**HEART DISEASE PREDICTION USING MACHINE LEARNING**

**Mr.R.U Rajaguru1, Dr. R. Senthil Kumar2**

1Student, B.Sc. Computer Science with Cognitive Systems, Dr N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India

2Professor, B.Sc. Computer Science with Cognitive Systems, Dr N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India

**ABSTRACT**

1. This study focuses on predicting heart disease using machine learning techniques, aiming to enhance model performance and explainability. The research explores various algorithms, data preprocessing techniques, and performance metrics. The proposed model improves prediction accuracy and provides interpretable results by integrating explainability tools and performance monitoring. The findings demonstrate that feature selection, hyperparameter tuning, and robust evaluation strategies significantly impact model effectiveness. The proposed approach outperforms traditional methods and enhances decision-making in clinical applications. Additionally, a comparative study of feature importance and model performance across different patient demographics provides deeper insights into model reliability.

**Keywords:** Machine Learning, Heart Disease Prediction, Model Performance, Explainability, Data Validation, Clinical Decision Support

1. **INTRODUCTION**

Heart disease remains a leading cause of mortality worldwide, necessitating accurate and early diagnosis. Machine learning models have demonstrated the potential to predict heart disease based on patient data. This paper presents an improved heart disease prediction model by optimizing performance, ensuring data validity, and incorporating explainability tools. The study highlights existing research, challenges, and methodologies to enhance predictive accuracy. Furthermore, this research examines the impact of imbalanced datasets on prediction outcomes and suggests strategies to mitigate bias in model training.

1. **METHODOLOGY**

The study utilizes a dataset containing patient health records with attributes such as age, cholesterol levels, and blood pressure. The workflow includes data preprocessing, feature selection, model training, and evaluation. Various machine learning algorithms, including logistic regression, decision trees, and neural networks, are assessed. Explainability techniques like SHAP (SHapley Additive Explanations) are integrated to enhance model transparency. Moreover, a synthetic data augmentation technique is applied to address data scarcity and class imbalance.

**2.1 Data Preprocessing**

* Handling missing values and outliers
* Normalization and feature engineering
* Oversampling and undersampling techniques to address class imbalance

**2.2 Model Training and Evaluation**

* Algorithms used: Logistic Regression, Random Forest, XGBoost, Neural Networks
* Performance metrics: Accuracy, Precision, Recall, F1-score, AUC-ROC, Matthews Correlation Coefficient (MCC)
* Hyperparameter tuning using Grid Search and Bayesian Optimization

**3. IMPLEMENTATION OF WEB APPLICATION**

To make the heart disease prediction model accessible to a wider audience, a web-based interface was developed. The front-end of the application was designed using HTML, CSS, and JavaScript, while the back-end was implemented using Flask and Python. The trained model was integrated into the web application to allow users to input patient health parameters and receive real-time predictions.

**3.1 Web Application Features:**

* **User-friendly Interface:** The web page provides a simple form for users to input required health parameters.
* **Model Integration:** The trained machine learning model is deployed using Flask APIs to process user input and generate predictions.
* **Interactive Data Visualization:** The results are displayed along with feature importance using SHAP analysis to help users understand key risk factors.
* **Security Measures:** The application ensures secure data transmission and follows privacy regulations for handling medical data.
* **Backend Processing:** The Flask API receives JSON input with patient parameters, normalizes the data using a preloaded scaler, and runs predictions using a trained **Random Forest model** (random\_forest\_model\_v1\_1\_3.pkl).
* **Real-time Predictions:** Users receive instant risk assessments based on their input, making it a practical tool for early diagnosis.
1. **MODELING AND ANALYSIS**

The model training results indicate that ensemble techniques and deep learning models perform better than traditional classifiers. Hyperparameter tuning using grid search improves the predictive capabilities. Explainability methods reveal key factors influencing predictions, aiding clinicians in decision-making.



**Figure 1: 4.1 Feature importance**

1. **RESULTS AND DISCUSSION**

The results indicate that deep learning-based approaches achieve the highest accuracy. However, simpler models like logistic regression remain interpretable and useful for practical applications. SHAP analysis highlights the importance of cholesterol levels, age, and blood pressure in predicting heart disease. The inclusion of explainability enhances trust in model predictions and assists healthcare professionals in understanding decision boundaries. Additionally, an ablation study is conducted to analyze the contribution of individual features in improving predictive accuracy, demonstrating that certain lifestyle factors significantly impact the risk assessment.

The web-based application provides an interactive and accessible means of utilizing the predictive model. Users can enter patient information and obtain real-time predictions, making it a practical tool for preliminary risk assessment before clinical consultation. The Flask backend processes the data and returns predictions dynamically, ensuring quick and efficient responses. The integration of explainability techniques like SHAP within the web application provides additional transparency regarding key risk factors for heart disease.

1. **CONCLUSION**

This study presents an improved heart disease prediction model with enhanced performance and interpretability. The integration of explainability techniques ensures transparency in predictions.

1. **REFERENCES**
2. T. Mohana Priya, Dr. M. Punithavalli & Dr. R. Rajesh Kanna, "Machine Learning Algorithm for Development of Enhanced Support Vector Machine Technique to Predict Stress," Global Journal of Computer Science and Technology, 2020.
3. Ganesh Kumar and P. Vasanth Sena, "Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit," International Journal of Computer Science and Network Security, 2015.
4. Gyusoo Kim and Seulgi Lee, "2014 Payment Research," Bank of Korea, 2015.
5. Chengwei Liu et al., "Financial Fraud Detection Based on Random Forest," International Journal of Economics and Finance, 2015.