Sains Malaysiana 54(1)(2025): 3741-3752
http://doi.org/10.17576/jsm-2025-5401-22

Optimizing Tuberculosis Treatment Predictions: A Comparative Study of XGBoost with Hyperparameter in Penang, Malaysia

(Mengoptimumkan Peramalan Rawatan Tuberkulosis: Suatu Kajian Perbandingan XGBoost dengan Hiperparameter di Penang, Malaysia)

YANIZA SHAIRA ZAKARIA[1], NUR AFIQAH ARIFFIN[2,*], AZIZUL AHMAD[3], RUSLAN RAINIS[2], AIDY M. MUSLIM[1] & WAN MOHD MUHIYUDDIN WAN IBRAHIM[2]

[1]Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
[2]Geography Section, School of Humanities, Universiti Sains Malaysia (USM), 11800 Pulau Pinang, Malaysia
[3]Centre for Spatially Integrated Digital Humanities (CSIDH), Faculty of Social Sciences & Humanities, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

Received: 24 April 2024/Accepted: 4 November 2024

ABSTRACT

The bacterium Mycobacterium tuberculosis causes a viral infection affecting the lungs and liver. Tuberculosis (TB) is a significant public health concern in developing countries, where it is often associated with poverty, poor living conditions, and limited access to healthcare services. According to the World Health Organization (2023), Tuberculosis continues to pose a substantial risk to public health on a global scale, with millions of people affected each year and around 1.5 million deaths in 2020. Healthcare providers often encounter significant challenges in addressing TB, leading to uncertain treatment outcomes. This study introduces a novel method for enhancing TB treatment using sophisticated machine learning techniques, particularly emphasizing the application of XGBoost and various predictive models in Penang State, Malaysia, to predict individual treatment outcomes based on clinical data. The models were trained using 2017 Penang data. Comparing predicted accuracy helps establish the optimum method. Clinical data was anonymized and analyzed. Decision tree accuracy is 63.7% using 2017 data. Logistic Regression is 63.3% accurate, while XGBoost is 66.3%. Hyperparameter-tuned XGBoost performs best at 68.1%. Comparing observed and expected results determines accuracy. TB result predictions are accurate using supervised learning. Calibrated ensemble models like XGBoost makes reliable predictions. Additional clinical characteristics may improve forecasts. The primary objective was to develop a reliable, clinically validated instrument that enhances TB treatments while optimizing resource efficiency across diverse healthcare environments.

Keywords: Classification; hyperparameter; logistic regression; prediction; random forest; tuberculosis

ABSTRAK

Bakteria Mycobacterium tuberculosis menyebabkan jangkitan virus yang menjejaskan paru-paru dan hati. Tuberkulosis (TB) adalah kebimbangan kesihatan awam yang signifikan di negara-negara membangun dan sering dikaitkan dengan kemiskinan, keadaan hidup yang buruk dan akses terhad kepada perkhidmatan kesihatan. Menurut Pertubuhan Kesihatan Sedunia (2023), TB terus menimbulkan risiko yang besar kepada kesihatan awam di peringkat global dengan berjuta-juta orang terjejas setiap tahun dan sekitar 1.5 juta kematian pada tahun 2020. Penyediaan penjagaan kesihatan sering menghadapi cabaran besar dalam menangani TB, yang membawa kepada hasil rawatan yang tidak menentu. Kajian ini memperkenalkan kaedah baharu untuk meningkatkan rawatan TB menggunakan teknik pembelajaran mesin yang canggih dengan penekanan khusus kepada aplikasi XGBoost dan pelbagai model ramalan di Pulau Pinang, Malaysia untuk meramalkan hasil rawatan individu berdasarkan data klinikal. Model-model tersebut dilatih menggunakan data Penang tahun 2017. Membandingkan ketepatan ramalan membantu menetapkan kaedah optimum. Data klinikal telah dianonimkan dan dianalisis. Ketepatan pokok keputusan adalah 63.7% menggunakan data 2017. Regresi Logistik adalah tepat 63.3%, manakala XGBoost adalah 66.3%. XGBoost yang diselaraskan dengan hiperparameter berprestasi terbaik pada 68.1%. Membandingkan hasil yang diperhatikan dan yang dijangkakan menentukan ketepatan. Ramalan keputusan TB adalah tepat menggunakan pembelajaran terawasi. Himpunan model yang dikalibrasi seperti XGBoost memberikan ramalan yang boleh dipercayai. Ciri klinikal tambahan mungkin dapat meningkatkan ramalan. Objektif utama adalah untuk membangunkan instrumen yang boleh dipercayai dan disahkan secara klinikal yang meningkatkan rawatan TB sambil mengoptimumkan kecekapan sumber pada pelbagai persekitaran penjagaan kesihatan.

Kata kunci: Hiperparameter; hutan rawak; pengelasan; ramalan; regresi logistik; Tuberkulosis

INTRODUCTION

Tuberculosis (TB) poses a significant challenge to global health, resulting in around 1.5 million deaths in 2020. In 2018, there were around 10 million newly reported cases, with 1.6 million deaths recorded (World Health Organisation 2022). In 2022, approximately 10.6 million individuals (95% uncertainty interval: 9.9-11.4 million) received a diagnosis of tuberculosis worldwide, marking an increase from the expected 10.3 million in 2021 and 10.0 million in 2020. A return to the decline observed before the pandemic could occur in 2023 or 2024 (World Health Organisation 2023). Despite the advancements achieved, addressing tuberculosis remains a considerable challenge. Transmission occurs through the inhalation of tuberculosis bacteria, typically due to actions like coughing, sneezing, or spitting. The symptoms include a persistent cough, blood in sputum, fatigue, unexplained weight loss, increased body temperature, and night sweats (Bukundi et al. 2021). If left unaddressed, tuberculosis can be fatal. The diagnostic process involves performing skin or blood tests to assess the immune response, followed by chest X-rays or sputum analysis to confirm the presence of active tuberculosis (Dheda et al. 2022).

Machine Learning techniques like Logistic Regression, XGBoost, Support Vector Machines, and random forests offer promising applications in tuberculosis to improve diagnosis, prognosis prediction, and decision support (Miotto et al. 2016). An example of applying Machine Learning models in healthcare is successfully predicting hospital readmission likelihood using electronic health records (Yang et al. 2021). Multiple predictive modelling techniques can be utilised to pinpoint patients who are at an elevated risk of having an unfavourable response to tuberculosis (TB) treatment (Hussain & Junejo 2018). Targeted interventions focused on high-risk groups can significantly improve outcomes (Abdullahi et al. 2019). The investigation carried out by (Xiong et al. 2018) utilised Machine Learning algorithms to predict the diagnosis of tuberculosis (TB) and the ensuing treatment outcomes. The findings of the study have been officially published in the Journal of Thoracic Disease. The study employed a dataset comprising of 430 individuals diagnosed with tuberculosis (TB). The findings of the study indicate that the Machine Learning model demonstrates a notable ability to predict treatment failure, achieving an accuracy rate of 87.2%.

To determine whether or not tuberculosis (TB) vaccination is successful, (Fayaz et al. 2024) conducted a study that used Machine Learning algorithms to examine data from a large-scale clinical trial (Gichuhi et al. 2023). The study successfully identified various factors that significantly elevated the likelihood of contracting tuberculosis (TB), encompassing age, gender, and HIV status. The study identified various factors that increased susceptibility to tuberculosis infection. The study utilized a sophisticated Machine Learning (ML) algorithm to analyze a comprehensive tuberculosis (TB) observation database, as reported in the prestigious journal PLOS Medicine. This study successfully identified several risk factors associated with TB infection, including age, gender, and smoking behavior. Moreover, according to Kouchaki et al. (2019) a publication in a technical or academic context. ML can potentially be employed for tuberculosis prevention and its various applications in diagnosis and treatment.

Janssens, Mourão-Miranda and Schnack (2018) found that ML models achieved an impressive accuracy rate of 90.1% in distinguishing TB patients from those without the disease. This research aims to develop ML models using clinical data from Penang State, Malaysia, to predict individual treatment outcomes based on local demographics. The 2017 dataset from a statewide tuberculosis surveillance program provides valuable insights, despite its limitations in accurately representing the current epidemiological landscape. This study focuses on the 2017 data because of the comprehensive clinical information collected that year, resulting in a robust and representative dataset for analysis. Additionally, using data from a single year minimizes potential biases and confounding variables that could emerge from multi-year analyses, thus enabling a clearer evaluation of the algorithms' performance. The effectiveness of these algorithms is assessed using metrics such as accuracy, precision, and recall, highlighting the potential of supervised learning to optimize patient treatment (Tiwari & Maji 2019). Notably, related research emphasizes the capacity of Machine Learning to identify patients at higher risk of relapse or mortality and suggests its application in developing personalized treatment strategies (Nicholson et al. 2023). Overall, these techniques significantly improve the prediction of treatment outcomes and enhance adherence tracking for tuberculosis patients.

MATERIALS AND METHODS

Jupyter Notebook effectively demonstrates the application of Machine Learning in classifying treatment outcomes for tuberculosis (TB) in individual patients. This investigation utilized a dataset consisting of 1,228 records of tuberculosis (TB) patients from Penang, Malaysia, gathered in 2017. The dataset encompasses various pertinent details, including age, diabetes diagnosis, smoking habits, tibial anatomy, and treatment outcomes, which reflect the success or failure of the treatment. Each record represents a distinct patient, with each column detailing specific information regarding their health status. The inclusion criteria are crucial for ensuring that only eligible volunteers participate in the study, which is vital for the validity and reliability of the outcomes (Figure 1). A tuberculosis (TB) treatment outcome study must establish clear inclusion criteria to ensure that only patients with a confirmed TB diagnosis are considered. Participation in the study requires a confirmed TB diagnosis, supported by medical records and relevant tests. Additionally, the efficacy of the therapy should be accurately assessed by incorporating comprehensive treatment records and follow-up data into the study.

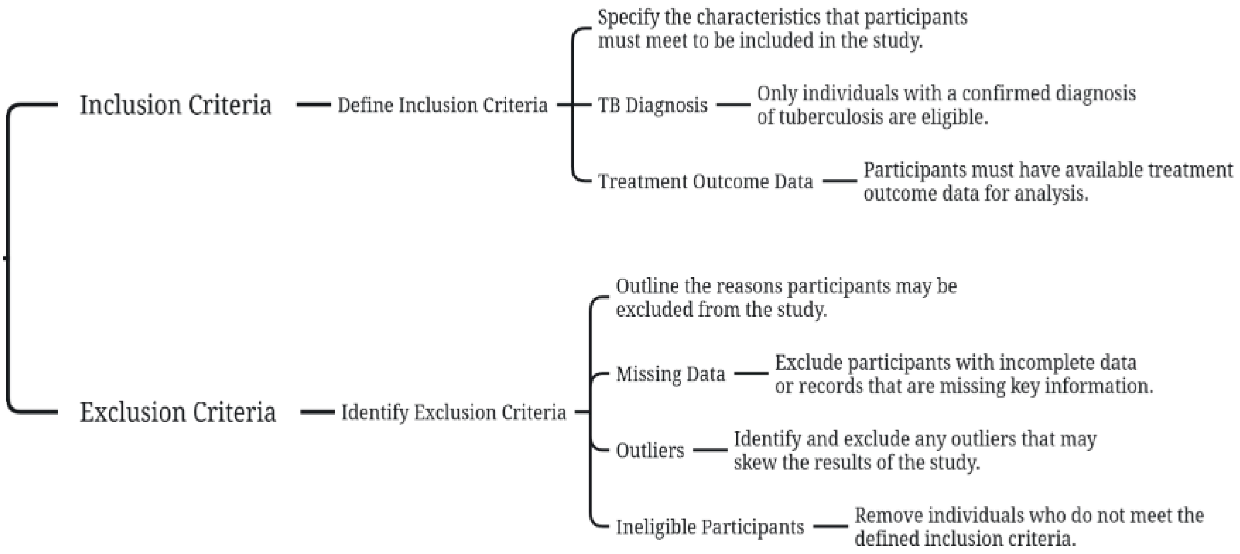

FIGURE 1. Flow chart illustrates the inclusion and exclusion criteria for the study

The exclusion criteria limit participation to those who meet the necessary inclusion standards, thus minimizing biases in the study. In research regarding TB treatment outcomes, it is essential to clearly define these exclusion criteria to ensure the inclusion of only eligible patients. Challenges such as missing data resulting from incomplete medical records, insufficient follow-up information, and inadequate documentation might bias the outcomes. Effectively managing outliers is vital to address extreme cases, statistical anomalies, and data integrity concerns that could affect the results. To maintain the study's focus, it is important to exclude ineligible participants, including those with non-TB conditions, insufficient treatment histories, and non-compliance issues. The dataset includes various variables associated with tuberculosis, such as diabetes mellitus, smoking behavior, chest X-ray results, treatment protocols, and other relevant factors (Gill et al. 2022). The treatment outcome is identified as the dependent variable, categorized into: 'Recovered', 'Dead', 'Multiple Causes', and 'Complete Treatment'. Factors including underlying conditions such as diabetes, lifestyle choices like smoking, HIV status, and tuberculosis diagnosis can significantly influence treatment effectiveness.

The dataset classifies individuals according to their smoking status, designating a numerical value of 1 for smokers and 0 for non-smokers. The negative impact of smoking on the immune system is extensively documented, resulting in increased vulnerability to numerous diseases, such as tuberculosis. Diabetes mellitus, defined by increased blood glucose levels resulting from impaired insulin production, function, or both, has been demonstrated to influence immune function (Ali et al. 2020). Researchers frequently utilize a dichotomous variable to indicate the presence or absence of diabetes and HIV. Research has demonstrated a link between diabetes and immune response, suggesting heightened susceptibility to tuberculosis in those affected. This study utilizes Machine Learning methods to classify treatment outcomes based on a dataset of tuberculosis patients. Determining an appropriate sample size is essential for ensuring the robustness and reliability of findings, especially for binary outcomes such as treatment success or failure. Predicting tuberculosis treatment outcomes requires advanced Machine Learning techniques that assess multiple patient characteristics to produce precise predictions. This study details the technical implementation and mathematical principles that form the basis of the key models utilized in this investigation. The 2017 data were imported into Pandas data frames and subsequently merged. All models were developed using data of 'Diabetes Mellitus', 'Smoking', 'Tibi Diagnosa', 'Chest X-ray status', 'Intensive Phase Treatment Regime', 'BCG scar', 'HIV Post-Diagnosa', 'DOT Intensive Phase', 'DOT Connection Phase', and 'Culture Status in Early Treatment'.

The determination of an appropriate sample size was crucial for ensuring statistical robustness in this Machine Learning study focused on tuberculosis treatment outcomes. Sample size calculation represents a critical component in research design, particularly for studies employing machine learning algorithms. The calculation was based on the following equation:

$$n = (Z^2\alpha/2 \times p \times (1-p)) / d^2$$

In this formula, 'n' represents the required sample size, while $Z^2\alpha/2$ corresponds to a 1.96 value, reflecting our chosen 95% confidence level. The expected proportion

(p) was set at 0.78, derived from previously documented tuberculosis treatment success rates in Malaysia. The study established a margin of error (d) of 5% (0.05) to ensure precise estimations. Upon applying these parameters to the formula:

$$n = (1.96^2 \times 0.78 \times 0.22) / 0.05^2$$

$$n = (3.8416 \times 0.78 \times 0.22) / 0.0025$$

$$n = 0.659955 / 0.0025$$

This calculation yielded a sample size of 1,228 participants. However, recognizing the potential for data loss and the need for robust machine learning analysis, we incorporated a 20% buffer to calculate sample size: Final sample size = 1,228 + (1,228 × 0.20) = 1474. This larger sample size provided several advantages for our Machine Learning analyses. First, it substantially reduced the margin of error in our predictions. Second, it enhanced the reliability of our predictive models by providing more comprehensive training data. Third, it allowed for appropriate data partitioning, with an 80:20 ratio for training and testing sets, respectively. The expanded dataset also enabled us to address potential class imbalances in treatment outcomes effectively. Furthermore, it provided sufficient data for thorough model validation and cross-validation procedures, essential components in machine learning methodology. This robust sample size ensured adequate statistical power to detect meaningful differences in treatment outcomes while maintaining the validity and reliability of our Machine Learning predictions. This comprehensive approach to sample size calculation and data collection has strengthened the foundation of this study, enabling more reliable and generalizable results in predicting tuberculosis treatment outcomes through machine learning applications.

The logistic function serves as an effective model for data analysis and evaluates the likelihood of an instance being classified into a specific category. The Logistic Regression model is constructed using the training data and subsequently employed to produce class predictions on the test data. The training process employs a maximum of 1000 iterations to mitigate convergence issues (Pedregopsa et al. 2011). The algorithms were assessed by comparing their label predictions with the actual labels from the 2017 dataset. The test data was utilized to produce predictions, which were then compared to the actual labels (y_test) for the calculation of evaluation metrics. The metrics accuracy_score, precision_score, and recall_score was utilized to evaluate essential metrics such as accuracy, precision, and recall. This facilitated a quantitative assessment of the efficacy of each model. Support Vector Machines (SVMs) are employed to identify the optimal hyperplanes. Hyperplanes effectively separate distinct classes and produce predictions for test data. This is achieved by training the SVM on the training data, focussing on optimizing the margin between the classes. A hyperplane is defined as a subspace with one dimension less than that of the original feature space. The method utilizes kernel techniques to distinguish between linear and nonlinear classifications. The decision tree algorithm predicts data point classifications by analyzing attribute values (Tiwari & Maji 2019).

The methodology entails dividing a dataset into smaller segments to improve its similarity or homogeneity. The random forest technique improves decision trees by generating multiple trees throughout the training process and establishing the output based on the most commonly occurring class (Lopez-Garnier, Sheen & Zimic 2019). Utilizing an ensemble method, in contrast to a single decision tree, diminishes variability and enhances accuracy AdaBoost, or adaptive boosting, functions as a meta-algorithm that constructs a robust learner by combining multiple weak learners. This method involves the iterative training of additive models, focussing more on instances misclassified by previous classifiers to correct those errors. The method exhibits superior noise resistance relative to a singular estimator. XGBoost, or Extreme Gradient Boosting, is a tree-boosting technique that exhibits enhanced effectiveness and accuracy relative to conventional gradient-boosting methods. GridSearchCV is employed to identify optimal hyperparameters for the AdaBoost and XGBoost models through an extensive search over a defined range of parameter values. This method enhances model performance by identifying optimal values for the number of estimators, learning rate, and maximum depth. The precision of different models in GridSearchCV was improved by fine-tuning hyperparameters using validation data. The dataset was divided into training (80%) and test (20%) sets to analyze tuberculosis treatment outcomes from 2017 and to develop predictive models for classifying treatment results. The training data established a basis for predicting treatment outcomes. A range of supervised learning algorithms was developed, and their effectiveness was rigorously evaluated to determine the optimal model for this clinical problem. This included Logistic Regression, Support Vector Classification (SVC), XGBoost (gradient boosting decision tree), AdaBoost, Random Forest, and GridSearchCV for hyperparameter optimization.

The study employs a randomized search method in conjunction with 5-fold cross-validation to optimize the model's hyperparameters. The assessment of the test set utilized metrics such as accuracy, precision, recall, and F1 score. The integration of these methodologies resulted in the creation of a Machine Learning pipeline designed to predict tuberculosis treatment outcomes (Figure 2). A thorough analysis was conducted through the training and evaluation of multiple models. The model with the highest accuracy score was identified as the optimal choice. The accuracy_score() function from the sklearn library was utilized to evaluate the precision of the predictions produced by this model for the target variable, based on hypothetical data points (Hrizi et al. 2022). The metrics

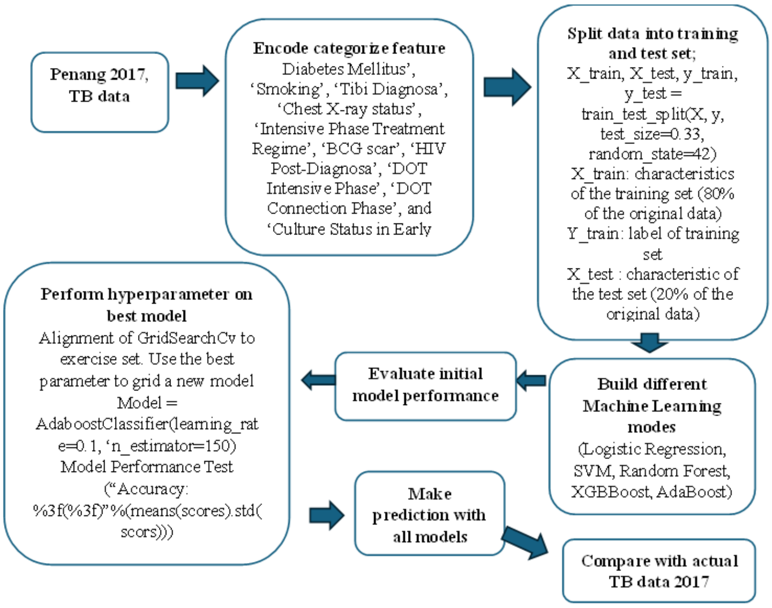

FIGURE 2. The flow chart of the main process in predicting for the year 2017

library computes the ratio of accurate forecasts to the total number of predictions. The prediction process included the extraction of feature values from samples, the multiplication of each feature value by its corresponding coefficient, the calculation of a weighted sum, and the conversion of this sum into a probability value (ranging from 0 to 1) through a sigmoid function, ultimately leading to predictions based on these probability values. The models sought to predict four key treatment outcomes: Recovery, Complete Treatment, Death, and Multiple.

RESULTS

The primary models employed comprised Logistic Regression, Support Vector Machine (SVM), XGBoost, Random Forest, and AdaBoost. The essence of the prediction process lies in the mathematical modelling of the input data and the anticipated results. For any specific patient, the input features can be represented as a vector X = [x_1, x_2, ..., x_n], where each x_i signifies a distinct patient characteristic or clinical measurement. The prediction target y represents the result of the treatment (Pedregosa et al. 2011). The primary model, XGBoost (eXtreme Gradient Boosting), achieved the highest accuracy of 68.1% after hyperparameter tuning, employing an ensemble of decision trees. The prediction formula for XGBoost can be expressed in the following manner: $\hat{y}_i = \sum_r f_r(x_i)$ Location: $\hat{y}_i$ represents the expected outcome for patient $_i$, whereas f_r indicates the particular decision trees. The feature vector corresponding to patient $_i$ is represented as x_i. The Logistic Regression model achieved an accuracy of 63.3%, and the prediction formula is; $P(y=1|X) = 1 / (1 + e^{\wedge}(-\theta^T X))$. The likelihood of achieving a successful treatment outcome is denoted as $P(y=1|X)$. θ represents the model parameters, while X functions as the input feature. The vector e signifies the natural exponential function. The optimization of the model involves minimizing the loss function: $L(\theta) = -1/m \sum_i[y_i \log(h\theta(x_i)) + (1-y_i) \log(1-h\theta(x_i))]$. In this context, m denotes the number of training examples, y_i signifies the observed outcome, and $h\theta(x_i)$ indicates the predicted probability. The process of optimizing hyperparameters for XGBoost involves carefully fine-tuning several key variables. Among these is the learning rate, which determines the size of each step taken during the gradient descent process.

Other crucial parameters include max_depth, which sets the maximum depth of the trees; n_estimators, referring to the total number of trees in the ensemble; and subsample, which indicates the percentage of samples used to construct each tree. This mathematical framework empowers models to uncover patterns in historical data and make predictions for new instances. The remarkable results achieved by the hyperparameter-tuned XGBoost model demonstrate its effectiveness in showing the complex relationships found in TB treatment outcome data. The prediction process concludes with a probability threshold, typically set at 0.5, where: If P(outcome) is greater than or equal to 0.5, a positive outcome is anticipated. Should the probability of the outcome be below 0.5, it is advisable to forecast a negative result. Table 1 provides a summary of the performance of these models. This comprehensive mathematical framework provides a solid foundation for predicting TB treatment outcomes, though continuous improvements to these models could yield better results moving forward. The models exhibited a satisfactory level of accuracy (60-70%), yet there is potential for improvement, particularly in terms of specificity and the detection of true negatives. This aspect highlights the importance of continuous investigation and advancement to enhance the predictive capabilities of these models and to fine-tune their effectiveness in identifying true negatives.

The Machine Learning algorithm underwent training with a dataset comprising X_train and y_train, and was then utilized to forecast treatment outcomes for the testing set (X_test). The accuracy of the model was assessed through the accuracy_score function, and the precision and recall metrics were obtained using the precision_score and recall_score functions, respectively. Among the models analyzed, Logistic Regression and Random Forest demonstrated the most significant test accuracies, approaching a value of 0.65. This suggests that the previously mentioned models exhibited a notable degree of accuracy in predicting treatment outcomes in approximately 65% of cases. The results indicate that the model demonstrates a significant degree of accuracy in forecasting treatment outcomes, with an estimated success rate of 65.93% observed in the test set. Precision and recall serve as essential metrics for evaluating the performance of a classification model.

Precision acts as a measure to assess the correctness of optimistic case predictions by determining the ratio of correctly predicted positive cases to the overall number of positive cases. The Logistic Regression model indicates an estimated precision of approximately 69.51%. The findings of the study show that approximately 69.51% of the anticipated positive treatment outcomes were effectively achieved. On the other hand, recall functions as a quantitative measure that assesses the proportion of correctly predicted positive instances relative to the total number of actual positive cases. The anticipated recall rate for the Logistic Regression model is approximately 65.93%. The results indicate that the model achieves a classification accuracy of around 65.93% in correctly identifying positive treatment outcomes (Figure 3). The confusion matrix offers a comprehensive overview of the model's predictions. A confusion matrix is instrumental in differentiating and categorizing true negatives (tn), false positives (FP), false negatives (fn), and true positives (tp). The specificity of the model is determined by calculating the ratio of true negatives to the total of true negatives and false positives. The Logistic Regression model produced a specificity of around 16.67%. At the outset, the accuracy of AdaBoost on the test set was recorded at 0.57, indicating the least effective performance among the algorithms analyzed. The Logistic Regression model exhibited superior performance for the specified dataset, as evidenced by the comparatively lower observed precision. The optimal model was selected due to its exceptional accuracy scores observed throughout the model selection process.

The analysis of the Confusion Matrix shows that the model demonstrated a significant accuracy in forecasting the treatment outcome 'Recovered', successfully identifying 176 out of the 285 total observed cases (Figure 4). Nonetheless, the model faced difficulties in precisely predicting the outcomes of 'Complete Treatment', achieving alignment in only 38 out of 122 forecasts with the actual results. The models demonstrate a modest level of specificity, estimated at around 16.7%. Although a test accuracy of 0.65 was attained, the Support Vector Machine (SVM) exhibited a significant drawback in specificity, resulting in a notably low value of 0.16. This indicates that the Support Vector Machine (SVM) demonstrated a considerable rate of erroneous predictions, as it incorrectly categorized negative events as positive. The SVM model achieved an accuracy of around 65.2% on the test set. The findings suggest that the model successfully forecasted treatment outcomes in around 65.2% of instances. The model demonstrates a precision of around 67.7% and a recall of about 65.2%. In the comparison of the two models, both Logistic Regression and SVM achieved similar levels of precision and recall.

Despite this, the Logistic Regression model was slightly more accurate than the SVM model. Predictions of treatment efficacy were only somewhat reliable using the Logistic Regression and Support Vector Machine (SVM) models (Figure 4). Nonetheless, there remains room for development, especially in terms of making accurate predictions about the results of 'Complete Treatment'. It is possible that the model's effectiveness could be improved with more research and development work. Implementing optimization techniques or investigating other algorithms is one possible path toward improving the model's performance. The probability distributions show that the models consistently overestimate the 'Recovered' category and underestimate the 'Complete Treatment' category. When comparing actual and predicted class sizes, we found that all the models struggled to forecast the smaller classes correctly. Both the Logistic Regression and random forest models consistently underestimate the prevalence of the 'Recovered' class while consistently overestimating the frequency of another significant class.

The XGBoost model demonstrated an accuracy of 0.648, suggesting that it successfully predicted the treatment outcomes for about 64.8% of the instances in the testing dataset. The precision score, denoting the model's capacity to detect affirmative cases accurately, was recorded as 0.681. The recall score, denoting the model's capacity to identify all positive examples correctly, was recorded as 0.648 (Table 1). These scores offer valuable data regarding the overall performance of the model and its capacity to categorize various treatment outcomes accurately. A Hyperparameter Optimization technique was employed to enhance the performance of the XGBoost model. Hyperparameters are a set of predetermined parameters that

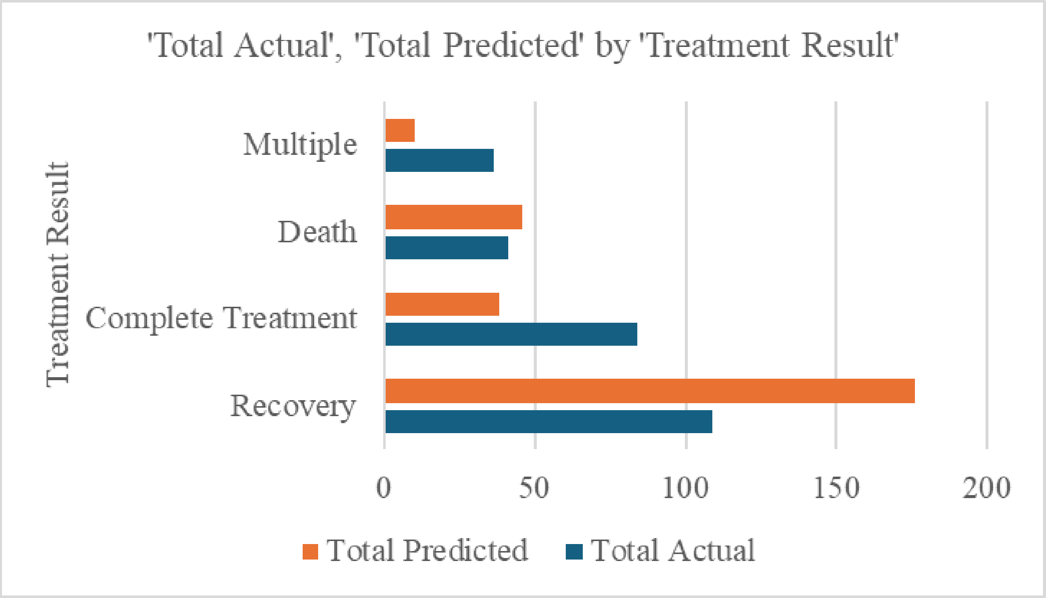

FIGURE 3. Predicting treatment result outcome using logistic regression model

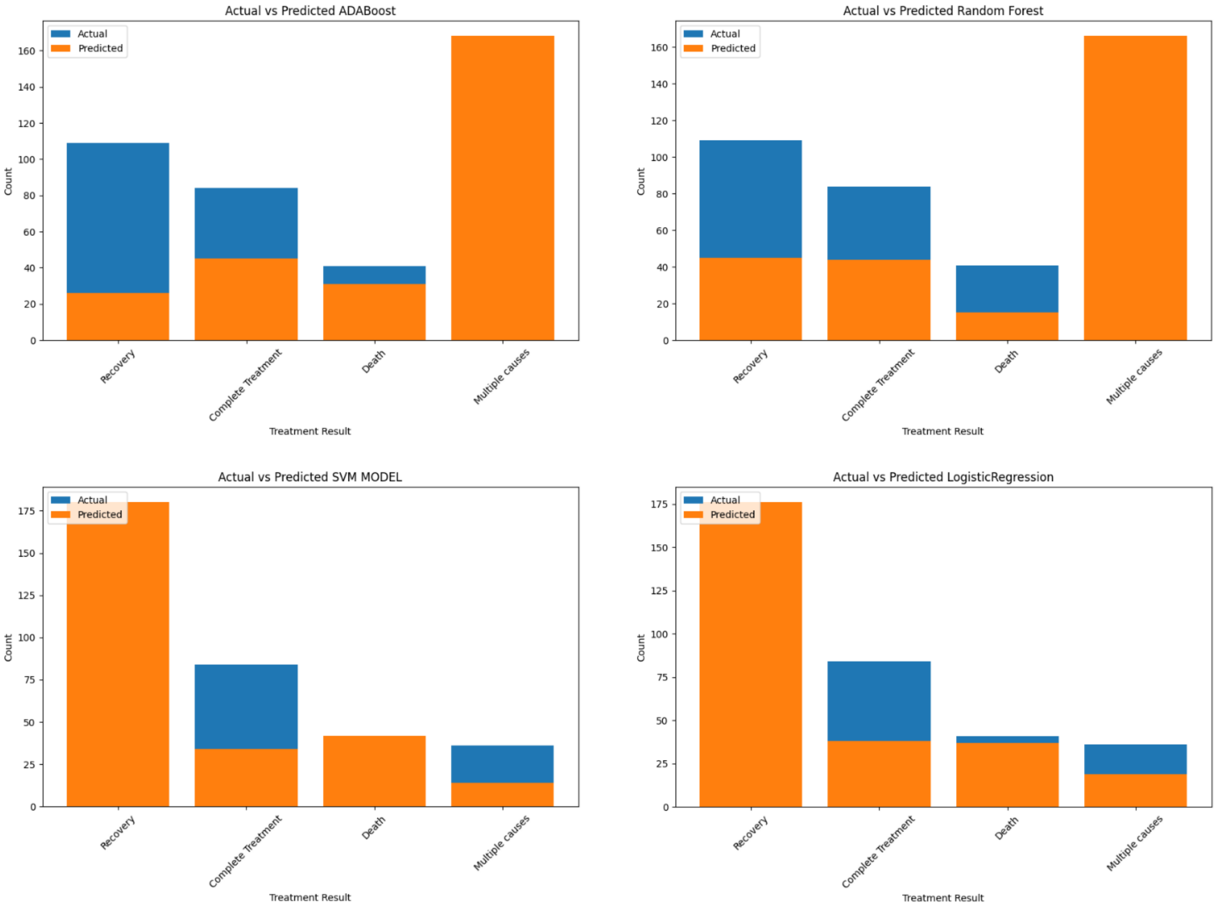

FIGURE 4. A comparative analysis of predictive accuracy in four models for predicting treatment outcomes in year 2017

TABLE 1. Model performance on hypothetical test data

Model	Accuracy	Precision	Recall	Specificity
Logistic Regression	66%	0.6950627075484054	0.6592592592	0.0
XGBoost	64% & 68% (hyper tunning)	0.6788911881	0.65788515552	0.1 & 0.61 (hyper tuning)
Random Forest	62%	0.627438	0.6333333	0.1
SVM	65%	0.6771250722511226	0.651851851851	0.1666
AdaBoost	57%	0.5398111351151	0.41	0.0

are established before the training process begins. Unlike other parameters, hyperparameters are not learned from the available data. The optimization of hyperparameters has the potential to significantly impact the performance of the model. The provided code sample showcases using the GridSearchCV class from the *scikit-learn* library for a grid search across a specified range of hyperparameters (Pedregosa et al. 2011). The hyperparameters that underwent optimization encompassed the learning rate, maximum depth of trees, and the number of estimators. The grid search method comprehensively assesses the model's performance by systematically testing all possible combinations of hyperparameters. The objective is to identify the optimal collection of hyperparameters that result in the highest level of performance. Following the hyperparameter optimization process, the XGBoost model exhibited an accuracy of 0.640, signifying a marginal performance improvement compared to the initial model (Figure 5). The precision score from the table improved to 0.679, however, the recall score stayed the same at 0.640. The study's findings suggest that the hyperparameter optimization technique had a notable impact on enhancing the model by improving both its overall accuracy and precision.

The model extracts feature values from a new sample before predicting. The coefficient of each feature value is multiplied to obtain a weighted sum. The feature's category 'score' is the sum of those weights. The total of all relevant attribute scores determines a sample's class score. The overall score is converted to a probability value from 0 to 1 using the sigmoid function. This function weights probabilities near 1 more. The anticipated likelihood increases with the score sum. The code may also emphasize model explanation and visualization above general predictions. The code uses 20 samples to show projected probability plots and explain the analysis. Visualization skills are as vital as modelling skills. The algorithm may have been designed to anticipate from a few real-world data points from the start. The code may additionally insist on teaching modelling and graphical methods. The model was not designed to make huge projections when it was installed and coded. We use small hypothetical samples to evaluate the code's predictions and visualizations while developing the prototype.

Model testing and improvement are needed before it may be utilized for numerous forecasts. Using a small number of samples for testing helps troubleshooting and finding code faults and restrictions easier. The algorithm only utilizes 20 fake instances to show its predictions and visualizations for numerous reasons. Explaining modelling and visualization approaches is prioritized over making general predictions. The core concepts can be explained more clearly with a smaller data set. Knowing that a 20-person sample has limits is crucial. This category includes overfitting, inadequate representativeness, and statistical significance. To verify model accuracy and conduct thorough performance evaluations. Model accuracy is satisfactory but might be better, showing room for improvement. The data has complex nonlinear interactions that Logistic Regression cannot represent. The models had precision and recall ratings of 60-70%, indicating good accuracy in identifying positive and negative situations. The models have low specificity, making it difficult to identify true negatives. Furthermore, the implementation of a hyperparameter tuning procedure to enhance XGBoost performance highlights a novel approach to refine model calibration, as such a targeted improvement strategy has the possibility to augment predictions and better determine patients who may be at heightened risk of undesirable outcomes like relapse or mortality (Xie et al. 2020).

DISCUSSION

Machine Learning models can enhance patient care and inform public health strategies in the treatment of tuberculosis (TB). By pinpointing patients who are at a higher risk for treatment failure, healthcare providers can implement enhanced monitoring protocols and tailored treatment strategies. This proactive approach increases success rates in TB treatment and minimizes the chances of drug resistance, leading to better patient outcomes. To implement this strategy more broadly, several actions can be taken, including the integration of predictive models with Electronic Health Records (EHR), expanding the

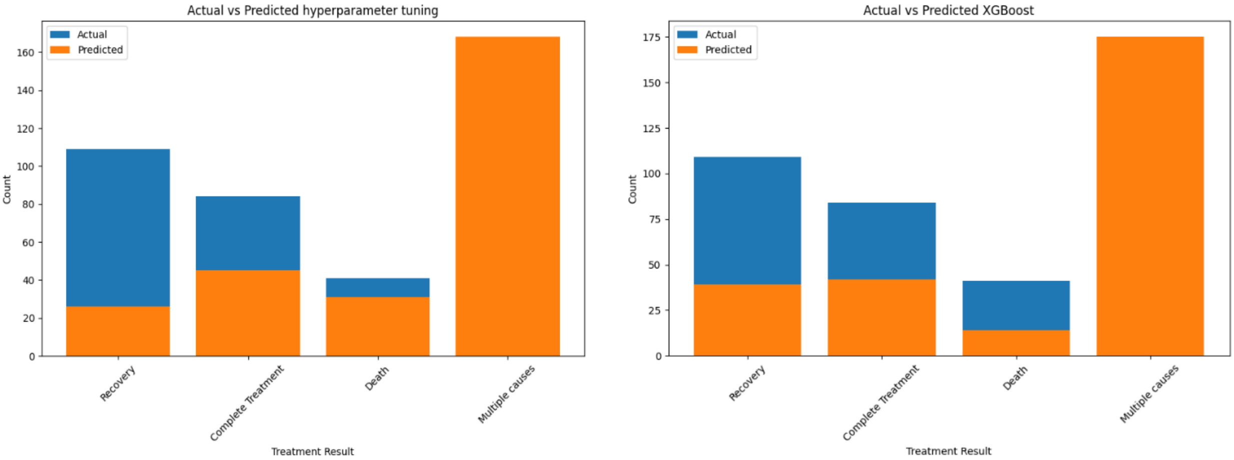

FIGURE 5. Comparison of prediction results for XGBoost Model: After hyperparameter tuning (left) and before hyperparametertuning (right)

investigation to include data from various healthcare facilities across different regions, and fostering collaboration among multiple centres to deepen our understanding of the factors influencing treatment outcomes. This study analysed four unique models: Logistic Regression, Random Forest, Gradient Boosting, and Neural Networks. Every model presents unique advantages: Logistic Regression offers clear interpretability, Random Forest shows remarkable performance, Gradient Boosting achieves the highest predictive accuracy, and Neural Networks are adept at managing large and intricate datasets. The results indicate that Gradient Boosting is the most effective model for predicting TB treatment outcomes, owing to its ability to comprehend complex, non-linear relationships within the data and its resilience against overfitting.

Moreover, this study also introduces an innovative method for enhancing TB treatment using sophisticated machine learning techniques, particularly emphasizing the application of XGBoost and various predictive models in Penang, Malaysia. The rise of drug-resistant TB strains and the intricacies of treatment protocols have required the implementation of more advanced strategies for patient management. The hyperparameter-tuned XGBoost model proved to be the most effective option, attaining a notable accuracy rate of 68.1%. This performance notably surpassed conventional methods, with Decision Trees attaining 63.7% accuracy and Logistic Regression achieving 63.3%. The effective execution of these models necessitates a thorough strategy for integrating healthcare systems, which involves building a strong infrastructure for data collection, setting up standardised reporting protocols, and designing user-friendly interfaces for healthcare providers.

Collaborative efforts across multiple centres create opportunities for enhancing the dataset and refining model accuracy among varied populations. When selecting a model, it is essential to consider various factors, including interpretability, computational resources, and the requirements of the clinical environment. The integration of Machine Learning models in tuberculosis treatment presents considerable opportunities for reducing costs and enhancing patient outcomes. Precise prediction models facilitate improved initial treatment decisions, which may lower the incidence of treatment failure and decrease the reliance on costly second-line therapies. Future initiatives ought to concentrate on expanding the range to include larger and more diverse datasets, conducting practical implementation studies, and facilitating the incorporation of these models into current healthcare frameworks. The primary objective was to develop a reliable, clinically validated instrument that enhances tuberculosis treatment outcomes while optimizing resource efficiency across diverse healthcare environments.

CONCLUSION

To effectively implement the potential benefits of using Machine Learning techniques to improve tuberculosis medication, researchers will need to carefully address key obstacles in future studies and undertake rigorous empirical evaluations in real-world settings. The primary problem that must be fixed is the absence of comprehensive datasets that contain all relevant data. The current study only used hypothetical datasets with a small number of individuals, even though numerous factors may influence the outcomes of therapy. Increasing the number of available real-world datasets and including a more diverse set of patient characteristics, risk factors, and diseases is crucial for enhancing model robustness and precision. Correcting the issue of insufficient external validation of models using distinct datasets encompassing varied demographics and circumstances is essential for ensuring accuracy and generalizability (Ahmad et al. 2024a, 2024b, 2024c, 2024d, 2024e, 2024f, 2024g, 2024h; Ariffin et al. 2024; Bismelah et al. 2024; Chabo et al. 2024; Jubit et al. 2023, 2024a, 2024b; Marzuki et al. 2023, 2024; Masron et al. 2024; Zakaria et al. 2023). Machine Learning systems can predict tuberculosis treatment success (Takarinda et al. 2017). Methodically tackling research challenges and undertaking rigorous empirical evaluations in real-world scenarios could bring Machine Learning's potential improvements to tuberculosis treatment to clinical practice. The main challenge is the unavailability of complete datasets with all relevant facts. Although several factors may affect therapy outcomes, this study only examined hypothetical datasets with few participants. To improve model stability and accuracy, more real-world datasets and more patient features, risk factors, and diseases are needed. Correcting poor external validation of models through separate datasets encompassing varied demographics and scenarios is critical for accuracy and generalizability.

ACKNOWLEDGMENTS

This study was approved by the National Medical Research Register of Malaysia (NMRR) under registration numbers NMRR-14-998-21596 and NMRR-19-447-447-447. The study is also associated with the Long-term Grant Scheme (LRGS): Evaluation of the National TB Prevention and Control Programme towards Achieving the Millennium Development Goals (MDG) (203.PSK.6722006). Their dedication to information sharing and the advancement of knowledge is genuinely admirable.

REFERENCES

Abdullahi, O.A., Ngari, M.M., Sanga, D., Katana, G. & Willetts, A. 2019. Mortality during treatment for tuberculosis; a review of surveillance data in a rural county in Kenya. *PLoS ONE* 14(7): e0219191. https://doi.org/10.1371/journal.pone.0219191

Ahmad, A., Kelana, M.H., Soda, R., Jubit, N., Mohd Ali, A.S., Bismelah, L.H. & Masron, T. 2024a. Mapping the impact: Property crime trends in Kuching, Sarawak, during and after the COVID-19 period (2020-2022). *Indonesian Journal of Geography* 56(1): 127-137. https://doi.org/10.22146/ijg.90057

Ahmad, A., Masron, T., Jubit, N., Redzuan, M.S., Soda, R., Bismelah, L.H. & Mohd Ali, A.S. 2024b. Analysis of the movement distribution pattern of violence crime in Malaysia's capital region-Selangor, Kuala Lumpur, and Putrajaya. *International Journal of Geoinformatics* 20(2): 11-26. https://doi.org/10.52939/ijg.v20i2.3061

Ahmad, A., Masron, T., Junaini, S.N., Barawi, M.H., Redzuan, M.S., Kimura, Y., Jubit, N., Bismelah, L.H. & Mohd Ali, A.S. 2024c. Criminological insights: A comprehensive spatial analysis of crime hot spots of property offenses in Malaysia's urban centers. *Forum Geografi: Indonesian Journal of Spatial and Regional Analysis* 38(1): 94-109. https://doi.org/10.23917/forgeo.v38i1.4306

Ahmad, A., Masron, T., Junaini, S.N., Kimura, Y., Barawi, M.H., Jubit, N., Redzuan, M.S., Bismelah, L.H. & Mohd Ali, A.S. 2024d. Mapping the unseen: Dissecting property crime dynamics in urban Malaysia through spatial analysis. *Transactions in GIS* 28(6): 1486-1509. https://doi.org/10.1111/tgis.13197

Ahmad, A., Masron, T., Kimura, Y., Barawi, M.H., Jubit, N., Junaini, S.N., Redzuan, M.S., Mohd Ali, A.S. & Bismelah, L.H. 2024e. Unveiling urban violence crime in the state of The Selangor, Kuala Lumpur and Putrajaya: A spatial–temporal investigation of violence crime in Malaysia's key cities. *Cogent Social Sciences* 10(1): 2347411. https://doi.org/10.1080/23311886.2024.2347411

Ahmad, A., Masron, T., Mohd Ali, A.S., Barawi, M.H., Nordin, Z.S., Abg Ahmad, A.I., Redzuan, M.S. & Bismelah, L.H. 2024f. Exploring the potential of geographic information system (GIS) application for understanding spatial distribution of violent crime related to United Nations sustainable development goals-16 (SDGS-16). *Journal of Sustainability Science and Management* 19(9): 35-63. https://doi.org/10.46754/jssm.2024.09.003

Ahmad, A., Masron, T., Mohd Ali, A.S., Kimura, Y. & Junaini, S.N. 2024g. Demographic dynamics and urban property crime: A linear regression analysis in Kuala Lumpur and Putrajaya (2015-2020). *Planning Malaysia: Journal of the Malaysian Institute of Planners* 22(4): 302-319. https://doi.org/10.21837/pm.v22i33.1550

Ahmad, A., Masron, T., Ringkai, E., Barawi, M.H., Salleh, M.S., Jubit, N. & Redzuan, M.S. 2024h. Analisis ruangan *hot spot* jenayah pecah rumah di negeri Selangor, Kuala Lumpur dan Putrajaya pada tahun 2015-2020. *Geografia-Malaysian Journal of Society and Space* 20(1): 49-67. https://doi.org/10.17576/geo-2024-2001-04

Ali, A., Alrubei, M.A.T., Hassan, L.F.M., Al-Ja'afari, M.A.M. & Abdulwahed, S.H. 2020. Diabetes diagnosis based on KNN. *IIUM Engineering Journal* 21(1): 175-181. https://doi.org/10.31436/iiumej.v21i1.1206

Ariffin, N.A., Wan Ibrahim, W.M.M., Rainis, R., Samat, N., Mohd Nasir, M.I., Abdul Rashid, S.M.R., Ahmad, A. & Zakaria, Y.S. 2024. Identification of trends, direction of distribution and spatial pattern of tuberculosis disease (2015-2017) in Penang. *Geografia-Malaysian Journal of Society and Space* 20(1): 68-84. https://doi.org/10.17576/geo-2024-2001-05

Bismelah, L.H., Masron, T., Ahmad, A., Mohd Ali, A.S. & Echoh, D.U. 2024. Geospatial assessment of healthcare distribution and population density in Sri Aman, Sarawak, Malaysia. *Geografia-Malaysian Journal of Society and Space* 20(3): 51-67. https://doi.org/10.17576/geo-2024-2003-04

Bukundi, E.M., Mhimbira, F., Kishimba, R., Kondo, Z. & Moshiro, C. 2021. Mortality and associated factors among adult patients on tuberculosis treatment in Tanzania: A retrospective cohort study. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases* 24: 100263. https://doi.org/10.1016/j.jctube.2021.100263

Chabo, D., Masron, T., Jubit, N. & Ahmad, A. 2024. Analisis corak ruangan keciciran murid sekolah menengah di Sarawak. *Malaysian Journal of Social Sciences and Humanities* 9(9): e002906. https://doi.org/10.47405/mjssh.v9i9.2906

Dheda, K., Perumal, T., Moultrie, H., Perumal, R., Esmail, A., Scott, A.J., Udwadia, Z., Chang, K.C., Peter, J., Pooran, A., von Delft, A., von Delft, D., Martinson, N., Loveday, M., Charalambous, S., Kachingwe, E., Jassat, W., Cohen, C., Tempia, S., Fennelly, K. & Pai, M. 2022. The intersecting pandemics of tuberculosis and COVID-19: Population-level and patient-level impact, clinical presentation, and corrective interventions. *The Lancet Respiratory Medicine* 10(6): 603-622. https://doi.org/10.1016/S2213-2600(22)00092-3

Fayaz, S.A., Babu, L., Paridayal, L., Vasantha, M., Paramasivam, P., Sundarakumar, K. & Ponnuraja, C. 2024. Machine learning algorithms to predict treatment success for patients with pulmonary tuberculosis. *PLoS ONE* 19(10): e0309151–e0309151. https://doi.org/10.1371/journal.pone.0309151

Gichuhi, H.W., Magumba, M., Kumar, M. & Mayega, R.W. 2023. A Machine Learning approach to explore individual risk factors for tuberculosis treatment non-adherence in Mukono district. *PLOS Glob Public Health* 3(7): e0001466. https://doi.org/10.1371/journal.pgph.0001466

Gill, C.M., Dolan, L., Piggott, L.M. & McLaughlin, A.M. 2022. New Developments in Tuberculosis Diagnosis and Treatment. *Breathe*, 18(1): 210149. https://doi.org/10.1183/20734735.0149-2021

Hrizi, O., Gasmi, K., Ben Ltaifa, I., Alshammari, H., Karamti, H., Krichen, M., Ben Ammar, L. & Mahmood, M.A. 2022. Tuberculosis disease diagnosis based on an optimized Machine Learning model. *Journal of Healthcare Engineering* 2022: 8950243. https://doi.org/10.1155/2022/8950243

Hussain, O. A., & Junejo, K. N. 2018. Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models. *Informatics for Health and Social Care* 44(2): 135–151. https://doi.org/10.1080/17538157.2018.1433676

Janssens, R.J., Mourão-Miranda, J. & Schnack, H.G. 2018. Making individual prognoses in psychiatry using neuroimaging and Machine Learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3(9): 798-808. https://doi.org/10.1016/j.bpsc.2018.04.004

Jubit, N., Masron, T., Ahmad, A. & Soda, R. 2024a. Investigating the spatial relation between landuse and property crime in Kuching, Sarawak through location quotient analysis. *Forum Geografi: Indonesian Journal of Spatial and Regional Analysis* 38(2): 153-166. https://doi.org/10.23917/forgeo.v38i2.4575

Jubit, N., Masron, T., Redzuan, M.S., Ahmad, A. & Kimura, Y. 2024b. Revealing adolescent drug trafficking and addiction: Exploring school disciplinary and drug issues in the Federal Territory of Kuala Lumpur and Selangor, Malaysia. *International Journal of Geoinformatics* 20(6): 1-12. https://doi.org/10.52939/ijg.v20i6.3327

Jubit, N., Masron, T., Puyok, A. & Ahmad, A. 2023. Geographic distribution of voter turnout, ethnic turnout and vote choices in Johor state election. *Geografia-Malaysian Journal of Society and Space* 19(4): 64-76. https://doi.org/10.17576/geo-2023-1904-05

Kouchaki, S., Yang, Y., Walker, T.M., Sarah Walker, A., Wilson, D.J., Peto, T.E.A., Crook, D.W., CRyPTIC Consortium & Clifton, D.A. 2019. Application of Machine Learning techniques to tuberculosis drug resistance analysis. *Bioinformatics* 35(13): 2276-2282. https://doi.org/10.1093/bioinformatics/bty949

Lopez-Garnier, S., Sheen, P. & Zimic, M. 2019. Automatic diagnostics of tuberculosis using convolutional neural networks analysis of MODS digital images. *PLoS ONE* 14(2): e0212094. https://doi.org/10.1371/journal.pone.0212094

Marzuki, A., Bagheri, M., Ahmad, A., Masron, T. & Akhir, M.F. 2024. Examining transformations in coastal city landscapes: Spatial patch analysis of sustainable tourism - A case study in Pahang, Malaysia. *Landscape and Ecological Engineering* 20: 513-545. https://doi.org/10.1007/s11355-024-00613-w

Marzuki, A., Bagheri, M., Ahmad, A., Masron, T. & Akhir, M.F. 2023. Establishing a GIS-SMCDA model of sustainable eco-tourism development in Pahang, Malaysia. *Episodes* 46(3): 375-387. https://doi.org/10.18814/epiiugs/2022/022037

Masron, T., Ahmad, A., Jubit, N., Sulaiman, M.H., Rainis, R., Redzuan, M.S., Junaini, S.N., Jamian, M.A.H., Mohd Ali, A.S., Salleh, M.S., Zaini, F., Soda, R. & Kimura, Y. 2024. *Crime Map Book*. Centre for Spatially Integrated Digital Humanities (CSIDH), Faculty of Social Sciences and Humanities, Universiti Malaysia Sarawak. https://www.researchgate.net/publication/384572873_Crime_Map_Book

Miotto, R., Li, L., Kidd, B.A. & Dudley, J.T. 2016. Deep patient: An unsupervised representation to predict patients' future from the electronic health records. *Scientific Reports* 6: 26094. https://doi.org/10.1038/srep26094

Nicholson, T.J., Hoddinott, G., Seddon, J.A., Claassens, M.M., van der Zalm, M.M., Lopez, E., Bock, P., Caldwell, J., Da Costa, D., de Vaal, C., Dunbar, R., Du Preez, K., Hesseling, A.C., Joseph, K., Kriel, E., Loveday, M., Marx, F.M., Meehan, S.A., Purchase, S., Naidoo, K., Naidoo, L., Solomon-Da, C.F., Sloot, R., Osman, M. 2023b. A systematic review of risk factors for mortality among tuberculosis patients in South Africa. *A Systematic Review* 12(1): 23. https://doi.org/10.1186/s13643-023-02175-8

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. 2011. *Scikit-learn*: Machine Learning in Python. *Journal of Machine Learning Research* 12: 2825-2830. https://dl.acm.org/doi/10.5555/1953048.2078195

Takarinda, K.C., Sandy, C., Masuka, N., Hazangwe, P., Choto, R.C., Mutasa-Apollo, T., Nkomo, B., Sibanda, E., Mugurungi, O., Harries, A.D. & Siziba, N. 2017. Factors associated with mortality among patients on TB treatment in the Southern Region of Zimbabwe, 2013. *Tuberculosis Research and Treatment* 2017: 6232071. https://doi.org/10.1155/2017/6232071

Tiwari, A. & Maji, S. 2019. Machine Learning techniques for tuberculosis prediction. *International Conference on Advances in Engineering Science Management & Technology (ICAESMT) - 2019, Uttaranchal University, Dehradun, India*. https://ssrn.com/abstract=3404486 or http://dx.doi.org/10.2139/ssrn.3404486

World Health Organization. 2023. *Tuberculosis*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/tuberculosis

World Health Organisation. 2022. *Global Tuberculosis Report 2022*. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022

Xie, Y., Han, J., Yu, W., Wu, J., Li, X. & Chen, H. 2020. Survival analysis of risk factors for mortality in a cohort of patients with tuberculosis. *Canadian Respiratory Journal* 2020: 1654653. https://doi.org/10.1155/2020/1654653

Xiong, Y., Ba, X., Hou, A., Zhang, K., Chen, L., & Li, T. 2018. Automatic detection of mycobacterium tuberculosis using artificial intelligence. *Journal of Thoracic Disease* 10(3): 1936–1940. https://doi.org/10.21037/jtd.2018.01.91

Yang, S., Zhu, F., Ling, X., Liu, Q. & Zhao, P. 2021. Intelligent health care: Applications of deep learning in computational medicine. *Frontiers in Genetics* https://doi.org/10.3389/fgene.2021.607471

Zakaria, Y.S., Ahmad, A., Said, M.Z., Epa, A.E., Ariffin, N.A., M Muslim, A., Akhir, M.F. & Hussin, R. 2023. GIS and oil spill tracking model in forecasting potential oil spill-affected areas along Terengganu and Pahang coastal area. *Planning Malaysia: Journal of the Malaysian Institute of Planners* 21(4): 250-264. https://doi.org/10.21837/pm.v21i28.1330

*Corresponding author; email: zarika27@gmail.com