International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 1, 2026

A Deep Learning Approach for Cardiovascular Risk Prediction

Suneetha Rikhari [1], E. Aravind [2], K Mohana Lakshmi [3]

[1]Dept. of Computer Science and Engineering, Chaitanya Deemed to be University, India.

[2] Dept. of Computer Science and Engineering, Chaitanya Deemed to be University, Hyderabad, India.

[3] ECE Dept., CMR Technical Campus,Hyderabad, India.

*Corresponding author. E-mail: *suneetha.rikhari@gmail.com*

Abstract

Cardio Vascular Diseases (CVDs) are known to cause high percentage of mortalities in the global population; hence, the necessity of reliable and timely systems of risk prediction has been highlighted in the literature. This paper introduces a new deep learning model, which is called Cardio-RiskNet, to predict cardiovascular risk and offers a rich set of clinical and lifestyle variables as inputs. The model developed in Python was tested on the popular dataset on Hearts Disease Health Indicators that could be obtained on Kaggle. The data includes various information including age, body mass index, blood pressure, cholesterol, glucose level, physical activity, and smoking behaviors. The class imbalance issue is addressed with the SMOTE-ENN hybrid resampling strategy, on the one hand, it gives an opportunity to better identify the minority classes, and on the other, it enhances the model generalization. Cardio-RiskNet is based on the integration of a Conv1D framework with Residual blocks and Squeeze-and-Excitation (SE) modules to perform effective hierarchical feature extraction and powerful representation learning. This was trained over 50 epochs and the model achieved a 84.21 percent accuracy and an AUC of 0.7074 which is very good predictive performance in the field of cardiovascular risk assessment. In addition, the Explainable AI (XAI) methods have been applied to present a more vivid understanding of the decisions made by the model, therefore, enabling the use of the model in a safe and reliable setting in terms of clinical decision support. Altogether, Cardio-RiskNet is an effective, easy-to-understand, and scalable application that will aid in the early prevention of cardiovascular risk, and it is highly likely to be adopted in preventive care and analytics in healthcare.

Key words: *SMOTE-ENN, Cardio-RiskNet, Squeeze-and-Excitation (SE) modules*

Introduction

Cardiovascular Diseases (CVDs) are a heterogeneous group of cardiovascular diseases that involve the heart and the circulation system, which includes coronary artery disease, cerebrovascular disease, rheumatic heart disease, congenital heart anomalies, among other vascular disorders. They are the leading cause of mortality and morbidity in the globe and they are not limited by geographical, economic, and demographic borders. The World Health Organization (WHO, 2021) has estimated that CVDs caused some 17.9 million deaths in 2019 or around 32% of all deaths in the world. Among them, 85 percent were related to heart attacks and strokes which highlights the enormous health and financial burden of the conditions. The increasing number of risk factors including obesity, diabetes, high blood pressure, lack of exercises, the use of tobacco and other unhealthy diets also contribute to the burden.

In addition to the personal burden, CVDs cost a great deal of socioeconomic burden. The direct medical treatment, hospitalization, and rehabilitation costs combined with the indirect costs comprised of the loss of productivity, absenteeism and long-term disability provide an immense burden on the healthcare systems-especially in the low and middle-income nations where early diagnosis and preventive care is usually inaccessible.

Timely intervention is important in reducing the effects of CVDs through early screening. Nevertheless, it is difficult to diagnose and predict risks because the genetic, behavioral, and environmental factors are very complex and interact. The pathophysiology of the heart and other cardiovascular diseases is multidimensional, nonlinear, which means that the relationships between such factors as blood pressure, lipid levels, heart rate variability, and the imaging aspects are dynamically changing. These are usually too complicated to be adequately explained by the traditional diagnostic instruments and human knowledge in isolation and require the incorporation of more advanced computational intelligence methods to help make clinical decisions.

This documentation aims at giving a detailed and organization system exploration of Deep Learning (DL) methods implemented in predicting heart diseases. With the growing access to multimodal medical data, including clinical data and imaging data as well as physiological data, DL is an encouraging direction towards the possibility of predicting disease with robust, automated, and

accurate predictions. The paper is designed in such a way that it discusses both theory and practice.

Related work

Machine learning (ML) and artificial intelligence (AI) have also been of significant importance in predicting and diagnosing cardiovascular diseases. The first work related to deep learning was done by Alizadehsani et al. with a data-mining approach to diagnosing coronary artery disease using clinical data and feature-selection techniques with the best diagnostic power [1].The idea of the neural networks as efficient in solving large pattern recognition problems was introduced by LeCun, Bengio, and Hinton [2]. The Deep patient framework by Miotto et al. was an application of unsupervised deep learning on electronic health records (EHRs) to make future clinical outcomes predictions [3]. The non-invasive cardiovascular diagnosis of the cardiovascular units became possible because Itu et al. proposed a machine-learning model of estimating fractional flow reserve using a coronary computed tomography image [4]. Choi et al. developed a patient-based, longitudinal, predictive, clinical-event recurrent, neural-network named Doctor AI [5]. Dawes et al. employed machine learning to forecast the results of pulmonary hypertension with reference to the right-ventricular motion through the cardiac MRI [6]. Proposed by Clifford et al. was automated single-lead short ECG recordings in atrial-fibrillation classification [7]. In medical AI systems, a framework proposed by Lundberg and Lee such as SHAP combines predictions of machine-learning models into a common language that became popular [8]. Bernard et al. studied the application of the deep learning as an automatic cardiac structure segmentation in MRI images [9]. As demonstrated by Zhang et al., the total automation of the echocardiogram interpretation can be achieved in the clinical workflow [10]. Tison et al. examined the concept of passive atrial-fibrillation-based detection of monitoring systems using smartwatches [11]. Attia et al. have revealed that AI-improved electrocardiograms had the capability to recognize dysfunction in the contraction of the heart [12]. In other places, Attia et al. developed AI-based algorithm on ECG to identify atrial fibrillation when in the sinus rhythm [13]. With the assistance of deep neural networks trained on ECG recordings, Hannun et al. were able to get to cardiologist-levels in detecting arrhythmia [14]. Obermeyer et al. identified an issue of bias in health care algorithms and the need of fairness in clinical AI systems [15]. The detectors suggested by Zreik et al. are a recurrent convolutional neural network to detect coronary artery plaque and stenosis with the CT angiography [16]. Rieke et al. discussed the issue of federated learning as a privacy-protecting algorithm of AI model training across institutions in healthcare [17]. Smistad et al. proposed deep learning algorithms to carry out automatic segmentation of left ventricle of cardiac MRI [18]. Ghorbani et al. developed interpretable deep-learning systems to process echocardiograms [19].The World Health Organization reports that cardiovascular diseases have taken the number one position as the leading cause of death in the whole world thus the necessity to have smart diagnostic systems [20]. Ota et al. applied deep learning in the detection of myocardial scars by MRI images [21]. The process of predicting chronic heart disease uses a hybrid feature-importance evaluation process, a feature that was proposed by Nasimov et al. [22]. The reliability of heart-disease classification was improved using class-balancing and feature-engineering by Sharma and Lalwani [23].

System architecture

The suggested cardiovascular disease prediction system is based on the deep-learning-basedarchitecture that would analyze clinical and lifestyle health indicators and predict the risk of heart-diseases with high accuracy. The system takes in the input dataset and balances the classes with SMOTE-ENN and trains a binary classification 1-D Convoluted Neural Network (Conv1D) model.

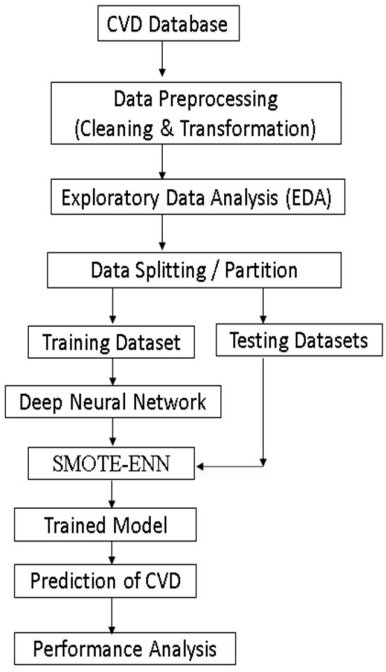

Figure 1: Block Diagram of Deep Learning based CVD Prediction System

The architecture does not require the manual feature engineering process and learns meaningful

representations of the data and the workflow of the proposed system by automatic means, as shown in Figure 1, and includes the following key components: Data Collection based on Kaggle Heart Disease Health Indicators Dataset. This is started by the acquisition of raw health related records in the CVD database. Attributes included in the dataset will be age, blood pressure, cholesterol, BMI, glucose, smoking habits, and other risk factors related to clinical risk. The data collected is then cleaned and transformed to achieve quality and consistencyNext, Exploratory Data Analysis (EDA) is done to determine the trend, correlations, distributions and anomalies in the data. Different methods of analysis are used such as visualizations, statistical summaries and feature correlation analysis to make meaningful conclusions and build models. Following preprocessing and EDA the dataset is further divided into two subsets i.e. training and testing dataset. The model is developed and optimized using the training data, and the testing data are left to objectively test the performance of the model. Deep Neural Network (DNN) model is used to extract the complex patterns and connections between the data. The model architecture which consists of layers, neurons and activation functions-is designed to be able to capture the hidden data relationships in an effective manner. The SMOTE-ENN method is used to solve the problem of class imbalance that is often witnessed in medical data. From using this, SMOTE creates artificial samples of the minority group whereas ENN eliminates the disturbed or incorrect samples which create a more balanced and valid set. After the model is trained on the improved data set, the model will be able to predict cardiovascular disease on the basis of the input features. Lastly, evaluation metrics would be used to determine the accuracy, reliability and the generalization ability of the model.

Deep Neural Network used in the proposed CVD Prediction System

The Figure 2 shows the architecture of the proposed Deep Convolutional Neural Network (DCNN) model that will be used to predict heart diseases. It makes use of a Deep 1D CNN architecture and incorporates Squeeze-and-Excitation (SE) modules and Residual Blocks to boost the feature learning capacity and network stability. The adaptively recalibrating feature responses of channels by the SE mechanism enhances the network selectivity to the most pertinent biomedical pattern of signals. In the meantime, the Residual Blocks are beneficial to address the problem of vanishing gradient and enable the further extraction of the features without information loss. A combination of these elements allows the model to automatically learn and extract discriminative temporal and spectral information on biomedical signals to support the correct prediction of heart disease risk and categorization of cardiac abnormalities.

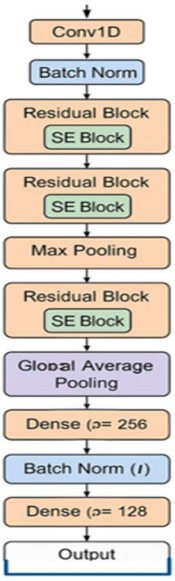

Figure 2: Deep Neural Network used in the proposed CVD Prediction System

The system begins with a publicly available heart disease dataset (Kaggle –Heart Disease Health Indicators). It contains patient health records such as age, BMI, blood pressure, cholesterol, smoking habits, and physical activity. In the Data Preprocessing (Cleaning&Transformation) block the missing or in consistent values are handled, data types are converted to the required format, features are scaled/normalized.Next, SMOTE-ENN technique is applied to balance the dataset and remove noisy samples.

A 1-D Convolutional Neural Network (Conv1D) architecture shown in Figure 2 is used. The layers include: conv1D, Batch Normalization, Residual Blocks with SE(Squeeze-and-Excitation) units, Max Pooling, Global Average Pooling, DenseLayers. This model model learns high-level features automatically. A Conv1D layer with 64 filters and kernel size 3 captures short-term local patterns in the signal, such as changes in heart rhythm or waveform morphology, followed by Batch Normalization to stabilize learning and improve convergence. The network includes multiple residual blocks, each consisting of two convolutional layers with batch normalization. A skip connection to prevent gradient vanishing and allow the model to learn deeper representations. A Squeeze-and-Excitation (SE) Block, which performs Global average pooling to capture channel-wise global information is employed. Then, two fully connected layers are included to learn inter-channel dependencies. Performing channel-

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

wise feature recalibration, allows the network to emphasize critical cardiac patterns (e.g., ST-segment elevation or arrhythmia spikes) relevant to heart disease. The model progressively increases the number of filters (64 → 128 → 256 → 512) across residual stages. MaxPooling1D layers between stages down sample the temporal resolution while preserving the most significant features. GlobalAveragePooling1D aggregates all temporal features into a compact global representation, summarizing overall signal characteristics indicative of cardiac health.

Two dense layers (256 and 128 neurons) with ReLU activation and Batch Normalization capture higher-level nonlinear relationships between extracted features. These layers integrate localized and global patterns, such as heart rate variability, waveform irregularities, and rhythm disruptions. The output layer contains a single sigmoid neuron that outputs a probability score between 0 and 1, representing the likelihood of heart disease presence.

Output ≈ 0: Healthy

Output ≈ 1: High risk of heart disease

Experimental setup and Results

The architecture eliminates the need for manual feature engineering and automatically learns meaningful representations from the data and the workflow of the proposed system, as illustrated in Figure 3.1, involves the following key components: Data collection from Kaggle Heart Disease Health Indicators Data set.The process begins with acquiring raw health-related records from the CVD database. The dataset typically contains attributes such as age, blood pressure, cholesterol levels, BMI, glucose levels, smoking habits, and other relevant clinical risk factors. The collected data is then cleaned and transformed to ensure quality and consistency. Next, Exploratory Data Analysis (EDA) is performed to identify trends, correlations, distributions, and anomalies within the data. Various analytical techniques, including visualizations, statistical summaries, and feature correlation analysis, are applied to gain meaningful insights and support model development. After pre-processing and EDA, the dataset is partitioned into two subsets: the training dataset and the testing dataset. The training data is used to develop and optimize the model, while the testing data is reserved to evaluate its performance objectively. A Deep Neural Network (DNN) model is employed to learn complex patterns and relationships from the data. The model architecture—comprising layers, neurons, and activation functions—is structured to effectively capture hidden data relationships. To address class imbalance commonly observed in medical datasets, the SMOTE-ENN technique is applied. SMOTE generates synthetic samples for the minority class, while ENN removes noisy or mis classified samples, resulting in a more balanced and reliable dataset.Once the model is trained using the enhanced dataset, it becomes capable of predicting cardiovascular disease based on input features. Finally, the performance of the model is assessed using evaluation metrics to ensure accuracy, reliability, and generalization capability.

Results and discussion

The dataset was highly imbalanced with far fewer heart disease cases. After applying SMOTE-ENN, the class distribution became more balanced, helping the deep learning model improve accuracy and recall during prediction.

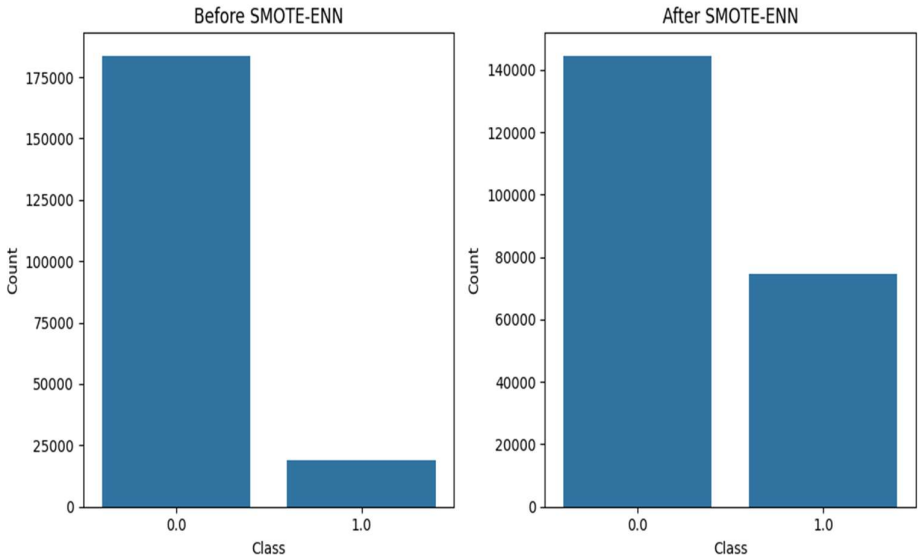

Figure 3: Data distribution before and after applying SMOTE-ENN

Figure 3 illustrates the class distribution of the dataset before and after applying the SMOTE-ENN

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

balancing technique. Initially, the dataset exhibits a significant class imbalance, with 183,830 samples belonging to Class 0 (No Heart Disease) and only 19,114 samples in Class 1 (Heart Disease). This imbalance may negatively impact model performance by biasing predictions toward the majority class. After applying the SMOTE-ENN hybrid resampling method, the class distribution becomes more balanced, resulting in 144,593 samples for Class 0 and 74,549 samples for Class 1. The balanced distribution ensures improved representation of minority class samples, enabling the model to learn more effectively and enhancing classification performance for heart disease prediction

.

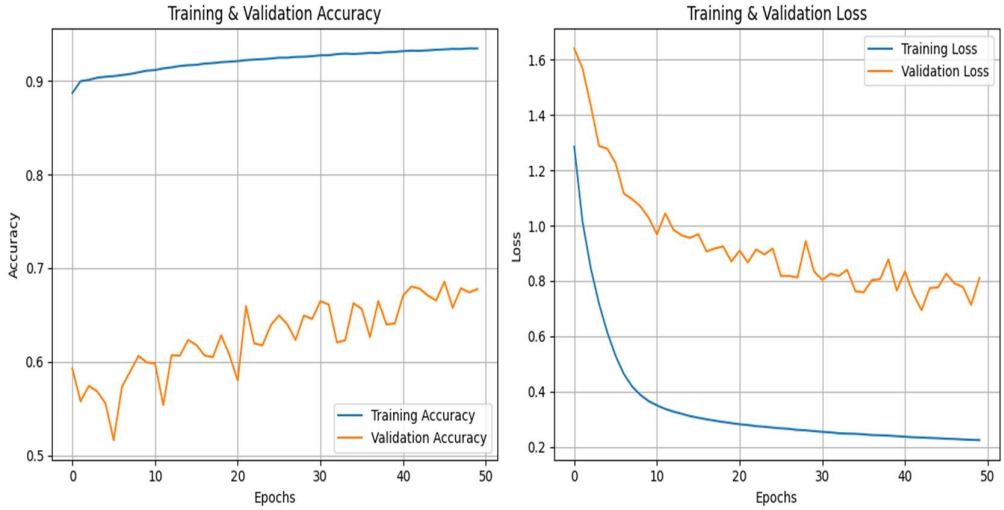

Figure 4: Model Training Performance

The Figure 4 illustrates the training and validation performance of the deep neural network over 50 epochs. The model achieved around 90% training accuracy with steadily decreasing training loss. Validation accuracy fluctuated slightly due to dataset variability, showing mild overfitting but overall strong learning performance.

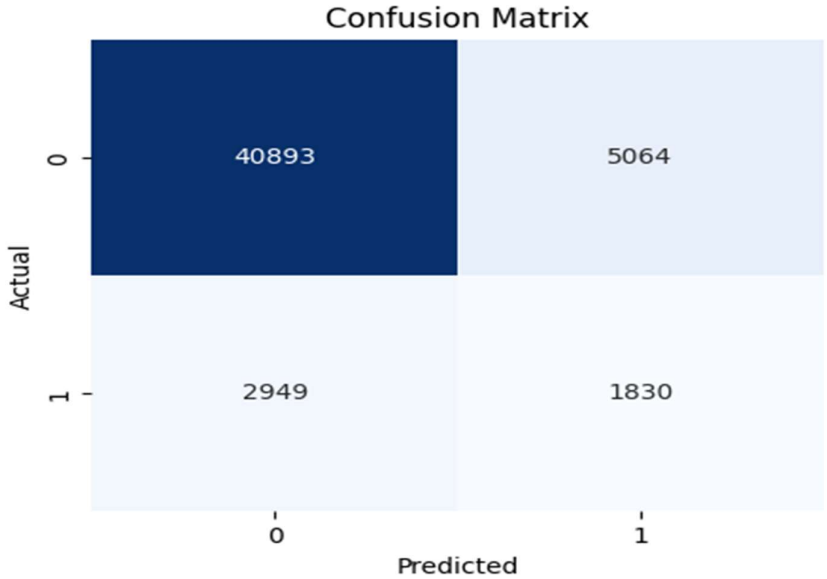

Figure 5: Confusion Matrix

Figure 5 presents the confusion matrix that evaluates the model's ability to differentiate between heart disease and non-disease cases. The results indicate that the model correctly identified 40,893 individuals without heart disease and accurately predicted 1,830 cases with heart disease.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

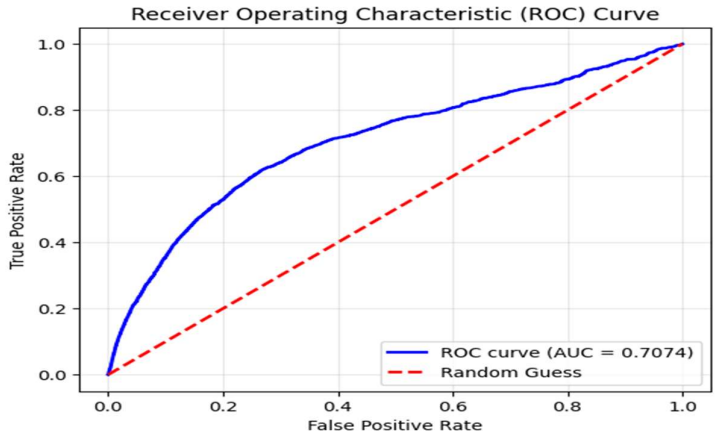

Figure 6: ROC Curve Analysis

Figure 6 illustrates the ROC Curve used to evaluate the model's ability to distinguish between heart disease and non-disease cases. In the graph, the blue curve represents the model's predictive performance, while the red dashed line serves as a baseline representing random guessing. The model achieved an AUC score of 0.7074, indicating a good level of classification discrimination. Since a higher AUC value reflects stronger prediction capability, the result confirms that the model performs significantly better than random chance in identifying heart disease cases.

Table 1: Comparative Analysis of existing method and proposed method

Method	Accuracy
ADASYN-Logistic Regression [1]	0.71
ADASYN-AdaBoost [1]	0.70
ADASYN-XGBoost [1]	0.75
ADASYN-Random Forest [1]	0.82
Weighted K-Nearest Neighbors [2]	0.743
SMOTEENN-CNN (The Proposed)	0.8451

Table 1 provides a comparative analysis between existing machine learning approaches and the proposed SMOTE-ENN-enhanced CNN model. The results indicate that earlier methods such as ADASYN-Logistic Regression, ADASYN-AdaBoost, and ADASYN-XGBoost achieved accuracies of 0.71, 0.70, and 0.75 respectively, reflecting moderate performance levels. The ADASYN-Random Forest model performed comparatively better with an accuracy of 0.82, while the Weighted K-Nearest Neighbors method obtained an accuracy of 0.743. In contrast, the proposed SMOTE-ENN–based Deep CNN model demonstrated the highest accuracy of 0.8451, outperforming all conventional approaches. This improvement highlights the effectiveness of the proposed model's architecture and preprocessing strategy in achieving superior prediction performance for heart disease classification.

Conclusion

The proposed deep learning-based model successfully predicts cardiovascular disease risk by analyzing key clinical and lifestyle factors. By applying the SMOTE-ENN technique, the class imbalance in the dataset was effectively handled, which improved the model's ability to identify heart disease cases more accurately. The model achieved an overall accuracy of 84.21% and an AUC score of 0.7074, indicating good performance in distinguishing between patients with and without heart disease. The evaluation metrics and confusion matrix show that the model performs strongly in detecting non-disease cases while moderately

identifying disease cases. Overall, this project demonstrates the potential of deep learning and Explainable AI in supporting early diagnosis and preventive healthcare, offering a reliable step toward intelligent cardiovascular risk prediction systems.

The future research directions involve the model can be enhanced using advanced hybrid architectures like CNN-LSTM and Attention Networks. Integration with IoT-based wearable devices can enable real-time heart health monitoring. A web or mobile application can be developed for easier access by doctors and patients. Improved Explainable AI (XAI) visualization tools like SHAP or LIME can increase model transparency. Using larger and more diverse healthcare datasets can further boost accuracy and generalization.

Abbreviations

Not Applicable.

Acknowledgment

None.

Author Contributions

The retrieval augmented generation model was designed, the experiments conducted and the manuscript ready by the first author. The second author oversaw the research work, gave technical advice, and went through the manuscript. The authors were content with the end version of the paper.

Conflict of Interest

The author declares no conflict of interest.

Ethics Approval Not applicable.

Funding

No funding received for the proposed work

References

1. Alizadehsani, R., Abdar, M., Roshanzamir, M., et al. Coronary artery disease detection using data-mining techniques and feature selection. Comput Methods Programs Biomed 111, 52–61 (2013).
2. LeCun, Y., Bengio, Y., Hinton, G. Deep learning. Nature 521, 436–444 (2015).
3. Miotto, R., Li, L., Kidd, B.A., Dudley, J.T. Deep Patient: An unsupervised representation to predict the future of patients from electronic health records. Scientific Reports 6, 26094 (2016).
4. Itu, L., Rapaka, S., Passerini, T., et al. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. Medical Image Analysis 44, 157–168 (2016).
5. Choi, E., Schuetz, A., Stewart, W.F., Sun, J. Doctor AI: Predicting clinical events via recurrent neural networks. In: Proceedings of Machine Learning for Healthcare, pp. 301–318 (2016).
6. Dawes, T.J.W., de Marvao, A., Shi, W., et al. Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension. Radiology 283, 381–390 (2017).
7. Clifford, G.D., Liu, C., Moody, B., et al. AF classification from short single-lead ECG recordings: The PhysioNet/Computing in Cardiology Challenge. Computing in Cardiology 44, 1–4 (2017).
8. Lundberg, S.M., Lee, S.-I. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 4765–4774 (2017).
9. Bernard, O., Lalande, A., Zotti, C., et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation. IEEE Transactions on Medical Imaging 37, 2514–2525 (2018).
10. Zhang, J., Gajjala, S., Agrawal, P., et al. Fully automated echocardiogram interpretation in clinical practice. Circulation 138, 1623–1635 (2018).
11. Tison, G.H., Sanchez, J.M., Ballinger, B., et al. Passive detection of atrial fibrillation using a smartwatch. Nature Medicine 24, 409–416 (2018).
12. Attia, Z.I., Kapa, S., Lopez-Jimenez, F., et al. Screening for cardiac contractile dysfunction using AI-enabled ECG. Nature Medicine 25, 70–74 (2019).
13. Attia, Z.I., Noseworthy, P.A., Lopez-Jimenez, F., et al. An AI-enabled ECG algorithm for atrial fibrillation detection during sinus rhythm. The Lancet 394, 861–867 (2019).
14. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., et al. Cardiologist-level arrhythmia detection using deep neural networks. Nature Medicine 25, 65–69 (2019).
15. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. Dissecting racial bias in health-care algorithms. Science 366, 447–453 (2019).
16. Zreik, M., Lessmann, N., van Hamersvelt, R.W., et al. Deep learning analysis of coronary CT angiography for plaque and stenosis detection. Medical Image Analysis 58, 101549 (2019).
17. Rieke, N., Hancox, J., Li, W., et al. The future of digital health with federated learning. npj Digital Medicine 3, 119 (2020).
18. Smistad, E., Falch, T.L., Bozorgi, M., et al. Automatic segmentation of the left ventricle in cardiac MRI using deep learning. Computerized Medical Imaging and Graphics 79, 101686 (2020).
19. Ghorbani, A., Ouyang, D., Abid, A., et al. Deep learning interpretation of echocardiograms. Nature Biomedical Engineering 4, 416–426 (2020).
20. World Health Organization. Cardiovascular diseases (CVDs). WHO Fact Sheet (2021).
21. Ota, H., Kidoh, M., Oda, S., et al. Deep learning-based myocardial scar detection using cardiac MRI. Journal of Magnetic Resonance Imaging 48, 1186–1195 (2018).
22. Nasimov, R., Nasimova, N., Muminov, B. Hybrid feature-importance evaluation for chronic heart-disease prediction. Procedia Computer Science 204, 678–685 (2022).

ISSN 2347–3657
Volume 14, Issue 1, 2026

23. Sharma, N., Lalwani, P. Increasing reliability of heart-disease classification models using class balancing and feature engineering. Biomedical Signal Processing and Control 84, 104777 (2023).