

PREDICTION OF HEART DISEASE USING ARTIFICIAL NEURAL NETWORK

Authors: SHAHID MEHMOOD AWAN, MUHAMMAD USAMA RIAZ², ABDUL GHAFAR KHAN³

Department of Software Engineering, University of Management and Technology, Lahore, Pakistan Email: shahid.awan@umt.edu.pk

Introduction:

The heart is the main organ of the human body. If it stops working correctly, our whole body can get infected, and in most cases, this disease leads to death. From 1990 to 2013, 41% cardiovascular disease increased globally. Life is not possible without a functioning heart, and so human beings cannot survive without it. It controls our blood pressure, body temperature, and many other vital health aspects including oxygen regulation in the blood. Some of the most common factors that contribute to heart disease include family genetic problems which come from generation to generation, high blood pressure, Cholesterol, Sex, Age mostly above 30, Poor diet, Calcium rate, a blood vessel of heart got over stretched, lack of exercise etc. Heart disease is a persistent apprehension of the whole world as it is frequently increasing in ratio per person. The heart is the main regulating organ of the human body. According to the Institute for Health Metrics and Evaluation (IHME), death rate due to this disease has shown an alarming 41% increase from 1990 to 2013 and a 9.83% further increase in 2014. Survey analysis of (WHO) shows 17 million deaths due to heart disease [5]. If any disease gets

predicted at earlier stages, then it becomes easier to find or apply a cure before it gets dangerous. Just like that, if a heart disease is predicted earlier, it becomes easier to find a cure. This type of prediction problem, related to medical diagnosis comes under the branch of computer science i.e., bioinformatics. In bioinformatics, for prediction, we need previous historical data of the patient and some pattern matching algorithms which can be trained on the respective data to generate results. In clinics and health centers, a lot of patients are being diagnosed daily so there is a lot of data collection, for example, in the form of Medical Reports. Collection of data from these reports can be used for extracting knowledge from data. This can be done by a well-known model of bioinformatics i.e., Knowledge Discovery from Data

Methodology:

1. Data Collection

- Use a publicly available dataset like the Cleveland Heart Disease dataset or a clinical dataset collected under ethical guidelines.
- Ensure the dataset contains relevant features such as:
 - Demographics (e.g., age, gender)
 - Medical history (e.g., blood pressure, cholesterol levels)
 - Symptoms (e.g., chest pain type, maximum heart rate achieved)
 - Diagnostic results (e.g., electrocardiographic results, thalassemia)

2. Data Preprocessing

- **Handle Missing Values:** Replace missing values using mean, median, or imputation techniques.

- **Feature Scaling:** Normalize or standardize features to ensure faster convergence.
- **Encoding Categorical Variables:** Convert categorical variables into numeric representations using one-hot encoding or label encoding.
- **Train-Test Split:** Divide the dataset into training, validation, and test sets (e.g., 70% training, 15% validation, 15% test).

3. Feature Selection

- Use feature importance techniques such as:
 - Correlation analysis
 - Recursive Feature Elimination (RFE)
 - Principal Component Analysis (PCA) for dimensionality reduction.

4. Neural Network Architecture

- Design a Multi-Layer Perceptron (MLP) for structured data:
 - Input Layer: Number of neurons equal to the number of features.
 - Hidden Layers: 2-3 layers with varying neurons (e.g., 64, 32).
 - Activation Functions: Use ReLU for hidden layers and Sigmoid for binary classification output.
 - Output Layer: One neuron with a Sigmoid activation for binary classification.
- Use dropout layers for regularization to prevent overfitting.

5. Training

- **Loss Function:** Binary Cross-Entropy for binary classification.
- **Optimizer:** Adam optimizer for faster convergence.

- **Batch Size:** Use a batch size of 32 or 64 for efficient training.
- **Learning Rate:** Start with a learning rate (e.g., 0.001) and use a scheduler for decay.
- **Epochs:** Train for 50-100 epochs with early stopping based on validation loss.

6. Evaluation

- Metrics:
 - Accuracy
 - Precision, Recall, and F1-Score
 - Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
- Validate the model on the test set to ensure generalization.

7. Hyperparameter Tuning

- Use techniques like grid search or Bayesian optimization to fine-tune:
 - Number of layers and neurons
 - Learning rate
 - Batch size
 - Dropout rate

8. Deployment

- Save the trained model using frameworks like TensorFlow or PyTorch.
- Deploy the model as an API for real-time predictions using Flask, FastAPI, or a cloud service like AWS, GCP, or Azure.

Conclusion:

- **Key Findings:**

- Neural networks, when properly trained, can achieve high accuracy in predicting heart disease.
- Performance is dependent on the quality of the dataset and feature engineering.

- **Advantages:**

- Ability to capture complex relationships in the data.
- Can be deployed for real-time or batch predictions in clinical applications.

- **Challenges:**

- Requires significant computational resources.
- Performance is sensitive to overfitting, especially with small datasets.

- **Future Work:**

- Incorporate ensemble methods or hybrid models to improve robustness.
- Expand the dataset with diverse populations for better generalization.
- Use explainable AI techniques (e.g., SHAP or LIME) to interpret predictions, enhancing clinical trust and adoption.