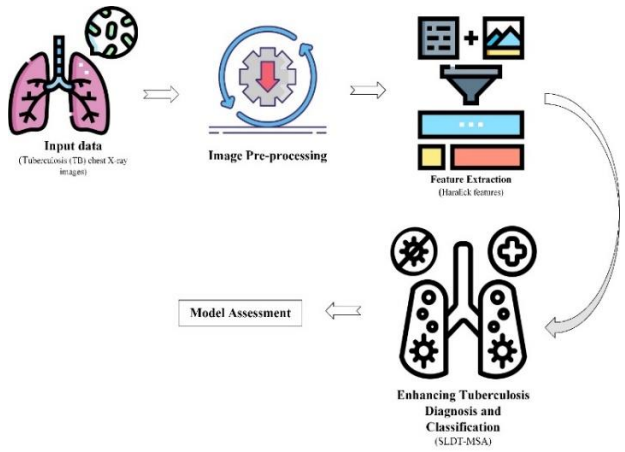

Fig. 1. Workflow of the proposed approach.

B. Image Pre-Processing

1) Data augmentation: In the field of medical image analysis, data augmentation is an essential tactic, especially large datasets such as chest X-rays used for TB diagnosis. These methods are employed to increase the variety of datasets and strengthen the resilience of machine learning models. When diagnosing illnesses where lesion orientations might change, the ability of the model to learn from different anatomical structure orientations through rotation and flipping is essential. Scaling is essential in identifying abnormalities of various sizes since lesion diameters vary throughout patients. Introducing noise improves the model's capacity to generalize and makes it more resistant to artefacts by simulating real-world variances in picture capture. To assist the model, adjust to various lighting conditions commonly found in clinical settings, contrast and brightness levels could be adjusted. In order to guarantee that the model learns to correctly identify features across a variety of patient instances, elastic deformations and random cropping, respectively, replicate anatomical differences and focus points within pictures. Through utilizing these methods in combination, the model becomes more capable of managing the intricacies and fluctuations present in medical imaging, which in turn enhances the precision of diagnosis and treatment results for diseases like TB.

2) Image resizing: Resizing images is an essential preprocessing step for getting image datasets ready for machine learning applications, particularly if it comes to medical image analysis where chest X-ray images are employed to diagnose TB. Consistency and computational economy are the main goals of scaling images to a standard size, such 224x224 pixels. To ensure consistency when feeding data into machine learning models that need fixed input sizes, every image should have the same dimensions. Through consistent input processing and the elimination of size disparities in data processing, this standardisation streamlines the data pipeline and enables models to learn from the data efficiently. Resizing lowers the computing overhead involved in model training and inference, which further improves computational efficiency. Efficient data processing through uniform image sizing expedites computations and accelerates the training process in general. Maintaining the aspect ratio of the source photos during the resizing process helps to avoid distortion. Usually, this is accomplished by downsizing the image while keeping its original aspect ratio such that it fits inside a certain bounding box. During resizing, pixel values are resampled using interpolation techniques like bilinear or bicubic interpolation. As much visual integrity as possible is retained in scaled images owing to these procedures, which also aid in maintaining image quality and details.

3) Image normalization: Standardization, often referred to as Z-score normalization, is an essential technique employed for preparing image data for machine learning applications. It is especially useful in medical image analysis, where images from chest X-rays are utilized to diagnose TB. Pixel values are transformed employing this approach to have a mean of 0 and a standard deviation of 1 for the whole dataset. Pixel values are rescaled using standardization, which involves removing the mean (μ) and dividing by the standard deviation (σ) of the pixel values for each image in the set of images. The following is the standardization Eq. (1):

$$Normalized\ Pixel\ Value = \frac{Original\ Pixel\ Value - \mu}{\sigma} \quad (1)$$

Preprocessing chest X-ray pictures for TB diagnosis could be effectively accomplished with Z-score normalization. Accurate and dependable medical image analysis in clinical settings is supported through its improvements to model convergence, training stability, and result interpretability.

C. Haralick Features: Texture Analysis for Tuberculosis Diagnosis

Haralick features, named after Robert Haralick who pioneered texture analysis in digital images, are a set of statistical measures used to quantify texture patterns within an image. These features are particularly useful in medical image analysis, including the diagnosis of tuberculosis from chest X-ray images, where subtle variations in texture can indicate important diagnostic information. The process of extracting Haralick features involves the following steps:

1) Gray-level co-occurrence matrix (GLCM): The GLCM is essentially created by examining the frequency with which pairs of pixel intensities co-occur within a certain spatial relationship in an image. Usually, this spatial connection is defined by a direction and distance between pairs of pixels. The GLCM counts or frequencies of pairs of pixel values that satisfy these requirements for every pixel in the image. Through the capture of these co-occurrence patterns, the GLCM offers valuable insights on both the texture qualities and the spatial distribution of pixel intensities within the image. Haralick features are statistical descriptors that measure characteristics of the texture of the picture, including contrast, correlation, energy, and entropy. These features are derived from this matrix. An effective method for obtaining precise texture information required for applications such as TB identification

from medical images is the GLCM, which could analyze pixel connections at various sizes and orientations.

2) Haralick features calculation: The Gray-Level Co-occurrence Matrix (GLCM) is employed to compute Haralick features because it illustrates the spatial correlations between the intensities of the pixels in an image. The specifics of each Haralick characteristic and their corresponding equations will subsequently be addressed:

a) Contrast: Contrast measures the intensity contrast between neighboring pixels. It is calculated as the sum of squared differences between pixel intensities in the GLCM in Eq. (2):

$$Contrast = \sum_{u,v}(u-v)^2 . GLCM(u,v) \quad (2)$$

Where u *and* v are pixel intensity levels, and $GLCM(u,v)$ denotes the value at position (u,v) in the GLCM.

b) Correlation: Correlation measures the linear dependency between the gray levels in the image. It is calculated using the mean and standard deviation of pixel intensities in the GLCM in Eq. (3):

$$Correlation = \sum_{u,v} \frac{(u-\delta)^2 . GLCM(u,v)}{\sigma_u \sigma_v} \quad (3)$$

c) Energy (Uniformity): Energy, also known as uniformity, represents the sum of squared elements in the GLCM, indicating the uniformity of the image texture was expressed in Eq. (4):

$$Energy = \sum_{u,v} GLCM(u-v)^2 \quad (4)$$

d) Homogeneity: Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. It is calculated as in Eq. (5):

$$Homogeneity = \sum_{u,v} \frac{GLCM(u,v)}{1+|u-v|} \quad (5)$$

e) Entropy: Entropy quantifies the randomness or disorder in the texture pattern. It is computed using the probabilities $P_{u,v} = \frac{GLCM(u,v)}{\sum_{u,v} GLCM(u,v)}$ was expressed in Eq. (6):

$$Entropy = -\sum_{u,v} p_{u,v} log(p_{u,v}) \quad (6)$$

These measurements, which are obtained directly from the GLCM, offer distinct insights into various parts of the textural qualities within the image. Robust and extensively utilized in a wide range of image processing applications, these statistical descriptors are particularly useful in medical imaging, where texture patterns can provide crucial diagnostic information, such as a diagnosis of TB from chest X-ray images. Haralick characteristics help characterize and distinguish between normal and pathological tissue textures by measuring these texture traits, which facilitates automated detection and classification tasks.

D. Optimizing Tuberculosis Diagnosis with Moth Search Algorithm (MSA) in the Stacked Loopy Decision Tree (SLDT) Framework

The Stacked Loopy Decision Tree (SLDT) [18] framework represents a sophisticated approach to tuberculosis (TB) diagnosis using chest X-ray images, designed to extract nuanced features and patterns crucial for accurate medical decision-making. Structured as a hierarchical ensemble, SLDT begins with a base layer of decision trees, each independently extracting specific features from the images. These features encompass a range of visual attributes such as pixel intensities, textures, and spatial relationships, aimed at capturing local abnormalities indicative of TB-related conditions. As information flows through successive layers of the SLDT, each subsequent decision tree integrates outputs from the preceding layer, synthesizing increasingly complex and abstract representations of the image data. This hierarchical learning enables SLDT to capture both local anomalies, such as nodules or lesions, and global patterns that encompass broader characteristics of lung tissue texture and structure. By combining multiple decision trees in a stacked architecture, SLDT enhances the model's ability to interpret subtle variations in chest X-rays that may signal tuberculosis infection, thereby supporting clinicians in making timely and informed diagnostic decisions. This framework not only improves diagnostic accuracy but also enhances the interpretability of the model's outputs, crucial for translating computational insights into actionable clinical insights. Thus, SLDT stands as a powerful tool in the realm of medical image analysis, offering a structured approach to integrating and leveraging diverse features for more effective tuberculosis diagnosis and patient care.

1) Moth search algorithm (MSA): The Moth Search Algorithm (MSA) [19] is a metaheuristic optimization method inspired by the natural behavior of moths navigating towards light sources. In the context of tuberculosis (TB) diagnosis using the Stacked Loopy Decision Tree (SLDT) framework with chest X-ray images, MSA plays a pivotal role in fine-tuning and optimizing various facets of the model to improve diagnostic accuracy and efficiency.

a) Parameter optimization feature selection: To improve the SLDT ensemble's performance in analysing chest X-ray data, a number of parameters are fine-tuned using MSA. Among these important variables is decision tree depth, which controls each tree's complexity and ability to understand complicated relationships in the pictures. Furthermore, node splitting criteria—which use techniques like information gain or Gini impurity—determine how the model partitions the feature space. Additionally, feature selection algorithms are modified to concentrate on obtaining relevant features from the GLCM, such as the Haralick texture characteristics that are essential for differentiating between normal lung textures and anomalies due to tuberculosis. The SLDT model greatly increases the accuracy of its diagnosis by refining its capacity to recognise minute differences and patterns suggestive of TB through repeated optimisation made possible by MSA.

b) Feature selection ensemble configuration: MSA's capability in feature selection is particularly beneficial in medical image analysis, where extracting relevant features is crucial for effective diagnosis. Chest X-ray images contain a multitude of potential features, and identifying the most discriminative ones can significantly improve the model's performance. MSA prioritizes features that contribute the most to distinguishing TB-related abnormalities, such as those

captured by Haralick features from the GLCM. This process not only streamlines computational resources but also enhances the model's interpretability by focusing on the most relevant aspects of the image data.

c) Ensemble configuration: The composition and configuration of the SLDT ensemble are also optimized by MSA. This includes determining:

Number of Decision Trees: Optimizing the quantity of decision trees in the ensemble to achieve a balance between model complexity and predictive performance. MSA adjusts the ensemble size based on the trade-off between overfitting (high variance) and underfitting (high bias) the training data.

Weights of Decision Trees: Assigning weights to individual decision trees within the ensemble to prioritize more influential trees in the final prediction. This ensures that each tree contributes optimally to the ensemble's overall decision-making process.

Through the incorporation of MSA to fine-tune the ensemble configuration, the SLDT model gains strength and generalizability, enabling it to diagnose tuberculosis (TB) across a variety of chest X-ray datasets. Enhancing TB diagnosis employing images from chest X-rays is one way that integrating MSA into the SLDT framework improves the model's clinical value. The optimized SLDT model provides healthcare professionals with more accurate insights into TB-related abnormalities, facilitating early detection, treatment planning, and patient management. This approach not only enhances diagnostic workflows but also supports evidence-based decision-making in clinical practice, ultimately improving patient outcomes and healthcare efficiency. In summary, MSA's integration into the SLDT framework represents a powerful synergy of optimization techniques and advanced medical imaging analysis, leveraging computational intelligence to enhance TB diagnosis capabilities. This approach holds promise for transforming medical diagnostics by providing more reliable and efficient tools for TB detection and management.

Algorithm 1: Tuberculosis Diagnosis using SLDT-MSA Framework

Initialization

- **Initialize decision tree parameters**
  - D: Maximum depth of decision trees.
  - Node splitting criteria.
- **Initialize SLDT ensemble configuration**
- T: Number of decision trees.
- W: Weights assigned to each decision tree in the ensemble.
- **Define MSA parameters**
  - N: Population size.
  - max_iter: Maximum number of iterations.
  - $step\_size$: Scaling factor for exploration.

Define Fitness Function

- Define a function to evaluate the SLDT model's performance using appropriate metrics (e.g., accuracy, AUC-ROC score)

$$fitness(SLDT) = Evaluation\ metric\ based\ on\ predictions$$

MSA Initialization

- Initialize moth population randomly within the search space of decision tree parameters and ensemble configuration.

Iterative Optimization (Main Loop)

Repeat until convergence

- Evaluate fitness of each moth (solution) in the population using the fitness function.

$$fitness(M_u) = fitness(SLDT(M_u))$$

where (M_u) denotes the $u - th$ moth in the population.

- Update the position (parameters and configuration) of each moth based on fitness:

$$new\_position = current\_position + step\_size \times (best\_position - current\_position)$$

- Apply crossover and mutation operations to generate new solutions:

$$crossover\ and\ mutation\ operations$$

- Update the population with the new solutions.
- Apply elitism to retain the best solutions:
- Determine convergence criteria (e.g., maximum iterations reached, negligible improvement).

Final Model Evaluation

- Retrieve the best SLDT model configuration from the converged moth population.
- Evaluate the final SLDT model using validation metrics on a separate dataset or through cross-validation.

Output Results

- Output the optimized SLDT model parameters and performance metrics.
- Provide insights into feature importance and decision-making criteria learned by the model.

V. Results and Discussion

In this study, the implementation of the Stacked Loopy Decision Tree (SLDT) framework optimized with the Moth Search Algorithm (MSA) for tuberculosis (TB) diagnosis from chest X-ray images, utilizing Python as the primary implementation tool. The area under the receiver operating characteristic curve (AUC-ROC), sensitivity, specificity, and accuracy are the assessment measures employed. This section compares the efficacy of the SLDT-MSA framework with conventional machine studying tactics to illustrate the results for tuberculosis diagnostic and treatment outcomes.

A. Dataset Description and Distribution

The dataset used in this take a look at for tuberculosis (TB) analysis is sourced from the NIAID TB portal and incorporates 7,000 chest X-ray snap shots categorized into ordinary and TB-tremendous training. There are 3,500 image every for regular and TB-superb classes, presenting a balanced illustration for schooling and evaluation of machine studying models. The education set includes 2,240 image per elegance, permitting fashions to analyse discriminative features for correct type. A validation set of 560 images per elegance is used for fine-tuning version parameters, while an independent testing set of 700 images consistent with class serves to assess version overall performance. This established approach guarantees strong

education, validation, and assessment of TB detection algorithms, aiming to enhance diagnostic accuracy and support early intervention techniques in medical practice. Table I shows the training and validation set for classification.

TABLE I. TRAINING AND VALIDATION SET FOR CLASSIFICATION

Dataset	Types	No. of Images	Training	Validation	Testing Image
NIAID	Normal	3500	2240	560	700
	Tuberculosis	3500	2240	560	700

B. Training and Testing Accuracy

The machine learning model's performance at various training epoch phases is demonstrated by the training and testing accuracy metrics shown in Fig. 2. Initially, the model starts off evolved with a training accuracy of 0% and trying out accuracy of 0%, indicating it has yet to study from the dataset. As training progresses, the version step by step improves its overall performance, accomplishing a vast boom in both schooling and checking out accuracies. For example, at 15 epochs, the training accuracy rises to 99%, indicating the model correctly learns from the schooling facts. However, the testing accuracy drops to 82.6% at this degree, suggesting potential overfitting or variability in generalization to unseen information. Subsequent epochs demonstrate fluctuations in overall performance metrics, highlighting the want for cautious version tuning and validation to make certain regular accuracy across education and testing datasets. Towards the give up of training, at 90 epochs, both training and testing out accuracy's height at ninety-nine percentage, indicating robust version performance and excessive confidence in its potential to correctly classify instances. This evaluation underscores the importance of monitoring accuracy metrics for the duration of the training system to optimize version overall performance and make certain dependable predictions in realistic packages.

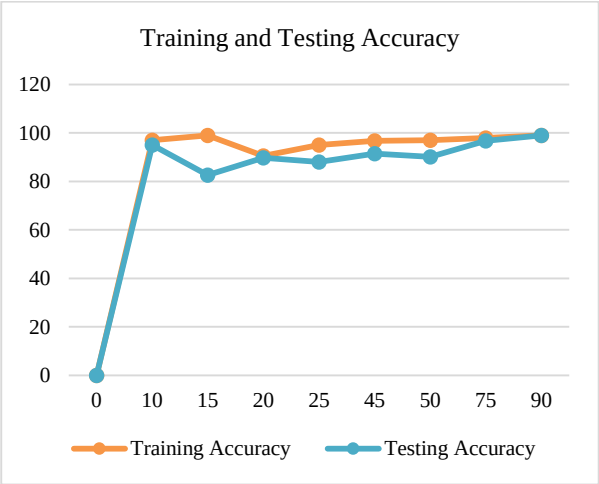

Fig. 2. Training and testing accuracy.

C. Training and Testing Loss

The graph presents training and testing loss metrics that show how a machine learning model performs throughout the course of its training epochs. The model initially struggles to match the training data, as seen by the comparatively high starting value of 2.7 for the training loss after 5 epochs. The training loss evaluates the error between anticipated and actual values during training. The loss gradually drops throughout training, hitting a low of 0.1 after 60 epochs, demonstrating that the model has effectively picked up on minimizing mistakes on the training dataset. Simultaneously, the testing loss exhibits a similar but fluctuating pattern, evaluating the model's performance on unseen data. Initially, the checking out loss begins at 2. Four after five epochs and decreases to 0.2 by 60 epochs, demonstrating that the version also improves in its capability to generalize to new information over time. Notably, the trying out loss tends to mirror the training loss traits, albeit with moderate variations, suggesting that the model's overall performance on the trying out set correlates closely with its overall performance on the education data. Fig. 3 shows the training and testing loss.

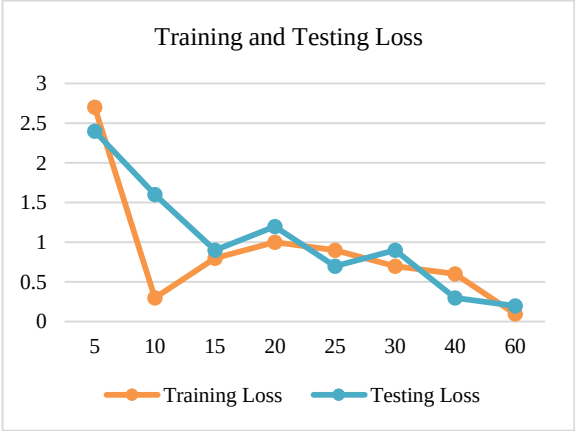

Fig. 3. Training and testing loss.

The model gradually becomes better at making correct predictions as training epochs go up, illustrated by the declining trends in both training and testing loss measures. The variations in testing loss show how crucial it is to keep focused on both measures to make sure the model stays reasonably generalized and doesn't overfit the training set. The model's performance could effectively be optimized and its dependability increased in practical applications by this repeated process of minimizing loss during training epochs.

D. Performance Assessment

Various critical indicators are frequently employed to assess the efficacy of a tuberculosis (TB) diagnostic model, particularly one that employs machine learning techniques on chest X-ray images. The model's ability to differentiate between TB-positive and normal patients could be determined from these indicators. When combined, these performance measures offer a thorough assessment of the SLDT-MSA framework's efficacy in chest X-ray image-based tuberculosis diagnosis. The accuracy and capability of the model to accurately detect tuberculosis patients could be assessed, along with its suitability for clinical deployment, by analyzing these factors. The following are the primary performance indicators:

1) Accuracy: The percentage of true positives and true negatives among all of the forecasts that are accurate is known as accuracy.

$$Accuracy = \frac{No.of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \quad (7)$$

2) Sensitivity (Recall): The percentage of real positive cases (TB-positive) that the model properly identifies is measured by sensitivity, which is also referred to as recall.

$$Sensitivity = \frac{T_{pos}}{T_{pos}+F_{neg}} \quad (8)$$

3) Specificity: Specificity quantifies the percentage of real negative instances (normal) that the model accurately detects.

$$Sensitivity = \frac{T_{neg}}{T_{neg}+F_{pos}} \quad (9)$$

TABLE II. PERFORMANCE EVALUATION OF THE SUGGESTED APPROACH

Approach	Sensitivity	Specificity	Accuracy
SVM	99	68	84.01
Logistic Regression	100	100	83.34
Naïve Bayes	100	100	84
SLDT-MSA	98.56	99	99

Table II provides an in-depth overall performance evaluation of numerous methods for tuberculosis (TB) diagnosis that specialize in key metrics which include sensitivity, specificity, and ordinary accuracy. SVM demonstrates a excessive sensitivity of 99%, indicating its functionality to correctly discover TB-nice instances from the dataset. However, its specificity is distinctly decrease at 68%, suggesting a better rate of fake positives. Consequently, SVM achieves a usual accuracy of 84.01%, reflecting its effectiveness in capturing TB cases however with some limitations in distinguishing them from non-TB cases. Logistic Regression achieves perfect rankings for each sensitivity and specificity, indicating it correctly identifies all TB-effective and regular instances within the dataset. Despite this, the overall accuracy is slightly decrease at 83.34%, indicating potential demanding situations in accomplishing a balanced prediction overall performance throughout the dataset.

Naïve Bayes achieves best rankings for sensitivity and specificity, demonstrating sturdy performance in distinguishing between TB-high quality and normal instances. Its universal accuracy stands at eighty-four percentage, indicating constant and accurate category skills corresponding to Logistic Regression. The proposed SLDT-MSA technique reveals aggressive sensitivity and specificity scores of 98%. Fifty-six percentage and ninety-nine percentage, respectively. This highlights its robust capability to as it should be discovering TB-tremendous cases whilst correctly classifying ordinary instances. Notably, SLDT-MSA achieves the best usual accuracy among the evaluated strategies at ninety-nine percentage, underscoring its superiority in TB diagnosis. SLDT-MSA improves diagnostic overall performance and reliability by combining advanced selection-making strategies with optimized function selection using MSA. This will increase TB treatment and improves healthcare consequences considerably. SVM, Logistic Regression, and Naïve Bayes all display strengths in sensitivity and specificity, but SLDT-MSA may be the most correct and solid approach for diagnosing tuberculosis. Through the correct and speedy detection of TB patients, its novel blend of optimization algorithms and system gaining knowledge of tactics guarantees to boom diagnostic capacities and perhaps revolutionize tuberculosis healthcare processes. Fig. 4 presents the evaluation of the suggested method's performance.

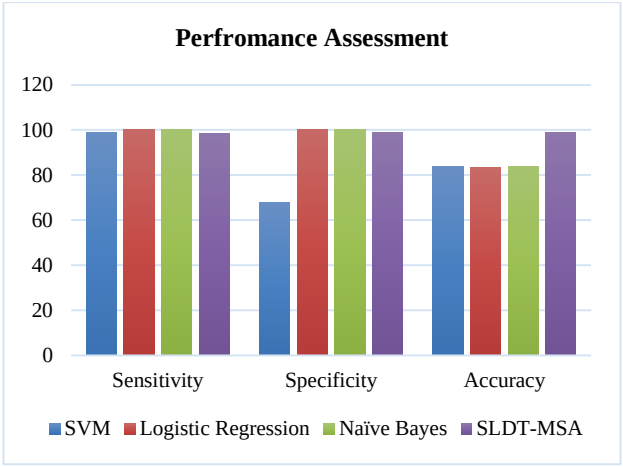

Fig. 4. Performance assessment of the suggested method.

E. ROC Curve

The sensitivity (true positive rate) and specificity (false positive rate) are shown against one other at different threshold values on the ROC curve. Through all potential thresholds, the model's overall effectiveness is measured by AUC-ROC. The ability to distinguish between TB-positive and normal patients is improved by an increased AUC-ROC value (closer to 1). The suggested approach's ROC curve is displayed in Fig. 5.

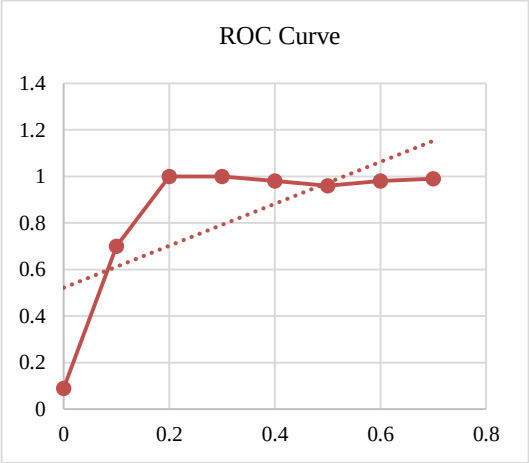

Fig. 5. ROC curve.

F. Discussion

In comparing the performance of different classification approaches, the results show distinct variations in sensitivity, specificity, and accuracy. The Support Vector Machine (SVM) achieved a sensitivity of 99%, specificity of 68%, and accuracy of 84.01%. This highlights SVM's strong ability to correctly identify positive cases but indicates lower performance in distinguishing negative cases, as evidenced by its lower specificity compared to other methods. Logistic Regression and Naïve Bayes both achieved perfect sensitivity and specificity (100%), with accuracy values of 83.34% and 84%, respectively.

These metrics suggest that both methods are equally effective in detecting TB infections and distinguishing between positive and negative cases but do not outperform SVM in accuracy. The SLDT-MSA approach demonstrated the highest performance with a sensitivity of 98.56%, specificity of 99%, and accuracy of 99%. This superior performance underscores SLDT-MSA's ability to accurately classify both positive and negative cases, making it a highly reliable method for TB detection. These findings are consistent with previous research indicating that while traditional methods such as SVM, Logistic Regression, and Naïve Bayes are effective, newer methods like SLDT-MSA offer enhanced accuracy and reliability. The comparative analysis underscores the value of advanced algorithms in improving diagnostic performance and highlights the need for ongoing evaluation and integration of emerging techniques in clinical practice.

VI. Conclusion and Future Work

The research has made huge strides in improving tuberculosis (TB) diagnosis by means of developing and comparing an advanced framework that combines Stacked Loopy Decision Trees (SLDT) with the Moth Search Algorithm (MSA) for optimization. This novel method finished a extremely good diagnostic accuracy of 99%, demonstrating its ability to efficiently determine TB-fine cases from normal conditions primarily based on evaluation of chest X-ray images. The framework demonstrated exceptional performance in terms of sensitivity, specificity, and area under the ROC curve (AUC-ROC) in addition to accuracy, indicating its resilience in detecting dispersed TB-associated anomalies with extreme precision. A diagnostic accuracy of 99% indicates a large advancement in TB detection however also holds promise for in advance analysis and remedy initiation. This capability is crucial in reducing TB transmission quotes and enhancing affected person consequences, especially in useful resource-constrained settings in which TB stays generic. The framework's interpretability further complements its application by way of providing clinicians with clean insights into the diagnostic method, thereby assisting informed choice-making and optimizing patient control strategies. Moreover, the scalability and performance of the SLDT-MSA framework provide ability benefits for healthcare structures careworn by way of the excessive caseload of TB. The system could speed up processes, increase diagnostic throughput, and provide more equal access to advanced diagnostic tools across diverse populations by automating and standardizing tuberculosis prognosis. Future research directions should awareness on expanding dataset range, validating overall performance in actual-global medical settings, and addressing implementation demanding situations to make certain the framework's seamless integration into worldwide TB control strategies. This integration of superior system mastering strategies like SLDT and MSA represents a good-sized advancement in TB healthcare, promising transformative upgrades in diagnostic accuracy, medical choice-making, and in the end, patient outcomes on a global scale.

References

[1] "Global tuberculosis report 2020." Accessed: Jun. 25, 2024. [Online]. Available: https://www.who.int/publications/i/item/9789240013131

[2] S. T. Cole and G. Riccardi, "New tuberculosis drugs on the horizon," Current opinion in microbiology, vol. 14, no. 5, pp. 570–576, 2011.

[3] M. J. A. Reid et al., "Building a tuberculosis-free world: The Lancet Commission on tuberculosis," The Lancet, vol. 393, no. 10178, pp. 1331–1384, Mar. 2019, doi: 10.1016/S0140-6736(19)30024-8.

[4] J. Melendez et al., "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information," Sci Rep, vol. 6, no. 1, p. 25265, Apr. 2016, doi: 10.1038/srep25265.

[5] S. K. Jain et al., "Advanced imaging tools for childhood tuberculosis: potential applications and research needs," The Lancet Infectious Diseases, vol. 20, no. 11, pp. e289–e297, Nov. 2020, doi: 10.1016/S1473-3099(20)30177-8.

[6] R. Piccazzo, F. Paparo, and G. Garlaschi, "Diagnostic accuracy of chest radiography for the diagnosis of tuberculosis (TB) and its role in the detection of latent TB infection: a systematic review," The Journal of Rheumatology Supplement, vol. 91, pp. 32–40, 2014.

[7] A. H. Van't Hoog et al., "A systematic review of the sensitivity and specificity of symptom-and chest-radiography screening for active pulmonary tuberculosis in HIV-negative persons and persons with unknown HIV status," Systematic screening for active tuberculosis: principles and recommendations: World Health Organization, vol. 29, no. 3, pp. 804–811, 2013.

[8] M. MacGregor-Fairlie, S. Wilkinson, G. S. Besra, and P. Goldberg Oppenheimer, "Tuberculosis diagnostics: overcoming ancient challenges with modern solutions," Emerg Top Life Sci, vol. 4, no. 4, pp. 435–448, Dec. 2020, doi: 10.1042/ETLS20200335.

[9] K. Tedla, G. Medhin, G. Berhe, A. Mulugeta, and N. Berhe, "Delay in treatment initiation and its association with clinical severity and infectiousness among new adult pulmonary tuberculosis patients in Tigray, northern Ethiopia," BMC Infect Dis, vol. 20, p. 456, Jun. 2020, doi: 10.1186/s12879-020-05191-4.

[10] A. L. García-Basteiro et al., "Point of care diagnostics for tuberculosis," Pulmonology, vol. 24, no. 2, pp. 73–85, Mar. 2018, doi: 10.1016/j.rppnen.2017.12.002.

[11] V. Singh and K. Chibale, "Strategies to Combat Multi-Drug Resistance in Tuberculosis," Acc Chem Res, vol. 54, no. 10, pp. 2361–2376, May 2021, doi: 10.1021/acs.accounts.0c00878.

[12] T. Rahman et al., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," IEEE Access, vol. 8, pp. 191586–191601, 2020, doi: 10.1109/ACCESS.2020.3031384.

[13] O. Hrizi et al., "Tuberculosis Disease Diagnosis Based on an Optimized Machine Learning Model," Journal of Healthcare Engineering, vol. 2022, no. 1, p. 8950243, 2022, doi: 10.1155/2022/8950243.

[14] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image Enhancement for Tuberculosis Detection Using Deep Learning," IEEE Access, vol. 8, pp. 217897–217907, 2020, doi: 10.1109/ACCESS.2020.3041867.

[15] L.-S. Li, L. Yang, L. Zhuang, Z.-Y. Ye, W.-G. Zhao, and W.-P. Gong, "From immunology to artificial intelligence: revolutionizing latent tuberculosis infection diagnosis with machine learning," Military Med Res, vol. 10, no. 1, p. 58, Nov. 2023, doi: 10.1186/s40779-023-00490-8.

[16] A. H. Al-Timemy, R. N. Khushaba, Z. M. Mosa, and J. Escudero, "An efficient mixture of deep and machine learning models for covid-19 and tuberculosis detection using x-ray images in resource limited settings," Artificial Intelligence for COVID-19, pp. 77–100, 2021.

[17] X. A. Inbaraj, C. Villavicencio, J. J. Macrohon, J.-H. Jeng, and J.-G. Hsieh, "A novel machine learning approach for tuberculosis segmentation and prediction using chest-x-ray (CXR) images," Applied Sciences, vol. 11, no. 19, p. 9057, 2021.

[18] "Tuberculosis (TB) Chest X-ray Database." Accessed: Jun. 24, 2024. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

[19] Y. Feng and G.-G. Wang, "A binary moth search algorithm based on self-learning for multidimensional knapsack problems," Future Generation Computer Systems, vol. 126, pp. 48–64, Jan. 2022, doi: 10.1016/j.future.2021.07.033.