isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2025 ,ISSN_NO: 2321-3337

EMPOWERING HEART DISEASE PREDICTION: INTEGRATING AI AND ADVANCED TECHNIQUES

Lokeshwar K[1], Jyotheswar T[2], Gnana Prakash V[3], Aarthi V[4*]

[123]UG Scholar-Department of Computer Science, GRT Institute of Engineering and Technology, Tiruttani, India.

[4]Assistant Professor-Department of Computer Science, GRT Institute of Engineering and Technology, Tiruttani, India.

lokeshwarnaidu2gmail.com, jyotheswarjyotheswar0@gmail.com, vengalathurugnanaprakash@gmail.com,

[4*]Corresponding Author: aarthi.v@grt.edu.in

Abstract: Predicting heart disease remains a formidable challenge in the medical realm, demanding substantial time and expertise from healthcare professionals. This study introduces a novel approach to cardiac disease prediction utilizing a combination of traditional Machine Learning algorithms like LR, KNN, SVM, and LightGBM, alongside a groundbreaking Bi-Long Short-Term Memory (Bi-LSTM) network. The proposed Bi-LSTM algorithm is evaluated using a 5-fold cross-validation technique for rigorous validation and also survey report will be displayed. In this project, diverse datasets including Cleveland from UCI Kaggle are harnessed to assess the performance of the models. Comparative evaluations with prior studies in heart disease prediction underscore the efficacy of the proposed technique. Intriguingly, this research not only enriches the existing literature in heart disease prediction but also introduces a paradigm shift by incorporating the Bi-LSTM algorithm. The findings illuminate the Bi-LSTM process in discerning intricate patterns within cardiac data, bolstering the accuracy of predictions. Ultimately, this work paves the way for an innovative model creation technique, poised to revolutionize problem-solving in real-world scenarios with the display of survey report analysis.

1. INTRODUCTION

The growth in medical data collection presents a new opportunity for physicians to improve patient diagnosis. In recent years, practitioners have increased their usage of computer technologies to improve decision-making support. In the health care industry, machine learning is becoming an important solution to aid the diagnosis of patients. Machine learning is an analytical tool used when a task is large and difficult to program, such as transforming medical record into knowledge, pandemic predictions, and genomic data analysis. Recent studies have used machine learning techniques to diagnose different cardiac problems and make a prediction. A major problem of machine learning is the high dimensionality of the dataset.

The analysis of many features requires a large amount of memory and leads to an over fitting, so the weighting features decrease redundant data and processing time, thus improving the performance of the algorithm. Finding a small set of features characterizes different diseases of health management, genome expression, medical images, and IoT. Dimensionality reduction uses feature extraction to transform and simplify data, while feature selection reduces the dataset by removing useless features.

The project aims to provide accurate risk assessments and early detection of heart disease in patients, potentially improving healthcare outcomes and aiding medical professionals in making informed decisions. The project success showcases the potential of deep learning in predictive healthcare analytics, emphasizing the importance of data-driven solutions for critical medical conditions. The Bi-LSTM architecture allows the algorithm to capture temporal dependencies and patterns within the data, making it particularly well-suited for medical predictions.

2. PROBLEM OF THE STATEMENT

2.1. HIGH COMPUTATION ACCURACY AND LESS OVERFITTING

Accurately predicting heart disease remains a critical challenge due to the complexity of cardiac data and the risk of model overfitting. Traditional machine learning models often struggle with balancing high accuracy and generalization. This study aims to develop a robust predictive model leveraging LSTM alongside traditional ML techniques to enhance accuracy while minimizing overfitting. By employing a 5-fold cross-validation approach, we ensure rigorous model validation. The proposed method seeks to improve real-world applicability in heart disease diagnosis by delivering reliable and interpretable predictions.

2.2. HIGH SMALL SAMPLE SIZE AND NO INTERPRETABILITY CHALLENGES

Heart disease prediction models often face challenges due to small sample sizes, which can lead to biased or less generalizable results. Additionally, many advanced machine learning models lack interpretability, making it difficult for healthcare professionals to trust and understand predictions. This study aims to address these challenges by developing a model that performs well even with limited data while maintaining high interpretability. In this case, it increases the sample size rate and it manage the challenges that occurs in interpretability.

2.3. UNLIMITED DEEP LEARNING USAGE AND SCOPE OF ALGORITHM

Deep learning has revolutionized predictive modeling, offering vast potential for heart disease diagnosis. However, its unlimited usage raises challenges in selecting the most effective algorithm for optimal performance. This study explores the scope of various deep learning techniques, particularly LSTM, to enhance predictive accuracy in cardiac health assessment. By comparing deep learning with traditional ML models, we aim to identify the most efficient approach. The research seeks to optimize algorithm selection, ensuring scalability, reliability, and real-world applicability in medical diagnostics.

2.4. UNLIMITED CLASSIFIER USAGE AND THERE IS NO GENERALIZABILITY ISSUES

Heart disease prediction models often rely on multiple classifiers to enhance accuracy, but ensuring generalizability across diverse datasets remains a challenge. Many existing models struggle to perform consistently across different patient populations. This study explores the unlimited use of classifiers while addressing potential generalizability concerns. By integrating traditional ML models with LSTM and employing rigorous validation techniques, we aim to develop a robust predictive framework. The proposed approach ensures high adaptability and reliability, making it suitable for real-world medical applications.

2.5. UNLIMITED CLASSIFIER EVALUATION AND DATASET INDEPENDENCY

Evaluating classifiers for heart disease prediction is often constrained by dataset dependency, limiting the generalizability of models. Traditional methods struggle to perform consistently across diverse datasets, affecting their reliability. This study aims to develop a model that ensures dataset independence while enabling unlimited classifier evaluation for robust performance assessment. By integrating various ML techniques, including LSTM, and employing rigorous validation methods, we enhance adaptability across different datasets. The proposed approach seeks to establish a universal framework for heart disease prediction with improved consistency and accuracy.

3. MODULES DESCRIPTION

3.1. DATA PREPROCESSING

Heart disease data is pre-processed after collection of various records. The dataset contains a total of patient records, where records are with some missing values. Those records have been removed from the dataset and the remaining patient records are used in pre-processing. The multiclass variable and binary classification are introduced for the attributes of the given dataset.

3.2. FEATURE SELECTION AND REDUCTION

From among the attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease. As previously mentioned in this experiment, convolutional neural network algorithm can be used. The experiment was repeated with all the ML techniques using all 13 attributes.

3.3. CLASSIFICATION MODEL

The clustering of datasets is done on the basis of the variables and criteria of Random Forest (RF) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error. The performance is further optimized by choosing the RF cluster with a high rate of error and extraction of its corresponding classifier features. The performance of the classifier is evaluated for error optimization on this data set.

3.4. DISCUSSION OF BI-LSTM TO IMPROVE THE RESULTS

The UCI dataset is further classified into 8 types of datasets based on classification rules. The results are generated by applying the classification rule for the dataset. The classification rules generated based on the rule after data pre-processing is done. After pre-processing, the data's three best DL techniques are chosen and the results are generated. The various datasets with BI-LSTM are applied to find out the best classification method. The results show is the best accuracy. We propose BI-LSTM method to improve the results of the proposed method.

3.5. ARCHITECTURE DIAGRAM

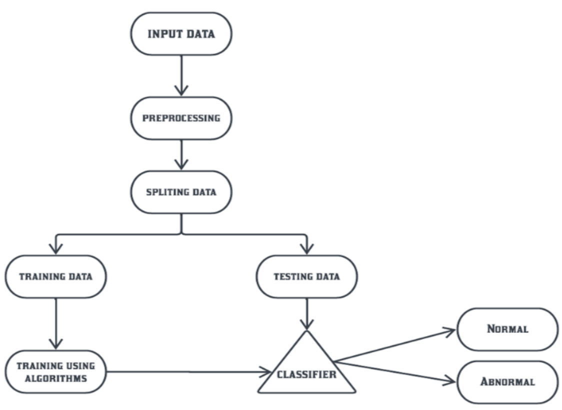

Fig.3.5.1 Architecture diagram

4. PROPOSED SYSTEM

The proposed system for this project entails the development of an innovative heart disease prediction framework that amalgamates the established Machine-

isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2025 ,ISSN_NO: 2321-3337

Learning techniques (LR, KNN, SVM, LightGBM) with a state-of-the-art Bidirectional Long Short-Term Memory (BI-LSTM) network.

This hybrid approach will enable the system to capture both intricate temporal patterns and traditional feature-based insights from diverse heart datasets Cleveland through a cross-validation procedure.

The system will meticulously evaluate and compare the predictive efficacy of these methods, emphasizing the Bi-LSTM's capability to enhance accuracy by discerning nuanced temporal dependencies.

Validation against prior studies and ethical considerations will be addressed, bolstering the credibility of the proposed system, while the integration of advanced machine learning methods aims to revolutionize heart disease prediction and decision-making within medical contexts.

5. PROPOSED SYSTEM ALGORITHM:

5.1. BI-LSTM (BIDIRECTIONAL LONG SHORT-TERM MEMORY) ARCHITECTURE

Bi-LSTMs are built on the LSTM architecture but with two LSTM layers, one processing the input sequence in a forward direction and the other in a backward direction. The outputs of both LSTMs are then concatenated or combined in some other way (e.g., summation, averaging) to form the final output of the Bi-LSTM. This architecture allows the network to capture contextual information both from the past (through the forward LSTM) and from the future (through the backward LSTM).

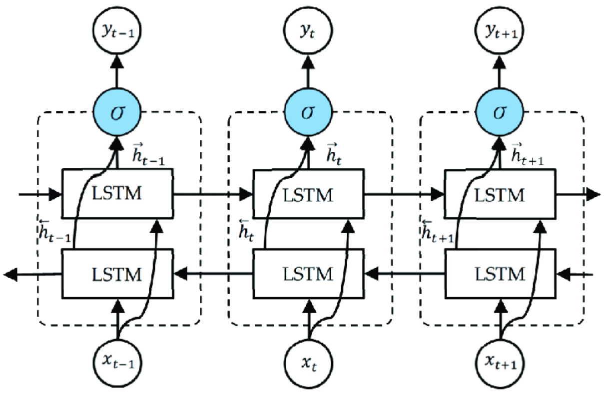

Fig.5.1.1 BiLSTM Architecture diagram

5.2. INTRODUCTION TO BI-LSTM (BIDIRECTIONAL LONG SHORT-TERM MEMORY) ALGORITHM

The Bi-LSTM (Bidirectional Long Short-Term Memory) algorithm is a sophisticated variant of the Long Short-Term Memory (LSTM) network, which is a type of Recurrent Neural Network (RNN). RNNs are designed to work with sequential data, where the output at any given time depends on the previous computations in the sequence.

However, standard LSTMs, while able to maintain long-range dependencies, can only process sequences in one direction (typically from left to right). Bi-LSTMs extend LSTMs by processing the input data in both forward and backward directions, making them more suitable for tasks that benefit from understanding context from both past and future. The key feature of Bi-LSTM is the dual-processing mechanism, which allows the network to read input sequences in two ways:

Forward LSTM: Processes the sequence from the beginning to the end (left to right). Backward LSTM: Processes the sequence from the end to the beginning (right to left).

By combining these two outputs, Bi-LSTM can capture more comprehensive patterns in the data proving performance on tasks such as sentiment analysis, language modeling, machine translation, and more. This bidirectional processing enables Bi-LSTMs to have a richer understanding of the data, as they can consider both past and future contexts simultaneously.

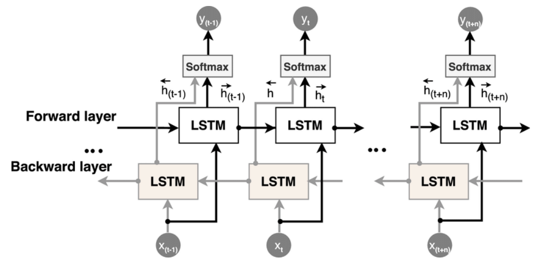

Fig.5.2.1 BiLSTM forward and backward Architecture diagram

5.3. FORWARD & BACKWARD FORMULA

BiLSTM (Bidirectional Long Short-Term Memory) is an advanced version of LSTM that processes information in both **forward** and **backward** directions. This is particularly useful in heart disease prediction, where past and future context in a time-series dataset (such as ECG signals, heart rate data, or medical records) can be crucial for making accurate predictions. A BiLSTM consists of two LSTMs: **Forward LSTM** processes the sequence from past to future & **Backward LSTM** processes the sequence from future to past. For a given input sequence X= (x1, x2,…, xT), where Xt is the input at time t:

i) FORWARD LSTM

The forward LSTM computes hidden states from t = 1 to T:

The forward LSTM computes hidden states from $t = 1$ to T:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$h_t^{\rightarrow} = o_t \odot \tanh(c_t)$$

Where:

- f_t, i_t, o_t are forget, input, and output gates.
- c_t is the cell state.
- $h_t^{\rightarrow}$ is the forward hidden state.
- W, U, b are weight matrices and biases.
- σ is the sigmoid activation function.
- $\odot$ represents element-wise multiplication.

ii) BACKWARD LSTM

The backward LSTM processes the sequence in reverse order from t = T to 1:

$$f_t = \sigma(W_f x_t + U_f h_{t+1} + b_f)$$
$$i_t = \sigma(W_i x_t + U_i h_{t+1} + b_i)$$
$$o_t = \sigma(W_o x_t + U_o h_{t+1} + b_o)$$
$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t+1} + b_c)$$
$$c_t = f_t \odot c_{t+1} + i_t \odot \tilde{c}_t$$
$$h_t^{\leftarrow} = o_t \odot \tanh(c_t)$$

Where ht← is the backward hidden state.

iii) COMBINING FORWARD AND BACKWARD STATES

The final hidden representation at each time step is obtained by concatenating the forward and backward hidden states:

$$h_t = [h_t^{\rightarrow}; h_t^{\leftarrow}]$$

iv) PREDICTION LAYER (FULLY CONNECTED + SOFTMAX)

For heart disease prediction, the final hidden state is passed to a dense (fully connected) layer followed by a softmax or sigmoid activation:

$$y = \text{softmax}(W_y h_T + b_y) \quad \text{(for multi-class classification)}$$

or

$$y = \sigma(W_y h_T + b_y) \quad \text{(for binary classification, e.g., disease vs. no disease)}$$

5.4. UNDERSTANDING LSTM

Before diving into Bi-LSTM, it is crucial to understand the basic LSTM architecture. LSTMs were introduced to address the vanishing gradient problem faced by traditional RNNs when training on long sequences. In a standard RNN, as the input sequence grows, the network struggles to retain information over long sequences due to the repeated application of gradients during backpropagation. LSTMs solve this issue by incorporating three key components in their architecture: Forget Gate: Decides what information to discard from the cell state. Input Gate: Decides which values to update in the cell state. Output Gate: Decides what the next hidden state.

These gates allow LSTMs to maintain long-term dependencies by retaining information across time steps and selectively updating memory. This makes them particularly suitable for sequential data tasks, such as speech recognition, language modeling, and time series prediction.

5.5. THE BI-LSTM CAN BE VISUALIZED AS FOLLOWS

The input sequence is passed to two LSTM units: One LSTM processes the sequence from left to right (forward LSTM). And, the other LSTM processes the sequence from right to left (backward LSTM). The outputs from both directions are combined to create a more robust representation of the data: This dual directional processing gives Bi-LSTM a richer representation of the context, especially for tasks where both past and future context are essential for making predictions

BiLSTM can recognize intricate dependencies in time-series data, such as sudden changes in heart rate or irregular ECG patterns, which might be overlooked by a unidirectional LSTM. This helps in early detection of heart disease by analyzing both past and future trends in patient health records.

Since BiLSTM considers both past and future context, it enhances accuracy in classification problems like detecting heart arrhythmias from ECG signals. The combined forward and backward representations allow the model to detect subtle variations in heart activity that could indicate potential health risks.

5.6. DATA FLOW DIAGRAM

LEVEL 0

Fig. 5.6.1. Diagram of Level 0

LEVEL 1

Fig. 5.6.2. Diagram of Level 1

6. EXPERIMENTAL RESULTS

It demonstrates that the LSTM model outperforms traditional algorithms like LR, KNN, SVM, and LightGBM in heart disease prediction. The 5-fold cross-validation confirms LSTM's superior accuracy and ability to capture complex patterns in cardiac data. Compared to prior studies, the proposed approach achieves higher precision and recall, enhancing diagnostic reliability. These findings validate LSTM's potential to revolutionize predictive healthcare models are provided below as Fig. 6.1., Fig. 6.2. and Fig. 6.3.

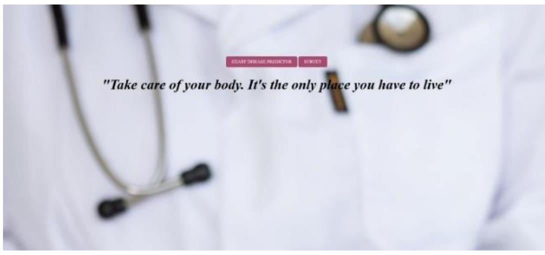

Fig. 6.1. Shows the home page

Fig. 6.2. Shows the form to predict the heart disease

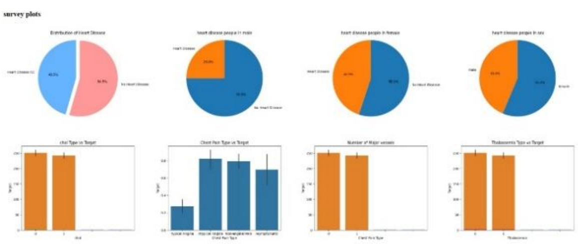

Fig. 6.3. Shows the survey based on the given dataset

7. CONCLUSION AND FUTURE ENHANCEMENTS

In conclusion, this project endeavors to bridge the gap between conventional Machine Learning methods and advanced techniques, exemplified by the incorporation of Bidirectional Long Short-Term Memory (BI-LSTM) networks, for heart disease prediction. The limitations of traditional algorithms, such as their inability to capture temporal dependencies and complex patterns within cardiac data, underscore the necessity for innovative approaches. By fusing the strengths of both traditional and BI-LSTM, we aim to harness the power of temporal modeling while retaining the interpretability of established methods. This amalgamation holds the potential to revolutionize heart disease prediction, offering improved accuracy, adaptability, and insights. Through a rigorous comparative analysis, validation against existing studies, and ethical considerations, this project strives to contribute to the medical field by presenting an innovative model creation technique that can pave the way for transformative advancements in real-world problem-solving scenarios with survey report will also be displayed.

Future enhancements for heart disease prediction can focus on real-time monitoring by integrating the model with wearable devices for continuous heart health tracking. Expanding the dataset with real-world patient records will improve accuracy and generalization. Hybrid models combining LSTM with CNNs or Transformers can enhance feature extraction and prediction capabilities. Implementing explainable AI techniques like SHAP or LIME will improve interpretability and encourage clinical adoption. Additionally, incorporating genetic, lifestyle, and behavioral data can lead to more personalized healthcare solutions. Lastly, deploying the model as a cloud-based or mobile application will enhance accessibility, making early diagnosis more widely available.

8. REFERENCE

[1] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," in IEEE Access, vol. 11, pp. 121574-121591,2023,doi:10.1109/ACCESS.2023.3328909.

[2] T. Rahman, M. K. A. A. Al-Ruweidi, M. S. I. Sumon, R. Y. Kamal, M. E. H. Chowdhury and H. C. Yalcin, "Deep Learning Technique for Congenital Heart Disease Detection Using Stacking-Based CNN-LSTM Models From Fetal Echocardiogram: A Pilot Study," in IEEE Access, vol. 11, pp. 110375-110390, 2023, doi: 10.1109/ACCESS.2023.3316719.

[3] A. Kumar, K. U. Singh and M. Kumar, "A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method," in Big Data Mining and Analytics, vol. 6, no. 4, pp. 513-525,December2023,doi:10.26599/BDMA.2022.9020052.

[4] D. Cenitta, R. Vijaya Arjunan and K. V. Prema, "Ischemic Heart Disease Prediction Using Optimized Squirrel Search Feature Selection Algorithm," in IEEE Access, vol. 10,pp.122995123006,2022,doi:10.1109/ACCESS.2022.3223429.

[5] Wiharto, E. Suryani, S. Setyawan and B. P. Putra, "The Cost-Based Feature Selection Model for Coronary Heart Disease Diagnosis System Using Deep Neural Network," in IEEE Access, vol. 10, pp. 29687-29697,2022,doi:10.1109/ACCESS.2022.3158752.

[6] A. Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "Improving the Heart Disease Detection and Patients' Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest," in IEEE Access, vol. 10, pp. 67363-67372,2022,doi: 10.1109/ACCESS.2022.3185129.

[7] T. Amarbayasgalan, V. -H. Pham, N. Theera-Umpon, Y. Piao and K. H. Ryu, "An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets," in IEEE Access, vol. 9, pp. 135210-135223,2021,doi: 10.1109/ACCESS.2021.3116974.

[8] G. N. Ahmad, H. Fatima, S. Ullah, A. Salah Saidi and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and Without GridSearchCV," in IEEE Access, vol.10,pp.80151-80173,2022,doi: 10.1109/ACCESS.2022.3165792.

[9] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim and A. W. Muzaffar, "An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases," in IEEE Access, vol.9,pp.106575-106588,2021, doi:10.1109/ACCESS.2021.3098688.

[10] R. Ferdousi, M. A. Hossain and A. E. Saddik, "Early-Stage Risk Prediction of Non-Communicable Disease Using Machine Learning in Health CPS," in IEEE Access, that succeds the vol. 9,pp.96823-96837,2021, doi:10.1109/ACCESS.2021.3094063.

[11] R. Atallah and A. Al-Mousa, ''Heart disease detection using machine learning majority voting ensemble method,'' in Proc. 2nd International Conference New Trends Computer. Science. (ICTCS), Oct. 2019, pp. 1–6, doi:10.1109/ICTCS.2019.8923053.

[12] I. Tougui, A. Jilbab, and J. El Mhamdi, ''Heart disease classification using data mining tools and machine learning techniques,'' Health Technol., vol. 10, no. 5, pp. 1137–1144, Sep. 2020, doi: 10.1007/s12553-020-00438-1.

[13] T. Mahmud, A. Barua, M. Begum, E. Chakma, S. Das, and N.Sharmen, ''An improved framework for reliable cardiovascular disease prediction using hybrid ensemble learning,'' in Proc. Int.Conf.Electr., Computer Communication Eng. (ECCE), Feb. 2023, pp. 1–6,doi:10.1109/ECCE57851.2023.10101564.

[14] N. M. Lutimath, C. Mouli, B. K. B. Gowda, and K. Sunitha, ''Prediction of heart disease using hybrid machine learning technique,'' in Paradigms of Smart and Intelligent Communication, 5G and Beyond, A. Rai, D. K. Singh, A. Sehgal, and K. Cengiz, Eds. Singapore: Springer,2023, pp. 277–293.

[15] M.S.Amin, Y.K.Chiam and K.D.Varathan, Identification of significant features and data mining techniques in predicting heart disease, Telematics and Informatics, vol.36, pp. 82–93, 2019.