JOURNAL OF DATA SCIENCE | Vol.2024:54
eISSN:2805-5160

Classification of Heart Disease Using a Stacking Framework of BiGRU, BiLSTM, and XGBoost

Haldi Budiman[1], Silvia Ratna[2], M. Muflih[3], Usman Syapotro[4], Muhammad Hamdani[5], M. Rezqy Noor Ridha[6]

[1,2,3,4,5,6]Faculty of Information Technology, Islamic University of Kalimantan Muhammad Arsyad Al-Banjari, Indonesia

Email: haldibudiman@uniska-bjm.ac.id[1], silvia.ratna@uniska-bjm.ac.id[2], mmuflihfti@uniska-bjm.ac.id[3], 05usman.syapotro.study@gmail.com[4], mhdhamdani.formal@gmail.com[5], rezqyridha@gmail.com[6]

Abstract

This study aims to develop a heart disease classification model using an ensemble approach by leveraging a Stacking framework that combines BiGRU, BiLSTM, and XGBoost models. In this research, the BiGRU and BiLSTM models are utilized as base models to extract temporal and spatial features from sequential data, while XGBoost is employed as a meta-model to perform the final classification based on the features generated by the two base models. The test results show that the BiGRU model achieves an accuracy of 0.77, while the BiLSTM model achieves an accuracy of 0.85. By applying the Stacking technique using XGBoost as the meta-model, the classification accuracy significantly increases to 0.92. These findings indicate that the Stacking framework can effectively enhance heart disease classification performance, making it a potentially powerful tool for medical applications in heart disease diagnosis.

Keywords

Classification of Heart Disease, Stacking, Bidirectional Gated Recurrent Unit (BiGRU), Bidirectional Long Short-Term Memory (BiLSTM), Extreme Gradient Boosting (XGBoost)

Introduction

Heart disease remains one of the leading causes of death worldwide, despite significant advancements in diagnosis and medical treatment(Rao et al., 2024). Accurate early diagnosis is crucial for reducing the risk of serious complications and improving recovery outcomes. However, conventional methods for diagnosing heart disease often fall short in leveraging the vast and complex volumes of medical data. In this digital era, machine learning and deep learning technologies have emerged as promising solutions(Almulihi et al., 2022). By utilizing these models' ability to recognize complex patterns in data, machine learning can help enhance the accuracy of classification and prediction in medical diagnoses. One increasingly popular approach is ensemble methods, which combine multiple models to improve prediction performance. In this context, the present study focuses on developing a more accurate and efficient heart disease classification model using a Stacking framework that integrates BiGRU,

Submission: 22 October 2024; **Acceptance**: 23 November 2024

BiLSTM, and XGBoost models.

Various previous studies have explored the use of ensemble techniques for the classification of cardiovascular diseases and diabetes. One relevant study employed a Stacking Classifier approach with Support Vector Machine (SVM) as the meta-classifier, demonstrating that this method can improve the accuracy of heart disease and diabetes diagnoses compared to individual models such as Naive Bayes, K-Nearest Neighbor (KNN), and Decision Tree. The study utilized several base algorithms, including Naive Bayes, KNN, Linear Discriminant Analysis (LDA), and Decision Tree, with results showing that the stacking model achieved an accuracy of 0.8871% for cardiovascular disease, higher than the individual models. This research highlights that the ensemble approach can significantly reduce the likelihood of classification errors and enhance prediction performance(Khan et al., 2022).

The novelty of this study lies in the integration of BiGRU and BiLSTM as feature extraction tools with XGBoost as the meta-classifier within a Stacking framework. This approach has not been widely applied in heart disease classification, particularly for sequential data, where BiGRU and BiLSTM excel in handling temporal and spatial dynamics. Additionally, XGBoost enhances classification performance by leveraging the strengths of gradient boosting. The significant accuracy improvement achieved by this method, reaching 0.9596, highlights the great potential of this combination in improving heart disease diagnosis accuracy. This model not only offers better accuracy compared to individual models but also provides a more robust solution for medical applications.

The rest of this paper is organized into the following sections: methodology, results and discussion, and conclusions and recommendations. The methodology section explains the research approach, covering data collection, processing, and the development of the framework. The results and discussion section presents the experimental findings, along with a thorough analysis to offer a deeper understanding. Lastly, the conclusions and recommendations section highlights the main findings of this research and provides suggestions for future work.

Methodology

Data Preparation

The dataset was imported from a CSV file using pd.read_csv from the pandas library(McKinney, 2022), consisting of 1,319 rows and 9 features. The Result_encoded column served as the binary target for classification. The data was reshaped for sequential modeling with GRU and LSTM models, treating each feature as a timestep. The dataset was then split into training and test sets with a 70:30 ratio using train_test_split, ensuring reproducibility with random_state=42(Barupal & Fiehn, 2019).

Preprocessing (Standardization and Normalization)

Features in both training and test sets were standardized using StandardScaler to eliminate scaling bias, accelerating model convergence. Afterward, features were normalized to a [0,1] range using MinMaxScaler (Barupal & Fiehn, 2019) to stabilize training and enhance model performance.

BiGRU Model

The BiGRU model was constructed using Keras' Sequential architecture and included several key components. First, an Input layer was adapted to match the dimensions of the training data. Then, a Bidirectional GRU layer with 16 units was added to process sequential information in both forward and backward directions. For binary classification, a Dense layer with a sigmoid activation function was incorporated. The model was compiled using the Adam optimizer, with binary_crossentropy as the loss function and accuracy as the evaluation metric. It was trained over 50 epochs with a batch size of 10.

BiLSTM Model

The BiLSTM model had a similar architecture to the BiGRU but used a Bidirectional LSTM layer with 16 units for better handling of long-term dependencies. The model was also trained for 50 epochs with the same optimizer, loss function, and evaluation metrics.

Model Evaluation

Both models generated probability predictions on the test set, which were converted into binary predictions using a 0.5 threshold. Accuracy was calculated using accuracy_score as the primary metric since the dataset was balanced.

Model Stacking

The probability predictions from the BiGRU and BiLSTM models were combined into stacked features and used to train a meta-model with XGBoost(Chen & Guestrin, 2016). XGBoost, chosen for its strong performance in preventing overfitting, was configured with n_estimators=1000, learning_rate=0.1, max_depth=3, and gamma=0.1. After training, predictions were made, and accuracy was computed as the performance metric.

Result and Discussion

This research focused on early diagnosing of heart disease by leveraging machine learning algorithm. This is because accurate early diagnosis is importan to minimize the risk of complications and to improve recovery outcome. In this research, the algorithm proposed is a framework of BiLSTM, BiGRU, stacked with XGBoost.

This reseach employs experimental approach where each model is trained and tested individually before stacked with XGBoost as a meta model. The dataset was acquired from Kaggle platform, which contains heart-related health data. Model performance will be evaluated using accuracy since the dataset classes is balanced.

After the execution of the models, result showed that the stacked model using XGBoost have the highest accuracy compared to standalone models, as shown in the next table:

Table 1 Model Performance Result

Model	**Accuracy**	**Recall**	**F1-Score**
BiGRU	0.77	0.78	0.77
BiLSTM	0.85	0.87	0.86
Stacking with XGBoost	0.92	0.93	0.92

Based on Table 1, the stacked model have a higher F1-Score, which mean generally a more balanced precision and recall.If we take a closer look using confusion matrix table, then we get the following result:

eISSN:2805-5160

Table 2 Confusion Matrix for BiGRU

Actual	**N**	121(TN)	34(FP)
	P	54(FN)	187(TP)
		N	**P**
		Predicted	

Table 3 Confusion Matrix for BiLSTM

Actual	**N**	147	8
	P	48	193
		N	**P**
		Predicted	

Table 4 Confusion Matrix for Stacked Model

Actual	**N**	146	9
	P	20	221
		N	**P**
		Predicted	

Based from the results that are shown in Table 2. Table 3, and Table 4, it can be seen that Stacked Model have a comparable False Positive and True Negative compared to BiLSTM, while outperformed other models in by having fewer False Negative and the highest number of True Negative.

Conclusion

In conclusion, the stacking framework of BiGRU and BiLSTM using XGBoost improved the result over using individual standalone models.Stacked models resulted in a comparable incorrect positive prediction and correct negative prediction, while having less incorrect negative prediciton and the highest number of correct negative prediciton. With accuracy of 0.92 stacked model outperformed BiGRU by 19.48% which have accuracy value of 0.77, and BiLSTM by 8.24% which have accuracy of 0.85.

Recommendations

Future research should try to use a larger dataset with a more varied features to provide a more accurate predictions. Explore a wider and more diverse combinations of algorithm, as well as leveraging optimization method such as grid search or Bayesian optimization to optimize the model further.

References

Almulihi, A., Saleh, H., Hussien, A. M., Mostafa, S., El-Sappagh, S., Alnowaiser, K., … Refaat Hassan, M. (2022). Ensemble Learning Based on Hybrid Deep Learning Model for Heart Disease Early Prediction. *Diagnostics*, *12*(12), 3215. https://doi.org/10.3390/diagnostics12123215

Barupal, D. K., & Fiehn, O. (2019). Generating the blood exposome database using a comprehensive text mining and database fusion approach. *Environmental Health Perspectives*, *127*(9), 2825–2830. https://doi.org/10.1289/EHP4713

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 785–794. https://doi.org/10.1145/2939672.2939785

Khan, A., Khan, A., Khan, M. M., Farid, K., Alam, M. M., & Su'ud, M. B. M. (2022). Cardiovascular and Diabetes Diseases Classification Using Ensemble Stacking Classifiers with SVM as a Meta Classifier. *Diagnostics*, *12*(11), 2595. https://doi.org/10.3390/diagnostics12112595

McKinney, W. (2022). Pandas: Powerful Python Data Analysis Toolkit. *Pandas - Powerful Python Data Analysis Toolkit*, 1–3743. https://pandas.pydata.org/pandas-docs/version/1.4.4/

Rao, G. M., Ramesh, D., Sharma, V., Sinha, A., Hassan, M., & Gandomi, A. H. (2024). AttGRU - HMSI : enhancing heart disease diagnosis using hybrid deep learning approach. *Scientific Reports*, 1–19. https://doi.org/10.1038/s41598-024-56931-4