JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
SICAIET
2024
Research Article
Collection: Artificial Intelligence and Emerging Technologies

ECG Based Heart Disease Diagnosis Using Machine Learning Approaches

Syed Ali Nawaz[1*], Muhammad Arshad[1], Tanveer Aslam[1], Abdul Haseeb Wajid[1]
Rizwan Ali Shah[1], and Mubashir H. Malik[4]

[1]Department of Information Technology, The Islamia University of Bahwalpur , Bahwalpur, 63100, Pakisatan
[2]Department of Computer Science, Institute of Southern Punjab, Multan, Pakistan.
[*]Corresponding Author: Syed Ali Nawaz. Email addresses: ali.nawaz@iub.edu.pk

Academic Editor: Salman Qadri **Published:** February 01, 2024

Abstract: The electrocardiogram (ECG) is crucial to monitor cardiac health, especially since its signal is the most important diagnostic tools to help in the detection of heart disease. ECG interpretation has largely been confined to manual analysis, which suffers from constraints such as expert availability in underserved areas, and diagnostic errors. Addressing these issues underwent research via machine learning through the fusion of ECG data for improved heartbeat classification. This study presents a novel approach that incorporates a Support Vector Machine (SVM) with Random Forest (RF), Logistic Regression (LR), Decision Tree (DT) models in a comprehensive method designed to classify heartbeats into normal, abnormal and COVID-19 affected. The individual performance of Decision Tree, Logistic Regression, Random Forest and Support Vector Machine models are evaluated on ECG image dataset. The respective accuracy rates were 77%, 82%, 78%, and 83%. The SVM model produced a superior accuracy of 84%. This comparative analysis thus identifies the potential for SVM model to the empower ECG signal interpretation and take the clinical depart towards remote diagnostics while ensuring early detection of cardiac anomalies.

Keywords: ECG; Support Vector Machine; Heart Disease; Diagnosis.

1. Introduction

Identification of abnormal electrocardiogram (ECG) patterns is of critical importance in early detection and management of cardiovascular diseases (CVDs), the leading cause of mortality worldwide [1]. The ECG is an important modality for assessing heart health, providing information about the electrical activity of the heart. Nevertheless, traditional methods of ECG analysis, which included manual interpretation by trained clinicians as well as automatic classification by algorithms, have generally been limited by low accuracy, consistency, and computational efficiency[2]. Because of its complex waveform morphology, inter-patient variability, and susceptibility to noise and artifacts, ECG signals have been a difficult and noisy signal that has often proven resistant to the best efforts of engineers and computer algorithms [3]. In response to the COVID-19 pandemic, there was an emergent need for advanced diagnostic assays that could differentiate abnormal from normal ECGs with demonstrated precision [1]. It was specifically investigated if the state-of-the-art discrimination capacity of SVM could be used in identifying intricate patterns inherent in ECG signals. A primary objective of this investigation was to set a bar for use of machine learning in cardiac diagnostic testing, with a view to such tools providing solutions for early detection and faithful monitoring of the stage of pathogenesis of Cardiovascular Diseases (CVD) [4]. In order to revolutionize ECG signal classification, to yield a scalable, accurate and efficient diagnostic model that could fundamentally transform the early-stage detection and longitudinal assessment of CVDs, further work is now required, towards investigating the generalization capabilities of RF, LR and DT models in a significantly broader manner, and their systematic comparison in ECG signal classification, building on the theoretical basis, algorithmic development, and comprehensive validation of the SVM model we have detailed in this paper, and the critical comparison with the performance of the RF, LR and DT models presented earlier in this section [5]. A ECG signal waveform is shown in figure 1 below.

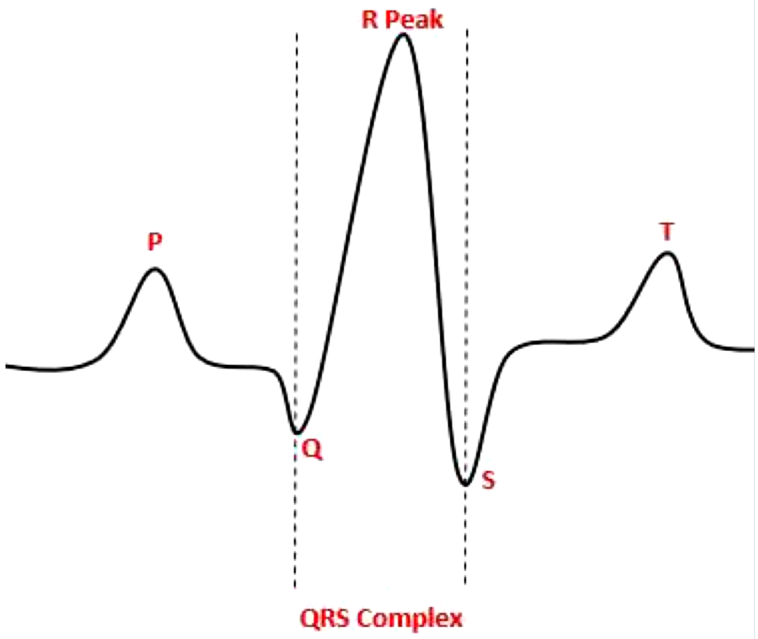

Figure 1. ECG signal waveform

2. Literature Review

Denoising ECG signals with the use of CNN, ANN, and DNN has seen a vast improvement in solving the challenges presented by way of noisy ECG signals. [6] have revealed the algorithm for noise reduction, which has garnered high accuracies in its attempt to preserve the integrity of ECG signals, thanks to the machine learning model, flexible filters and decomposition of the signal itself.

Detecting the abnormal events in an ECG signal that correspond to arrhythmia is also a critical aspect of the phenomenon of arrhythmia. (Wang et al., 2023) have made extensive efforts to diagnose the types of (different) arrhythmias. The machine learning model for diagnosing arrhythmia with the use of ICA, for instance, has been combined with CART, RF, KNN, and ANN and has undoubtedly paved the way in matching the computational costs with the actual diagnosis -to a tune of fast and dependable diagnosis of arrhythmia in real-time[8].

ECG classification have been remarkably advanced by exploiting deep learning technologies as discussed by [9]. The deployment of DNNs, CNNs, and hybrid models have made it feasible to distinguish between various arrhythmia types with unprecedented accuracy [10]. An exemplary technique is the usage of continuous wavelet transform (CWT) and CNNs, for both feature extraction and classification, reflecting the ability to carry out very accurate ECG signal analysis. [11]

[12] however, also underscore a big stride toward making ECG signal analysis even more reliable as they put forward the potential of ensemble learning algorithms for further boosting classification accuracy and robustness. These ensemble learning methods that exploit multiple ML and DL models (e.g., bagging, boosting, stacking) complement the power that many have individually shown in outclassing the capabilities of separate algorithms, as they endeavour to handle the most complex ECG classification tasks [13].

Research gaps, however, still exist as evidenced by the comparison analysis of ML techniques in ECG data classification and a deeper understanding of the distinguishing characteristics of physiological systems as distinct from artificial ones in advancing classification methods and feature extraction techniques is needed. This realization is is key to the development of more refined and accurate ECG analysis systems [2]

In conclusion, the convergence of ML, DL and ensemble learning techniques in ECG signal analysis opens a new frontier to significant enhancements in heart disease diagnosis [11]. Future research should seek in the same vein as to further exploit these synergies to bring forth more precise, efficient and reliable ECG analysis systems.

3. Materials and Methods

This section describes the overview of the complete methodology of this research study. The methodology comprises a number of knowledge discovery and data mining steps, which includes dataset collection, preprocessing, features extraction and classification, model tuning and evaluation**.** Complete methodology diagram of our proposed model is presented in Figure 2.

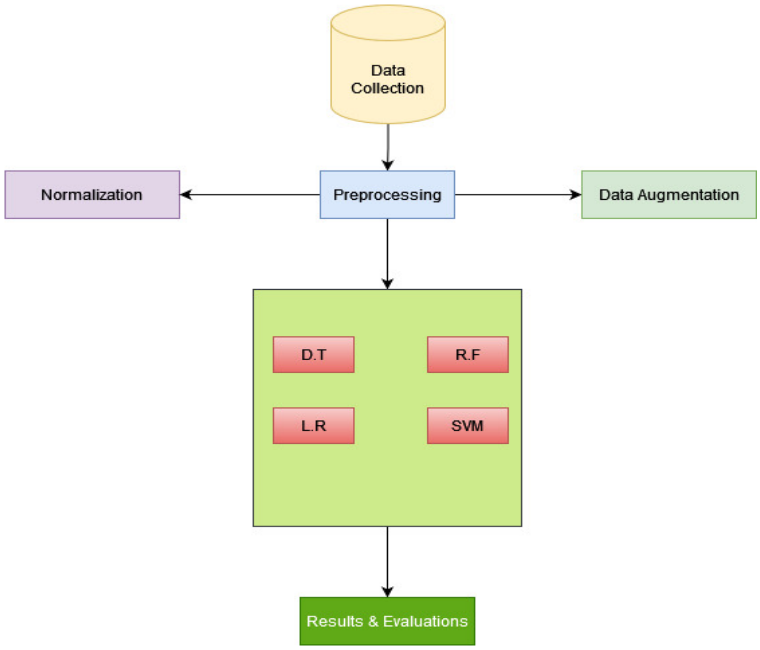

Figure 2. Methodology diagram of our proposed model

3.1 Dataset Description

The study employed the ECG images dataset from Kaggle, collated by Khan et al. This dataset contains ECG images of patients with cardiac and COVID-19 illness. It encompasses 250 images of COVID-19 patients, 859 of normal heartbeats and 548 of non-normal heartbeats for probing their potential diagnostic values for COVID-19 and cardiovascular disorders, respectively.

Table 1. Dataset Description

Class	No. of Images
Normal HB	859
Abnormal HB	250
COVID-19 Patients HB	548

3.2. Preprocessing

The ECG images underwent several preprocessing steps to prepare them for analysis which include image resizing and reshaping: All images were resized to 224×224 pixels using nearest-neighbor interpolation, such that the original signals would in turn remain unaltered. First and foremost, this would ensure uniformity among the images, providing compatibility with the neural network input layer and minimizing the loss of data. It also avoids excessively large matrix sizes that would decrease computational efficiency. Normalization: Pixel values across the images were normalized such that they would fall within the range [0,1]. These bounded values accelerated convergence during training as a result. Data Augmentation: Data augmentation strategies such as rotation, zoom, and horizontal flipping were implemented to address both overfitting as well as to help the model better generalize to newer data. The sample preprocessed ECG images are shown in Figure 3.

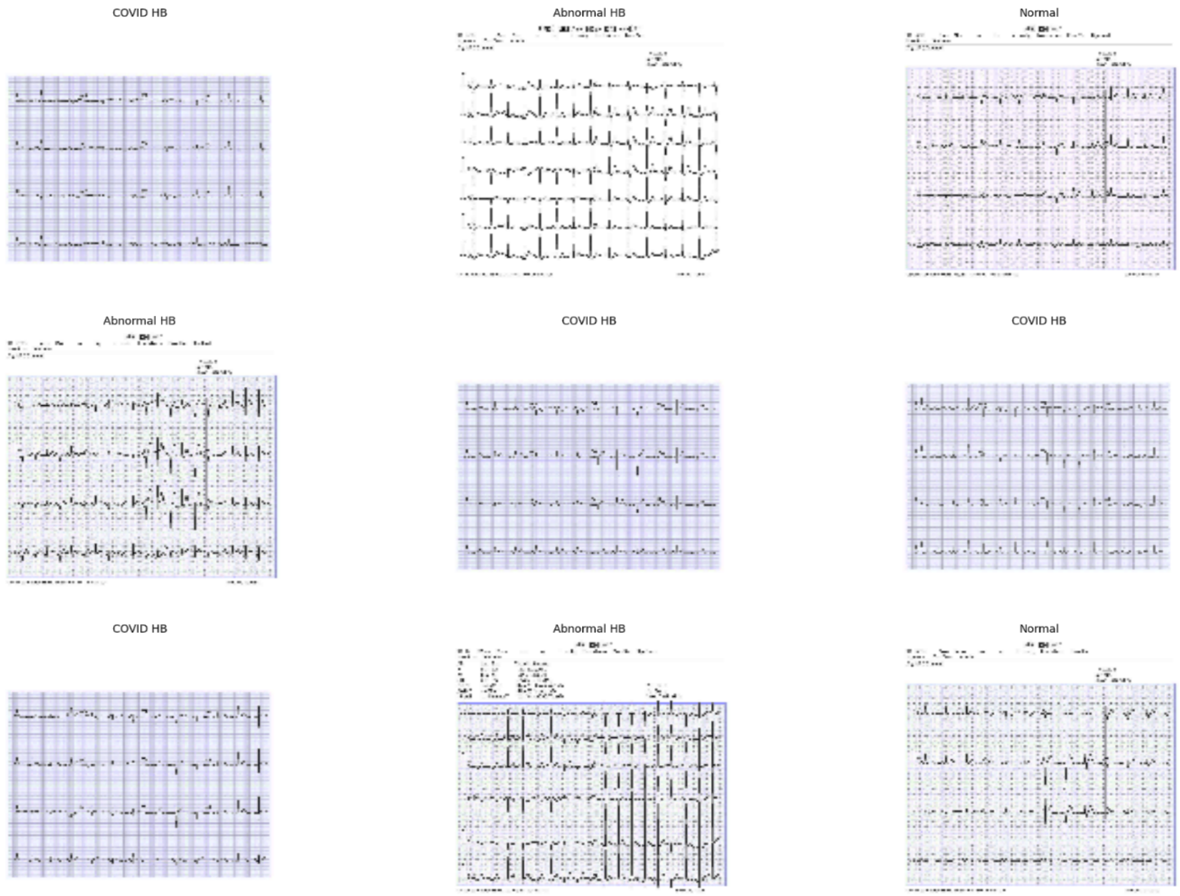

Figure 3. The sample preprocessed ECG images

3.3. Machine Learning Models

Four supervised machine learning algorithms were considered in this study: Decision Tree (DT), multinomial Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM). Each algorithm was selected from the point of view of how it handled a classification problem which allows a rather comprehensive understanding of how these classifications approaches are applied to ECG signal classification. Decision Trees (DT): A decision tree maps data items to a class label from which a decision can be made. These are particularly suited to classification as they allow the resultant to be explained in terms of binary features. These can be used to binary classification or multi-classification. Logistic Regression (LR): A statistical model that estimates probabilities using a logistic function; it was built for binary classification and extended to multi-class classification using the one-vs-rest scheme. Random Forest (RF): An ensemble of decision trees; designed to improve classification through bagging and feature randomness. Support Vector Machine (SVM): A powerful classifier that attempts to find an optimal hyper-plane that tries to best separate the two classes in feature space. It was used with the radial basis function (RBF) kernel to allow for non-linear data.

3.3.1. Hyperparameter Tuning

Hyperparameter tuning was carried out in order to optimise the model performance. Grid search was used, evaluating a range of values for key parameters:

· DT: Maximum depth of the tree, minimum samples split and minimum samples leaf

· LR: Regularization strength and type of solver

· RF: Number of trees, maximum depth, minimum samples split

· SVM: C (regularization parameter), gamma (kernel coefficient)

3.3.2. Training and Evaluation

To evaluate the performance of the models, the dataset was divided in to training (80%) and testing (20%) sets. Models were trained on the training set and their accuracy, precision, recall, and F1 score was calculated based on their performance on the test set. These evaluation metrics give a comprehensive understanding of the strengths and weaknesses of the model in classifying the ECG images into their respective categories.

4. Results and Discussion

The evaluation of the machine learning models on the ECG dataset yielded the following results, summarized in Table 2, and illustrated in the accompanying accuracy graph.

Table 2. Comparison of Machine Learning Model Results

Models	Accuracy (%)	Training	Time (sec)
Decision Tree	77	41.4024	0.028162
Logistic Regression	82	101.072	0.07035
Random Forest	78	26.8203	0.102554
SVM	83	102.974	15.8303

The overall performance graph of the four machine learning classifiers is shown below in the Figure 4. The graph illustrates the accuracy of each model. This shows that the Support Vector Machine (SVM) model is the most accurate with the accuracy of 83%. While Logistic Regression (LR) and Random Forest (RF) are at cascade by 82% and 78% respectively. The accuracy of Decision Tree (DT) is 77%.

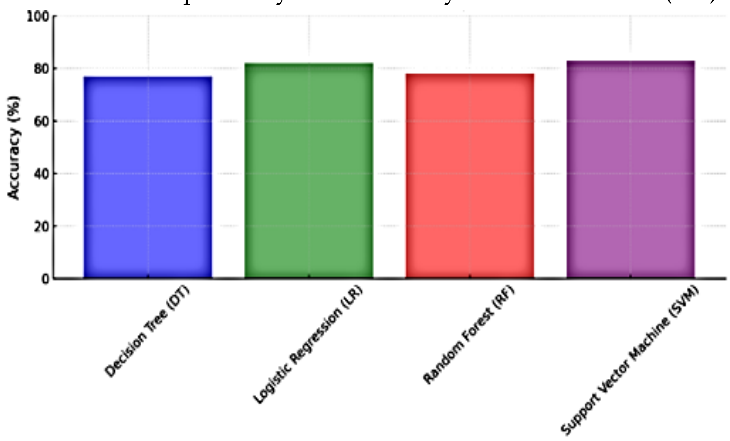

Figure 4. Performance Graph of Four Machine Learning Classifiers

Support Vector Machine (SVM) model yielded maximum accuracy among the models evaluated, indicating its efficacy in classifying ECG signals into normal, abnormal, and COVID-19 afflicted categories. Logistic Regression model also yielded a good performance, making it another alternative for ECG signal classification. Decision Tree and Random Forest models exhibited slightly lesser accuracy compared to the previous two models, with the tradeoff of interpretability and computational efficiency for their use in applications.

The training and prediction times varied across models, where the Decision Tree model was the fastest for both training and prediction. This is inherent to the tradeoff between accuracy and computational efficiency when using different models in real-world applications. The superior accuracy of the SVM model, however, could enrich ECG based heart disease diagnosis that could potentially,

The following table compares this proposed study's Support Vector Machine (SVM) model's accuracy to that of previous studies´. The table shows how much each of these different works have advanced the state of the art in ECG signal classification through machine learning.

Table 3. Comparison With Other State-of-The-Art Techniques

Reference	Technique	Accuracy (%)
[14]	CNN	71
[15]	CNN	62
[16]	NB	82
Proposed Study	**SVM**	**83**

On comparison, the SVM model in the present work displayed remarkable classification accuracy in the ECG signal classification that is quite at par with the machine learning and deep learning techniques available in literature. Especially aforesaid performance of SVM matches the maximum accuracy contributed as micro-average by the technique among all the previous studies. It can be concluded that SVM still provides robust and efficient model in ECG signal classification. It also indicates interest and importance to develop such machine learning models that can plausibly convert the ECG data to become certified and reliable towards improved diagnosis of heart diseases.

5. Conclusion

In this study, a comprehensive comparative analysis has been presented to classify the ECG signals as normal, abnormal, and COVID-19 affected through various ML models. The evaluated models (i.e., Decision Tree, Logistic Regression, Random Forest, and Support Vector Machine) showed that the SVM outperformed the other models with 83% accuracy and thus has the capability to improve the ECG based heart disease detection accuracy. In addition, the SVM has been compared with the related previous techniques, which confirmed that the proposed SVM has a distinctive effectiveness, and competitiveness over the existing ECG signal classification techniques.

The research showcases the critical role of machine learning in addressing the challenges of traditional ECG analysis, which include errors in manual interpretation and the need for an expert, especially in underserved areas. In doing so, it contributes to a broader movement that is bringing advanced computational methods to cardiac diagnostics, which is likely to result in far more accurate, efficient and available heart disease detection.

5.1 Future Work

While the current study provides valuable insights into the application of machine learning for ECG signal classification, some future directions are applications of ensemble learning and deep learning techniques.

Data Availability: The dataset will be provided on demand for research study.

Conflicts of Interest: The authors declare no conflicts of interest.

Acknowledgments: This work was supported by *IUB* and *MNS-UAM*. The authors would like to thank Dr. Kamran (BVH), Mr. Tanveer Aslam, Mr. Umer (BVH), and Dr. Umair (BVH) for technical and field assistance.

Funding Statement : This Research did not receive any specific grant from any public, commercial or non-profit able funding agencies.

References

1 L. Mhamdi, O. Dammak, F. Cottin, and I. Ben Dhaou, “Artificial Intelligence for Cardiac Diseases Diagnosis and Prediction Using ECG Images on Embedded Systems,” Biomedicines, vol. 10, no. 8, Aug. 2022, doi: 10.3390/biomedicines10082013.

2 N. Musa et al., “A systematic review and Meta-data analysis on the applications of Deep Learning in Electrocardiogram,” J Ambient Intell Humaniz Comput, vol. 14, no. 7, pp. 9677–9750, Jul. 2023, doi: 10.1007/s12652-022-03868-z.

3 A. Arefeen, A. Akbari, S. I. Mirzadeh, R. Jafari, B. A. Shirazi, and H. Ghasemzadeh, “Inter-Beat Interval Estimation with Tiramisu Model: A Novel Approach with Reduced Error,” ACM Trans Comput Healthc, Jan. 2023, doi: 10.1145/3616020.

4 A. Dourson et al., “PulseNet: Deep Learning ECG-signal classification using random augmentation policy and continous wavelet transform for canines,” May 2023, [Online]. Available: http://arxiv.org/abs/2305.15424

5 A. Y. Hannun et al., “Publisher Correction: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network (Nature Medicine, (2019), 25, 1, (65-69), 10.1038/s41591-018-0268-3),” Nature Medicine, vol. 25, no. 3. Nature Publishing Group, p. 530, Mar. 01, 2019. doi: 10.1038/s41591-019-0359-9.

6 H. Cho et al., “Angiography-based machine learning for predicting fractional flow reserve in intermediate coronary artery lesions,” J Am Heart Assoc, vol. 8, no. 4, 2019, doi: 10.1161/JAHA.118.011685.

7 X. Wang et al., “A Two-Stage Automatic System for Detection of Interictal Epileptiform Discharges from Scalp Electroencephalograms,” eNeuro, vol. 10, no. 11, Nov. 2023, doi: 10.1523/ENEURO.0111-23.2023.

8 G. Van Steenkiste, G. van Loon, and G. Crevecoeur, “Transfer Learning in ECG Classification from Human to Horse Using a Novel Parallel Neural Network Architecture,” Sci Rep, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-019-57025-2.

9 W. Yang, Q. Feng, J. Lai, H. Tan, J. Wang, and L. Ji, “Practical cardiac events intelligent diagnostic algorithm for wearable 12-lead ECG via self-supervised learning on large-scale dataset,” 2022, doi: 10.21203/rs.3.rs-1796360/v1.

10 H. IEEE Engineering in Medicine and Biology Society. Annual International Conference (40th : 2018 : Honolulu, IEEE Engineering in Medicine and Biology Society, and Institute of Electrical and Electronics Engineers, 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society : Learning from the Past, Looking to the Future : July 17-21, 2018, Hawaii Convention Center, Honolulu, Hawaii.

11 Mohebbanaaz, L. V. R. Kumari, and Y. P. Sai, “Classification of ECG beats using optimized decision tree and adaptive boosted optimized decision tree,” Signal Image Video Process, vol. 16, no. 3, pp. 695–703, Apr. 2022, doi: 10.1007/s11760-021-02009-x.

12 J. Mohana et al., “Application of Internet of Things on the Healthcare Field Using Convolutional Neural Network Processing,” J Healthc Eng, vol. 2022, 2022, doi: 10.1155/2022/1892123.

13 M. Mohebbanaaz, Y. P. Sai, and L. V. Rajani Kumari, “Removal of Noise from ECG Signals using Residual Generative Adversarial Network,” in 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2021, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/UPCON52273.2021.9667595.

14 T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, “Automatic ECG classification using continuous wavelet transform and convolutional neural network,” Entropy, 2021, doi: 10.3390/e23010119.

15 T. F. Romdhane, H. Alhichri, R. Ouni, and M. Atri, “Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss,” Comput Biol Med, 2020, doi: 10.1016/j.compbiomed.2020.103866.

16 M. Mahbubur Rahman, M. Rahman Rana, N. A. Alam Munna, M. Saikat Islam Khan, M. Nur-A-Alam, and K. Mohammad Mohi Uddin, “A web-based heart disease prediction system using machine learning algorithms,” 2022. [Online]. Available: www.iaees.org