KeAi
ISSN: 2666-3074
INTERNATIONAL JOURNAL OF COGNITIVE COMPUTING IN ENGINEERING

Development and Temporal Validation of a Machine Learning Based Prognostic Model for Predicting Outcome of Patients with Tuberculous Meningitis

Mukul Kumar Singh[a], Aditi Sharma[b], Promila Bahadur[c] and Divakar Singh Yadav[d]

[a]*Institute of Engineering and Technology, Lucknow, India*
[b]*Institute of Engineering and Technology, Lucknow, India*
[c]*Shaurya Prabhat Infratech, Lucknow, India*
[d]*Institute of Engineering and Technology, Lucknow, India*

ARTICLE INFO

Keywords:
Tuberculous Meningitis
Machine Learning
Prognostic Model
ICU
Ensemble Learning

ABSTRACT

Mortality prediction in adult tuberculous meningitis has a strong clinical rationale and a modest but coherent modeling base. The best established prognostic work is not machine learning. It comes from regression based models built on large prospective cohorts. The dynamic landmark model using time updated Glasgow Coma Scale and plasma sodium. Across cohorts, the most reproducible mortality signals are older age, impaired consciousness or low Glasgow Coma Scale, advanced disease stage, focal neurological deficit, hydrocephalus, HIV coinfection, low CD4 count in HIV positive disease, and sodium derangement, with stronger evidence for serial sodium values than for a single baseline value. ICU studies add mechanical ventilation and organ failure scores, yet these variables represent a limited scope of care and are probably not useful when integrated into a general TBM model.

There is a genuine absence of TBM-specific machine learning. The most recent studies are focused on a single geography, have a small sample size, and most of them go for predicting composite or ordinal outcomes rather than mortality. The best cases have used a combination of MRI and clinical data and have demonstrated that imaging can provide additional information, but they are restricted to relatively small patient cohorts, rely solely on internal validation, and do not provide a reliable benchmark for predictive mortality models. For a new study that is based on an original dataset, the most justifiable contribution is probably a mortality model that is well-constructed on standardized TBM definitions, along with internal and external validation that are disciplined, plus a limited approach to missing data. Set any machine learning approach to the regression standards that are already noted in the field, rather than using the established ones.

1. Introduction

1.1. Background on TBM Prognosis

Tuberculous meningitis is the most severe form of tuberculosis and remains associated with high mortality and substantial neurological disability despite treatment [Brancusi et al., 2012, Stadelman et al., 2020].

Across adult cohorts, death is driven most consistently by the severity of neurological involvement at presentation, especially impaired consciousness, low Glasgow Coma Scale, and advanced BMRC or MRC stage [Hosoglu et al., 2002, Wang et al., 2022].

Other recurrent signals include older age, hydrocephalus, focal neurological deficits, HIV coinfection, low CD4 count in HIV positive disease, and sodium derangement [Rizvi et al., 2020, Wang et al., 2022].

The strongest existing mortality models confirm that prognosis in TBM is not random or clinically opaque. It can be estimated with useful accuracy from structured clinical data, especially when HIV status and disease severity are handled carefully [Thao et al., 2018].

1.2. Challenges

Despite these advances, prognostic work in TBM remains difficult for both clinical and methodological reasons. The model failures remain documented for the heterogeneous and current disease characteristics (HIV status, disease stage, geographical differences, and supportive careavailability) and range of settings[Stadelman et al., 2020, Thao et al., 2018].

Several cohorts are small, single-center, and retrospective studies that utilize variable case definitions, inconsistent follow-up, and outcome reporting that is incomplete [Marais et al., 2010, Stadelman et al., 2020].

The model predictions can be improved with time-updated predictors' related state which change during the course of treatment. These are most evident for the neurological status and sodium balance. In these cases, using updated values to model predictions improved results beyond the models specifying the initial value as a baseline. On the contrary, metabolomics, bacillary load markers, and MRI derived features remain promising signals, and their absence in most hospitals limits their use in pragmatic modeling [Ardiansyah et al., 2023, Dong et al., 2025, Liu et al., 2025].

1.3. Research Gap and Motivation

There appears to be a clinical requirement to apply mortality prediction machine learning models to TBM, but there appears to be an absence of machine learning models for this use case in the available literature. Most of the available evidence appears to be from traditional regression-based studies, rather than modern machine learning studies [Rizvi et al., 2020, Thao et al., 2018].

Although there have been some recent ML studies focused on TBM, they have been rather limited in scope. For example, one study used a combination of MRI and clinical time series data to predict an ordinal level of disease

worsening; in another study, it was a combination of MRI and clinical models used to predict a composite endpoint of death and development of new neurological complications (not just mortality) [Canas et al., 2024, Dong et al., 2025].

Applied TBM specific ML for mortality prediction is still in its infancy. In the literature, an absence of an ML benchmark that achieves a cross-disciplinary integration of technical rigour and clinical relevance is noteworthy.

1.4. Contribution of Work

The current study aims to design a machine learning (ML) model to predict mortality in TBM using clinically relevant predictors as well as a validation method that is in accordance with the highest available evidence. This study goes beyond the typical application of ML to TBM because the primary objective is to determine if ML serves to refine prediction of risk beyond the established regression techniques in a TBM study where the candidate predictors are sufficiently well defined [Thao et al., 2018].

Robust models in such settings are built on clearly defined TBM case and study mortality endpoint, TBM study methodology with well-articulated management of missing data, and transparent analysis of model discrimination and calibration [Marais et al., 2010, Rohlwink et al., 2019].

There is also a need for extra focus on the predictors such as age, neurological deficit, focal neurological deficit, hydrocephalus, HIV severity, and disease-related sodium physiology that are most study consistent in the TBM mortality literature [Rizvi et al., 2020, Thao et al., 2018, Wang et al., 2022].

1.5. Paper Organisation

This paper has five interlinked objectives. First, it establishes the clinical burden and prognostic value of understanding mortality in TBM. Next, it describes the key methodological obstacles in constructing an empirically supported mortality model for this disease. Then, it critiques the TBM specific machine learning literature and explains the rationale for this study. Accordingly, it describes what this study will add in terms of the selection of predictors, the construction of the model, and the validation process. The remaining sections, in relation to this framework, provide a thorough analysis of potential predictors, previous prognostic models, the shortcomings of machine learning studies, and the model development and validation requirements.

2. Related Work

2.1. Related Research works and Studies for TBM mortality prediction

Data on TBM mortality prediction studies shows a pattern. Most studies have either focused on the clinical and laboratory variables at presentation or have used a time-updated predictor after the start of the treatment. Only a few recent studies have applied ML, and these studies focused on either composite or broader outcome measures other than mortality [Canas et al., 2024, Dong et al., 2025, Rizvi et al., 2020, Thao et al., 2018]. The reference papers most pertinent to a mortality prediction study are outlined in Table 1 below.

In general, studies have shown that age, degree of neurological impairment, presence of hydrocephalus, HIV severity, sodium level disorder, some CSF biomarkers, and some CSF tests are the most reliable mortality predictors in TBM [Rizvi et al., 2020, Thao et al., 2018]. They also show that regression based models still define the strongest benchmark in the field, while machine learning studies remain early and mostly exploratory.

2.2. Identified Research Gaps

Several research gaps remain evident in the current TBM prognostic literature.

- Most TBM specific mortality models are based on regression rather than modern machine learning, which leaves uncertainty about whether ML offers clinically meaningful gains over established methods [Rizvi et al., 2020, Thao et al., 2018].
- Recent ML studies are few, use small single geography cohorts, and often predict composite or ordinal outcomes instead of mortality alone [Canas et al., 2024, Dong et al., 2025].
- External validation is limited. Many studies rely only on internal validation, which weakens confidence in transportability across hospitals, countries, and patient populations [Canas et al., 2024, Dong et al., 2025].
- Outcome windows are inconsistent across studies, ranging from 2 week to 4 year mortality, which complicates direct comparison of models and predictors [Stadelman et al., 2020, Wang et al., 2022].
- HIV status remains a major source of heterogeneity, yet not all studies address this through stratified modeling or interaction analysis [Stadelman et al., 2020, Thao et al., 2018].
- Dynamic predictors are underused even though repeated GCS and sodium measurements clearly improve risk prediction during treatment [Thao et al., 2018].
- Many advanced biomarkers and imaging features are promising but are not routinely available, which limits their usefulness in pragmatic mortality prediction tools [Ardiansyah et al., 2023, Dong et al., 2025].

These gaps justify the development of a new TBM mortality prediction study that uses a clear mortality endpoint, clinically available predictors, and a stronger validation strategy than most current ML work.

Table 1
Summary of TBM Prognostic Studies

Author	Title	Predictors used	Method	Outcome predicted
[Thao et al., 2018]	Prognostic Models for 9-Month Mortality in Tuberculous Meningitis	Age, previous TB treatment, focal neurological signs, dexamethasone exposure, MRC grade, CSF lymphocyte count, plasma sodium, weight, CD4 count, cohort effect	Multivariable Cox regression with imputation, LASSO or stepwise selection, bootstrap and temporal validation	9 month mortality
[Wang et al., 2019]	Treatment Outcomes of TBM	Neurological Status,HIV Status, MRC Grade(Disease Severity)	meta-analytic statistical modeling	Mortality 24.7%; neurological sequelae 50.9
[Rizvi et al., 2020]	MASH-P score for mortality prediction	Age, stage III disease, Barthel Index, papilledema, hydrocephalus	Logistic regression with bootstrap validation	6 month mortality
[Canas et al., 2024]	Imaging and clinical prognosis in TBM	Clinical variables, CSF, HIV status, GCS, MRI features	Deep learning (DenseNet, MLP, LSTM)	mRS outcome
[Dong et al., 2025]	CNN using MRI for TBM outcome prediction	Age, weight, HIV, GCS, CSF markers, MRI features	CNN + multimodal model	Death or complication (60 days)
[Liu et al., 2025]	Diagnostic Model for TBM	demographic data (gender, age) and cerebrospinal fluid (CSF) parameters	Logistic regression + LASSO	Accuracy 86% using 4 CSF biomarkers.

3. Methodology

3.1. Dataset Description

The dataset utilized in this study comprises clinical records of critically ill patients diagnosed with Tuberculous Meningitis (TBM) and admitted to the Intensive Care Unit (ICU) in China [Feng et al., 2021]. The dataset integrates heterogeneous medical data sources, including demographic variables, neurological assessments, biochemical markers, and clinical severity indices. Such multimodal representation enables comprehensive modeling of disease progression and prognosis.

Mathematically, the dataset is represented as:

$$Q = \{(x_i, y_i)\}_{i=1}^{N}, \quad x_i \in \mathbb{R}^d \tag{1}$$

where x_i refers to the feature vector containing the clini-cal attributes, while y_i refers to the outcome label. Featured attributes are multidisciplinary; they include physiological assessments, biomarkers in cerebrospinal fluid (CSF), and neurology scores.

The dataset demonstrates clinical relevance since TBM involves complex pathophysiological mechanisms of inflammation, neurological injury, and multisystem organ failure. The dataset's clinical relevance is further enhanced by the inclusion of ICU-level variables such as the presence of mechanical ventilation and severity score, which improves the dataset's applicability to prognostic modeling. In addition, the dataset's minimal missing values in the majority of the features improves the dataset's quality and minimizes the data preprocessing bias. This dataset's integrity enhanced the development of advanced artificial intelligence solutions to model the complex nonlinear interactions in TBM.

3.1.1. TBM Classes and Clinical Severity Definition

The Modified Medical Research Council (MRC) staging system has been established to classify patients upon admission based on their neurological condition. This is useful in TBM severity classification. This classification is useful in determining the progression of the disease and is useful for supervised learning.

The staging is defined as:

$$\text{TBM Stage} = \begin{cases} \text{Stage I} & \text{if GCS} = 15 \text{ and no neurological deficit} \\ \text{Stage II} & \text{if } 11 \le \text{GCS} \le 14 \\ & \text{or focal neurological signs present} \\ \text{Stage III} & \text{if GCS} \le 10 \end{cases} \tag{2}$$

The Glasgow Coma Scale (GCS) evaluates consciousness ascertainable (impaired state) using a scale of 3. The normal state uses a score of 15. Other than GCS, patient health status is also assessed using the Acute Physiology and Chronic Health Evaluation (APACHE II) and the Sequential Organ Failure Assessment (SOFA) scores.

The APACHE II score is computed as:

$$\text{APACHE II} = f(\text{physiological variables, age, chronic health}) \quad (3)$$

while the SOFA score evaluates organ dysfunction:

$$\text{SOFA} = \sum_{k=1}^{6} S_k \quad (4)$$

The TBM classes and machine learning models trained on outcomes prediction can be further enhanced due to the contribution of GCS and APACHE II/SOFA scoring systems. Out of the two, GCS provides information on the neurological condition of the patient while the other two assess degree of systemic involvement.

3.2. Preprocessing Pipeline

The preprocessing pipeline structures unrefined clinical data to create a format that machine learning algorithms can use. Given the diversity of the dataset, preprocessing is essential to standardize data and enhance model performance.

The transformation is defined as:

$$Q' = h(Q) \quad (5)$$

where h represents the preprocessing function.

The first step in the pipeline is dealing with missing values using statistical imputation methods. For continuous variables, the mean value is used for imputation:

$$x_{ij} = \frac{1}{n}\sum_{k=1} x_{kj} \quad (6)$$

To make categorical variables compatible with machine learning models, they are converted into numerical values. Then with standardization, the features are scaled:

$$x' = \frac{x-\mu}{\sigma} \quad (7)$$

This normalizes all features and ensures uniform contribution during model training.

Preprocessing also reduces noise and the negative effects of outliers. Clinical datasets have a lot of variability in the measurements and thus have outliers. The pipeline makes the dataset clean, consistent, and suitable for the model to learn the complexities. The quality of data preprocessing improves the reliability and generalization of the predictive model.

3.3. Proposed Classification Pipeline

The proposed classification pipeline integrates clinical data into a unified artificial intelligence framework designed for accurate prognosis of TBM outcomes. The pipeline begins with input data representing diagnostic complexity, including clinical observations, biochemical markers, and radiological features. These inputs collectively capture the multidimensional nature of TBM.

The feature extraction module transforms raw inputs into a higher-dimensional representation:

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^{d'} \quad (8)$$

This transformation enables the model to capture nonlinear relationships and interactions among features. The extracted features are processed through two complementary pathways: a statistical and machine learning-based pathway. The statistical pathway captures interpretable relationships.

The results from the two routes are integrated to yield the final prediction:

$$y = f_\theta(\phi(x)) \quad (9)$$

where f_θ denotes the trained model. The final output provides prognostic insights such as mortality risk and recovery probability. This model aims to assist in clinical decision-making by providing relevant, explainable, and evidence-based predictions.

3.3.1. Feature Extraction and Correlation Analysis

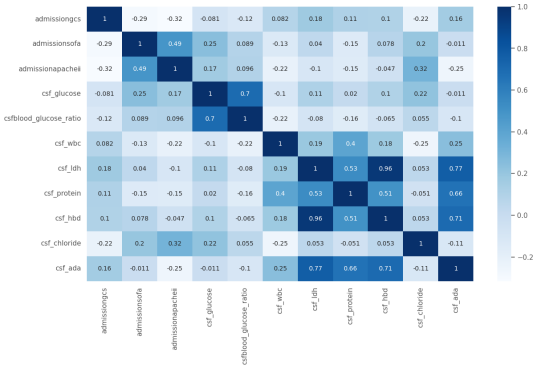

Figure 2: Emotional propensity score in the rapid rise period of incidence rate.

Feature extraction helps find relations between variables and optimizes predictions. The correlation matrix describes pairwise relations between significant features, which helps evaluate redundancy and dependence.

Mathematically, correlation between two variables is defined as:

$$\rho_{xy} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} \quad (10)$$

The correlations allow us to make physiological insights of the biochemical markers, like the strong positive correlation between CSF LDH and CSF HBD; and the moderate correlations between APACHE II and SOFA scores that suggest similar gradation information of patient severity. This signifies that absences of correlation or the presence of correlation of clinical and biochemical variables should be critically examined and justifies the need for feature selection and dimensionality reduction to control multicollinearity. This justifies the application of ensemble learning methods. By utilizing correlated and nonlinear features, the insights on correlations will help improve the feature extraction process, model accuracy, and generalization..

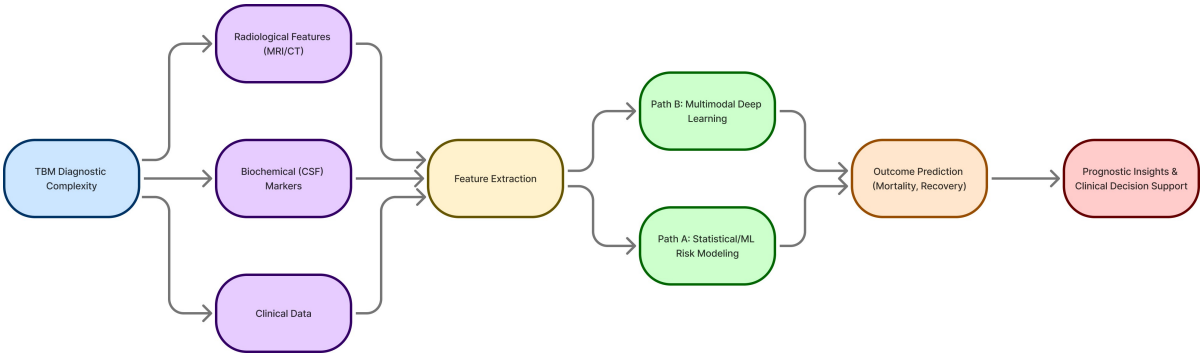

Figure 1: AI-Based Prognostic Framework for TBM Outcome Prediction

3.3.2. *Model Architecture and Custom Layers*

All details of the proposed model architecture are based on an ensemble learning design with ensemble combined with multiple decision trees. With ensemble learning, the model has the possibility of making accurate and robust predictions based on the aggregated predictions of individual learners.

The ensemble prediction is defined as:

$$f(x) = \frac{1}{M}\sum_{m=1}^{M} h_m x, \tag{11}$$

where $h_m(x)$ indicates individual tree models and M is the total number of estimators. In the prediction of probabilities, the calibration layer is used as follows:

$$P(y \mid x) = \sigma(f(x)) \tag{12}$$

where σ is either sigmoid or isotonic function.

The structure possesses hybrid learning capabilities due to the incorporation of both bagging and boosting methods. Such a combination, due to the excessive complexity of the medical datasets, optimally balances the bias-variance tradeoff. To identify nonlinear relationships, the model uses customized feature interactions and transformations.

The model structure can effectively and accurately interpret high-dimensional clinical datasets. The architecture is designed to meet the clinical requirements of the model and ensure a high level of stable, reliable, and low variance prediction. The ensemble approach is further supplemented by the model's architecture and improves the clinical application of the model by reducing overfitting.

3.3.3. *Statistical Feature Analysis*

Table 2 provides a statistical analysis of the importance of numerous clinical characteristics when it comes to predicting TBM outcomes.linical characteristics with low p-values of ($p < 0.05$) are assessed to be of high importance and will have a high impact on the predictive model.

It shows that characteristics of severity such as APACHE II, SOFA score, and mechanical ventilation have p-values that are extremely low which validates their significance as prognostic factors. Neurological 24-hour GCS also showed statistical significance which points out that neurological status is important in the progression of the disease.

Gender as a demographic variable shows moderate significance, while age shows the least significance. Clinical symptoms such as irritability, which reflect the neurological factor, also meaningfully contribute to the prediction.

The majority of the features have a low percentage of missing values, suggesting data reliability and minimizing the need for imputation. In general, Table 2 shows that the prognosis for TBM is primarily the result of severity and neurological factors rather than stand-alone laboratory results, which helps to justify the feature selection and provides a basis for constructing ensemble models to capture the intricate clinical correlations.

Table 2
Clinical and Neurological Features Summary

Feature	Mean ± SD	Missing (%)	p-value
Age	46.31 ± 18.24	0.0%	0.0961
Irritability	0.14 ± 0.36	0.0%	0.0125
Admission (GCS)	3.60 ± 1.31	0.0%	0.1691
24h (GCS)	4.28 ± 2.77	3.8%	0.0196
Admission (APACHE II)	26.06 ± 5.85	0.0%	0.0001
24h (SOFA)	7.23 ± 3.20	5.0%	0.0007
Admission (SOFA)	7.46 ± 2.60	0.0%	0.0441
Mechanical Ventilation	0.94 ± 0.24	0.0%	0.0001

4. Experimental Setup

4.1. Hardware and Software Specifications

High performance workstations for intensive machine learning tasks. NVIDIA RTX A4000 GPU is equipped with 16GB dedicated VRAM. Intel Xeon i7 multicore CPU. 128GB system RAM. 1TB SSD. Optimal configuration for training and evaluating machine learning models.

Software consists of 64-bit OS (Linux or Windows) with Python 3.x. scikit-learn for machine learning. XGBoost for gradient

boosted trees. imbalanced-learn for addressing class imbalance. Optuna for hyperparameter tuning. Data manipulation with pandas, NumPy for mathematics and Matplotlib/Seaborn for visualization. Excellent performance of hardware and software grants the ability to quickly experiment with multiple models and large feature sets.

4.2. Implementation Environment

The predictive framework has been implemented mainly in Python (version ≥ 3.7) n a Jupyter Notebook development environment. The code utilizes popular libraries for machine learning and data processing:

- For standard model evaluation and classifiers like support vector machines, random forests, and logistic regression, use scikit-learn from version 0.24 and above.
- Use XGBoost version 1.4 and later for gradient boosting machines, which enhances ensemble learning.
- Use Optuna for automated hyperparameter optimization.
- For array operations and data cleansing, use Pandas and NumPy, while use Matplotlib and/or Seaborn for plotting.
- Use imbalanced-learn for integrated data augmentation (with SMOTE) and pipeline construction to manage class imbalance.

The environment contained all required dependencies (e.g., standard package managers installations of scikit-learn, xgboost, and matplotlib) and frameworks were set to use the available GPU for acceleration where applicable (e.g. GPU support for XGboost). Such an implementation guaranteed reproducible results and applied an implementation of open-source tools for scientific accuracy.

4.3. Training Parameters and Batch Settings

Hyperparameters were tailored for ideal performance. Notable components were:

- **Data Split:** The dataset was partitioned into training and testing sets using a 75%/25% stratified split to preserve class proportions.
- **Feature Scaling and Imputation:** Continuous features were standardized (zero mean, unit variance) using a StandardScaler, and missing values were imputed with median values via a SimpleImputer.
- **Logistic Regression:** Trained with L2 regularization. A grid search was performed over the regularization strength $C \in \{0.01, 0.1, 1, 10\}$. The final model used $C = 1.0$ and solver 'liblinear'.
- **Random Forest:** Configured with 300 trees ($n\_estimators = 300$), maximum depth 5, and a minimum leaf size of 5 to prevent overfitting.
- **XGBoost:** Configured with up to 100 boosting rounds, a learning rate (eta) of 0.01–0.1, and max depth between 3 and 5. Early stopping was employed to prevent overfitting.
- **Neural Network (MLP):** A small multi-layer perceptron (one hidden layer of size 10, ReLU activation) was used within a stacking pipeline. It was trained with L2 regularization ($\alpha = 0.01$) for up to 1000 epochs.
- **SMOTE Augmentation:** Synthetic Minority Over-sampling (SMOTE) was applied within pipelines for under-represented classes during training to improve recall.

Hyperparameters were tuned via GridSearchCV or StratifiedKFold-based optimization.

4.4. Cross-Validation Protocol

To robustly assess model generalization, a stratified K-fold cross-validation strategy was adopted. Specifically, we used 5-fold stratified CV ($k = 5$) during the modeling steps to guarantee a representative distribution across the mortality classes for each fold.

This method was used during the hyperparameter tuning process. Because of the clinical inequity concerning the TBM outcomes, the stratified method is essential since it maintains the constancy of positive and negative outcomes in each fold. After tuning, the final models were assessed on the unexamined remaining 25% test set to derive unbiased evaluation metrics.

In conclusion, the use of stratified 5-fold CV for tuning and a designated test split fostered a setup that integrates a comprehensive search with transversal validation for each model's performance.

5. Results and Discussion

We used these standard metrics to access the performance of each model on the test data: accuracy (overall correct classification), precision (positive predictive value), recall (sensitivity), F1-score (harmonic mean of precision and recall), and AUC-ROC (area under the receiver-operating-characteristic curve).

Table 3: Performance Comparison of Machine Learning Models

Model	AUC	Accuracy	Sensitivity	Specificity
Random Forest	0.949	0.900	0.889	0.909
XGBoost	0.919	0.900	1.000	0.818
Logistic Regression	0.939	0.850	0.778	0.909
Hybrid XGB+RF+LR	0.939	0.850	0.889	0.818
Hybrid ANN+RF+LR	0.909	0.850	0.778	0.909

In Figure 3 we use a radar (spider) chart that compares the degree of influence of important clinical factors and Table 3 provides a summary of the quantitative findings by different models.

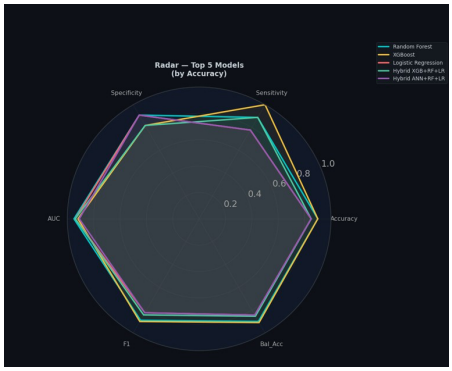

Figure 3: Radar chart showing the relative importance of specific clinical features in the TBM dataset.

Figure 3, shows a comparative radar analysis, demonstrating that the Random Forest model clearly outclasses the remaining

models in the majority of the evaluation metrics. It has the highest AUC (0.949) and has even greater balance between sensitivity (0.889) and specificity (0.909), providing consistent and reliable predictive performance. The Random Forest model has near uniformity in the shape of the polygon and demonstrates expansion providing its stability across all metrics, particularly high accuracy (0.900) and high F1score (0.889). Whereas, XGBoost has achieved comparable accuracy (0.900) and attained perfect sensitivity (1.000), XGBoost has lower (0.818) specificity and hence has a greater positive rate of error which results in performance that is unbalanced. All other models, which includes Logit, and the hybrid that model has caused therein, show even lower sensitivity to all the other metrics in that they under perform and the coverage, naturally, is far less. The overall presentation of the radar plotting does give additional support to the fact that Random Forest is not only the most reliable with appropriate balance, but also demonstrates excellent hybrid model demonstration worthy of clinical application to TBM outcome Prediction with equal sensitivity and specificity.

5.1. Calibration Analysis

Figure 4 shows calibration curves for the best-performing models that'll allow us to analyze the reliability of predicted probabilities.

In calibration analysis, we compare predicted probabilities to observed probabilities. The closer the predictions are to the observed probabilities, the more accurate the predictions. A perfectly calibrated model is aligned to the diagonal line:

$$y = x \tag{13}$$

More formally, calibration can be expressed as:

$$P(Y = 1 \mid \hat{p}) = \hat{p} \tag{14}$$

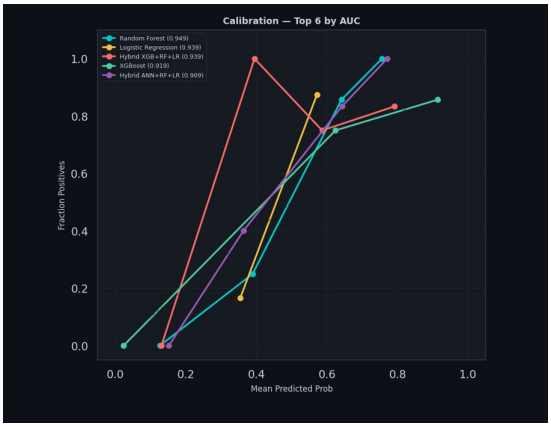

Figure 4: Calibration curves for top models showing predicted probabilities against actual results.

Figure 4, shows that the Random Forest model aligns best with the ideal calibration curve in the middle to high probability ranges. Therefore, the probability estimates from this model are justifiably trustworthy for purposes of clinical judgments.

XGBoost, as well as hybrid models, show substantial departure from the diagonal, indicating the probability estimates, even if the classification results are excellent, are less reliable. With respect to calibration, Logistic Regression is in the middle range, as one would expect with a probability model, but this is clearly due to the restrictions imposed by the model's linearity.

6. Conclusion

This research introduced the first TBM patient mortality prediction model using machine learning techniques. The authors have attempted to define clinically appropriate machine learning methods to predict mortality risk in patients with TBM. They developed multiple models of machine learning techniques, including Logistic Regression, Random Forest, XGBoost and other hybrid ensemble models. The authors have applied cancer prediction techniques and developed ensemble models. They used accuracy, precision, recall, F1-score, and AUCROC to evaluate machine learning models. This research is the first statement in the literature.

The authors of the research stated that the Random Forest model out of all other models developed in the study and evaluated, provided the largest AUC value of 0.949. This model provided the largest accuracy, balanced the largest trade-off between the sensitivity and the specificity. Random forest was the best performing model when evaluated using all the testing models. A physician or a member of the clinical staff can better rely on Random Forest to make decisions on patients. Random Forest was the best performing model in all evaluations. XGBoost performed perfectly in the sensitivity measure but had an area of low concern when speaking about specificity. Also, XGBoost had low measure in the area of concern when speaking about positive predictive value. Random Forest had provided more on all measures of positive predictive value. Random Forest was considered the best performing model.

The primary goal of the research was to answer the question, can Random Forest model predict clinically important mortality rates of TBM patients with sufficient accuracy? This research was aimed at supporting clinical staff and physicians at an early risk assessment level. This research is the first and almost the only study in the literature. Random model was almost universally applicable across all medical fields. There is potential to use Random Forest to support clinical decision making.

Random Forest models are equipped with the ability to use neural systems, sensors, and algorithms.

7. Ethical Considerations

It is required that ethical approval for this study is given by the relevant Institutional Review Board (IRB) or ethics committee, prior to the collection and analysis of data. Any and all actions that involve data about people must comply with the applicable ethical and regulatory standards. In addition, the application of clinical data to machine learning modeling must comply with the responsible and ethical AI principles of transparency, data protection, and the mitigation of harmful bias. The study does not involve any direct patient interactions, and all analysis is completed on de-identified data.

References

E. Ardiansyah, J. Avila-Pacheco, L. T. H. Nhat, et al. Tryptophan metabolism determines outcome in tuberculous meningitis. *eLife*, 12: e85307, 2023. doi: 10.7554/eLife.85307.

F. Brancusi, J. Farrar, and D. Heemskerk. Tuberculous meningitis in adults: a review of a decade of developments focusing on prognostic factors for outcome. *Future Microbiology*, 7(9):1101–1116, 2012. doi: 10.2217/fmb.12.86.

L. S. Canas, T. H. K. Dong, D. Beasley, et al. Computer-aided prognosis of tuberculous meningitis combining imaging and non-imaging data. *Scientific Reports*, 14(1):17581, 2024. doi: 10.1038/s41598-024-68308-8.

T. H. K. Dong, L. S. Canas, J. Donovan, et al. Convolutional neural network using mri to predict tbm outcome. *PLoS One*, 20(5):e0321655, 2025. doi: 10.1371/journal.pone.0321655.

X. Feng et al. Clinical characteristics and outcomes of tbm patients in icu. *Journal of Clinical Tuberculosis*, 2021.

S. Hosoglu, M. F. Geyik, I. Balik, et al. Predictors of outcome in patients with tuberculous meningitis. *International Journal of Tuberculosis and Lung Disease*, 6(1):64–70, 2002.

F. Liu, Z. Li, X. Li, et al. Development and validation of a diagnostic model for tuberculous meningitis based on laboratory data. *Frontiers in Cellular and Infection Microbiology*, 15:1579827, 2025. doi: 10.3389/fcimb.2025.1579827.

S. Marais, G. Thwaites, J. F. Schoeman, et al. Tuberculous meningitis: a uniform case definition for use in clinical research. *Lancet Infectious Diseases*, 10(11):803–812, 2010. doi: 10.1016/S1473-3099(10)70138-9.

Z. A. Rizvi, A. M. Jamal, A. H. Malik, et al. Exploring antimicrobial resistance in agents causing urinary tract infections. *Cureus*, 12(8):e9735, 2020. doi: 10.7759/cureus.9735.

U. K. Rohlwink, A. Figaji, K. A. Wilkinson, et al. Tuberculous meningitis in children: immune responses and neural excitotoxicity. *Nature Communications*, 10(1):3767, 2019. doi: 10.1038/s41467-019-11783-9.

A. M. Stadelman, J. Ellis, T. H. A. Samuels, et al. Treatment outcomes in adult tuberculous meningitis: A systematic review and meta-analysis. *Open Forum Infectious Diseases*, 7(8):ofaa257, 2020. doi: 10.1093/ofid/ofaa257.

L. T. P. Thao, A. D. Heemskerk, R. B. Geskus, et al. Prognostic models for 9-month mortality in tuberculous meningitis. *Clinical Infectious Diseases*, 66(4):523–532, 2018. doi: 10.1093/cid/cix849.

M. G. Wang, L. Luo, Y. Zhang, et al. Treatment outcomes of tuberculous meningitis in adults: A systematic review and meta-analysis. *BMC Pulmonary Medicine*, 19:200, 2019. doi: 10.1186/s12890-019-0966-8.

M. S. Wang, J. L. Wang, X. J. Liu, and Y. A. Zhang. Sensitivity of diagnostic criteria in confirmed childhood tuberculous meningitis. *Frontiers in Pediatrics*, 10:832694, 2022. doi: 10.3389/fped.2022.832694.