**Heart Disease Prediction Using Artificial Neural Networks (ANN) - Dashboard Report**

**Ashish Michael Chauhan**

**055007**

**1. Introduction**

This project involves the development of a **Heart Disease Prediction Dashboard** using **Streamlit** and an **Artificial Neural Network (ANN)**. The dashboard allows users to upload the heart disease dataset, select hyperparameters, and train a deep learning model to predict the presence of heart disease. The model is interactive, enabling users to visualize the training performance through loss and accuracy plots.

**2. Dataset Overview**

The dataset (heart\_disease.csv) consists of **clinical and cardiovascular attributes** representing patient health records. It contains the following features:

* **Age**: Continuous – Age of the patient (in years)
* **Sex**: Categorical – Gender (1 = Male, 0 = Female)
* **CP (Chest Pain Type)**: Categorical – Chest pain type (0-3)
* **Trestbps**: Continuous – Resting blood pressure (mm Hg)
* **Chol**: Continuous – Serum cholesterol level (mg/dl)
* **Fbs**: Binary – Fasting blood sugar > 120 mg/dl (1 = yes)
* **Restecg**: Categorical – Resting electrocardiographic results
* **Thalach**: Continuous – Maximum heart rate achieved
* **Exang**: Binary – Exercise-induced angina (1 = yes)
* **Oldpeak**: Continuous – ST depression induced by exercise
* **Slope**: Categorical – Slope of peak exercise ST segment
* **Ca**: Continuous – Number of major vessels colored
* **Thal**: Categorical – Thalassemia type
* **Target**: Binary – Heart disease presence indicator (1 = disease, 0 = no disease)

**3. Preprocessing Steps**

* **Removing Unnecessary Columns:**
  + The dataset did not require column removal, as all attributes were relevant.
* **Handling Missing Values:**
  + Categorical features: Missing values were filled with the **mode**.
  + Continuous features: Missing values were filled with the **mean**.
* **Encoding Categorical Variables:**
  + LabelEncoder was applied to categorical features like Sex, CP, Restecg, Exang, Slope, and Thal.
* **Feature Scaling:**
  + **StandardScaler** was used to normalize numerical features for uniform model input.
* **Train-Test Split:**
  + The dataset was divided into **80% for training** and **20% for testing** using **Stratified Sampling** to maintain class balance.

**4. ANN Model Architecture**

* **Input Layer:**
  + Defined based on the number of features (13).
* **Hidden Layers:**
  + User-defined number of layers (1-5)
  + Customizable neurons per layer
  + Selectable activation functions: ReLU, Tanh, Sigmoid
  + **Dropout layers** to prevent overfitting.
* **Output Layer:**
  + A **single neuron** with sigmoid activation (binary classification: Disease or No Disease).
* **Loss Functions:**
  + binary\_crossentropy, mean\_squared\_error, or hinge (user-selected)
* **Optimizers:**
  + adam, sgd, rmsprop (user-selected)
* **Hyperparameters:**
  + **Dropout Rate:** Adjustable (0.0 – 0.5)
  + **Epochs:** User-defined (10 – 200)
  + **Batch Size:** User-defined during Streamlit execution

**5. Model Training and Evaluation**

* **Training the Model:**
  + The ANN model was trained on the **user-defined settings** via the Streamlit interface.
  + The model used **validation data (X\_test, y\_test)** to monitor performance in real-time.
* **Performance Visualization:**
  + **Loss Curve:** Comparison of **Training vs. Validation Loss** over epochs.
  + **Accuracy Curve:** Comparison of **Training vs. Validation Accuracy** over epochs.
* **Final Model Evaluation:**
  + Displays **Final Validation Accuracy** after training completion.
  + Uses model evaluation metrics (Accuracy, Loss) to measure performance.

**6. Experiments**

✅ **Configuration 1:**

* **Parameters:** 2 Hidden Layers (16, 16 neurons), ReLU, Adam, Binary Crossentropy, Dropout 0.2, 30 Epochs
* **Result:**
  + **Validation accuracy**: 0.8811
  + Good learning curves with minor overfitting.

✅ **Configuration 2:**

* **Parameters:** 3 Hidden Layers (32, 32, 16 neurons), ReLU, Adam, Binary Crossentropy, Dropout 0.2, 25 Epochs
* **Result:**
  + **Validation accuracy**: 0.9024
  + Highest accuracy achieved with moderate overfitting.
  + **Best configuration** based on accuracy and stability.

✅ **Configuration 3:**

* **Parameters:** 4 Hidden Layers (32, 32, 16, 16 neurons), Sigmoid, Adam, Binary Crossentropy, Dropout 0.2, 30 Epochs
* **Result:**
  + **Validation accuracy**: 0.7643
  + Significantly lower accuracy, poor generalization.
  + **Sigmoid activation** resulted in lower performance.

✅ **Configuration 4:**

* **Parameters:** 2 Hidden Layers (16, 16 neurons), ReLU, RMSprop, Binary Crossentropy, Dropout 0.3, 20 Epochs
* **Result:**
  + **Validation accuracy**: 0.8429
  + Slightly lower accuracy, indicating underfitting.
  + **RMSprop** performed worse than Adam.

**7. Overall Interpretations**

📊 **ReLU is Superior:**

* The **ReLU activation function** consistently outperformed Sigmoid.
* Sigmoid in **Configuration 3** resulted in poor accuracy, indicating it might not be suitable for this type of problem.

📊 **Overfitting is a Key Challenge:**

* Several configurations showed signs of **overfitting**, where the model performed better on the training data than on the validation data.
* This was evident from the **diverging training and validation accuracy curves**.
* **Dropout layers** were used to mitigate overfitting.

📊 **More Layers Don't Always Mean Better:**

* Increasing the number of hidden layers did not consistently improve performance.
* The **3-layer model** achieved the **best validation accuracy**, while the 4-layer model performed worse, indicating diminishing returns.

📊 **Optimizer Selection Matters:**

* The **Adam optimizer** consistently performed better in terms of both speed and accuracy.
* **RMSprop** showed signs of underfitting in some configurations.

📊 **Hyperparameter Tuning is Critical:**

* Small changes in hyperparameters (such as dropout rate, epochs, and optimizer) significantly impacted model accuracy.

**8. Key Takeaways**

✅ The **optimal ANN architecture** for this heart disease prediction task is:

1. **Number of Hidden Layers:** 3
2. **Neurons per Layer:** 32, 32, and 16
3. **Activation Function:** ReLU for all hidden layers
4. **Optimizer:** Adam
5. **Loss Function:** Binary Crossentropy
6. **Dropout Rate:** 0.2
7. **Epochs:** 25

✅ **Insights:**

* The model achieves **high accuracy (~90.24%)** with this configuration.
* **Overfitting mitigation** with dropout is essential.
* Further improvements could involve more advanced techniques, including:
  + **Learning rate adjustments**
  + **Feature engineering**
  + **Hybrid models** combining ANN with other ML algorithms.

**9. Conclusion**

This project successfully demonstrates how deep learning can be applied to **heart disease prediction**. The **best configuration** achieves a validation accuracy of **~90.24%**, making it a **powerful diagnostic tool** for healthcare analytics. The Streamlit interface enables real-time hyperparameter tuning and performance visualization, making the dashboard **user-friendly and effective** for medical data analysis.