Heart Disease Prognosis with Artificial Intelligence and Neural Networks

Dr.M.Navaeetha Krishnan
Professor
Computer science and Engineering
St.Joseph College of Engineering

J Jeni Nenatha (Student)
Jenijebaraj2001@gmail.com
ME Computer science and Engineering
St. Joseph College of engineering

***Abstract*— Heart disease is universally recognized as a serious health concern. Over the years, the number of patients diagnosed with this life-threatening condition has continued to rise rather than decline, as evidenced by the increasing death rates nationwide. The mortality rate increases all over the world on a daily basis. One significant reason for this rise is the increasing number of patients with cardiovascular diseases. When considering death rates and the large number of people suffering from heart disease, the importance of early diagnosis becomes evident. Traditionally, heart disease prediction involves doctor's examinations or numerous medical tests such as ECG Stress Tests and Heart MRI. Nowadays, the healthcare industry contains a vast amount of healthcare data, which includes hidden information crucial for making effective decisions. Computer-based information combined with advanced data mining techniques can yield appropriate results. Neural networks are widely used tools for predicting heart disease diagnoses. In this project, a heart disease prediction system using an artificial neural network (ANN) with a backpropagation algorithm is proposed. The system uses 14 clinical features as input for the neural network. The neural network is trained using the backpropagation algorithm to predict the absence or presence of heart disease.**

***Keywords* — heart disease, machinelearning, prediction accuracy, artificial neural network.**

I. Introduction

Cardiovascular diseases (CVD) are a major cause of death globally. Effective diagnosis and treatment of these diseases are critical in reducing mortality rates. The healthcare industry is increasingly turning to advanced technologies to improve patient care and outcomes. Among these technologies, machine learning and artificial intelligence (AI) have shown significant promise. These methods can analyze vast amounts of data to uncover patterns and makepredictions that might not be immediately apparent to human practitioners. One of the critical areas where these technologies are making an impact is in the prediction and diagnosis of heart disease, which remains a leading cause of death worldwide. A cardiovascular risk factor is defined as any factor that increases the likelihood of developing cardiovascular disease. These factors include smoking, physical inactivity, unhealthy diet, high blood pressure, diabetes, and obesity. Understanding and mitigating these risk factors are crucial for the prevention and management of heart disease.

Heart disease prediction using artificial neural networks (ANN) represents a significant advancement in medical diagnostics. ANNs, which are designed to simulate the way the human brain processes information, are capable of learning from large datasets. This learning capability allows them to identify complex patterns and relationships within the data, which can then be used to make accurate predictions about a patient's health.

Also, the traditional methods of diagnosing heart disease involve a combination of physical examinations, patient history, and various diagnostic tests such as electrocardiograms (ECGs), stress tests, and imaging studies. While these methods are effective, they are also time-consuming, expensive, and require specialized equipment and expertise. Moreover, these tests can sometimes lead to false positives or negatives, adding to the uncertainty in diagnosis and treatment.

The machine learning approaches, particularly those using ANNs, offer a more streamlined and potentially more accurate alternative. By training on historical data, an ANN can learn to recognize the subtle indicators of heart disease that might be missed by traditional diagnostic methods. This capability can lead to earlier detection and intervention, which is crucial for improving patient outcomes. A key advantage of using ANNs in heart disease prediction is their ability to handle a large number of input variables. In the context of heart disease, these variables can include demographic information, medical history, lifestyle factors, and results from various medical tests. The ANN processes all this information simultaneously to arrive at a prediction, taking into account the complex interrelationships between different risk factors.

In developing an effective ANN for heart disease prediction is the quality and completeness of the data used for training the model. High-quality data is essential for training an accurate model, and this includes not only clinical data but also comprehensive patient histories and lifestyle information. Data preprocessing steps such as normalization, missing value treatment, and feature selection are critical to ensure the model's robustness and reliability.

Moreover, the architecture of the ANN itself plays a crucial role in its performance. Factors such as the number of layers, the number of neurons in each layer, the activation functions used, and the training algorithm all influence the model's ability to learn and generalize from the data. The backpropagation algorithm, a common training method for ANNs, involves adjusting the weights of the network based on the error rate of each output compared to the expected result. This iterative process continues until the model achieves a satisfactory level of accuracy.

Although the integration of artificial neural networks into heart disease prediction represents a significant step forward in medical diagnostics. By leveraging the power of deep learning, these systems can provide more accurate and timely predictions, ultimately leading to better patient outcomes. However, to fully realize the potential of these technologies, ongoing research and development are necessary to address current limitations and ensure their reliable and ethical application in clinical practice.

This study will gather variations of heart disease prediction studies using ANN that have been done from the year until now, in addition to analyzing the level of accuracy of the results that have been obtained.

II. Previous Works

This literature review focuses on heart disease prediction algorithms that utilize machine learning and techniques. Key studies in this field include:

Prediction with Machine Learning

To justify the involvement of ML in the prognosis of heart disease, discusses the analysis that summarizes the research work related to the useof variations of machine learning in the process of prognosis and classification of heart disease, then asserts the most effective and accurate algorithm. The system is expected to be able to produce accurate heart disease prognosis results compared to other systems. A system that uses Cleveland and Statlog (heart disease) datasets for training and testing, further improves classifier performance by using feature subset selection. The results show the Artificial Neural Network performs best in predicting heart disease from large medical data.

One of the critical strengths of ML in heart disease prediction lies in its ability to utilize diverse datasets. These datasets can include demographic information, medical histories, lifestyle factors, genetic information, and results from diagnostic tests such as blood tests, ECGs, and imaging studies. By integrating these varied sources of data, ML models can create a comprehensive profile of each patient, enabling more accurate risk assessments.

The research highlights various data detection methods employed in heart disease prediction and their standard levels of accuracy. By reviewing these methods, the project aims to enhance both the detection and prediction of heart disease with higher precision and to facilitate more effective medication recommendations based on predicted outcomes.

Feature selection is a crucial step in developing effective ML models. In heart disease prediction, relevant features might include age, gender, cholesterol levels, blood pressure, smoking status, physical activity levels, and family history of heart disease. Advanced ML techniques can automatically select the most significant features from the data, enhancing the model's predictive power and reducing the risk of overfitting.

Several ML algorithms are commonly used for heart disease prediction, each with its strengths and weaknesses. Decision trees, for example, provide a straightforward and interpretable method for classification and regression tasks. Random forests, an ensemble learning method, improve predictive accuracy by combining multiple decision trees. Support vector machines (SVM) are effective for binary classification problems, such as predicting the presence or absence of heart disease.

Artificial neural networks (ANNs) and deep learning models are particularly noteworthy due to their ability to model complex, non-linear relationships in the data. These models can handle large datasets with many features, making them well-suited for heart disease prediction. Convolutional neural networks and recurrent neural networks are specialized deep learning models that have shown promise in analysing medical images and sequential data, respectively.

Training an ML model involves feeding it a large amount of labelled data so that it can learn to make accurate predictions. In the case of heart disease prediction, labelled data might consist of patient records where the presence or absence of heart disease is known. The model adjusts its internal parameters to minimize prediction errors, typically using optimization algorithms such as gradient descent.

Validation is a critical step to ensure that the ML model generalizes well to new, unseen data. Common validation techniques include cross-validation, where the data is divided into multiple subsets, and the model is trained and tested on different combinations of these subsets. This process helps in identifying potential overfitting and ensures that the model performs well across different patient groups.

Despite the potential benefits, several challenges must be addressed when using ML for heart disease prediction. One significant challenge is data quality and availability. Medical datasets can be noisy, incomplete, and biased, which can negatively impact the model's performance. Ensuring data privacy and security is also paramount, given the sensitive nature of medical information.

Ethical considerations are crucial in the deployment of ML models in healthcare. These models must be transparent and explainable, allowing healthcare professionals to understand the basis of their predictions. Bias in ML models is another concern, as models trained on unrepresentative data can lead to disparities in healthcare outcomes. Continuous monitoring and updating of ML models are necessary to maintain their accuracy and fairness. Accountability for model-driven decisions must be established to address potential adverse outcomes. Ensuring inclusivity and representation in training datasets helps avoid perpetuating existing biases and supports equitable care. Human oversight remains essential to complement ML predictions, ensuring that final decisions are made with professional judgment. Ethical AI development guidelines should be flowed to align model design and deployment with ethical standards.

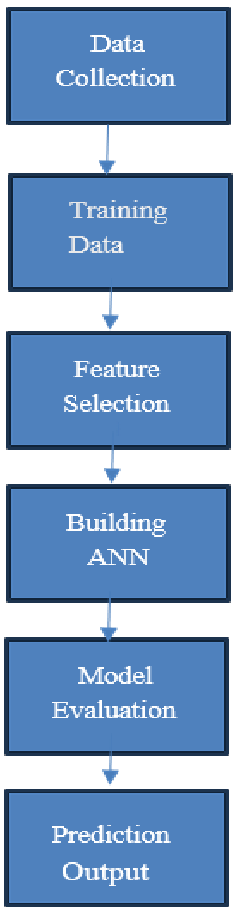
ML-driven heart disease prediction tools are becoming a standard in clinical settings. They enable doctors to quickly identify high-risk patients and enhance preventive care, while also providing swift evaluations in emergencies for faster treatment of cardiac issues.

The future of ML in heart disease prediction is promising, with ongoing research focused on improving model accuracy and interpretability. The integration of ML with other technologies, such as wearable devices and mobile health applications, is expected to enhance real-time monitoring and personalized care. Advances in explainable AI are also paving the way for more transparent and trustworthy ML models, fostering greater acceptance and adoption in the healthcare community.

In conclusion, ML holds significant potential for improving heart disease prediction and patient care. By leveraging advanced algorithms and diverse data sources, ML models can provide accurate and timely risk assessments, ultimately contributing to better health outcomes and reduced mortality rates. However, careful attention to data quality, model transparency, and ethical considerations is essential to fully realize the benefits of ML in healthcare.

III. RESEARCH METHODOLOGY

There are several steps that have been taken to collect data throughout this study. The steps are as below:



1. Data Collection

Data collection is the foundational step in any machine learning project. In the context of heart disease prediction, this involves gathering comprehensive datasets that include patient information and relevant clinical features. These datasets can be sourced from medical records, health surveys, hospital databases, and public health repositories. Key data points might include:

- **Demographic Information**: Age, gender, ethnicity, etc.
- **Medical History**: Previous heart conditions, family history of heart disease, diabetes, hypertension, etc.
- **Lifestyle Factors**: Smoking status, alcohol consumption, physical activity levels, diet, etc.
- **Clinical Measurements**: Blood pressure, cholesterol levels, heart rate, ECG results, and other diagnostic test results.

Ensuring the data is accurate, complete, and representative of the target population is critical. Data collection may involve obtaining necessary permissions and ensuring compliance with data privacy regulations.

2. Data preprocessing

Data preprocessing is essential to prepare the raw data for analysis. This step involves several sub-tasks:

- **Data Cleaning**: Identifying and correcting errors or inconsistencies in the data. This includes handling missing values, removing duplicates, and correcting erroneous entries.
- **Normalization/Standardization**: Scaling the data so that features with different units or magnitudes do not disproportionately influence the model. Common techniques include min-max normalization and z-score standardization.
- **Encoding Categorical Variables**: Converting categorical data (e.g., gender, smoking status) into numerical format using methods like one-hot encoding or label encoding.
- **Splitting the Data**: Dividing the dataset into training, validation, and test sets. Typically, the data is split in a ratio such as 70-20-10, where 70% is used for training, 20% for validation, and 10% for testing.

3. Feature selection

Feature selection involves identifying the most relevant features that contribute to the prediction of heart disease. This step is crucial for enhancing the model's performance and reducing complexity. Several techniques can be used:

- **Expert Knowledge**: Leveraging domain expertise to select features known to be associated with heart disease.
- **Correlation Analysis**: Assessing the correlation between each feature and the target variable. Features with high correlation to the target and low correlation to each other are typically selected.

- **Statistical Tests**: Using tests like chi-square, ANOVA, or t-tests to determine the significance of each feature.
- **Dimensionality Reduction**: Applying techniques like Principal Component Analysis (PCA) to reduce the number of features while retaining most of the variance in the data.

4. Building ANN

Building an ANN involves several key steps:

- **Designing the Network Architecture**: Determining the structure of the ANN, including the number of layers (input, hidden, and output layers), and the number of neurons in each layer. A typical architecture for heart disease prediction might include:
    1. **Input Layer**: Corresponds to the number of features selected.
    2. **Hidden Layers**: One or more layers with a specified number of neurons, using activation functions such as ReLU (Rectified Linear Unit).
    3. **Output Layer**: A single neuron with a sigmoid activation function for binary classification (presence or absence of heart disease).
- **Initializing Weights**: Setting initial weights for the network, often using methods like Xavier initialization to facilitate efficient training.
- **Forward Propagation**: Passing the input data through the network to obtain predictions. Each neuron calculates a weighted sum of its inputs and applies an activation function.
- **Backpropagation**: Adjusting the weights based on the error between predicted and actual outcomes. This process involves calculating gradients and updating weights using optimization algorithms such as stochastic gradient descent (SGD) or Adam.
- **Training the Network**: Iteratively feeding the training data through the network and updating weights until the model converges to a minimumerror.

5. Model Evaluation

Evaluating the model's performance is crucial to ensure it generalizes well to new data. Common evaluation metrics for binary classification in heart disease prediction include:

- **Accuracy**: The proportion of correctly predicted instances out of the total instances.
- **Precision**: The proportion of true positive predictions out of all positive predictions.
- **Recall (Sensitivity)**: The proportion of true positive predictions out of all actual positive instances.
- **F1 Score**: The harmonic mean of precision and recall, providing a balance between the two.
- **ROC-AUC Curve**: The Receiver Operating Characteristic curve plots true positive rate againstfalse positive rate. The Area Under the Curve (AUC) measures the model's ability to distinguish between classes.

Cross-validation techniques, such as k-fold cross-validation, are often used to assess the model's performance on different subsets of the data, providing a more robust evaluation.

6. Prediction

Once the ANN model is trained and evaluated, it can be used to make predictions on new, unseen data. The process involves:

1. **Inputting New Data**: Feeding the new patient's data (after preprocessing) into the trained ANN model.
2. **Generating Output**: The model processes the input data through the network and generates a probability score indicating the likelihood of heart disease.
3. **Interpreting Results**: Based on the output probability, a decision threshold (commonly 0.5) is used to classify the prediction as positive (presence of heart disease) or negative (absence of heart disease).

In clinical practice, these predictions can assist healthcare professionals in making more informed decisions, enabling early intervention and personalized treatment plans.

In conclusion, the process of predicting heart disease using machine learning involves careful data collection, thorough preprocessing, strategic feature selection, meticulous model building, rigorous evaluation, and practical prediction deployment. Each step is integral to developing an accurate and reliable prediction system that can significantly enhance patient care and outcomes.

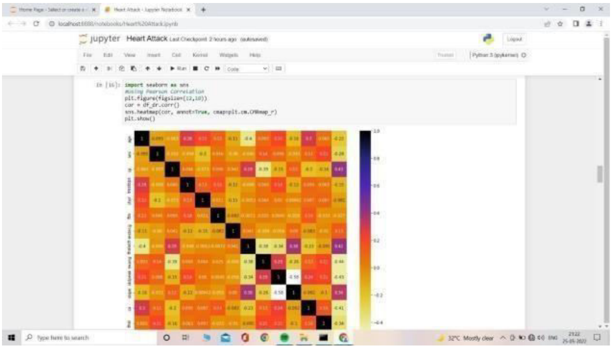
The application of machine learning, particularly artificial neural networks, in heart disease prediction represents a significant advancement in medical diagnostics. By leveraging diverse datasets and sophisticated algorithms, these models can offer accurate and timely risk assessments, enabling earlier intervention and personalized treatment strategies. Despite the challenges of data quality, model transparency, and ethical considerations, ongoing research and development are paving the way for more reliable and interpretable ML models. As these technologies continue to evolve, they hold immense potential to improve patient outcomes, reduce healthcare costs, and ultimately save lives.

IV Result and Discussion

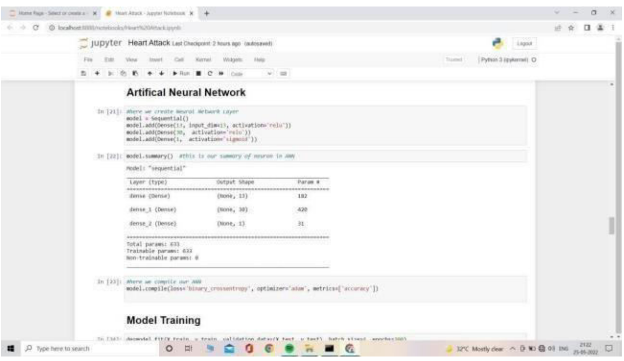
From this study follow a structured methodology that incorporates multiple stages to ensure robust and accurate heart disease prediction using Artificial Neural Networks (ANNs).

Initially, the necessary modules are imported, and a sample dataset is retrieved for visualization. The dataset's attributes are described in detail to provide context for the analysis. Preprocessing steps include dropping duplicate entries to clean the data, followed by visualizing the attribute values using histograms. These steps are critical for understanding the data distribution and ensuring its quality.

A correlation matrix is then plotted to visualize relationships between attributes. This step is essential for identifying which attributes have significant correlations with the target variable, aiding in effective feature selection. The correlation matrix helps in understanding the interplay between various features and their collective impact on heart disease prediction.

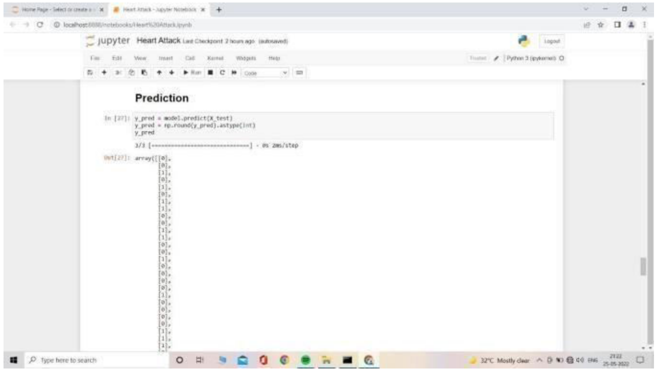


Feature selection is performed using the Chi-square algorithm. This step identifies the most relevant features that contribute to accurate predictions, thereby enhancing model performance. By selecting only the most significant attributes, the model's complexity is reduced, leading to improved efficiency and accuracy.



After training, the model's performance is evaluated through various metrics. A classification report is generated to provide a detailed analysis of the model's precision, recall, F1 score, and overall accuracy.

Performance evaluation also includes plotting the loss and accuracy over the training epochs to visualize the model's learning process. Additionally, the Receiver Operating Characteristic (ROC) curve is plotted to assess the model's ability to distinguish between classes, providing a comprehensive evaluation of its predictive power.



The final step involves acquiring the required output through custom input, demonstrating the model's practical applicability. This step ensures that the model is not only theoretically sound but also practically useful in real-world scenarios.

V Conclusion

After examining the research and development of heart disease prediction using artificial neural networks (ANNs), it is evident that machine learning and deep learning offer numerous advantages for the healthcare field. A meticulous and intelligent approach in these technologies can yield results comparable to the expertise of health professionals. By leveraging existing data, ANNs can uncover hidden patterns and relationships, facilitating early disease prediction and aiding healthcare workers in making informed decisions.

The variety of machine learning and deep learning options available today provides ample opportunities to develop cutting-edge predictive systems. These technologies offer diverse methodologies, allowing researchers to select algorithms that best suit their objectives and to enhance model performance by integrating additional techniques.

Studies have shown that deep learning models, including ANNs, are capable of achieving high levels of accuracy, often exceeding 85%. Moreover, the integration of techniques such as feature selection within machine learning models can push accuracy levels even higher, reaching up to 97%. This demonstrates that while deep learning models tend to perform exceptionally well, both machine learning and deep learning algorithms remain highly relevant. The choice between them depends on the specific goals and requirements of the predictive model being developed.

In conclusion, the implementation of ANNs for heart disease prediction represents a significant advancement in medical diagnostics. These models not only enhance the early detection and treatment of heart disease but also contribute to more personalized and effective patient care. As research continues to evolve, the potential for these technologies to revolutionize healthcare becomes increasingly apparent, promising better outcomes and improved quality of life for patients.

REFERENCES

[1] American Heart Association (Ed.). (2022, May 4). What is cardiovascular disease? www.heart.org. Retrieved December 14, 2022, from https://www.heart.org/en/health- topics/consumer-healthcare/what-is-cardiovascular-disease

[2] Online, T. S. (2022, September 29). Heart disease leading cause of death globally, treatment cost expected to rise. The Star. Retrieved December 14, 2022, from https://www.thestar.com.my/news/nation/2022/09/29/heart- disease-leading-cause-of-death-globally-treatment-cost- expected-to-rise

[3] CK Gomathy, & MAR Naidu. (2021). THE PREDICTION OF DISEASE USING MACHINE LEARNING. International Journal of Scientific Research inEngineering and Management (IJSREM), 05(1). Retrieved from https://www.researchgate.net/publication/357449131_THE_ PREDICTION_OF_DISEASE_USING_MACHINE_LEAR NING

[4] Yu, Z., Wang, K., Wan, Z., Xie, S., & Lv, Z.(2022). Popular deep learning algorithms for Disease prediction: A review. Cluster Computing. https://doi.org/10.1007/s10586-022-03707-y

[5] Rajamhoana, S. P., Devi, C. A., Umamaheswari, K., Kiruba, R., Karunya, K., & Deepika, R. (2018, July). Analysis of neural networks based heart disease predictionsystem. In 2018 11th international conference on human system interaction (HSI) (pp. 233-239). IEEE.

[6] Cherian, R. P., Thomas, N., & Venkitachalam, S. (2020). Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm. Journal of Biomedical Informatics, 110, 103543.

[7] Patil, P. S. (2021). Automated Heart Disease Recognition System. International Journal of Research inEngineering and Science, 9, 18-23.

[8] Dutta, A., Batabyal, T., Basu, M., & Acton, S. T. (2020). An efficient convolutional neural network for coronary heart disease prediction. Expert Systems with Applications, 159, 113408.

[9] Nandy, S., Adhikari, M., Balasubramanian, V., Menon, V. G., Li, X., & Zakarya, M. (2021). An intelligent heart disease prediction system based on swarm-artificial neural network. Neural Computing and Applications, 1-15.

[10] Jeyaranjani, J., Rajkumar, T. D., & Kumar, T. A. (2021). Coronary heart disease diagnosis using the efficient ANN model. Materials Today: Proceedings.

[11] Islam, R., Beeravolu, A. R., Islam, M. A. H., Karim, A., Azam, S., & Mukti, S. A. (2021, December). A Performance Based Study on Deep Learning Algorithms inthe Efficient Prediction of Heart Disease. In 2021 2nd International Informatics and Software Engineering Conference (IISEC) (pp. 1-6). IEEE.

[12] Harkulkar, N., Nadkarni, S., & Patel, B. (2020). Heart Disease Prediction using CNN Deep Learning Model.International Journal for Research in Applied Science and Engineering Technology, 8(12), 875-881.

[13] Manimurugan, S., Almutairi, S., Aborokbah, M. M., Narmatha, C., Ganesan, S., Chilamkurti, N., ... & Almoamari, H. (2022). Two-Stage Classification Model forthe Prediction of Heart Disease Using IoMT and Artificial Intelligence. Sensors, 22(2), 476.

[14] MAlnajjar, M. K., & Abu-Naser, S. S. (2022). Heart Sounds Analysis and Classification for Cardiovascular Diseases Diagnosis using Deep Learning.

[15] Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., & Masood, M. (2021). Prediction of heart disease using deep convolutional neural networks. Arabian Journal for Science and Engineering, 46(4), 3409- 3422.

[16] Sharma, S., & Parmar, M. (2020). Heart diseases prediction using deep learning neural network model. Interna-tional Journal ofInnovative Technology and Exploring Engineering (IJITEE), 9(3), 124-137.

[17] Verma, K., Bartwal, A. S., & Thapliyal, M. P. A Genetic Algorithm based Hybrid Deep Learning Approachfor Heart Disease Prediction.

[18] Raihan, M., Mandal, P. K., Islam, M. M., Hossain,T., Ghosh, P., Shaj, S. A., ... & More, A. (2019, February). Risk prediction of ischemic heart disease using artificial neural network. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-5). IEEE.

[19] Brunese, L., Martinelli, F., Mercaldo,F., & Santone, A. (2020). Deep learning for heart disease detection through cardiac sounds. Procedia ComputerScience, 176, 2202-221.

[20] Ahmed, Rizgar Maghdid, and Omar Qusay Alshebly. "Prediction and factors affecting of chronic kidney disease diagnosis using artificial neural networks model and logistic regression model." Iraqi Journal of Statistical Sciences 16.28 (2019): 140-159.

[21] Burse, Kavita, et al. "Various pre-processing methods for neural network based heart disease prediction." Smart innovations in communication and computational sciences. Springer, Singapore, 2019. 55-65.

[22] Pasha, S. N., Ramesh, D., Mohmmad, S., & Harshavardhan, A. (2020, December). Cardiovascular disease prediction using deep learning techniques. In IOP conference series: materials science and engineering (Vol.981, No. 2, p. 022006). IOP Publishing.