Artificial Intelligence Lab
CSE-3636

Project Report

Diabetes Prediction Using MLP Algorithm

Submitted To

Dr. Abu Nowshed Chy

Adjunct Faculty,Dept. OF CSE

- **Team Members**
- MD. Faisal Hoque Rifat C221076
- Parba Das Gupta C221049
- Nazrul Islam Sajib C221074
- Istiaque Ahmed C221073
- Miftahul Hoque Siyam C221075

Diabetes Prediction Using MLP Algorithm

Introduction:

Artificial Intelligence (AI) has significantly transformed the healthcare industry by providing advanced tools for diagnosis and decision-making. One such application is diabetes prediction, which is crucial for early detection and management of this chronic disease. Diabetes prediction involves analyzing various health parameters to assess the likelihood of a person developing diabetes. This project implements a Multi-Layer Perceptron (MLP) algorithm using TensorFlow and Keras to build a predictive model. The model is trained on medical datasets containing patient information, enabling accurate classification and assisting healthcare professionals in risk assessment and preventive care.

Objective:

- Develop a machine learning model to predict the likelihood of diabetes in individuals.
- Train the model using a Multi-Layer Perceptron (MLP) algorithm on relevant health datasets.
- Analyze health metrics such as glucose levels, blood pressure, BMI, and other factors for prediction.
- Evaluate the model's performance using metrics like accuracy, precision, and recall.
- Provide a reliable tool for early diabetes diagnosis and preventive healthcare interventions.

Data Overview:

General Information:

- **Number of Instances**: 442
- **Number of Attributes**: 10 numerical predictive attributes and 1 target variable.
- **Target Variable**: A quantitative measure of diabetes disease progression one year after baseline.

Attribute Information:

The dataset contains the following features:

1. **Age**: Age of the individual (in years).
2. **Sex**: Gender of the individual.

3. **BMI**: Body Mass Index, calculated as weight in kg/(height in m)^2.
4. **BP**: Average blood pressure.
5. **S1**: Total serum cholesterol (tc).
6. **S2**: Low-density lipoproteins (ldl).
7. **S3**: High-density lipoproteins (hdl).
8. **S4**: Total cholesterol to HDL ratio (tch).
9. **S5**: Possibly the log of serum triglycerides level (ltg).
10. **S6**: Blood sugar level (glu).

Target Variable:

- **Column 11**: A quantitative measure indicating the progression of diabetes after one year from baseline.

Methodology:

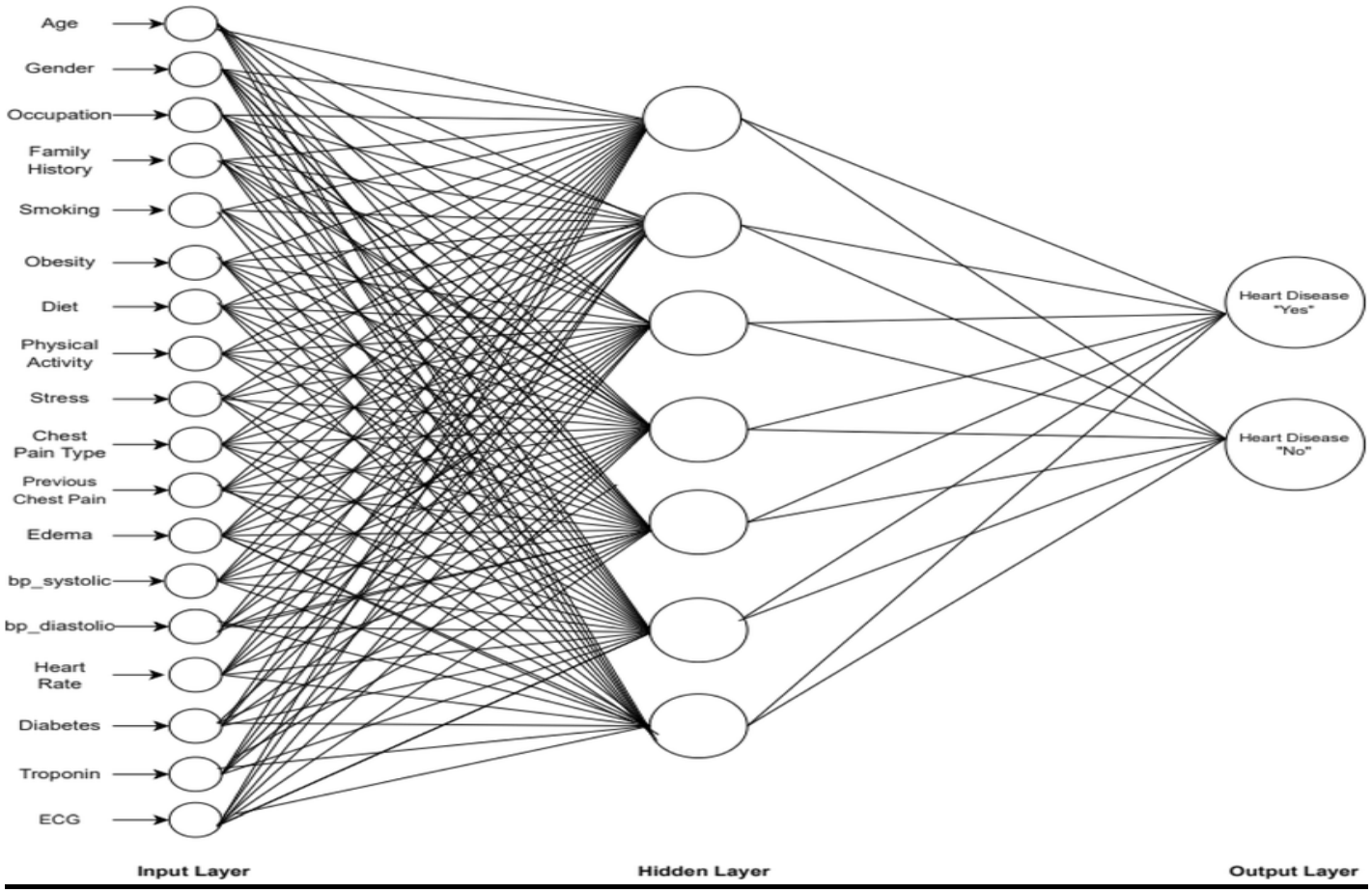
1. **Dataset**: The Diabetes dataset from Scikit-learn is used for this project. It contains **442** samples with **10** predictive attributes and **one** target variable representing the progression of diabetes.
2. **Preprocessing**:
    - Features were normalized using StandardScaler.
    - Target labels were encoded numerically using LabelEncoder.
3. **Model Architecture:** A Multi-Layer Perceptron (MLP) neural network is used with the following structure:
    - **Input Layer**: Accepts 10 input features representing patient attributes.
    - **Hidden Layer 1**: Dense layer with 128 neurons and ReLU activation.
    - **Hidden Layer 2**: Dense layer with 64 neurons and ReLU activation.
    - **Output Layer**: Dense layer with 1 neuron and no activation (for regression tasks).
4. **Model Compilation and Training:**
    - **Optimizer:** Adam
    - **Loss Function:** Binary Crossentropy
    - **Metrics:** Accuracy
    - **Epochs:** 20
5. **Evaluation**:
    - After training, the model was evaluated on the test dataset, achieving a high accuracy rate.

Results and Evaluation:

The trained model achieved a test accuracy of approximately 76%, demonstrating effective learning of diabetes prediction.

Conclusion:

The implemented MLP model effectively predicts diabetes progression using health metrics, achieving reliable performance on the test dataset. This tool can assist in early diagnosis and preventive healthcare interventions.


Age
Gender
Occupation
Family History
Smoking
Obesity
Diet
Physical Activity
Stress
Chest Pain Type
Previous Chest Pain
Edema
bp_systolic
bp_diastolic
Heart Rate
Diabetes
Troponin
ECG
Heart Disease "Yes"
Heart Disease "No"
Input Layer
Hidden Layer
Output Layer