# Predictive Analytics in Cardiovascular Health: Leveraging Deep Learning Algorithms for Early Cardiac Disease Identification

**Salman Muneer[1*], Syeda Anmol Jameel Habibi[2], Hammad Raza[3]**

[1]University of Central Punjab, Lahore, 51000, Pakistan
[2]Business and Marketing, Anglia Ruskin University Uk
[3]The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan
[*]Corresponding Author: Author's Name. Email: salman.muneer@ucp.edu.pk

Received: XX Month 202X; Accepted: XX Month 202X

**Abstract:** Cardiac diseases have remained a major issue in today's era; if diagnosed at some early stages, not only can human lives be saved at the initial level of disease, but a proactive approach can also be employed worldwide accordingly. Nowadays, cardiac diseases are frequent, increasing so rapidly in humans due to improper diet, smoking, lack of exercise, diabetes, people having stress, high blood pressure, and, more specifically, deficient knowledge about the disease occurrence. Most healthcare units lack classification and decision-making techniques to anticipate the disease, consequently unable to perform necessary precautionary measures to decrease the disaster impacts of disease; therefore, it is required to work on such effective approaches having the projection of prior identification of disease and present more reliable decision-making results. The proposed model will provide a reliable Recurrent Neural Network (RNN) approach toward cardiac disease prediction, presenting the improvement in the previous research's success ratio and decreasing the possible loss and execution time. This proposed model achieves more than an accuracy of 97% in predicting cardiac disease at an early stage.

**Keywords:** Cardiac disease; intelligent model; disease prediction; deep learning

## 1 Introduction

Cardiac diseases have been increasing day by day, which is an alarming situation for the people of the whole world. According to the World Health Organization (WHO), the patient-dying ratio has increased to approximately 10 million every year due to cardiac disease [1]. It occurs primarily due to unhealthy lifestyles like poor diet, stressful life, and lack of knowledge concerning taking precautionary measures against the disease. Many people living in urban areas are more affected than in rural areas. Moreover, it is not so easy to identify the disease as numerous factors are involved, including age, gender, cholesterol, blood pressure, glucose level, ECG, status of heartbeat, angina, and patient's condition during peak exercise, etc. [2]. People exceeding 65 years of age are major common victims of heart failure, more specifically with a prior heart attack(s) in their lives. Doctors analyze the medical data of the patients as well as the family background to know the likelihood of the disease, physical checkups, medical test reports, etc. An immense amount of individual patient data is being generated, specifically at some early stages, which demands careful devotion towards performing an accurate, efficient, and trusted diagnosis of the disease, specifically with this substantial data. As medical data of the patients are required to diagnose heart failure disease, a proactive approach is being presented in this research article that possesses the prospects to present the prior disease diagnoses and prediction mechanism to minimize the impacts of disease risks and save human lives [3].

According to the WHO Index 2019, 55% of the 55.4 million demises that occurred globally were due to the top 10 leading causes. In terms of the overall number of lives lost, the world's leading causes of mortality fall into three categories: cardiovascular (like ischemic cardiac disease, stroke), respiratory (like chronic obstructive pulmonary disease, lower respiratory infections), and newborn diseases. Congenital asphyxia, as well as congenital trauma, neonatal sepsis, and infection, especially prenatal problems, are all examples of these. Communicable (infectious and parasitic illnesses, as well as maternal, congenital, and nutritional disorders), non-communicable (chronic), and injury problems are the three categories of deaths. In 2019 Worldwide, 7 of the top 10 reasons for death were not contagious. 44% of all deaths happened due to these 7 main reasons, or 80% of the top 10 demised, but in 2019, no infectious disease accounted for 74% of all deaths worldwide. Ischemic disease is the world's biggest cause of death, accounting for 16 per cent of all deaths worldwide, according to WHO figures. Since 2000, the illness has caused the greatest increase in fatalities, exceeding 2 million in 2019 to 8.9 million entire demises: paralysis and chronic obstructive pulmonary disease account for 11 and 6% of the second and third dominant deaths, respectively [4].

Lower respiratory infections are the worst infectious illness in the world, accounting for the fourth most common source of demise. However, there has been a considerable reduction in the number of fatalities: in the year 2019, 2.6 million individuals were required to breathe, which is 460 thousand less than in 2000. The condition of the newborn is ranked fifth. Neonatal mortality, on the other hand, is one of the groups that have had the greatest reduction in absolute fatalities during the last two decades: In 2019, 1.2 million fewer infants and small children died because of these illnesses than in 2000. Diarrhoea is the most common cause of death reduction globally; fatalities have decreased from 2.6 million in 2000 to 1.5 million in 2019 [5].

Diabetes is growing rapidly by 70% since 2000 as one of the 10 leading diseases that are harmful. The diseases that were shown in 2000 are not on the list of leading diseases because diseases change from time to time. Like HIV / AIDS, which was at its peak in 2000. In the previous 20 years, HIV/AIDS-related deaths have decreased by 51%, and it has been ranked 19th in the world from 2000 to 2019. The number of deaths increased in 2000 from 813,000 to 1.3 million in 2019. It is noted that people who spend their lives in low-income countries are more likely to suffer from infection problems than non-communication disease problems. Although the worldwide crisis, in low-income countries, primarily deaths occurred due to 6 leading diseases in 10 diseases. Malaria, TB, and HIV/AIDS treatment are among the top ten illnesses. All three, though, are rapidly dwindling. HIV/AIDS is the leading cause of mortality, with 59 percent fewer fatalities in 2019 than that in 2000, 161,000 people and 395,000 people, correspondingly [6].

Diarrhea is one of the leading causes of death in low-income nations, accounting for one of the top five causes of death in this group. However, diarrhea will be down in low-income countries, with the second largest deaths in the leading 10, 231,000, with low mortality. Chronic obstructive pulmonary disease mortality is lower than in other low-income groups, especially in low-income countries, as shown in Figure 1. It does not appear among the top ten limited-income countries but in the top five income categories. People from affluent nations can expect to live 18 years longer than their poor neighbours, and affluence can determine access to health care in individual countries and cities [7].

Data mining techniques, like regression, classification, clustering, etc., when implemented with various machine learning-based models on a large-scale using health-care data, the hidden patterns in the data can possibly be detected at some level, and the concerned authorities can develop such policies that are based on the processed data blueprint. Clinical Support Systems (CSS) and Decision Support Systems (DSS) are used as intelligent models in the medical field for healthcare purposes [8].

In the medical field, for analysis, support, and diagnosis purpose for medical specialists and healthcare units, a system is introduced called the Clinical Decision Support System (CDSS) that's the application of DSS in many healthcare centres, DSS is used for the collection of important information regarding the healthcare. Three main categories of machine learning algorithms include supervised, unsupervised, and reinforcement learning algorithms. Supervised learning algorithms are implemented on a known dataset having inputs and outputs (label) values, pattern in data is identified with algorithms, analysis of the data and its predictive outcomes, and the process continues until better accuracy is achieved. Examples of supervised algorithms are Artificial Neural Networks (ANN), K-nearest neighbors (KNN), Support Vector Machine (SVM), Linear Regression, Decision Trees, etc [18-22]. In unsupervised learning algorithms, data is not labeled, just having input values and no output values. Examples of unsupervised algorithms are the Apriori algorithm and K-Means Clustering etc. A list of algorithms is provided in Figure 2, including the proposed model. The proposed model discusses supervised learning algorithms with ensemble learning techniques, Bagging, and Boosting. In the end, it will be concluded that the proposed algorithm XGboost performs better on big data [9].

## 2 Literature Review

The authors [10] investigated a training function of backpropagation through a neural network that achieved 91% accuracy on the training section of data and 76% for testing the data. Here dataset consists of 180 samples, and 4 attributes are selected for analysis purposes. The dataset is taken from the UCI machine learning repository. In this technique, the researcher used 70% of the data for training purposes, 15% for validation, and 15 % for testing purposes with the help of the MATLAB tool.

The authors [11] proposed a fuzzy logic method using the RBFL prediction algorithm, achieving an accuracy of 76.51%. Here, the Dataset is taken from the UCI machine learning repository, having Cleveland, Hungarian, and Switzerland datasets. Here, two key components are used: Feature reduction and disease prediction with a fuzzy classifier.

The authors [12] of this research presented that a cardiac disease prediction system was developed using two machine learning algorithms: Support Vector Machine (SVM) and Naïve Bayes. The study combined three datasets: Cleveland, Hungary, and Switzerland, and used these to train the algorithms. The results showed that the accuracy of the proposed approach was 87% for SVM and 86% for Naïve Bayes. This indicates that the prediction system could correctly predict cardiac disease in most cases. SVM is an algorithm that uses a boundary, known as a hyperplane, to separate data into different categories. In this case, it separated individuals with cardiac disease from those without it. Naïve Bayes, on the other hand, is a probabilistic algorithm that predicts the class of a data point based on the likelihood of each class. The algorithm assumes that the features are independent of each other. In this study, both algorithms were able to provide good results, with SVM slightly outperforming Naïve Bayes. The combination of these algorithms and the use of multiple datasets improved the accuracy of the prediction system, making it a helpful tool for predicting cardiac disease.

The authors proposed an analysis of the dataset in the Anaconda Python language and trained the network using a gradient descent algorithm. The accuracy is 85.074 for the backpropagation neural network (BNN) and 92.58% for logistic regression (LR). BNN is used in multilayer feed-forwarding networks, while Logistic regression (LR) is used for the categorical value. Cleveland Dataset is occupied by UCI machine learning, which consists of 303 instances, and 13 attributes out of these 270 instances are used for analysis.

The authors [14] proposed a machine learning model using the multi-layer perception algorithm, here the Cleveland dataset from UCI machine learning is used having 303 instances and 13 attributes, 245 instances are tested out of 303. 80.85% accuracy has been achieved from 245 records.

The authors [15] proposed a cardiac disease prediction system using the Naïve Bayes algorithm, Artificial Neural Network (ANN), Classifier, and Support Vector Machine as an ensemble model. They achieved 93% overall accuracy, and bagging was implemented on this ensemble model with the help of the Fast Fourier transform method and combined.

The authors [16] employed Machine Learning (ML) proved more effective for prediction and decision-making for large health care datasets. Due to Machine Learning advantages its techniques are used in different areas of the Internet of Things (IoT). Here Researchers proposed a model for the prediction of cardiac disease prediction by using significant features of Machine Learning techniques to improve the accuracy. The model has many combinations of features and classification techniques. They obtained 88.7% accuracy from a Hybrid Random Forest with a linear model (HRFLM) for the prediction model of cardiac disease.

The authors proposed a parameter tuning framework first training the algorithm, after transforming the function, then selecting hidden nodes performing its simulation, and checking its validation result has higher accuracy. This dataset was taken from the UCI machine learning repository in Cleveland, and the Star log cardiac disease dataset was used to investigate cardiac disease using a feed-forward back propagation artificial neural network. 80% of the data was used for training, 10% for validation, and 10% for testing. The average accuracy achieved from the Cleveland dataset is 90.9%, and for the Star log, it is 90% [17].

## 3 Proposed Methodology

The early detection and diagnosis of cardiac disease is paramount in the healthcare industry, as it can help save lives and prevent further progression of the disease. To achieve this goal, this research proposes using deep learning to classify cardiovascular disease datasets and accurately predict cases of cardiac disease with a minimum number of attributes. The proposed model is depicted in Figure 1 and has the potential to be a significant step forward in the fight against cardiac disease.

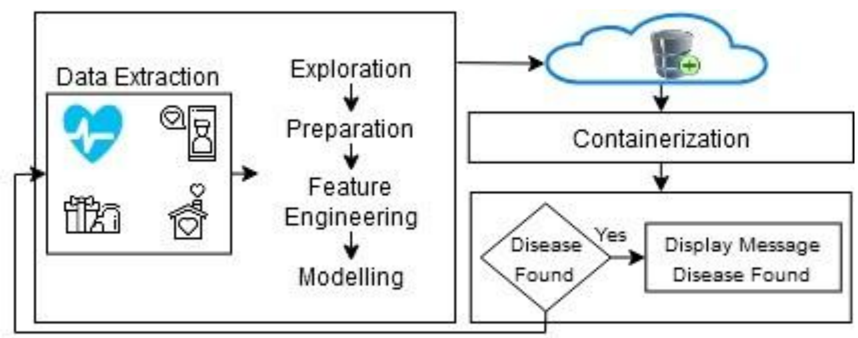


**Figure 1**: Proposed Methodology

The proposed model for predicting cardiac disease is a deep-learning approach consisting of three crucial steps. The first step of the model is data extraction, which is the process of collecting and retrieving diverse types of information related to cardiac disease from multiple sources. These sources may be poorly organized or unstructured digital devices, making the data extraction process challenging. The extracted data is then sent for preparation, which involves gathering, combining, structuring, and organizing it to make it usable for analytics and data visualization.

The prepared data undergoes feature engineering in the second step, where the raw data is transformed into features suitable for supervised learning. Feature engineering is crucial in this process as it involves selecting and manipulating the data to extract meaningful features that can significantly impact the accuracy of the disease prediction. The designed features are then sent for modelling, where machine learning algorithms are applied to predict the presence of cardiac disease based on the set of parameters.

In the final step, the model is stored in the cloud, and the trained data is imported from the cloud for containerization. Containerization is a crucial aspect of the model, as it packages the code and its dependencies into a standard unit, enabling the application to run efficiently and reliably across different computing environments. After containerization, the model is used to check if the disease prediction is present, and if yes, a message is displayed, and the result is stored in the cloud. In the case of a negative prediction, the model is retrained, and the process continues until an accurate prediction is made. The proposed model for predicting cardiac disease is a deep learning approach consisting of three critical steps: data extraction, preparation, feature engineering, and modelling. The model's outcome is stored in the cloud, and the trained data is containerized for efficient and reliable deployment. This proposed model is expected to provide an accurate prediction of the presence of cardiac disease, and the results of the study show that it achieved an accuracy rate of 87% and 86% using SVM and Naive Bayes algorithms, respectively, on the combined Cleveland, Hungarian, and Switzerland datasets.

**4 Conclusion**

Implementing an intelligent model for early-stage cardiac disease prediction through deep learning is significant in preventive healthcare. Advanced algorithms enable the accurate analysis of diverse patient data, ranging from medical history to diagnostic parameters. This model enhances the efficiency of disease detection and facilitates timely intervention, potentially saving lives. By harnessing the power of deep learning, we have created a sophisticated tool that complements traditional diagnostic methods, offering a more proactive and personalized approach to cardiac care. The model's ability to adapt and learn from new data ensures continuous improvement and adaptability to evolving healthcare landscapes. As we stride into the era of predictive medicine, this intelligent model stands as a beacon of hope in the quest to reduce the global burden of cardiac diseases. Its integration into healthcare systems has the potential to revolutionize early intervention strategies, promoting a healthier and more resilient society.

**5 Limitations and Future Recommendations**

Today, Multiple ML approaches have been utilized in the healthcare sector to forecast diseases, particularly cardiac disease. Several traditional healthcare systems are made for accurate cardiac disease prediction at an early stage, but some limitations exist in the accuracy prediction. For example, in this research, the authors applied the Convolutional Neural Network (CNN) approach to predict heart disease. They achieved an accuracy of 97%. In this research, an intelligent model is developed by using a Recurrent Neural Network (RNN) to predict cardiac disease by using a deep learning approach. This proposed model overcomes these types of issues and shows better accuracy, more than 97%, as compared to the based approach. In the future, this proposed model may be effective for lung disease prediction.

## *Reference:*


1. Ihnaini, B., Khan, M. A., Khan, T.A., Abbas, S., Daoud, M.S., Ahmad, M., Khan, Muhammad Adnan, 2021. A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning. Comput. Intell. Neurosci. 2021.
2. Benaddi, H., Ibrahimi, K., Rashid Abdulqadir, H., R. M. Zeebaree, S., M. Shukur, H., Mohammed Sadeeq, M., Wasfi Salim, B., Abid Salih, A., Fattah Kak, S., Huang, P., Fan, K., Yang, H., Zhang, K., Li, H., Yang, Y., Raja, G., Manaswini, Y., Vivekanandan, G.D., Sampath, H., Dev, K., Bashir, A.K., Wei, P.C., Wang, D., Zhao, Y., Tyagi, S.K.S., Kumar, N., Shrimali, B., Patel, H.B., Wang, H.H., Zhang, J., Yue, D., Li, R., Zhang, Y., Tian, W., Huang, Y., Zhu, L., Wu, Y., Gai, K., Choo, K.R.K.K.R., Zyskind, G., Nathan, O., Pentland, A.S., Ashik, M.H., Maswood, M.M.S., Alharbi, A.G., Mayuranathan, M., Murugan, M., Dhanakoti, V., Velmurugadass, P., Dhanasekaran, S., Shasi Anand, S., Vasudevan, V., Qi, W., Sun, M., Hosseini, S.R.A., Zou, J., He, D., Zeadally, S., Kumar, N., Wang, H.H., Choo, K.R.K.K.R., Wilczyński, A., Kołodziej, J., Rimba, P., Tran, A.B., Weber, I., Staples, M., Ponomarev, A., Xu, X., Turnbull, S., Divya Mounika, R., Naresh, R., 2020. Blockchain-based

verification framework for data integrity in edge-cloud storage. Futur. Gener. Comput. Syst. 7, 164996–165006.
3. Prianga, S., Sagana, R., Sharon, E., 2018. Evolutionary Survey on Data Security in Cloud Computing Using Blockchain. 2018 IEEE Int. Conf. Syst. Comput. Autom. Networking, ICSCA 2018 1–6.
4. Krishnan, A.S., Clive, D.R., Bhat, V., Ramteke, P.B., Koolagudi, S.G., 2018. A Transfer Learning Approach for Diabetic Retinopathy Classification Using Deep Convolutional Neural Networks, in: INDICON 2018 - 15th IEEE India Council International Conference. Institute of Electrical and Electronics Engineers Inc.
5. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, Justin, Prasadha, M.K., Pei, J., Ting, M., Zhu, J., Li, C., Hewett, S., Dong, Jason, Ziyar, I., Shi, A., Zhang, R., Zheng, L., Hou, R., Shi, W., Fu, X., Duan, Y., Huu, V.A.N., Wen, C., Zhang, E.D., Zhang, C.L., Li, O., Wang, X., Singer, M.A., Sun, X., Xu, J., Tafreshi, A., Lewis, M.A., Xia, H., Zhang, K., 2018. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 172, 1122-1131.e9.
6. Alhaidari F, Almotiri SH, Al Ghamdi MA, Khan MA, Rehman A, Abbas S, Khan KM, Rahman AU. Intelligent software-defined network for cognitive routing optimization using deep extreme learning machine approach. Computers, Materials & Continua. 2021 Jan 1;67(1):1269-85.
7. Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., Singh, P., 2021. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Comput. Intell. Neurosci. 2021.
8. Zimmerman, M., Sable, C., 2020. Congenital heart disease in low-and-middle-income countries: Focus on sub-Saharan Africa. Am. J. Med. Genet. Part C Semin. Med. Genet. 184, 36–46.
9. Wang, X., Wen, J., Zhang, Y., Wang, Y., 2014. Real estate price forecasting based on SVM optimized by PSO. Optik (Stuttg). 125, 1439–1443.
10. Tabassum N, Ditta A, Alyas T, Abbas S, Alquhayz H, Mian NA, Khan MA. Prediction of cloud ranking in a hyperconverged cloud ecosystem using machine learning. Computers, Materials & Continua. 2021 Jan 1;67(3):3129-41.
11. Ahmad, M., Aftab, S., 2017. Analyzing the Performance of SVM for Polarity Detection with Different Datasets. Int. J. Mod. Educ. Comput. Sci. 9, 29–36.
12. Malviya, S., Tiwari, A.K., Srivastava, R., Tiwari, V., 2020. Machine Learning Techniques for Sentiment Analysis: A Review. SAMRIDDHI A J. Phys. Sci. Eng. Technol. 12, 72–78.
13. Sebastian Raschka, V.M., 2019. Python Machine Learning, 3rd Edition, third. ed. Packt Publishing Ltd.
14. Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y.R., Suraj, R.S., 2021. Heart Disease Prediction using Hybrid machine Learning Model. Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021 1329–1333.
15. Rashid, M.M., Kamruzzaman, J., Hassan, M.M., Imam, T., Gordon, S., 2020. Cyberattacks detection in iot-based smart city applications using machine learning techniques. Int. J. Environ. Res. Public Health 17, 1–21.
16. Saleem, M., Abbas, S., Ghazal, T.M., Adnan Khan, M., Sahawneh, N., Ahmad, M., 2022. Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques. Egypt. Informatics J.
17. Patel, Jaydutt, Khaked, A.A., Patel, Jitali, Patel, Jigna, 2021. Heart Disease Prediction Using Machine Learning. Lect. Notes Networks Syst. 203 LNNS, 653–665.
18. Ahmed, F., Asif, M. and Saleem, M., 2023. Identification and Prediction of Brain Tumor Using VGG-16 Empowered with Explainable Artificial Intelligence. International Journal of Computational and Innovative Sciences, 2(2), pp.24-33.
19. Saleem, M., Khan, M.S., Issa, G.F., Khadim, A., Asif, M., Akram, A.S. and Nair, H.K., 2023, March. Smart Spaces: Occupancy Detection using Adaptive Back-Propagation Neural Network. In 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-6). IEEE.
20. Athar, A., Asif, R.N., Saleem, M., Munir, S., Al Nasar, M.R. and Momani, A.M., 2023, March. Improving Pneumonia Detection in chest X-rays using Transfer Learning Approach (AlexNet) and

Adversarial Training. In 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-7). IEEE.
21. Abualkishik, A., Saleem, M., Farooq, U., Asif, M., Hassan, M. and Malik, J.A., 2023, March. Genetic Algorithm Based Adaptive FSO Communication Link. In 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-4). IEEE.
22. Sajjad, G., Khan, M.B.S., Ghazal, T.M., Saleem, M., Khan, M.F. and Wannous, M., 2023, March. An Early Diagnosis of Brain Tumor Using Fused Transfer Learning. In 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-5). IEEE.