FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 7, July, 2025, pp 226 – 234
DOI: https://doi.org/10.33003/fjs-2025-0907-3067

MITIGATING CLASS IMBALANCE IN TUBERCULOSIS DETECTION: COMBINING SMOTE AND TOMEK LINK WITH MODIFIED FOCAL LOSS AND CLASS WEIGHTING IN A TRANSFER LEARNING FRAMEWORK

[1]Sulaiman Hashim Ibrahim, [2]Ishola D. Muraino, [3]Sabiu Danlami and *[4]Olalekan Joel Awujoola

[1]Department of Information and Communication Technology, Federal University of Technology Minna
[2]Department of Computer Science, Yusuf Maitama Sule University, Kano
[3]Department of Computer Science, Airforce Institute of Technology, Kaduna
[4]Department of Computer Science, Nigerian Defence Academy, Kaduna

*Corresponding authors' email: ojawujoola@nda.edu.ng

ABSTRACT

Tuberculosis (TB) remains a major global health concern, and early detection is crucial for effective treatment. This study addresses the challenge of class imbalance in existing machine learning models for TB prediction, which often leads to biased results. A novel approach combining the hybrid SMOTE-Tomek Links technique, modified focal loss, and class weighting was developed and applied to a dataset of X-ray images categorized into normal and TB classes. The hybrid SMOTE-Tomek Links method generates synthetic samples for the minority class while removing ambiguous samples, ensuring a balanced dataset. The modified focal loss and class weighting help focus on misclassified cases and address class disparities. The model was evaluated against benchmark models, including EfficientNetB3, Random Forest, and XGBoost, with and without SMOTE. The developed model achieved a remarkable accuracy of 99.7%, outperforming the benchmark models (92.72%–99.1%). These results demonstrate the effectiveness of the proposed approach in improving TB prediction accuracy and handling class imbalance. The study's findings provide valuable insights into medical image classification and offer a robust framework for enhancing diagnostic tools, with potential applications beyond TB detection. This research could significantly improve TB management and diagnosis in clinical settings.

Keywords: Machine Learning, SMOTE, Tomek Link, Tuberculosis

INTRODUCTION

Tuberculosis (TB), caused by Mycobacterium tuberculosis, persists as a formidable global public health challenge due to its highly infectious nature and capacity to precipitate widespread outbreaks. Early detection, particularly in clustered cases, proves essential for timely intervention and containment of transmission, as emphasized by Fox et al. (2021). Transmitted through airborne droplets, TB demands substantial public health resources, with prevention of outbreaks proving far more cost-effective than managing established ones, especially in resource-limited settings where financial burdens exacerbate morbidity and mortality (Mitruka et al., 2011; Althomsons et al., 2022). Recent studies underscore the influence of environmental and socio-economic factors, such as meteorological changes and air pollution, on TB transmission dynamics, further complicating control efforts (Tang et al., 2023; Wang et al., 2021). Despite medical advancements, TB remains the second deadliest infectious disease globally, claiming millions of lives annually, with treatment failure contributing significantly to high mortality rates and necessitating advanced diagnostic tools (Makam and Matsa, 2021). Machine learning has emerged as a transformative approach to enhance TB prediction and inform treatment decisions, offering potential to address these challenges (Ahmed et al., 2020; Chekroud et al., 2021). However, dataset imbalance in medical imaging, particularly in chest X-ray (CXR) analysis, where TB-positive cases are underrepresented, poses a significant obstacle to achieving high model accuracy and sensitivity (Kieu et al., 2020).

This work builds upon the methodology and techniques proposed by Nafisah and Mohammad (2024), who developed an automatic TB detection system for chest X-rays using deep learning and segmentation techniques, achieving an impressive 99.1% accuracy with EfficientNetB3, highlighting the critical role of segmentation in improving detection accuracy. However, certain limitations in their approach warrant attention to further enhance performance. The benchmark study does not explicitly address class imbalance, a prevalent issue in medical datasets where TB-positive cases are significantly fewer than TB-negative ones. Without targeted techniques to mitigate this imbalance, models risk bias toward the majority class (TB-negative), resulting in reduced sensitivity for detecting TB-positive cases, which is paramount for early diagnosis. To overcome this limitation, a hybrid approach is adopted, integrating Synthetic Minority Over-sampling Technique (SMOTE) with Tomek Links, alongside class weighting and Modified Focal Loss (MFE). SMOTE generates synthetic TB-positive samples to balance the dataset, while Tomek Links eliminates noisy or ambiguous samples to ensure data quality. Class weighting adjusts the loss function to prioritize the minority class (TB-positive), and MFE enhances focus on challenging examples, collectively improving sensitivity to TB-positive cases compared to the benchmark's approach, which lacks these mechanisms.

Furthermore, the benchmark study relies predominantly on EfficientNetB3 for feature extraction, achieving high accuracy but without incorporating multiple architectures to bolster robustness. Dependence on a single model may constrain the diversity of extracted features, potentially overlooking complementary patterns critical for complex medical images like chest X-rays. To address this, two pre-trained convolutional neural networks (CNNs), EfficientNetB0 and DenseNet121, are employed for feature extraction. EfficientNetB0 offers computational efficiency, while DenseNet121's densely connected layers facilitate enhanced feature propagation by allowing each layer to receive inputs from all preceding layers. By combining features from both architectures, a broader and more

comprehensive feature set is captured from chest X-ray images, improving robustness and generalization compared to the benchmark's single-model approach.

This methodology leverages transfer learning frameworks with EfficientNetB0 and DenseNet121 for feature extraction, followed by a Random Forest classifier for robust prediction. Data augmentation and preprocessing, facilitated by ImageDataGenerator, ensure consistent input, while comprehensive evaluation using metrics such as the confusion matrix and Receiver Operating Characteristic (ROC) curve verifies reliable performance. By addressing the identified limitations in the benchmark study, this approach advances machine learning-based diagnostic tools, contributing to more effective early detection strategies for TB, particularly in resource-constrained settings.

Modified Focal Loss

Modified Focal Loss extends the standard Focal Loss to better handle class imbalance in machine learning tasks, particularly in medical image classification and object detection (Yeung et al., 2022; Chen et al., 2022). It is effective in datasets with a small number of positive samples relative to a large number of negative samples, preventing model bias toward the majority class.

Focal Loss enhances Cross-Entropy loss by emphasizing hard-to-classify examples and down-weighting easy ones (Xi et al., 2023; Guo, 2024). It is mathematically defined as:

The standard Focal Loss function is represented as:

$$FL(pt) = -\alpha_t(1-pt)^{\gamma}\log(pt) \quad (1)$$

Where:

pt is the predicted probability for the correct class.

α_t is a weighting factor that adjusts the importance of positive versus negative samples.

γ is the focusing parameter that controls the modulating factor. When γ = 0, Focal Loss reduces to Cross-Entropy Loss, and when γ > 0, it emphasizes harder examples.

Modified Focal Loss enhances this by incorporating class weighting strategies, improving model focus on minority class instances, such as TB-positive or cancerous cases, which are often harder to classify in medical datasets. The modified function is:

$$MFL(pt) = -w_c \,.\, \alpha_t(1-pt)^{\gamma}\log(pt) \quad (2)$$

Where:

w_c is the class weight, introduced to give more importance to the minority class.

Other terms (pt , α_t, and γ) remain as in the original Focal Loss function.

This adaptation improves model sensitivity to rare cases, reducing false negatives, which is crucial in healthcare to prevent undetected diseases. For tuberculosis detection, Modified Focal Loss enhances deep learning models' ability to correctly identify TB-positive cases despite their rarity. When combined with oversampling techniques like SMOTE and Tomek Links in a transfer learning framework, this approach strengthens model performance, leading to more reliable diagnostic predictions and improved healthcare outcomes.

Tomek Link

Tomek Links refine oversampling techniques like SMOTE by addressing class overlap in imbalanced datasets, particularly in medical diagnostics (Alamri & Ykhlef, 2024; Leng et al., 2024). While oversampling generates synthetic samples to balance classes, it may introduce noise or place samples near the decision boundary, leading to misclassification (Rao et al., 2024; Wen et al., 2024; Xie et al., 2023). Tomek Links help mitigate this by identifying and removing ambiguous pairs of samples from different classes that are each other's nearest neighbors (Swana et al., 2022).

A Tomek Link exists when a minority class instance (e.g., a disease-positive case) and a majority class instance (e.g., a healthy case) are closer to each other than to any other sample in their respective classes. These pairs typically lie near the decision boundary and may introduce classification errors. Removing Tomek Links sharpens class separation, reduces noise, and enhances model performance (Viadinugroho, 2023; Przybyła-Kasperek, 2022; Ai-jun & Peng, 2020).

For example, in tuberculosis (TB) detection datasets, a Tomek Link might form between a TB-positive and TB-negative sample due to their proximity in feature space. Removing such pairs after applying SMOTE improves dataset clarity, reducing misclassification risks.

Tomek Links are commonly used as a post-processing step following oversampling to refine synthetic data, remove outliers, and enhance model robustness. This is particularly crucial in medical applications where noisy data can lead to incorrect diagnoses and adverse patient outcomes. When combined with SMOTE and methods like Modified Focal Loss, Tomek Links create a more effective framework for handling class imbalance, improving the detection of minority cases, and enhancing diagnostic accuracy

SMOTE (Synthetic Minority Oversampling Technique)

SMOTE is a widely used technique for addressing class imbalance in medical datasets, particularly in diagnostic applications such as disease detection (Awujoola et al., 2020). Imbalanced datasets, where the number of positive cases is significantly lower than negative cases, can bias machine learning models toward the majority class, leading to poor sensitivity in detecting diseases (Awujoola et al., 2021). Traditional oversampling methods that duplicate minority class samples often result in overfitting, reducing model generalization.

SMOTE mitigates this issue by generating synthetic data points through interpolation between existing minority class samples and their k-nearest neighbors. For example, in a TB detection dataset, where TB-positive chest X-ray images are underrepresented, SMOTE enhances minority class distribution without mere duplication. This approach helps models generalize better, improving sensitivity to rare but crucial cases such as early-stage diseases.

To mathematically represent how SMOTE generates synthetic instances, consider a minority class data point x_i in the feature space, and its k-nearest neighbors $x_{i1}, x_{i2}, \ldots, x_{ik}$. SMOTE creates a synthetic data point x_{new} by interpolating between x_i and one of its nearest neighbors, say x_{ij} using the following equation:

$$x_{new} = x_i + \lambda \,.\, (x_{ij} - x_i) \quad (3)$$

where:

x_i is the original minority class instance,

x_{ij} is one of its k-nearest neighbors,

λ is a random number between 0 and 1, which controls the interpolation between x_i and x_{ij}

This equation ensures that the synthetic instance x_{new} lies along the line segment between x_i and x_{ij} in the feature space. By repeating this process for different neighbors and original instances, SMOTE generates diverse synthetic samples, helping to better represent the minority class and mitigate overfitting in imbalanced datasets.

Review of Related Works

Viswanatha et al. (2023) developed a machine learning model using K-Nearest Neighbour (K-NN) and Histogram of

Oriented Gradients (HOG) for the early detection of tuberculosis (TB). Their model, trained on TB-related symptoms and demographic data, achieved 98% accuracy, providing a more precise detection method than manual diagnosis, potentially reducing TB-related deaths.

Rodrigues et al. (2024) focused on predicting loss to follow-up (LTFU) during anti-tuberculosis treatment (ATT) using national registry data. They tested three machine learning models—Logistic Regression, Random Forest, and Light Gradient Boosting—finding that the latter achieved the best predictive performance with an AUC of 0.72, aiding healthcare workers in identifying at-risk patients.

Silva et al. (2024) examined TB clusters in Brazil's Amazonian region, utilizing Moran's I and the Getis-Ord GI* method to identify geographical clusters. A Random Forest model trained on six surveillance variables achieved an AUC of 0.81, helping to predict high-incidence areas for targeted TB prevention and control.

Sharma et al. (2024) developed a deep learning model for TB diagnosis from chest X-rays, incorporating UNet for segmentation and Xception for classification. Their model achieved an accuracy of 99.29% and an AUC of 0.999, with Grad-CAM heatmaps offering interpretable insights into TB lesions, improving diagnostic precision.

Kumar et al. (2024) proposed a system using VGG16 and machine learning classifiers to diagnose TB from chest X-rays. The Adaboost classifier achieved the best accuracy at 95.11%, demonstrating the potential for improved TB diagnosis with large datasets and machine learning.

Ou et al. (2024) employed multiple deep learning models, including U-Net and Attention U-Net, to detect and segment TB lesions in chest X-rays. Their ensemble model outperformed individual models, achieving high accuracy and a mean intersection-over-union (MIoU) of 0.70, offering reliable tools for TB lesion detection.

Jonathan et al. (2024) investigated the use of trained TB-detection rats, applying machine learning models like SVM and random forest to predict rat performance. The SVM model achieved 83.39% accuracy, improving the effectiveness of rat-based TB detection by analyzing patterns in detection behavior.

Nafisah and Mohammad (2024) developed an automatic TB detection system for chest X-rays using deep learning and segmentation techniques. EfficientNetB3 achieved 99.1% accuracy, outperforming other models, and highlighting the importance of image segmentation for improving TB detection accuracy.

MATERIALS AND METHODS

The research methodology for this study focuses on the development and evaluation of a deep learning framework aimed at mitigating class imbalance in tuberculosis (TB) detection using chest X-ray images. This approach leverages advanced techniques such as Synthetic Minority Over-sampling Technique (SMOTE), Tomek Links, Modified Focal Loss (MFL), and class weighting within a transfer learning architecture to enhance prediction accuracy, particularly for the minority class of TB-positive cases. The study employs transfer learning models like EfficientNetB0 and DenseNet121 for feature extraction and utilizes Random Forest as the classifier to achieve robust performance in TB detection. Figure 1 depicts the research methodology flow.

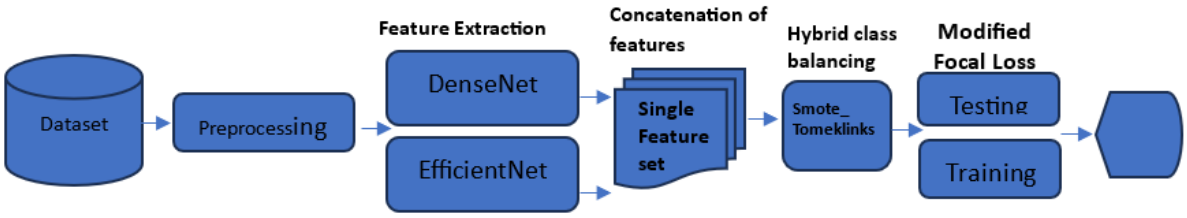

Figure 1: Research methodology Flow

Research Model Description

The model developed for tuberculosis (TB) detection using chest X-ray images incorporates advanced techniques to tackle the inherent challenge of class imbalance, a common issue in medical datasets where the number of TB-positive cases is significantly smaller than the TB-negative cases. To address this, a hybrid approach is employed, combining the Synthetic Minority Over-sampling Technique (SMOTE) with Tomek Links, along with Modified Focal Loss (MFE) and class weighting loss. The purpose of this approach is to enhance the model's performance in detecting minority classes, specifically TB-positive cases. For feature extraction, two deep learning architectures, EfficientNetB0 and DenseNet121, are utilized, and Random Forest is used as the final classifier for prediction. This multi-faceted methodology not only improves the balance in the dataset but also enhances the model's overall accuracy.

The model begins by employing the ImageDataGenerator to handle data augmentation and preprocessing. This function normalizes the pixel values in the chest X-ray images by scaling them between 0 and 1, which is critical for ensuring consistent input to the model during training. Furthermore, ImageDataGenerator efficiently loads and batches the data, preventing memory overload during training and evaluation. Normalizing and batching large datasets help improve the model's convergence and performance.

For feature extraction, two pre-trained convolutional neural networks (CNNs), EfficientNetB0 and DenseNet121, are employed. Both networks are initially trained on the ImageNet dataset, allowing them to learn generalized image features. EfficientNetB0, known for its computational efficiency, is particularly useful in this model due to its ability to balance performance and resource usage. On the other hand, DenseNet121 enhances feature propagation through its densely connected layers, where each layer receives input from all preceding layers. This architecture improves the model's ability to extract deep, relevant features from medical images, which is critical for TB detection. By leveraging these two models, the feature extraction process is both efficient and comprehensive.

The challenge of class imbalance is addressed through a combination of SMOTE and Tomek Links. SMOTE is a widely-used technique for generating synthetic examples of the minority class, thereby balancing the dataset and allowing the model to learn from more instances of TB-positive cases. Tomek Links complement SMOTE by identifying and removing noisy or ambiguous samples, ensuring that the synthetic samples generated are representative and realistic. This process reduces the risk of overfitting, which can occur

when the model learns patterns from noise rather than genuine features.

To further address class imbalance, class weighting is applied during model training. The class weighting technique adjusts the loss function, giving more importance to the minority class, in this case, TB-positive instances. This ensures that the model focuses on hard-to-classify examples and does not become biased toward the majority class. Class weighting works in tandem with SMOTE and Tomek Links to improve the model's sensitivity to TB-positive cases, enhancing its ability to detect early signs of the disease in chest X-ray images.

Following feature extraction, a Random Forest classifier is used for the final prediction stage. Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to form the final output. This method is particularly effective in handling large datasets and offers a robust solution to overfitting. By averaging the predictions of many decision trees, Random Forest enhances the model's ability to generalize well to new, unseen data, thus improving the reliability of the TB detection system.

For evaluating the model's performance, key metrics such as the confusion matrix and Receiver Operating Characteristic (ROC) curve are used. The confusion matrix provides detailed insights into the model's classification performance, indicating the number of true positives, false positives, true negatives, and false negatives. This information is critical for understanding how well the model distinguishes between TB-positive and TB-negative cases. Additionally, the ROC curve visualizes the trade-off between sensitivity and specificity across different classification thresholds, with the area under the curve (AUC) providing a summary measure of the classifier's performance. A high AUC value indicates that the model is proficient in distinguishing between positive and negative cases, reinforcing its utility in medical diagnostics.

The tuberculosis detection model therefore, integrates multiple advanced techniques to overcome the challenges posed by class imbalance and the complexity of medical image processing. By employing EfficientNetB0 and DenseNet121 for feature extraction, and Random Forest for classification, the model is able to achieve high accuracy and robustness. The hybrid approach using SMOTE, Tomek Links, Modified Focal Loss, and class weighting ensures that the model is sensitive to the minority class without overfitting, leading to a more effective diagnostic tool for TB detection. The comprehensive evaluation of the model through metrics like the confusion matrix and ROC curve demonstrates its capability to make accurate and reliable predictions, providing valuable assistance in the early detection of tuberculosis. Figure 3.1 depicts the research methodology.

Dataset Source

In this research, the dataset used for tuberculosis (TB) detection was born out of collaborative team of researchers from Qatar University in Doha, Qatar, and the University of Dhaka in Bangladesh, alongside their partners from Malaysia, has developed an extensive database of chest X-ray images specifically curated for the study of Tuberculosis (TB). This project, carried out in conjunction with medical doctors from Hamad Medical Corporation in Qatar and healthcare professionals from Bangladesh, aims to provide a valuable resource for the advancement of TB detection through medical imaging.

The dataset contains chest X-ray images of both TB-positive cases and healthy, normal cases, offering a robust foundation for the development of diagnostic models using artificial intelligence and machine learning. In its current public release, the database includes 700 readily accessible TB-positive X-ray images. Additionally, an expanded set of 2,800 TB-positive images is available for download from the National Institute of Allergy and Infectious Diseases (NIAID) TB portal. Access to these additional images requires signing a data-sharing agreement to ensure ethical use and data privacy. Complementing this, the dataset also contains 3,500 X-ray images of healthy individuals, providing a balanced set of normal images to facilitate comparative studies and model training.

This database represents a significant step forward in creating open-access medical datasets that can support research efforts in TB detection, enabling the development of more accurate diagnostic tools and improving the global response to TB through early and reliable detection. The dataset can be sourced through the NIAID TB portal program dataset [Online]. Available:https://tbportals.niaid.nih.gov/download-data.

Data Preprocessing

The preprocessing of the Tuberculosis (TB) image dataset is a crucial step in the development of the model, ensuring that the images are properly prepared for training and evaluation. The preprocessing begins with the use of the ImageDataGenerator, a powerful function that handles data augmentation and normalization. One of its primary functions is to normalize the pixel values of the chest X-ray images by scaling them between 0 and 1. This normalization is essential for maintaining consistency in the input data and helps ensure that the deep learning model can interpret the image data effectively. By standardizing pixel values, the model can focus on identifying patterns within the images, improving its ability to detect TB.

In addition to normalization, ImageDataGenerator performs data augmentation, which is a technique used to artificially expand the size of the training dataset by applying random transformations such as rotations, zooms, and flips. This step is particularly valuable in medical imaging, where obtaining large and diverse datasets can be challenging. Augmentation introduces variability into the training data, enhancing the model's robustness and helping it generalize better to unseen data, ultimately improving its performance in real-world scenarios.

Another important function of ImageDataGenerator is its efficient handling of large datasets. The function loads and batches the data in real-time during training, rather than loading the entire dataset into memory at once. This prevents memory overload, especially when dealing with high-resolution medical images, and ensures that the training process remains smooth. By organizing the data into manageable batches, the model can process the images more effectively, leading to faster convergence during training.

The preprocessing of the TB dataset using ImageDataGenerator involves normalization to ensure consistency in pixel values, data augmentation to improve model robustness, and efficient data batching to manage memory usage. These preprocessing steps collectively contribute to enhancing the model's performance and its ability to accurately detect TB from chest X-ray images.

Performance Evaluation

Performance evaluation is a critical component of this study, providing insight into the effectiveness of the developed model in detecting Tuberculosis (TB) from chest X-ray images. The evaluation process involves several metrics and visual tools that help assess how well the model performs in distinguishing between TB-positive and normal cases.

The first step in evaluating the model is the use of a confusion matrix, a tabular representation of the true positives, true negatives, false positives, and false negatives. The confusion matrix is especially useful in understanding the distribution of correct and incorrect predictions. In the context of TB detection, it provides valuable insights into how often the model correctly identifies TB-positive cases (true positives) and how often it mistakenly classifies normal cases as TB-positive or vice versa. By analyzing this matrix, one can assess the model's sensitivity (or recall), which reflects its ability to detect TB cases, and specificity, which measures how well it avoids false positives.

Accuracy
Measures the proportion of correct predictions made by the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Where:

a. TP = True Positives
b. TN = True Negatives
c. FP = False Positives
d. FN = False Negatives

Precision (or Positive Predictive Value)
Measures how many of the predicted positive cases are actually positive.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Recall (or Sensitivity, True Positive Rate)
Measures how many actual positive cases the model correctly identified.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Specificity (or True Negative Rate)
Measures how well the model identifies true negatives.

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

F1-Score
The harmonic means of precision and recall, used when there is an uneven class distribution (imbalanced data).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

False Positive Rate (FPR)
Measures the proportion of negative instances that were incorrectly classified as positive.

$$FPR = \frac{FP}{FP + TN} \quad (9)$$

False Negative Rate (FNR)
Measures the proportion of positive instances that were incorrectly classified as negative.

$$FNR = \frac{FN}{FN + TP} \quad (10)$$

The model's ability to distinguish between TB-positive and normal cases is further evaluated using the Receiver Operating Characteristic (ROC) curve. The ROC curve plots the true positive rate against the false positive rate at various threshold settings. The area under the ROC curve (AUC) is a key metric, where a higher AUC indicates a better model. A high AUC value reflects the model's strong discriminatory power in differentiating between TB and non-TB cases across different decision thresholds.

RESULTS AND DISCUSSION

This section analyzes the outcomes of the implemented hybrid approach, which integrates Synthetic Minority Over-sampling Technique (SMOTE), Tomek Links, Modified Focal Loss (MFE), and class weighting loss within a transfer learning framework. The performance of the proposed methods is benchmarked against baseline approaches, offering insights into their contributions to model robustness and sensitivity in detecting TB-positive cases, which represent the minority class.

Table 1 presents the classification report obtained from the experiment, while Figures 2, 3, 4, 5 and 6 illustrate the imbalanced class distribution, balanced class distribution, training and validation accuracy, training and validation loss, and the confusion matrix, respectively.

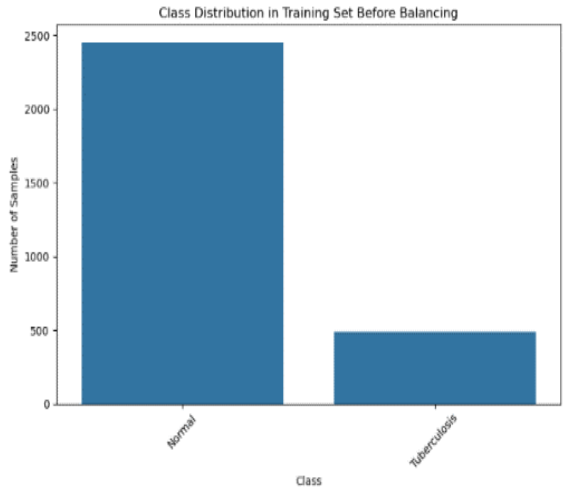

Figure 2: Imbalance Data

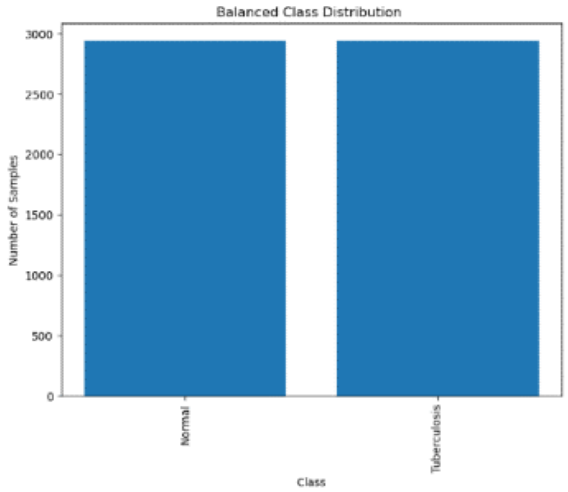

Figure 3: Balanced Data

Figures 2 and 3visually compare the class distributions before and after applying the hybrid technique, respectively. These visualizations highlight the effectiveness of the SMOTE-Tomek Links method in mitigating class imbalance, thereby creating a solid foundation for training robust and unbiased models for tuberculosis detection

Table 1: Classification Report for the Hybrid model

Class	Precision	Recall	F1-Score	Support
Normal	1.00	1.00	1.00	525
Tuberculosis	1.00	0.98	0.99	106
Accuracy			1.00	631
Macro Avg	1.00	0.99	0.99	631
Weighted Avg	1.00	1.00	1.00	631

The classification report presented in Table 4.1 demonstrates the performance of the developed tuberculosis detection model in classifying X-ray images into two categories: Normal and Tuberculosis. The results, as reflected in the precision, recall, F1-score, and support metrics, reveal the robustness and effectiveness of the model in addressing the challenges associated with this classification task.

Precision, which indicates the proportion of true positive predictions among all positive predictions for a given class, was recorded at 1.00 for the Normal class and 1.00 for the Tuberculosis class. This shows that the model effectively minimized false positive errors, accurately distinguishing between the two classes. The recall values, measuring the proportion of actual positive instances correctly identified, were similarly high, with the Normal class achieving 1.00 and the Tuberculosis class achieving 0.98. These recall scores highlight the model's capability to correctly identify a majority of the instances in both categories, particularly for the minority class, Tuberculosis, which is critical for medical diagnostics. The F1-score, a balanced metric combining precision and recall, was 1.00 for the Normal class and 0.99 for the Tuberculosis class, demonstrating the model's ability to maintain an excellent trade-off between precision and recall. The weighted average F1-score of 1.00 further underscores the model's overall effectiveness, even in the presence of class imbalance. The support metric, representing the number of samples in each class, indicates that the dataset included 525 samples of Normal images and 106 samples of Tuberculosis images, reflecting an imbalanced distribution.

The accuracy of the model across the entire dataset was 0.9968 approximated to be 1.00, signifying that the model correctly classified all samples with near-perfect precision. This exceptional performance can be attributed to the hybrid SMOTE-Tomek Links class balancing technique, which addressed the issue of class imbalance by generating synthetic samples for the minority class while removing ambiguous samples to improve decision boundaries. Additionally, the integration of the Modified Focal Loss function further enhanced the model's ability to handle class imbalance by penalizing misclassification of minority class samples. The classification report, therefore highlights the effectiveness of the developed model in distinguishing between Normal and Tuberculosis cases with remarkable precision, recall, and overall accuracy. This performance establishes the model as a reliable tool for tuberculosis detection, capable of mitigating class imbalance issues and providing robust predictions for real-world medical applications.

Figure 4: Training and Validation Accuracy

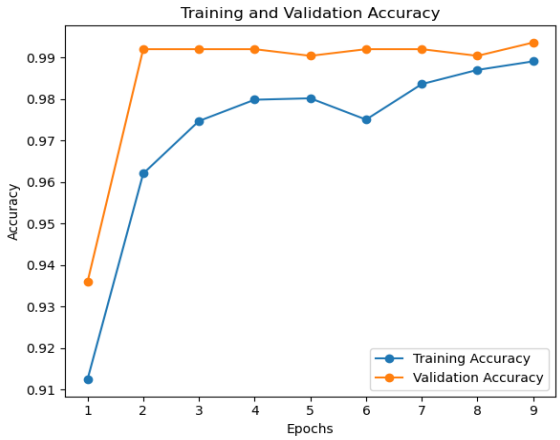

Figure 5: Training and Validation Error

Figure 4 depicts the trends in training and validation accuracy over successive epochs during the model training process. The plot reveals a rapid improvement in accuracy within the first few epochs, indicating that the model was able to learn key features from the dataset early in the training process. The training accuracy shows a consistent upward trend, stabilizing around the 5th epoch, and continues to improve gradually towards the 9th epoch. Similarly, validation accuracy follows a similar trajectory, demonstrating a high level of alignment with the training accuracy, which signifies minimal overfitting. By the end of training, both training and validation accuracy surpass 99%, showcasing the model's excellent generalization ability on unseen data

In Figure 5, the training and validation error are presented as complementary measures to the accuracy. The training loss decreases sharply within the first few epochs, reflecting the model's ability to minimize errors through the optimization process. A similar trend is observed in the validation loss, which also drops significantly early on and stabilizes after the 3rd epoch. Notably, the gap between training and validation loss remains minimal throughout the epochs, further reinforcing the robustness of the model and its capability to generalize well across the training and validation datasets. Finally, the analysis of both figures demonstrates that the model training process was effective, with the hybrid SMOTE-Tomek Links balancing technique and Modified Focal Loss function likely contributing to this outstanding performance. These results validate the model's suitability for detecting tuberculosis in X-ray images with high reliability and precision.

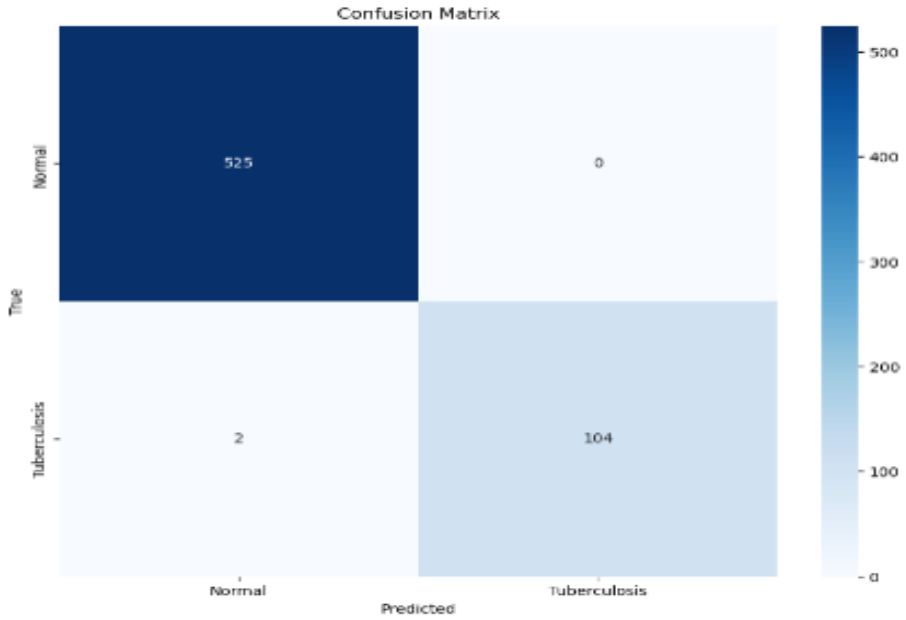

Figure 6: Confusion Matrix

The confusion matrix presented in Figure 4.5 provides a comprehensive summary of the classification performance of the developed model on the test dataset. The matrix comprises two rows corresponding to the actual classes and two columns representing the predicted classes. The diagonal elements indicate the instances that were correctly classified, while the off-diagonal elements represent misclassified cases. For the Normal class, the model accurately classified 525 instances as Normal, showcasing a perfect match between the actual and predicted labels for this category. There were no misclassified instances of the Normal class, highlighting the model's exceptional precision and recall for this group.

For the Tuberculosis class, the model demonstrated high accuracy by correctly classifying 104 out of 106 actual Tuberculosis cases. However, two instances from this class were misclassified as Normal. These misclassifications suggest a minor limitation in the model's ability to distinguish certain Tuberculosis cases, potentially due to subtle feature overlaps, noise in the data, or challenging edge cases present in the dataset.

The model's performance is remarkably robust, with a total of 629 correctly classified instances out of 631, resulting in an impressive accuracy of approximately 99.7%. While the two misclassified Tuberculosis cases point to areas where the model could be refined, its ability to correctly identify the majority of cases in both categories demonstrates its reliability for tuberculosis detection in medical imaging. Further investigation of the misclassified cases could provide valuable insights for enhancing the model's performance and addressing any underlying challenges.

CONCLUSION

In conclusion, this study successfully developed an advanced machine learning model for Tuberculosis prediction using medical X-ray images. By addressing the prevalent challenge of class imbalance in medical datasets, the research introduced innovative techniques, including the hybrid SMOTE-Tomek Links method, a modified focal loss function, and class weighting. These approaches enhanced the model's ability to accurately classify both majority and minority classes, ensuring a balanced learning process.

The developed model demonstrated superior performance, achieving an impressive accuracy of 99.7%, surpassing established benchmark models such as EfficientNetB3, Random Forest, and XGBoost. The hybrid SMOTE-Tomek Links technique effectively balanced the dataset, while the modified focal loss and class weighting prioritized learning from underrepresented classes. These advancements significantly improved the model's predictive accuracy and generalization capability.

The results affirm the effectiveness of the proposed methodology in addressing class imbalance and enhancing predictive performance in medical image classification tasks. The findings have practical implications for improving early and accurate detection of Tuberculosis, potentially aiding in better clinical decision-making and patient management. Furthermore, this research underscores the importance of integrating innovative machine learning techniques to overcome limitations in traditional approaches, setting a foundation for future advancements in the field of healthcare and medical diagnostics.

REFERENCES:

Althomsons, S. P., Winglee, K., Heilig, C. M., Talarico, S., Silk, B., Wortham, J., ... & Navin, T. R. (2022). Using machine learning techniques and national tuberculosis surveillance data to predict excess growth in genotyped tuberculosis clusters. *American journal of epidemiology*, *191*(11), 1936-1943.

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, *2020*, baaa010.

Ai-jun, L., & Peng, Z. (2020, June). Research on unbalanced data processing algorithm base tomeklinks-smote. In *Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition* (pp. 13-17).

Awujoola, O. J., Ogwueleka, F., Odion, P. O., & Irhebhude,(2020), M. An Intelligent Homogenous Model For

Prediction Of Network Intrusion Detection Using Synthetic Minority Over Sampling Technique and Local Outlier Factor.

Awujoola, O. J., Ogwueleka, F. N., Irhebhude, M. E., & Misra, S. (2021). Wrapper based approach for network intrusion detection model with combination of dual filtering technique of resample and SMOTE. In *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities* (pp. 139-167). Cham: Springer International Publishing.

Alamri, M., & Ykhlef, M. (2024). Hybrid undersampling and oversampling for handling imbalanced credit card data. *IEEE Access*.

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., ... & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154-170.

Chen, J., Fu, C., Xie, H., Zheng, X., Geng, R., & Sham, C. W. (2022). Uncertainty teacher with dense focal loss for semi-supervised medical image segmentation. *Computers in Biology and Medicine*, *149*, 106034.

Fox, G. J., Johnston, J. C., Nguyen, T. A., Majumdar, S. S., Denholm, J. T., Asldurf, H., ... & Velen, K. (2021). Active case-finding in contacts of people with TB. *The International Journal of Tuberculosis and Lung Disease*, *25*(2), 95-105.

Gopalaswamy, R., Shanmugam, S., Mondal, R., & Subbian, S. (2020). Of tuberculosis and non-tuberculous mycobacterial infections–a comparative analysis of epidemiology, diagnosis and treatment. *Journal of biomedical science*, *27*, 1-17

Guo, Y. (2024). Multimodal Multilabel Classification by CLIP. *arXiv preprint arXiv:2406.16141*.Jonathan, J., Barakabitze, A. A., Fast, C. D., & Cox, C. (2024). Machine Learning for Prediction of Tuberculosis Detection: Case Study of Trained African Giant Pouched Rats. *Online Journal of Public Health Informatics*, *16*, e50771

Kumar, S., Mishra, A. K., Mishra, R. K., Shrivastava, A., Chhabra, P., & Chhabra, G. (2023, July). Identification of Mycobacterium Tuberculosis Employing VGG-16 Feature Extraction and Classification Using Prominent Machine Learning Classifiers on X-rays. In *International Conference on Data Science and Applications* (pp. 119-130). Singapore: Springer Nature Singapore.

Kieu, S. T. H., Bade, A., Hijazi, M. H. A., & Kolivand, H. (2020). A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions. *Journal of imaging*, *6*(12), 131.

Leng, Q., Guo, J., Tao, J., Meng, X., & Wang, C. (2024). OBMI: oversampling borderline minority instances by a two-stage Tomek link-finding procedure for class imbalance problem. *Complex & Intelligent Systems*, 1-18.

Mitruka, K., Oeltmann, J. E., Ijaz, K., & Haddad, M. B. (2011). Tuberculosis outbreak investigations in the United States, 2002–2008. *Emerging infectious diseases*, *17*(3), 425.

Makam, P., & Matsa, R. (2021). "Big Three" infectious diseases: tuberculosis, malaria and HIV/AIDS. *Current Topics in Medicinal Chemistry*, *21*(31), 2779-2799.

Nafisah, S. I., & Muhammad, G. (2024). Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. *Neural Computing and Applications*, *36*(1), 111-131.

Ou, C. Y., Chen, I. Y., Chang, H. T., Wei, C. Y., Li, D. Y., Chen, Y. K., & Chang, C. Y. (2024). Deep Learning-Based Classification and Semantic Segmentation of Lung Tuberculosis Lesions in Chest X-ray Images. *Diagnostics*, *14*(9), 952

Przybyła-Kasperek, M. (2022). Study of selected methods for balancing independent data sets in k-nearest neighbors classifiers with Pawlak conflict analysis. *Applied Soft Computing*, *129*, 109612.

Rao, G. M., Ramesh, D., Sharma, V., Sinha, A., Hassan, M. M., & Gandomi, A. H. (2024). AttGRU-HMSI: enhancing heart disease diagnosis using hybrid deep learning approach. *Scientific Reports*, *14*(1), 7833.

Rodrigues, M. M., Barreto-Duarte, B., Vinhaes, C. L., Araújo-Pereira, M., Fukutani, E. R., Bergamaschi, K. B., ... & Andrade, B. B. (2024). Machine learning algorithms using national registry data to predict loss to follow-up during tuberculosis treatment. BMC Public Health, 24(1), 1385.

Silva, L., Motta, L. G. D., & Eberly, L. (2024). Prediction of tuberculosis clusters in the riverine municipalities of the Brazilian Amazon with machine learning. *Revista Brasileira de Epidemiologia*, *27*, e240024.

Sharma, V., Gupta, S. K., & Shukla, K. K. (2024). Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images. *Intelligent Medicine*, *4*(2), 104-113.

Swana, E. F., Doorsamy, W., & Bokoro, P. (2022). Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset. *Sensors*, *22*(9), 3246.

Tang, N., Yuan, M., Chen, Z., Ma, J., Sun, R., Yang, Y., ... & Zhou, J. (2023). Machine Learning Prediction Model of Tuberculosis Incidence Based on Meteorological Factors and Air Pollutants. *International journal of environmental research and public health*, *20*(5), 3910.

Viadinugroho, R. A. A. (2023). Imbalanced classification in python: SMOTE-Tomek Links method combining SMOTE with Tomek Links for imbalanced classification in python.

Viswanatha, V., Ramachandra, A. C., Togaleri, A. R., & Gowda, N. S. (2023). Tuberculosis Prediction using KNN Algorithm. *International Journal of Engineering and Management Research*, *13*(4), 58-71.

Wang, W., Guo, W., Cai, J., Guo, W., Liu, R., Liu, X., ... & Zhang, S. (2021). Epidemiological characteristics of tuberculosis and effects of meteorological factors and air pollutants on tuberculosis in Shijiazhuang, China: A distribution lag non-linear analysis. Environmental research, 195, 110310

Wen, D., Soltan, A., Trucco, E., & Matin, R. N. (2024). From data to diagnosis: skin cancer image datasets for artificial

intelligence. *Clinical and Experimental Dermatology*, llae112.

Xi, Y., Li, M., Zhou, F., Tang, X., Li, Z., & Tian, J. (2023). SE-Inception-ResNet Model with Focal Loss for Transmission Line Fault Classification Under Class Imbalance. *IEEE Transactions on Instrumentation and Measurement*.

Xie, Y., Wan, Q., Xie, H., Xu, Y., Wang, T., Wang, S., & Lei, B. (2023). Fundus image-label pairs synthesis and retinopathy screening via GANs with class-imbalanced semi-supervised learning. *IEEE Transactions on Medical Imaging*, *42*(9), 2714-2725.

Yeung, M., Sala, E., Schönlieb, C. B., & Rundo, L. (2022). Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, *95*, 102026.