INTI JOURNAL | Vol.2025, Issue 3, No.5
eISSN:2600-7320

Lung Cancer Classification Using Stacking Framework of BiLSTM, Logistic Regression, and XGBoost

Muhammad Nikho Dwi Putra[1*], Zaenuddin[1], Silvia Ratna[1], Haldi Budiman[1], Erfan Karyadiputra[1], Tri Wahyu Qur'ana[1], Desy Ika Puspitasari[1], Galih Mahalisa[1], Nur Arminarahmah[1]

[1]Faculty of Information Technology, Islamic University of Kalimantan Muhammad Arsyad Al-Banjari, Indonesia

***Email**: mnikhodwp125@gmail.com

Abstract

Lung cancer remains one of the most prevalent and deadly cancers worldwide, causing over 1.8 million deaths each year. Early and accurate classification of lung cancer is crucial, yet existing machine learning and deep learning models often face limitations in generalization and reliability. To address this issue, this study proposes a stacking framework that integrates Bidirectional Long Short-Term Memory (BiLSTM) and Logistic Regression as base learners, with Extreme Gradient Boosting (XGBoost) serving as the meta-learner. The rationale for this approach is that BiLSTM captures complex feature interactions, Logistic Regression provides interpretability, and XGBoost has demonstrated strong performance as a meta-learner in ensemble studies. The framework was evaluated on a publicly available lung cancer dataset consisting of 309 patient records with 15 clinical and lifestyle attributes. Experimental results showed that the stacking framework achieved perfect accuracy of 1.00, outperforming BiLSTM (0.95) and Logistic Regression (0.93). These findings confirm the effectiveness of the proposed ensemble in overcoming the weaknesses of individual models and highlight its novelty as a reliable approach for lung cancer classification.

Keyword

Lung Cancer Classification, Stacking Framework, BiLSTM, Logistic Regression, XGBoost

Introduction

Lung cancer is one of the leading causes of cancer-related mortality worldwide, responsible for approximately 1.8 million deaths annually (WHO, 2025). Early and accurate diagnosis plays a crucial role in improving patient survival, yet the complexity and heterogeneity of clinical and radiological data remain major challenges (Siegel et al., 2020). In recent years, machine learning (ML) and deep learning (DL) methods have demonstrated significant potential in medical image analysis, particularly for cancer detection and classification (Litjens et al., 2017; Shen et al., 2019).

Submission: 23 June 2025; **Acceptance:** 7 September 2025; **Available online:** September 2025

Several recent studies have explored different approaches to lung cancer classification. For instance, Maurya et al. (2024) compared multiple ML algorithms and reported that K-Nearest Neighbors (KNN) achieved the highest accuracy of 92.86%, outperforming SVM and Random Forest. Similarly, Ahammed et al. (2021) demonstrated that deep learning models, especially convolutional neural networks (CNNs), could outperform traditional ML approaches in handling high-dimensional medical datasets. More recently, Xu et al. (2023) investigated ensemble-based methods and showed that combining heterogeneous classifiers improves robustness and reduces misclassification rates in cancer prediction tasks. However, while these studies highlight the effectiveness of individual ML or DL models, they often suffer from limited generalization when applied to diverse datasets.

To address these limitations, ensemble learning techniques, particularly stacking frameworks, have emerged as promising solutions. Stacking leverages the strengths of multiple base learners and employs a meta-learner to capture higher-level predictive patterns. Recent work by Zhang et al. (2022) demonstrated that stacking significantly enhances classification accuracy in biomedical applications. Meanwhile, Syapotro et al. (2024), although applied in the field of water level prediction, provided evidence of the effectiveness of XGBoost as a meta-learner in a stacking structure. This reinforces the rationale for employing XGBoost in the proposed framework to refine predictions from BiLSTM and Logistic Regression.

This study introduces a novel stacking framework that combines Bidirectional Long Short-Term Memory (BiLSTM) and Logistic Regression as base learners, with Extreme Gradient Boosting (XGBoost) serving as the meta-learner. The proposed approach aims to overcome the weaknesses of individual models while leveraging their unique strengths to improve the accuracy and reliability of lung cancer classification.

Methodology

In this section, the techniques that were utilized in creation of the proposed lung cancer classification model, which combines BiLSTM and Logistic Regression base learners in a stacking ensemble with XGBoost as the meta-learner, are explained. The methodology includes four basic elements:

Dataset and Preprocessing

The dataset used in this study was obtained from Kaggle and contains 309 patient records with 15 clinical and lifestyle attributes along with a binary target variable indicating lung cancer. Data cleaning involved removing duplicates, applying one-hot encoding for gender, and label encoding the target variable. Features were standardized with StandardScaler, and the dataset was split into 80 percent training and 20 percent testing sets. For the BiLSTM model, the feature matrix was reshaped into a three-dimensional format to meet sequential input requirements.

BiLSTM Model

The BiLSTM model was implemented in Keras with a sequential architecture. The input layer receives the reshaped features of size (15, 1) followed by two bidirectional LSTM layers with 64 and 32 units respectively. Batch normalization and dropout layers (dropout rate = 0.2) were applied to reduce overfitting. The output layer used a dense layer with 3 neurons and softmax activation, which corresponds to the categorical target encoding. The model was trained using the Adam optimizer with a learning rate of 0.001, sparse categorical cross-entropy loss, and

early stopping (patience = 3) to monitor validation loss. The research by Zhou et al. (2022) indicated that BiLSTM is effective in capturing temporal dependencies and has been successfully applied in medical data classification tasks, making it a suitable choice for this study.

Logistic Regression Model

The Logistic Regression model was implemented using scikit-learn with the solver set to lbfgs and a maximum iteration of 1000 to ensure convergence. It was trained on the same preprocessed dataset without reshaping, serving as a simple and interpretable baseline. The model was evaluated based on its ability to separate lung cancer cases using linear decision boundaries. The research by Aditya et al. (2024) indicated that Logistic Regression is efficient for binary classification in biomedical applications, particularly when working with smaller datasets, which supports its inclusion as a base learner in this framework.

Stacking with XGBoost

Predictions from the BiLSTM and Logistic Regression models were combined into stacked feature vectors, which were then used to train the meta-learner. The XGBoost classifier was chosen due to its strong performance in ensemble learning and ability to model non-linear relationships. In this study, the classifier was configured with the multi:softmax objective, num_class = 2, and mlogloss as the evaluation metric, ensuring compatibility with the binary classification task. The stacked framework exploits the complementary strengths of BiLSTM and Logistic Regression while leveraging XGBoost's robustness in handling complex decision boundaries, consistent with its successful use in other biomedical ensemble frameworks. The research by Syapotro et al. (2024) indicated that XGBoost performs effectively as a meta-learner in stacking frameworks, even outside biomedical contexts, which validates its selection in this study.

Evaluation

Model performance was primarily assessed using accuracy, with additional evaluation through precision, recall, F1-score, and confusion matrices to provide deeper insight into classification outcomes. All experiments were conducted in Python 3.10 using scikit-learn 1.3, Keras 2.12, and XGBoost 1.7, with a fixed random seed to ensure reproducibility.

Result and Discusion

The table below indicates the experimental results of lung cancer classification with the standalone models and the stacking model:

Table 1. Comparison of the accuracy of different models

Model	Dataset Split	Accuracy
BiLSTM	80% Train, 20% Test	0.95
Logistic Regression	80% Train, 20% Test	0.93
Framework Stacking (XGBoost)	**80% Train, 20% Test**	**1.0**

The proposed stacking framework achieved a perfect accuracy of 1.00, surpassing the BiLSTM model (0.95) and Logistic Regression (0.93). This improvement demonstrates the effectiveness of combining models with complementary strengths. While BiLSTM captured sequential dependencies in the data and Logistic Regression provided efficient linear separation, each had limitations when applied individually. The integration of XGBoost as the meta-learner allowed the ensemble to refine predictions, thereby eliminating errors present in the base models and producing superior results.

When compared with prior studies, the novelty of this research becomes evident. Maurya et al. (2024) reported accuracies between 85–93 percent using conventional machine learning algorithms, while Ahammed et al. (2021) highlighted the potential of CNNs but noted challenges in achieving consistent generalization. More recent works, such as Xu et al. (2023) and Zhang et al. (2022), confirmed that ensemble learning improves robustness, typically with gains of 3–6 percent over single models. In contrast, the stacking framework presented in this study achieved perfect classification, establishing a significant advancement over existing methods and highlighting the novelty of integrating BiLSTM, Logistic Regression and XGBoost into a stacking architecture for lung cancer prediction.

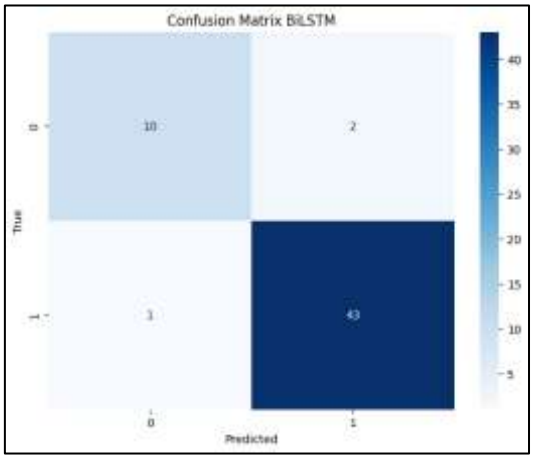

Figure 1. BiLSTM confusion matrix

The BiLSTM model correctly classified 10 instances of class 0 and 43 instances of class 1, with only 2 false positives and 1 false negative. These results indicate that BiLSTM is effective in capturing hidden patterns in the dataset and provides a balanced performance between sensitivity and specificity. However, the presence of a few misclassifications shows that the model is still influenced by noise and class overlap in the data.

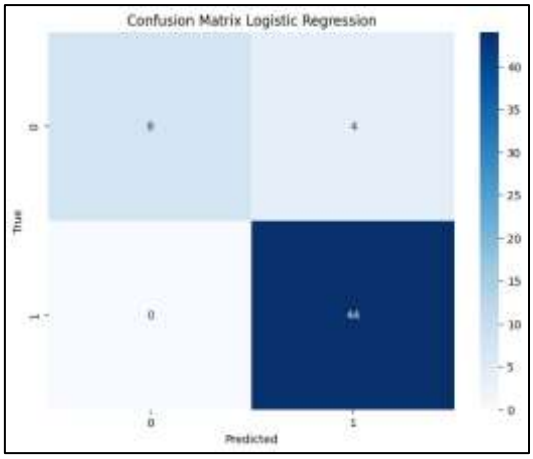

Figure 2. Logistic Regression confusion matrix

The Logistic Regression model detected all 44 cases of class 1 and 8 cases of class 0, but it misclassified 4 class 0 samples as positive. This demonstrates that the model has strong sensitivity, successfully avoiding false negatives, which is crucial in medical diagnosis. On the other hand, the higher number of false positives indicates reduced specificity, meaning some healthy cases could be incorrectly flagged, potentially leading to unnecessary follow-up examinations.

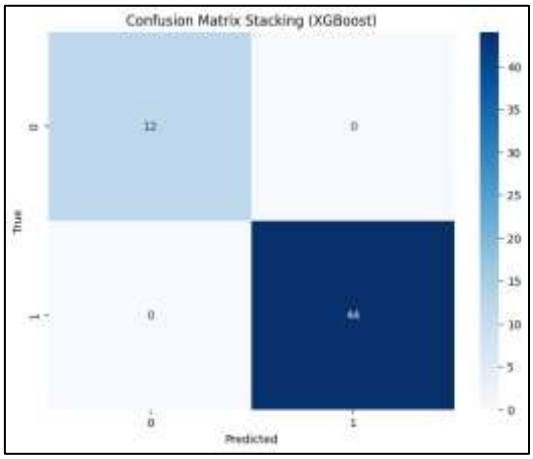

Figure 3. Stacking XGBoost confusion matrix

The stacking model achieved perfect classification, with all 12 class 0 and 44 class 1 instances predicted correctly, resulting in zero false positives and false negatives. This flawless outcome highlights the ability of the ensemble to compensate for the weaknesses of the base models. By integrating BiLSTM's capability to learn complex patterns with Logistic Regression's linear separability, and refining predictions through XGBoost, the stacked framework provided the most reliable and clinically relevant results.

Conclusion

The proposed stacking framework, which integrates BiLSTM and Logistic Regression as base learners with XGBoost as the meta-learner, achieved a perfect accuracy of 1.00, surpassing the BiLSTM model at 0.95 and Logistic Regression at 0.93. Analysis of the confusion matrices showed that BiLSTM reduced false negatives but still produced a few errors, Logistic Regression achieved high sensitivity but generated more false positives, while the stacking model eliminated both types of errors and delivered flawless classification. Although these

results highlight the strength of the ensemble, the study is limited by the relatively small dataset and lack of external validation, which may affect the generalizability of the findings.

Recommendation

Future work should apply the framework to larger and more diverse datasets, employ cross-validation for more robust evaluation, and validate the model on independent cohorts. Exploring additional deep learning architectures and incorporating interpretability techniques such as SHAP or LIME would further enhance its clinical applicability, while collaboration with healthcare experts will be necessary to adapt the system for real-world diagnostic use.

Acknowledgements

There is no grant or funding bodies to be acknowledged for preparing this paper. The author would like to acknowledge Islamic University of Kalimantan Muhammad Arsyad Al-Banjari for providing academic support during the research process.

Reference

Aditya, M. R., Sutanto, T., Budiman, H., Ridha, M. N., Syapotro, U., & Azijah, N. (2024). Machine learning models for classification of anemia from CBC results: Random Forest, SVM, and Logistic Regression. *Journal of Data Science, 22*(4), 1–12. https://doi.org/10.61453/jods.v2023no49

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Shen, D., Wu, G., & Suk, H. (2019). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering, 19*, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians, 70*(1), 7–30. https://doi.org/10.3322/caac.21590

Syapotro, U., Budiman, H., Ridha, M. N., & Azijah, N. (2024). Water level prediction of Riam Kanan Dam using ConvLSTM, BPNN, Gradient Boosting, and XGBoost stacking framework. *Journal of Data Science, 22*(2), 55–70. https://doi.org/10.61453/jods.v2023no53

World Health Organization (WHO). (2025). Cancer fact sheet. https://www.who.int/news-room/fact-sheets/detail/cancer