

ABSTRACT

Tuberculosis (TB) is a pulmonary infectious disease that can affect all age groups, from children to the elderly, and its prevalence remains high in Indonesia. The disease is caused by *Mycobacterium tuberculosis* and is characterized by typical symptoms such as a prolonged productive cough lasting more than two weeks, sometimes accompanied by hemoptysis, weight loss, loss of appetite, and shortness of breath. This study aims to develop and evaluate a TB classification model using a Gradient Boosting algorithm based on clinical patient data at the Gayamsari Community Health Center. The dataset consisted of 255 patient observations with two classification classes, namely TB Positive and TB Negative. The factors used in the classification process included sex, age, fever, weight loss, loss of appetite, shortness of breath, chest pain, fatigue, cough duration, and type of cough. The dataset was divided into training and testing sets using an 80:20 ratio. The modeling process involved hyperparameter optimization using Grid Search and Optimization Using Tree-structured Parzen Estimator (OPTUNA). Model evaluation was conducted using a confusion matrix and ROC–AUC. The Gradient Boosting algorithm was selected due to its ability to handle nonlinear relationships among variables in clinical data. Model interpretability was performed using SHapley Additive exPlanations (SHAP) to identify the global contribution of variables and Local Interpretable Model-Agnostic Explanations (LIME) to provide local explanations for a single patient observation. The results showed that the Gradient Boosting model achieved an initial F1-score of 96%, after hyperparameter optimization using Grid Search and OPTUNA, the F1-score decreased to 95% for both methods, indicating that the baseline model parameters had already provided optimal performance. SHAP analysis revealed that cough duration, age, loss of appetite, and weight loss were the variables with the greatest contribution to the TB classification process. These findings indicate that the use of the Gradient Boosting algorithm combined with SHAP and LIME produces an accurate and interpretable classification model, with the potential to support clinical decision-making in TB diagnosis.

Keywords: Tuberculosis, Gradient Boosting, Hyperparameter Optimization, SHAP, LIME, Clinical Classification