p–ISSN: 2723 - 6609 e-ISSN: 2745-5254
Vol. 5, No. 5 Mei 2024 http://jist.publikasiindonesia.id/

Optimization of Early Detection of Tuberculosis: Use of Multilayer Perceptron and Extreme Learning Machine with Clinical Data

Ammar Waliyuddin Jannah[1*], Berlian Al Kindhi[2]
Institut Teknologi Sepuluh Nopember, Indonesia
Email: ammarwd@gmail.com
*Correspondence

Keywords: Tuberculosis; Early Detection; Machine Learning; Multilayer Perceptron; Extreme Learning Machine.

ABSTRACT

This research takes an innovative step in the fight against Tuberculosis (TB), one of Indonesia's prominent public health challenges, by developing and evaluating Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM) models in machine learning for early detection of TB using clinical data. The main result of this study was the discovery that the MLP model, when applied without the Synthetic Minority Over-sampling Technique (SMOTE), achieved an impressive accuracy of 95.00%, signalling significant progress in TB early detection efforts. This discovery not only highlights the great potential of applying machine learning technology in improving the accuracy of TB diagnosis but also paves the way for the possible application of advanced technology in the health sector to deal with infectious diseases. This research illustrates how machine learning technology can be integrated into clinical practice to detect TB cases early, thus enabling faster and more precise treatment, ultimately reducing the spread of the disease. This is particularly important given TB's significant impact on public health, especially in developing countries. The results also open up opportunities for further research into applying machine learning techniques to other infectious diseases, promising a paradigm shift in how we detect and manage various health conditions.

Introduction

Tuberculosis (TB) remains one of the deadliest infectious diseases in the world, with millions of new cases and deaths each year. According to a World Health Organization (WHO) report, TB is included in the top 10 leading causes of death globally (Alwarthan, Aslam, & Khan, 2022). In Indonesia, TB conditions are very concerning, with the third rank in the world for the number of TB cases. Factors such as HIV prevalence, poor socioeconomic conditions, and limited access to health services exacerbate the situation. Low awareness and limited knowledge about TB among the public often lead to inappropriate or late treatment. In addition, increasing resistance to TB drugs is an additional challenge in controlling this disease (Al Amien, Rizki, &

Nasution, 2022). This situation requires innovative and effective detection and treatment strategies to reduce the burden of TB in Indonesia.

Early detection of TB is essential to reduce the spread of the disease and ensure effective treatment. Challenges in the early detection of TB are often associated with the need for more accurate and affordable diagnostic tools, especially in developing countries (Arumnisaa & Wijayanto, 2023). Many TB cases go undiagnosed or diagnosed late due to nonspecific early symptoms. Therefore, advanced technologies such as machine learning can help overcome these challenges, enabling more accurate and rapid analysis of patient clinical data (Baharuddin, Azis, & Hasanuddin, 2019). This approach can help identify cases of TB at an early stage, which is crucial to prevent transmission and initiate appropriate treatment. Implementing this technology could be a paradigm shift in how we detect and manage TB, especially in regions with limited resources (World Health Organization, 2022).

Machine learning, as a branch of artificial intelligence, offers the possibility of complex health data analysis. In the context of TB, machine learning algorithms can identify patterns in clinical data that may not be seen with traditional methods (Ahmed & Prakasam, 2023). Techniques such as Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM) hold promise in TB research, processing non-linear data effectively. Machine learning can potentially improve TB diagnosis accuracy, reduce the time it takes to detect cases and facilitate TB identification at an earlier stage compared to conventional methods (Charles et al., 2016). The utilisation of this technology in research and clinical practice could revolutionise how we deal with TB (Yeo, Balakrishnan, Selvaperumal, & Nor, 2022).

Multilayer Perceptron (MLP) is a model of feedforward artificial neural networks that relies on a layered structure that includes an input layer, one or more hidden layers, and an output layer (Bikku, 2020). MLP's main advantage lies in its ability to model complex non-linear relationships between variables, a common characteristic of clinical data. MLP's multi-layered structure and adaptability make it suitable for analysing and interpreting complex and varied patient data (Huang, Wang, & Lan, 2011).

In contrast, the Extreme Learning Machine offers a different approach. With its main characteristic of randomly generating weights and biases for hidden layers, ELM eliminates the need for weight adjustment during the learning process. The speed and efficiency of ELM learning are key advantages, especially in the context of limited datasets (Erlin, Desnelita, Nasution, Suryati, & Zoromi, 2022). ELM's ability to produce fast and accurate results, albeit with smaller datasets, makes it an invaluable method in this study (Peralez-González, Pérez-Rodríguez, & Durán-Rosal, 2023).

These two methods were chosen to compare the effectiveness of two different approaches in processing and analysing limited data. MLP and ELM, each with its strengths and uniqueness, offer valuable insights into how machine learning can be applied in this research.

In machine learning research for TB, the available data is often biased. To overcome this, this study will use data synthesis techniques to balance the dataset (Rashidi et al., 2022). This will allow machine learning models to learn from more balanced datasets, reduce prediction bias, and improve accuracy (Kavvas et al., 2018). Data synthesis not only helps in providing a better representation of minority cases but also enables model testing across multiple scenarios, improving model reliability in real-world conditions (Gao et al., 2023).

This research aims to develop an effective machine-learning model for TB prediction, with a particular focus on classifying two categories: TB positive and TB negative. Researchers will collect patient clinical data from lung poly in hospitals, then develop and test Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM) models, with and without data balancing techniques. The evaluation will focus on the accuracy of classifying these two labels, with the hope that this study will make an important contribution to the early detection of TB.

Table 1
Previous Research

Research 1	
Heading	Evaluating the performance of multilayer perceptron algorithm for tuberculosis disease Raman data
Year	2022
Researchers	Rahatullah, Suranjam Khan, Zahra Ali, Heena Ali, Adnan Ahmed, & Iftikhar Ahmed
Research Review	This study evaluated the use of Multilayer Perceptron (MLP) in analysing Raman spectroscopic data for tuberculosis (TB) patients.
	The study examined peak Raman intensity related to the concentration of specific biomolecules in blood serum samples.
Research Results	MLP's performance in classifying samples from active TB cured TB and healthy control patients showed sensitivity and specificity of 62-92% and 81-88%, respectively.
Relationship with Research to be Researched	This study provides insight into the potential use of MLP in TB diagnosis. In the context of the researchers' analysis, which aimed to detect early TB using machine learning by comparing ELM and MLP on 115 patient clinical datasets, Ullah et al.'s study offers valuable insights into MLP implementation in a health context. Researchers can utilise their methodology and results as a reference to strengthen

	aspects of data analysis in this study, specifically in comparing the accuracy of MLP with ELM based on various clinical parameters such as temperature, weight loss, pulse, appetite, cough, shortness of breath, etc. This will help in evaluating the effectiveness of these two methods in diagnosing TB, as well as enable more accurate identification of positive/negative TB
Improvements in Research	Ullah et al., 2022 Ullah et al., 2022

Based on the problem formulation that has been set previously, the research objectives are as follows:

1. To improve disease detection and management, Build and test machine learning models to accurately classify two TB diagnostic labels.
2. Assess and compare the effectiveness of Multilayer Perceptron and Extreme Learning Machine in TB diagnosis using clinical data of patients in hospitals.
3. Develop models that can overcome the problem of data imbalance in TB studies using data synthesis techniques.
4. Determine the most efficient and effective machine learning approach for TB detection, considering factors such as accuracy and data variability.

Research Methods

The researchers' research methodology aims to detect cases of pulmonary TB using machine learning models. They adopt a framework, as shown in Figure 3.1, "IBM Data Science Methodology," which consists of ten important stages. In this chapter, they will detail how they apply these stages in the context of this study.

This research was conducted with the understanding that early detection of TB can play a crucial role in limiting the spread of the disease, improving treatment outcomes, and reducing the economic burden associated with advanced disease care. By referring to clinical data from the Lung Poly of RSUD Prof. Dr W Z Johannes Kupang, the study aims to develop data-based solutions to optimise TB detection. The proposed machine learning model seeks to recognise patterns indicative of TB from clinical data, enabling faster and more accurate detection, which could facilitate timely and effective medical interventions.

This research will investigate how existing clinical data can be leveraged to inform and train machine-learning models. This will involve collecting, cleaning, and analysing data to identify significant variables that might predict the presence of TB. The success of this approach is measured not only in terms of the technical performance of the resulting model but also in its effect on improving early detection of TB and decreasing TB disease prevalence and mortality rates in the region.

Analytic Approach

In the context of this study, the analytics approach is the primary focus in determining how patient clinical data will be used to develop predictive models using machine learning techniques. The model to be developed will utilize a dataset consisting of a basic physical examination of the patient that includes the following variables: temperature, Age, Sex, Weight, Pulse, Oxygen Saturation, Presence of Cough, Cough with Phlegm, Ronchi, Wheezing, Alcohol Consumption, Presence of Fever, Weight Loss, and Cigarettes Consumed.

Data Requirements

This research requires comprehensive and specific data collection for an accurate classification model. The data required should reflect various aspects of the patient's essential physical examination related to TB symptoms and risk factors.

This study's data collection was done ethically and with patient privacy in mind. All data is anonymised before analysis to ensure the safety of patients' personal information. This data is also maintained in quality to ensure the model's integrity. This means that the data must be complete, consistent, and accurate. Data checks will include validation for unreasonable values or outliers, handling missing data, and verifying documented values.

Data was collected at the Lung Poly of RSUD Prof. Dr W Z Johannes Kupang. The main objective was to collect comprehensive and high-quality data, which supported the development of machine-learning classification models in the early detection of tuberculosis. Figure 3.2 above shows the assessment form used at the Lung Poly of RSUD Prof. Dr W Z Johannes Kupang, which is the main instrument in collecting patient clinical data.

In the data collection process, selecting the right subject is the key to ensuring the validity and reliability of the research results. Therefore, inclusion and exclusion criteria are carefully set to choose the most suitable subjects; here are the inclusion and exclusion criteria of this study

Inclusion Criteria:

1. Patients undergoing assessment at the Lung Poly of RSUD Prof. Dr W Z Johannes Kupang in 2023.
2. Patients who seek outpatient treatment, are sick or check related to lung disease at the poly.
3. Patients diagnosed or suspected of tuberculosis and those who have recovered from tuberculosis only do routine examinations or regular checks.

Exclusion Criteria:

1. Patients who do not undergo an assessment at the Lung Poly of RSUD Prof. Dr W Z Johannes Kupang in 2023.
2. Patients with medical conditions unrelated to lung disease, such as cognitive impairment or severe comorbidities that may affect assessment results.

The researchers used purposive sampling to select patients who met the inclusion criteria. Subjects were selected based on their relationship with the study's purpose, namely, early detection of tuberculosis. This ensures that the sample obtained is highly relevant to the research objectives and produces data that can provide significant insight into the research problem.

Results and Discussion

Data Understanding Phase

This sub-chapter will describe the data understanding phase, an integral part of the data analysis process. This phase concerns the collection, examination, and initial understanding of the dataset used in this study. The main focus of this phase is to gain deep initial insight into the characteristics and structure of the data, which will help in the analysis and interpretation of subsequent results.

Data Description

This study relied on data from the assessment form provided by the Lung Poly of RSUD Prof. Dr W Z Johannes Kupang. This assessment form is specifically designed to record extensive and detailed clinical information from patients visiting pulmonary poly. The data collected includes:

1. Clinical Information: This includes symptoms experienced, history of smoking and alcohol, and results of physical examination. This information helps to obtain a comprehensive picture of the patient's lung health status and represents an essential variable in the clinical assessment of lung conditions.
2. Relevance to Lung Conditions: These data include common symptoms such as coughing or shortness of breath and specific physical examination results such as pulmonary auscultation revealing Ronchi or wheezing. Information about weight loss or appetite is also collected, which can indicate chronic lung diseases such as Tuberculosis.

Table 2
Variables and data types

Variable	Data Type	Relevance to Early Detection of Tuberculosis
Gender	Demographic	Gender affects the susceptibility and manifestations of TB
Age	Demographic	Age can affect the risk and type of TB
Fever	Symptom	Fever is a common symptom of TB
The presence of a cough	Symptom	Cough is one of the main symptoms of TB

Coughing up phlegm	Symptom	Coughing up phlegm is common in TB
Coughing up blood	Symptom	It is one of the severe and characteristic clinical symptoms of TB, mainly when TB infects the lungs and causes damage to lung tissue and blood vessels.
Nyeri Dada	Symptom	It can indicate infection or inflammation of the lungs, which is often associated with conditions such as TB.
Weight Loss	Symptom	Weight loss is a symptom of TB
Decreased appetite	Symptom	Decreased appetite is common in TB
Shortness of breath	Symptom	Shortness of breath can be a symptom of advanced TB
Ronchi	Symptom	*Ronchi* may indicate a lung abnormality
Wheezing	Symptom	*Wheezing* can be related to airway obstruction in TB
Weight	Physical Examination	Low body weight can be an indication of TB
Weight	Physical Examination	Low body weight can be an indication of TB
Pulse	Physical Examination	Pulse changes may indicate complications of TB
Oxygen Saturation	Physical Examination	Low oxygen saturation can be a sign of TB
Body Temperature	Physical Examination	Fever is a common symptom of TB
Blood pressure	Physical Examination	Changes in blood pressure can be related to lung conditions
Alcohol Consumption	Medical History	Alcohol consumption can worsen the condition of TB
Cigarette Consumption	Medical History	Cigarette consumption increases the risk of TB

Table 1 presents the variables analysed in the study and their relevance to early detection of Tuberculosis. These variables include:

1. Gender and Age (Demographic)
2. Fever, Presence of Cough, Cough with Phlegm/Blood, Chest Pain, Weight Loss/Appetite, Shortness of Breath, Ronchi, and Wheezing (Symptoms)
3. Body Weight, Pulse, Oxygen Saturation, Body Temperature, and Blood Pressure (Physical Examination)
4. Alcohol and Cigarette Consumption (Medical History)

Each of these variables has a vital role in understanding and identifying TB early, increasing the chance for successful treatment and reducing the spread of the disease. Gender and age can determine which groups are more susceptible to TB. Symptoms such as coughing up phlegm and fever are the leading clinical indicators that trigger further examination. Physical examination provides objective data about the patient's physical condition, while medical history, such as alcohol and cigarette consumption, provides context regarding risk factors.

Data Transformation

Transformation is the process of transforming data from one form or representation to another form or representation for better analysis. The following data transformations are carried out:

1. Age Grouping

Age grouping in clinical data aims to help machine learning models recognise patterns related to tuberculosis risk. By reducing variability in age data, models can learn from general trends related to specific age groups without being distracted by irrelevant individual fluctuations. This facilitates learning and improves the accuracy of model predictions.

Age grouping in researchers' clinical datasets is simplified into four categories to help machine learning algorithms identify important patterns more efficiently. By excluding unnecessary age variations, models are more accessible to train and tend to be more accurate in their predictions. These age categories are:

1. 0 = Children and Adolescents (0-18 years)": These are children and adolescents in a phase of growth and development in which TB's immune response and manifestations differ from those of adults.
2. 1 = Young Adults (19-35 years)": This group was studied to evaluate the influence of social and behavioural risk factors in early adult life.
3. 2 = Adults (36-60 years)": This phase is analysed by considering working conditions, lifestyle, and the presence of comorbidities that can affect TB risk.
4. 3 = Elderly (61 years and over)": This age group focuses on treating TB in older people, decreased immune function and potential complications.

Categorical Variables

The categorical variables in the researchers' dataset were originally in text form. They needed to be converted into a numeric format to allow machine learning algorithms to process and learn the data more effectively. Here are the changes made:

1. Gender was changed using '0' for females and '1' for males.
2. Change features that indicate the presence or absence of a condition, such as 'cough', 'fever', 'chest pain', 'Ronchi', or 'wheezing', and confirm that there are only two possible values (0 for 'no' and 1 for 'yes').
3. Ordinal variables, such as "Weight Loss" with sequential categories, are coded according to the order shown in Table 2.

Table 3
Ordinal Variable Encoding

Weight Loss	Code
No drop	0
There is a decrease of 1-5 kg	1
There is a decrease of 6-10 kg	2
There is a decrease of 11-15 kg	3
There is a decrease of more than 15 kg	4

Data distribution before and after SMOTE

The application of Synthetic Minority Over-sampling Technique (SMOTE) in this study is aimed at overcoming the issue of class imbalance discussed in subchapter 4.1.2. The total dataset includes 101 data divided into TB- and TB+ classes. The SMOTE technique was chosen for its effectiveness in multiplying synthetic samples from minority classes, which helped achieve a more balanced distribution of courses for model training. Figure 4.3.3 shows the distribution of data on the first fold before and after the implementation of SMOTE, illustrating the balance obtained between the two classes in the training set after SMOTE.

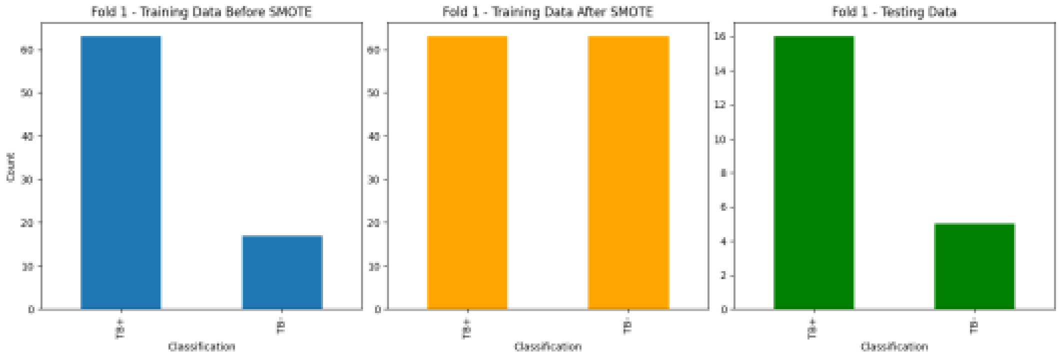

Figure 1 Stratified cross-validation k-fold data distribution

For k-fold stratified cross-validation performed in five iterations, the data was divided into groups, with 80% of the data used as training sets and 20% as test sets in each fold. This implies that in each fold, approximately 80 data (63 tb+ and 17 tb-) are used as training sets and 21 (16 tb+ and five tb-) as test sets. SMOTE is applied only to training sets, increasing minority class representation without changing the natural distribution of test sets. This approach ensures that model evaluation can reflect predictive capabilities on data not involved in the training process, confirming the accuracy and reliability of the model under realistic test conditions.

Modelling

In this chapter, researchers will explore the process of creating models using two different methods: Extreme Learning Machine (ELM) and Multi-Layer Perceptron (MLP). Both methods will be tested using synthesis data and without synthesis data and then compared to determine which is more effective in the early detection of tuberculosis.

Implementasi Extreme Learning Machine (ELM)

In this study, an implementation of Extreme Learning Machine (ELM) using Python was applied to a dataset consisting of 101 samples with 20 features. To determine the optimal configuration of neurons in the hidden layer, extensive testing was carried out on a range of 5 to 50 neurons. Aims to determine the configuration that provides the best balance between the capacity of the model to learn complex patterns in the data and the risk of overfitting and the need to maintain adequate learning capacity, especially in the face of the complexity of classification tasks with two different classes. This determination of the number of neurons ensures that the model has sufficient capacity to identify meaningful patterns in the data.

Evaluation

Evaluation of model results is a critical step in measuring the effectiveness and accuracy of the algorithms applied in predicting data. In this sub-chapter, we will analyze and compare the performance of the two methods used in this study: Extreme Learning Machine (ELM) and Multi-Layer Perceptron (MLP), both with the application of data synthesis techniques using SMOTE and without data synthesis.

Comparison and Overall Analysis

Tabel 4

Perbandingan ELM dan MLP tanpa dan dengan SMOTE

	Accuracy (%)	
	Without SMOTE	**SMOTE**
ELM	67.33	81.33
MLP	95.00	94.00

This sub-chapter presents a synthesis of evaluations that have been performed on Extreme Learning Machine (ELM) and Multi-Layer Perceptron (MLP) in the context of using SMOTE data synthesis techniques and without SMOTE. This analysis aims to explore more profound insights into how these two algorithms react to class imbalance handling techniques and their implications for classification performance.

In Figure 4.8, ELM shows a significant improvement in accuracy from 67.33% to 81.33% after the implementation of SMOTE, which confirms the effectiveness of SMOTE in overcoming class imbalances. This is reflected in the minority class's increased precision (TB—) and a more balanced F1 Score between the two classes, signaling an improvement in fairer classification capabilities.

Conversely, MLP, which showed excellent performance without SMOTE with an accuracy of 95.00%, only experienced a slight decrease in accuracy to 94.00% after the implementation of SMOTE. This suggests that MLP is more resistant to class imbalance and that SMOTE does not provide significant improvements in this context. However, the decrease in recall for class TB- following SMOTE adoption requires further consideration in the context of medical applications where the detection of false negatives is critical.

From the comparative analysis that has been carried out, it becomes clear that the decision of model selection depends not only on the evaluation metrics but also on the context of the practical application in which the model will be applied. ELM, with the application of SMOTE, showed significant improvements in addressing class imbalances, while MLP showed robustness and high performance without the support of data synthesis techniques. In clinical contexts, where precision and reliability of detection are priorities, the model with the highest accuracy will take precedence for deployment. Therefore, given the results that have been obtained, MLP without SMOTE, which shows the highest accuracy, will be the top choice for further implementation. This step reflects a commitment to the utilization of models that are not only statistically superior but also that best fit real, immediate diagnostic needs.

Deployment

Deployment is an essential step in the machine learning model development cycle, where the trained model is integrated into a production environment for use in real applications. In the context of this research, the deployment process will focus on deploying machine learning models developed using Python for early detection of tuberculosis.

Deployment Preparation

Before the deployment process, it is necessary to ensure that the model has been comprehensively tested and tuned to achieve optimal performance. Models selected based on the evaluations in the previous subchapter should be serialized or saved in a reusable format, such as using the pickle or joblib libraries in Python.

Deployment Techniques

The deployment technique will be creating an API (Application Programming Interface) using the Flask framework in Python. This API will enable easy and flexible integration with various front-end applications, both web-based and mobile.

The steps include:

1. API creation: Build an API that accepts input data, processes that data using a saved model, and returns prediction results.
2. Integration with Models: This involves loading serialized models into the API so that they can make predictions based on the inputs provided.
3. API testing: Performs comprehensive testing to ensure that the API works appropriately, including error handling and invalid input.

API Publication

Once the API has been successfully created and tested, the next step is to publish or host it so that users can access it. This is done using cloud services.

Issuance steps include:

1. Server Setup: Configure and set up servers in the cloud service.
2. API Deployment: Upload API code to the server and configure the server environment to run it.
3. Security and Access Settings: Set up security protocols such as HTTPS and set up API access.
4. Monitoring and Maintenance: Monitor API performance and perform maintenance regularly to ensure that the API continues to operate optimally.

Integration with Client Applications

Published APIs can be integrated with client applications, which are website platforms. This integration allows users to input data that the model will predict and receive prediction results directly through the user interface.

Conclusion

Machine Learning Model Development (MLP and ELM): This research successfully developed two machine learning models, Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM), for the classification of Tuberculosis (TB) based on clinical data. The MLP model shows awe-inspiring performance with an accuracy of 95.00% without SMOTE implementation and 94.00% with SMOTE. On the other hand, the ELM model showed a significant improvement in its performance after the implementation of SMOTE, with accuracy increasing from 67.33% to 81.33%. These findings signal success in developing effective models for early detection of TB.

Model Effectiveness in TB Detection: In the context of TB detection effectiveness, MLP models without SMOTE show superior results. This model successfully achieves high precision and recall balance for both classes (TB- and TB+), which is very important in medical contexts. ELM, although showing improvement with SMOTE, remains less effective than MLP. These findings underscore the importance of selecting appropriate models in clinical applications for optimal outcomes.

Data Synthesis to Address Data Imbalance: This study also explores the effectiveness of the Synthetic Minority Over-sampling Technique (SMOTE) technique in addressing data imbalance. SMOTE managed to significantly improve the performance of ELM models, demonstrating its effectiveness in overcoming class imbalances in datasets. Although the application of SMOTE to MLP models resulted in a slight decrease in accuracy, it still showed outstanding performance, confirming the importance of techniques such as SMOTE in the development of machine learning models. MLP and ELM Performance Comparison: In the comparison between MLP and ELM models, the findings show that MLP without SMOTE has the best performance with 95.00% accuracy, followed by MLP with SMOTE (94.00%) and ELM with SMOTE (81.33%). Meanwhile, ELM without SMOTE showed the lowest accuracy (67.33%), which highlights the importance of addressing class imbalances in model development. These conclusions clarify the advantages of MLP in TB classification and its potential for implementation in clinical environments.

Bibliography

Ahmed, S. Nafees, & Prakasam, P. (2023). A systematic review on intracranial aneurysm and hemorrhage detection using machine learning and deep learning techniques. *Progress in Biophysics and Molecular Biology*.

Al Amien, Januar, Rizki, Yoze, & Nasution, Mukhlis Ali Rahman. (2022). Implementasi Adasyn Untuk Imbalance Data Pada Dataset UNSW-NB15 Adasyn Implementation For Data Imbalance on UNSW-NB15 Dataset. *Jurnal CoSciTech (Computer Science and Information Technology)*, *3*(3), 242–248.

Alwarthan, Sarah, Aslam, Nida, & Khan, Irfan Ullah. (2022). An explainable model for identifying at-risk student at higher education. *IEEE Access*, *10*, 107649–107668. https://doi.org/10.1109/ACCESS.2022.3211070

Arumnisaa, Ressa Isnaini, & Wijayanto, Arie Wahyu. (2023). Comparison of Ensemble Learning Method: Random Forest, Support Vector Machine, AdaBoost for Classification Human Development Index (HDI). *Sistemasi: Jurnal Sistem Informasi*, *12*(1), 206–218.

Baharuddin, Mus Mulyadi, Azis, Huzain, & Hasanuddin, Tasrif. (2019). Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, *11*(3), 269–274.

Bikku, Thulasi. (2020). Multi-layered deep learning perceptron approach for health risk prediction. *Journal of Big Data*, *7*(1), 50.

Charles, M. Katherine, Lindegren, Mary Lou, Wester, C. William, Blevins, Meridith, Sterling, Timothy R., Dung, Nguyen Thi, Dusingize, Jean Claude, Avit-Edi, Divine, Durier, Nicolas, & Castelnuovo, Barbara. (2016). Implementation of tuberculosis intensive case finding, isoniazid preventive therapy, and infection control (" three I's") and HIV-tuberculosis service integration in lower income countries. *PloS One*, *11*(4), e0153243.

Erlin, Erlin, Desnelita, Yenny, Nasution, Nurliana, Suryati, Laili, & Zoromi, Fransiskus. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(3), 677–690.

Gao, Cong, Killeen, Benjamin D., Hu, Yicheng, Grupp, Robert B., Taylor, Russell H., Armand, Mehran, & Unberath, Mathias. (2023). Synthetic data accelerates the development of generalizable learning-based algorithms for X-ray image analysis. *Nature Machine Intelligence*, *5*(3), 294–308.

Huang, Guang Bin, Wang, Dian Hui, & Lan, Yuan. (2011). Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, *2*, 107–122.

Kavvas, Erol S., Catoiu, Edward, Mih, Nathan, Yurkovich, James T., Seif, Yara, Dillon, Nicholas, Heckmann, David, Anand, Amitesh, Yang, Laurence, & Nizet, Victor. (2018). Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. *Nature Communications*, *9*(1), 4306.

Peralez-González, Carlos, Pérez-Rodríguez, Javier, & Durán-Rosal, Antonio M. (2023). Boosting ridge for the extreme learning machine globally optimised for classification and regression problems. *Scientific Reports*, *13*(1), 11809.

Rashidi, Hooman H., Khan, Imran H., Dang, Luke T., Albahra, Samer, Ratan, Ujjwal, Chadderwala, Nihir, To, Wilson, Srinivas, Prathima, Wajda, Jeffery, & Tran, Nam K. (2022). Prediction of tuberculosis using an automated machine learning platform for models trained on synthetic data. *Journal of Pathology Informatics*, *13*, 100172. https://doi.org/10.4103/jpi.jpi_75_21

Yeo, Yee Jin, Balakrishnan, Arun Seeralan, Selvaperumal, Sathish Kumar, & Nor, Illanur Muhaini Binti Mohd. (2022). Android Controlled Fire Fighter Robot Using IoT. *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 32–42. IEEE.