

Deep Learning-Based Mobile Application for Cardiovascular Disease Prediction Integrating Environmental Factors

Bilal Ahmed

chbilalahmed0059@gmail.com

Control Automotive and Robotics Lab, Mirpur University of Science and Technology (MUST), Pakistan

Faryal Amin Jarra

faryalamin27@gmail.com

College of Physics and Optoelectronics Engineering, Shenzhen University, Shenzhen 518060, China

Mubeen Javaid Khan

Mubeenkhan270396@gmail.com

Birmingham City University, Birmingham, UK

Hafiz Muhammad Sulman

Birmingham City University, Birmingham, UK

Sobia Sher

sobia.csit@must.edu.pk

Department of CS&IT, MUST, Mirpur, Pakistan

Muhammad Fahad Latif

mr.muhammadfahadlatif@gmail.com

Department of CS&IT, MUST, Mirpur, Pakistan

Corresponding Author: Bilal Ahmed chbilalahmed0059@gmail.com

Received: 18-02-2026

Revised: 10-03-2026

Accepted: 17-03-2026

Published: 27-03-2026

Abstract

Cardiovascular disease (CVD) remains a leading cause of death worldwide. Conventional machine-learning methods such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression have been widely used for CVD prediction, but their performance can vary across clinical datasets. In addition, many prediction systems do not incorporate environmental determinants such as air pollution. This study presents a deep-learning framework based on an Artificial Neural Network (ANN) and integrates real-time Air Quality Index (AQI) data into a mobile application. After preprocessing and comparative evaluation, the ANN trained on a 303-record clinical dataset achieved a test accuracy of 88.81%. A React Native application provides location-based AQI alerts, while a FastAPI backend handles prediction requests. The resulting system combines clinical prediction with environmental context to support accessible, real-time cardiovascular risk assessment.

Index Terms— Cardiovascular disease, artificial neural networks, mobile applications, Air Quality Index, machine learning, deep learning.

INTRODUCTION

Cardiovascular disease (CVD) is a major global cause of mortality and disability [1], [2]. Common forms include coronary heart disease, hypertension, and cerebrovascular disease [3]. Established risk factors include age, sex, smoking, diet, physical inactivity, high blood pressure, diabetes, and elevated cholesterol [4]. Large international studies also show that many cardiovascular events are associated with modifiable behavioural and lifestyle factors [5].

Recent research has broadened this view to include environmental determinants. Air pollution, temperature variation, noise, and other exposures can contribute to cardiovascular risk through mechanisms including inflammation, oxidative stress, and autonomic imbalance [6]–[9]. Air pollution is of particular interest because particulate matter, nitrogen dioxide (NO₂), and ozone (O₃) can affect cardiovascular function, with both short- and long-term exposure associated with adverse outcomes [9]–[13]. The Air Quality Index (AQI) provides a practical way to communicate ambient air-quality conditions to the public [10].

Artificial intelligence (AI) has also become increasingly important in cardiovascular risk prediction. Traditional classifiers, ensemble methods, and neural networks have been evaluated on structured clinical datasets, with performance depending strongly on feature selection, preprocessing, dataset size, and validation strategy [17]–[24], [27]–[39]. However, most published prediction systems focus primarily on clinical variables and do not provide environmental information directly to end users.

This study therefore develops an ANN-based CVD prediction system and integrates it with real-time AQI information in a cross-platform mobile application. The main contributions are:

- A deep-learning CVD prediction model trained and evaluated on a clinical dataset.
- Integration of real-time AQI information to provide environmental context.
- A React Native mobile application connected to a FastAPI prediction service.
- An age- and AQI-based alert mechanism for users in predefined higher-risk age groups.

Related Work

LITERATURE REVIEW

Recent literature shows three closely related research directions: machine-learning prediction of CVD, integration of environmental variables, and deployment of prediction models in interactive or real-time systems. Reviews and comparative studies report extensive use of Logistic Regression, KNN, SVM, Random Forest, boosting methods, and neural networks [17]–[24], [27], [32], [36]–[39]. These studies generally emphasize preprocessing, feature selection, hyperparameter optimization, and model comparison rather than a single universally superior classifier.

Environmental variables are increasingly recognized as relevant predictors. Sianga et al. [28] reported improved predictive performance when geographical and environmental variables were added to clinical and demographic features. Reviews of environmental cardiovascular risk also identify particulate matter, ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide, heavy metals, pesticides, temperature, and other exposures as potentially important factors [9], [14], [41], [43]. These findings support incorporating environmental context into user-facing cardiovascular risk systems.

Hybrid and ensemble approaches have produced strong results on several benchmark datasets [33]–[39]. However, reported accuracies vary substantially because studies use different datasets, preprocessing procedures, feature sets, and evaluation protocols. Consequently, accuracy values from separate studies should not be interpreted as directly comparable. Real-time systems have also demonstrated the feasibility of combining monitoring, machine learning, and user interfaces [25]. Cross-platform frameworks such as React Native and Flutter further reduce the development burden for mobile health applications [14].

The literature therefore indicates a gap between high-performing offline prediction studies and practical systems that combine clinical prediction with real-time environmental information. This work addresses that gap by linking an ANN model to a mobile interface and an external AQI service. The emphasis is

on system integration and usability rather than on claiming universal superiority over all existing algorithms.

METHODOLOGY

Dataset Description

Three datasets were initially considered: two from the UCI repository and one from Kaggle. The final model used a UCI dataset containing 303 records and 14 attributes. The dataset contained 165 positive cases (54.3%) and 138 negative cases (45.7%). Exploratory analysis and correlation inspection were performed before model training.

Table 1

CVD status		Proportion
CVD present (1)	165	54.3%
CVD absent (0)	138	45.7%
Total	303	100%

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Figure 1

Model Comparison and Hyperparameter Tuning

Logistic Regression, KNN, SVM, and ANN were evaluated using the same clinical dataset. Baseline classifiers were first tested without tuning and then re-evaluated after hyperparameter optimization.

TABLE II. MACHINE-LEARNING ACCURACY BEFORE TUNING MACHINE-LEARNING ACCURACY BEFORE TUNING TUNING		
Model	Train accuracy	Test accuracy
Logistic Regression	85.78%	86.81%
KNN	88.62%	86.81%
SVM	92.41%	87.91%
TABLE III. MACHINE-LEARNING ACCURACY AFTER TUNING MACHINE-LEARNING ACCURACY AFTER TUNING TUNING		
Model	Train accuracy	Test accuracy
Logistic Regression	85.30%	87.91%
KNN	81.04%	84.61%
SVM	86.25%	87.91%

Data Preprocessing

Missing values and duplicate records were removed. Categorical variables were numerically encoded, and numerical features were standardized using StandardScaler to obtain zero mean and unit variance. The processed dataset was divided into training and testing subsets using an 80:20 split. The workflow is shown in Fig. 3.

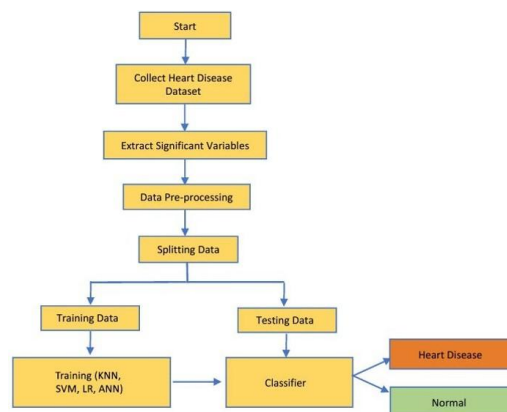


Fig. 2. Workflow of the predictive model.

Artificial Neural Network Architecture

The ANN was implemented using TensorFlow and Keras. The input layer represented 13 clinical characteristics, followed by two hidden layers using ReLU activation and an output layer using sigmoid activation for binary classification. The Adam optimizer and binary cross-entropy loss were used.

Hyperparameters included the learning rate, batch size, and number of neurons in the hidden layers. The selected ANN achieved 88.18% training accuracy and 88.81% test accuracy.

Hyperparameter Optimization

A grid-search strategy evaluated learning rates of 0.0005 and 0.001, batch sizes of 16 and 32, and different hidden-layer configurations. Training was limited to 100 epochs, with early stopping used to reduce overfitting.

TABLE IV. HYPERPARAMETER CONFIGURATIONS AND VALIDATION ACCURACY				
Learning rate	Batch size	Hidden layers	Validation accuracy	
0.001	16	[32, 16]	85.3%	
0.001	32	[16, 8]	87.2%	
0.0005	32	[16, 8]	88.1%	

System Architecture and Mobile Application Integration 1) Backend Architecture

The backend was developed in Python using FastAPI. The trained TensorFlow/Keras model was exported and served through a REST API. During development, the API was hosted on a local machine and accessed by the mobile application through the machine's local IP address.

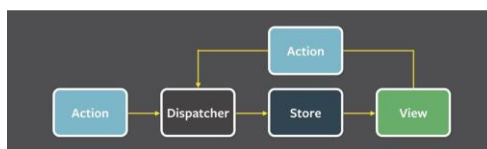


Fig. 3. Application architecture.

Frontend and Mobile Integration

A React Native application allows users to enter clinical parameters and receive predictions. React Navigation manages screen transitions, while the Fetch API handles HTTP requests. The interface also presents environmental alerts when the predefined AQI and age conditions are satisfied.



Fig. 4. Mobile application user interface.

Environmental Risk Comparator

The system retrieves AQI information through an external air-quality service. A warning is triggered when AQI exceeds 150 and the user's age is below 15 or above 60. The rule is intentionally simple and is used as an environmental alert rather than as a clinical diagnosis.

TABLE V.

ENVIRONMENTAL RISK TRIGGER CONDITIONS

AQI level	Age group	Risk flag
$AQI \leq 100$	Any	Normal
$100 < AQI \leq 150$	<15 or >60	Moderate risk
$AQI > 150$	<15 or >60	High risk (alert)
$AQI > 150$	15–60	Moderate risk

Risk rule: High risk if $AQI > 150$ and ($Age < 15$ or $Age > 60$); otherwise, Normal.

Application and Model Flow

The mobile application sends clinical inputs to the FastAPI service as JSON. The server preprocesses the data, retrieves AQI information, invokes the ANN model, and returns the prediction to the application. The sequence is illustrated in

Fig. 5

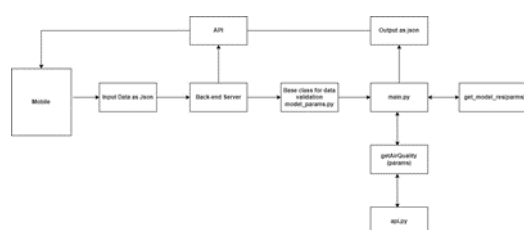
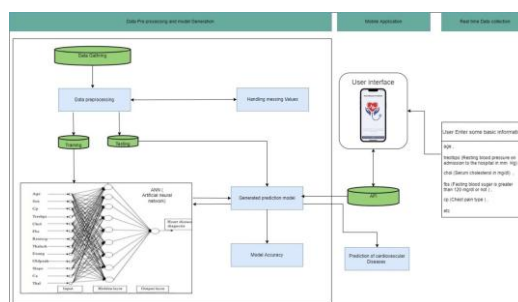


Fig. 6



PRODUCT TESTING AND EVALUATION

Testing focused on input validation, backend communication, model inference, network configuration, and end-to-end operation. The application verifies data types and required fields before submitting requests. Early communication problems caused by using localhost from an Android device were resolved by using the host machine's local IP address and configuring the required Android manifest setting.

The model was also tested using known dataset samples. The application returned predictions consistent with the expected labels for the tested samples. AQI alerts were validated by simulating locations with AQI values above the configured threshold.

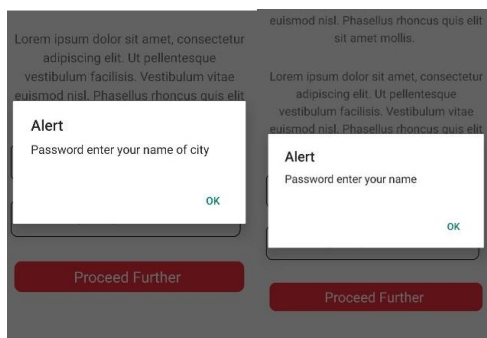


Fig. 7. Input-validation alerts.

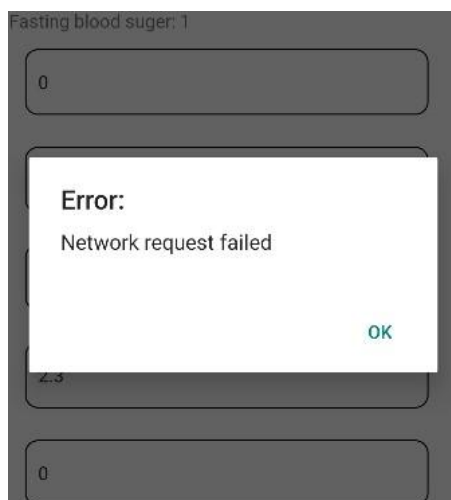


Fig. 8. Network error caused by an incorrect localhost configuration.

Input Validation

TABLE VI. INPUT VALIDATION RULES			
	Field	Validation rule	
Age		Integer, > 0	
Sex		Binary: 0 (female), 1 (male)	

Cholesterol	Integer, non-negative
Chest pain type	Integer (0–3)
Exercise-induced angina	Boolean (0 or 1)

Backend Response Verification

POST requests from the mobile application were checked using backend logs. JSON payloads were parsed successfully, preprocessing was executed, and predictions were returned with an average response time of 2.1 s.

End-to-End Functionality

Test ID	Condition	Expected output	Result
T1	Normal clinical + AQI ≤ 150	Low risk, no alert	Pass
T2	High-risk clinical + AQI ≤ 150	High risk, no alert	Pass
T3	Low-risk clinical + AQI > 150	Low risk + AQI alert	Pass
T4	High-risk clinical + AQI > 150	High risk + AQI alert	Pass
T5	Missing age	Input error alert	Pass

Mobile Performance

The application loaded in under 2 s during testing, and prediction results were displayed with minimal delay. The interface was responsive across Android devices with different screen resolutions.

```
new_model = tf.keras.models.load_model('content/my_model.h5')
result = new_model.predict(sc.transform([[164, 1, 0, 1, 30, 0, 0, 0, 0, 240, 136.5, 85.0, 20.420000, 70.0, 77]]))

if result[0] >= 0.50 :
    print('We are sorry, according to the system prediction, you are living with heart disease')
else :
    print('We are happy to announce, currently you are living with healthy heart')
We are happy to announce, currently you are living with healthy heart
```

Fig. 9. Model validation using a dataset sample entered through the mobile application.

RESULTS AND DISCUSSION

Model Evaluation Metrics

The ANN achieved a test accuracy of 88.81%, with precision of 86.70%, recall of 91.20%, and an F1-score of 88.90%.

TABLE VIII. ANN MODEL PERFORMANCE METRICS

Metric	Value	
Accuracy	88.81	%
Precision	86.70	%
Recall	91.20	%

Comparison with Traditional Classifiers

For consistency, the final comparison below reports the tuned test accuracies from Table III alongside the selected ANN result. This avoids mixing results from different experimental configurations.

TABLE IX. FINAL TEST ACCURACY COMPARISON

Classifier	Test accuracy
ANN	88.81 %
SVM	87.91 %
KNN	84.61 %
Logistic Regression	87.91 %

Under the reported experimental configuration, the ANN obtained the highest test accuracy among the evaluated models. Because the datasets, preprocessing procedures, and validation protocols differ across published studies, these results should be interpreted as an internal comparison rather than as evidence of superiority over all prior work.

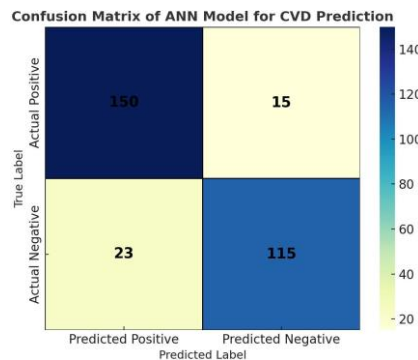


Fig. 10. Confusion matrix for the ANN CVD classifier.

Integration of Environmental Data

The application supplements the clinical prediction with real-time AQI information. When the configured AQI threshold is exceeded, the application presents an environmental alert according to the age-based rule in Table V. This provides contextual information alongside the model prediction rather than treating AQI as a replacement for clinical assessment.

Application Alert Response

After the user submits clinical parameters, the FastAPI service preprocesses the data and returns the ANN prediction. The mobile interface displays the result as an alert containing the prediction outcome and a recommendation to seek medical advice when appropriate.

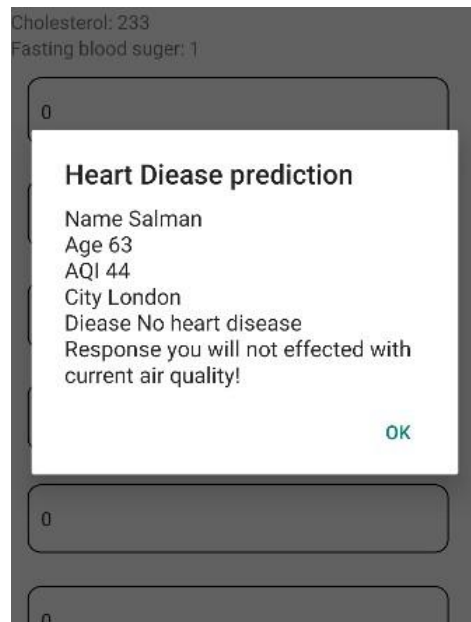


Fig. 11. Mobile application showing the model prediction and alert.

LIMITATIONS

The study has several limitations. First, the clinical dataset contains only 303 records, which limits generalizability to larger and more diverse populations. Second, the environmental component is limited to AQI; additional variables such as temperature, physical activity, and smoking status were not integrated because of data-availability constraints. Third, the application has not been evaluated in prospective clinical trials. Further validation is therefore required before conclusions about real-world clinical effectiveness can be drawn.

CONCLUSION AND FUTURE WORK

This study presented a deep-learning-based mobile application for CVD risk prediction using clinical parameters and real-time AQI information. The ANN achieved a test accuracy of 88.81% on the selected 303-record dataset, while the application provided an integrated interface for clinical prediction and environmental alerts. The system demonstrates how a prediction model can be connected to a mobile interface and an external environmental data service.

Future work will focus on larger and more diverse datasets, additional lifestyle and environmental predictors, prospective validation in real-world clinical settings, and integration with wearable health-monitoring devices. Multilingual and region-specific interfaces could also improve accessibility.

REFERENCES

- Gaziano, T. A., Reddy, K. S., Paccaud, F., et al. (2006). Cardiovascular disease. In *Disease Control Priorities in Developing Countries* (2nd ed., pp. 645–662). The World Bank.
- World Health Organization. (2025). *Cardiovascular diseases (CVDs)*. <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- Fuchs, F. D., & Whelton, P. K. (2020). High blood pressure and cardiovascular disease. *Hypertension*, 75(2), 285–292.

- McCreadie, C. (2003). Coronary heart disease risk factors. *British Journal of Cardiac Nursing*, 18(6), 214–218.
- Yusuf, S., et al. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study). *The Lancet*, 364(9438), 937–952.
- Francula-Zaninovic, S., & Nola, I. A. (2018). Management of measurable variable cardiovascular disease risk factors. *Current Cardiology Reviews*, 14(3), 153–163.
- Kreatsoulas, C., & Anand, S. S. (2010). The impact of social determinant on cardiovascular disease. *Canadian Journal of Cardiology*, 26(Suppl. C), 8C–13C.
- Lee, B. J., Kim, B., & Lee, K. (2014). Air pollution exposure and cardiovascular disease. *Toxicological Research*, 30(2), 71–75.
- Dominici, F., et al. (2019). Air pollution and health: a review of the evidence. *JAMA*, 322(14), 1351–1352.
- European Environment Agency. (2021). *Air Quality in Europe 2021* (EEA Rep. No. 15/2021).
- Kelly, F. J. (2003). Oxidative stress: its role in air pollution and adverse health effects. *Occupational and Environmental Medicine*, 60(8), 612–616.
- Green Car Congress. (n.d.). *Ultrafine particles and heart disease*. <https://www.greencarcongress.com>
- NHS Inform. (2012). *Air pollution and cardiovascular health*. NHS. <https://www.nhsinform.scot>
- Jost, G., & Taneski, V. (2025). State-of-the-art cross-platform mobile application development frameworks: a comparative study of market and developer trends. *Informatics*, 12(2), 45.
- scikit-learn. (n.d.). *RandomForestClassifier — scikit-learn documentation*. <https://scikit-learn.org>
- Muller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Media.
- Sharma, S., & Parmar, M. (2020). Heart diseases prediction using deep learning neural network model. *International Journal of Innovative Technology and Exploring Engineering*, 9(3), 2245–2248.
- Arthy, R. S., et al. (2018). Review of heart disease prediction models using data mining techniques. *International Journal of Computer Applications*.
- Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*, 1022(1).
- Akhil, M., et al. (2013). Heart attack prediction using KNN and logistic regression. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*.
- Sharma, V., et al. (2020). Early heart disease detection using machine learning. *Journal of Medical Systems*.
- Kavitha, M., et al. (2021). Hybrid models for heart disease risk prediction. *International Journal of Biomedical Research*.
- Ramprakash, P., et al. (2020). Combining deep neural networks with statistical models for heart disease prediction. *Computational Intelligence in Healthcare*.

- Jagtap, A., et al. (2019). Web-based heart disease prediction using machine learning. *International Journal of Advanced Research in Computer Science*.
- Otoom, A. F., Kefaye, A., Ashour, M., Shanti, Y., & Al-Majali, M. (2015). Real-time monitoring of patients with coronary artery disease. *International Journal of Future Computer and Communication*, 4(3), 207–210.
- Purushottam, S., et al. (2015). Heart disease prediction using C4.5 and KEEL framework. *International Journal of Computer Applications*.
- Karna, V. V. R., Karna, V. R., Janamala, V., Rao Devana, V. N. K., Ravi Sankar, Ch. V., & Tummala, A. B. (2024). A comprehensive review on heart disease risk prediction using machine learning and deep learning algorithms. *Archives of Computational Methods in Engineering*, 32(8).
- Sianga, B. E., Mbago, M. C., & Msengwa, A. S. (2025). Predicting the prevalence of cardiovascular diseases using machine learning algorithms. *Intelligence-Based Medicine*, 11.
- Tasnim, A. F., et al. (2025). Explainable machine learning algorithms to predict cardiovascular strokes. *Engineering, Technology & Applied Science Research*, 15(1), 20131–20137.
- Qi, X., et al. (2025). Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biology*, 79.
- Kamble, V. B., Gulabani, B., Narkhede, S., & Godse, S. (2025). Predicting heart disease with machine learning: enhancing accuracy through algorithmic approach. *International Journal of Multidisciplinary on Science and Management*, 2(2), 36–56.
- Ahmed, M., & Husien, I. (2024). Heart disease prediction using hybrid machine learning: a brief review. *Journal of Robotics and Control*, 5(3), 884–892.
- Mandava, M., & Vinta, S. R. (2024). MDensNet201–IDRSRNet: efficient cardiovascular disease prediction system using hybrid deep learning. *Biomedical Signal Processing and Control*, 93.
- Navita, et al. (2025). Advanced hybrid machine learning model for accurate detection of cardiovascular disease. *International Journal of Computational Intelligence Systems*, 18.
- Tripathy, B., Beborita, S., Tripathy, S. S., & Tran, T. A. (2024). An application of hybrid machine learning framework to predict the heart diseases. In *Smart Healthcare Systems: AI and IoT Perspectives*.
- Rasheed, S., et al. (2024). Heart disease prediction using GridSearchCV and Random Forest. *EAI Endorsed Transactions on Pervasive Health and Technology*, 10(1).
- Khan, M. A., et al. (2024). Optimal feature selection for heart disease prediction using modified Artificial Bee Colony (M-ABC) and K-nearest neighbors (KNN). *Scientific Reports*, 14.
- Hoque, R., Billah, M., Debnath, A., Hossain, S. M. S., & Sharif, N. B. (2024). Heart disease prediction using SVM. *International Journal of Science and Research Archive*, 11(2), 412–420.
- Elsedimy, E. I., AboHashish, S. M. M., & Algarni, F. (2024). New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization. *Multimedia Tools and Applications*, 83, 23901–23928.

- Parizad, R., et al. (2025). Emerging risk factors for heart failure in younger populations: a growing public health concern. *World Journal of Cardiology*, 17(4).
- Kumar, V., et al. (2025). Environmental pollutants as emerging concerns for cardiac diseases: a review on their impacts on cardiac health. *Biomedicines*, 13(1).
- Bradley, M., Land, D., Thompson, D. A., & Cwiertny, D. M. (2025). A critical review of a hidden epidemic: examining the occupational and environmental risk factors of chronic kidney disease of unknown etiology. *Environmental Science: Advances*, 4, 57–76.
- Joseph, P., et al. (2025). Cardiovascular disease in the Americas: the epidemiology of cardiovascular disease and its risk factors. *The Lancet Regional Health – Americas*, 42.