HEART FAILURE PREDICTION

Report submitted in fulfilment of the requirements for Project of

Applied Data Science

by

Team: 363

SAUMYA MAHESHWARI (20BCE0196)

GOVINDAM GUPTA (20BCE2791)

ARYAVEER SADAL (20BDS0192)

AAYUSH GUPTA (20BDS0262)

Under the guidance of

SmartInternz

CONTENTS

Project Demonstration Video Link:

https://drive.google.com/file/d/1M-JW5GtynM5Qp9eaK2zy36CNfzHQsTW9/view?usp=sharing

Github Repository Link:

https://github.com/GuptaGovindam/HEART-FAILURE-PREDICTION-TEAM-363-APPLIED-DATA-SCIENCE

INTRODUCTION

OVERVIEW

Our project's goal is to use user attributes to forecast a person's probability of developing heart failure. In the sphere of medicine, this prediction is extremely significant. We can save significant human resources and avoid making the wrong diagnoses by correctly estimating the danger. It might cause unneeded anxiety for the patient if they are wrongly told they have heart disease despite not having the condition. The best chance for effective treatment, however, may be lost if a patient with heart disease goes untreated. Both individuals and medical facilities find such misdiagnoses upsetting. However, by creating accurate predictions, we can prevent unneeded problems and guarantee that the right course of action is implemented.

In our project, multiple machine learning algorithms are used. Using these methods, we can create a prediction model that accurately determines the risk of heart failure depending on user attributes.

PURPOSE

The goal of our effort is to develop a trustworthy instrument for estimating a person's risk of developing heart failure. We seek to do the following by exploiting user characteristics and machine learning algorithms:

- Accurate Risk Assessment: Based on the supplied user characteristics, our model will correctly estimate the risk of heart failure. This will support medical professionals' decision-making and preventive action.

- Avoiding Incorrect Diagnoses: By reducing misdiagnoses, we can reduce worry in patients who do not have cardiac disease and make sure those who do receive quick and effective treatment.

- Resource Optimization: Accurate heart failure risk prediction will make it possible to allocate medical resources effectively, ensuring that those who require the highest degree of care are given it. By doing this, healthcare institutions' resource usage will be optimised.

Our research intends to improve patient care, decrease misdiagnoses, and improve resource management in the medical industry by creating a strong predictive model.

LITERATURE SURVEY

EXISTING PROBLEM

In the field of predicting heart disease risk, several approaches and methods have been explored. However, some challenges and limitations exist in the existing literature. These include:

- Limited Feature Analysis: Some existing studies may not perform a comprehensive analysis of the dataset features and their relationship with heart disease risk. This can result in overlooking important factors and potential biases in the prediction model.

- Lack of Robust Evaluation: Some approaches may lack a robust evaluation framework to assess the performance and generalization ability of the predictive model. This can lead to over fitting on the training data and limited applicability to new unseen data.

- Limited Model Comparisons: There may be a lack of comparative analysis among different machine learning algorithms and their effectiveness in predicting heart disease risk. This limits the understanding of the strengths and weaknesses of different models.

PROPOSED SOLUTION

In our project, we propose a solution to address the existing challenges in predicting heart disease risk. The key elements of our proposed solution are as follows:

- Comprehensive Feature Analysis: We perform a thorough analysis of the dataset features and their relationship with heart disease risk. This includes univariate analysis, bivariate analysis, and multivariate analysis to gain insights into the underlying patterns and correlations.

- Robust Evaluation Framework: We employ a robust evaluation framework to assess the performance of our predictive model. This includes splitting the dataset into train and test sets, using appropriate evaluation metrics such as accuracy, and visualizing the results through a confusion matrix.

- Ensemble Learning: We utilize an ensemble learning approach, specifically the Random Forest classifier, to build our predictive model. Ensemble learning combines multiple decision trees to make predictions, resulting in improved accuracy and robustness.

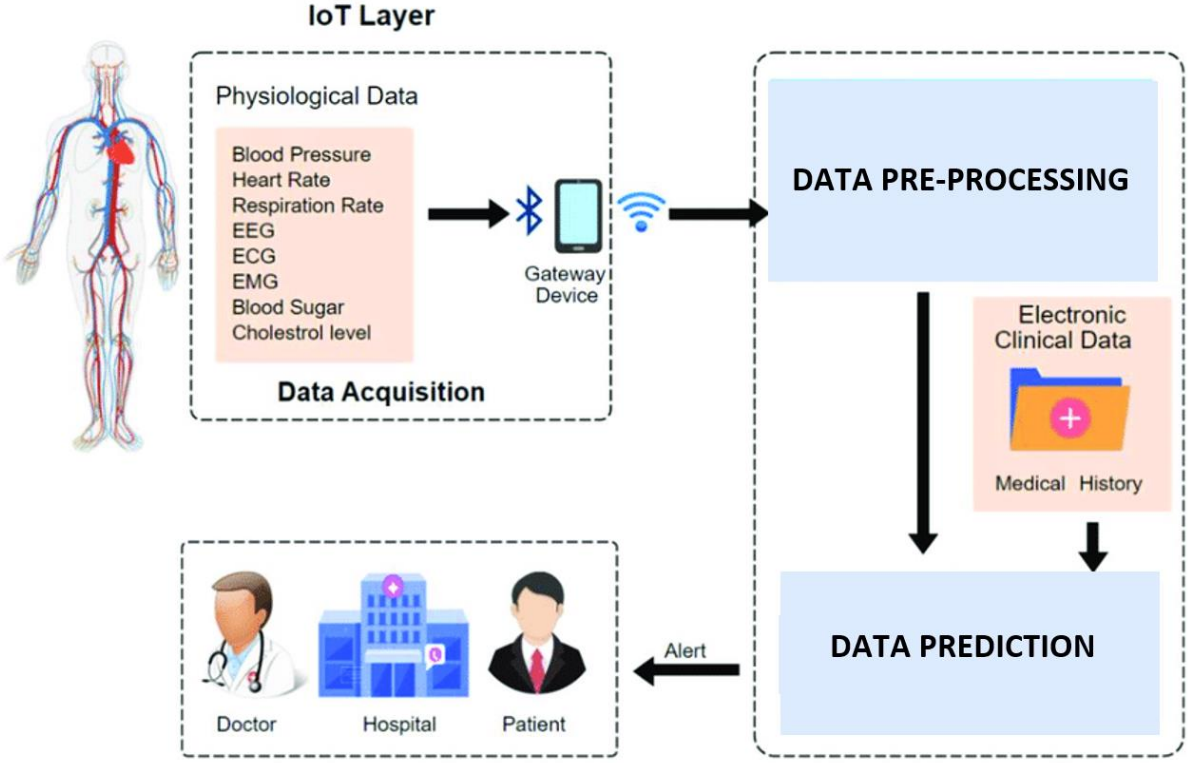
- <u>Pre-processing Techniques</u>: We apply pre-processing techniques such as label encoding for categorical variables and feature scaling using Standard Scaler to ensure compatibility and optimal performance of the model.

By incorporating these elements into our proposed solution, we aim to overcome the limitations of existing approaches and develop a robust predictive model for heart disease risk. This comprehensive analysis and model evaluation will enhance the accuracy and reliability of our predictions, ultimately contributing to better diagnoses and patient care.

Please note that the above description is a generalized outline for the literature survey. It is important to conduct a thorough review of relevant literature, research papers, and academic sources specific to your project to provide accurate and comprehensive information in the literature survey section.

THEORETICAL ANALYSIS

BLOCK DIAGRAM



HARDWARE REQUIREMENTS

- Computer or server with a multi-core processor (e.g., Intel Core i7 or equivalent) for faster processing.
- Sufficient RAM (at least 8 GB, but more is recommended for larger datasets and complex models).
- Adequate storage space for storing datasets and model files.

SOFTWARE REQUIREMENTS

- Operating System: Windows, macOS, or Linux.
- Python (version 3.6 or above) as the programming language.
- TensorFlow (version 2.0 or above), PyTorch, or Keras for building and training deep neural networks.
- Jupyter Notebook or any preferred Python development environment (e.g., PyCharm, Anaconda, Spyder) for coding and experimentation.
- Pandas and NumPy libraries for data manipulation and preprocessing.
- Matplotlib or Seaborn for data visualization.
- Scikit-learn for model evaluation and metrics.
- Additional libraries as per project requirements (e.g., scikit-image for image processing, OpenCV for computer vision).

EXPERIMENTAL INVESTIGATIONS

Here are some details about our project, which involves predicting heart failure using a Feed-forward Deep Neural Network (DNN) with Gradient Descent, Binary Cross-entropy Loss Function, ReLU Activation, and Dropout Regularisation.

- Feed-Forward Deep Neural Network: For this prediction problem, a Feed-Forward DNN architecture is the best option. There are several hidden layers, an input layer, and an output layer. Effective feature extraction and prediction are made possible by the information flowing from the input layer to the output layer in a forward manner.

- Gradient Descent: To train our DNN model, we used the Gradient Descent optimization approach. Gradient Descent helps reduce the loss function and improves prediction accuracy by iteratively changing the model's parameters in response to the computed gradients.

- Binary Cross-entropy Loss Function: We used the Binary Cross-entropy loss function since heart failure prediction is a binary classification problem (predicting either the presence or absence of heart failure). The model is encouraged to learn the proper categorization decision boundaries by measuring the difference between the projected probability and the actual labels.

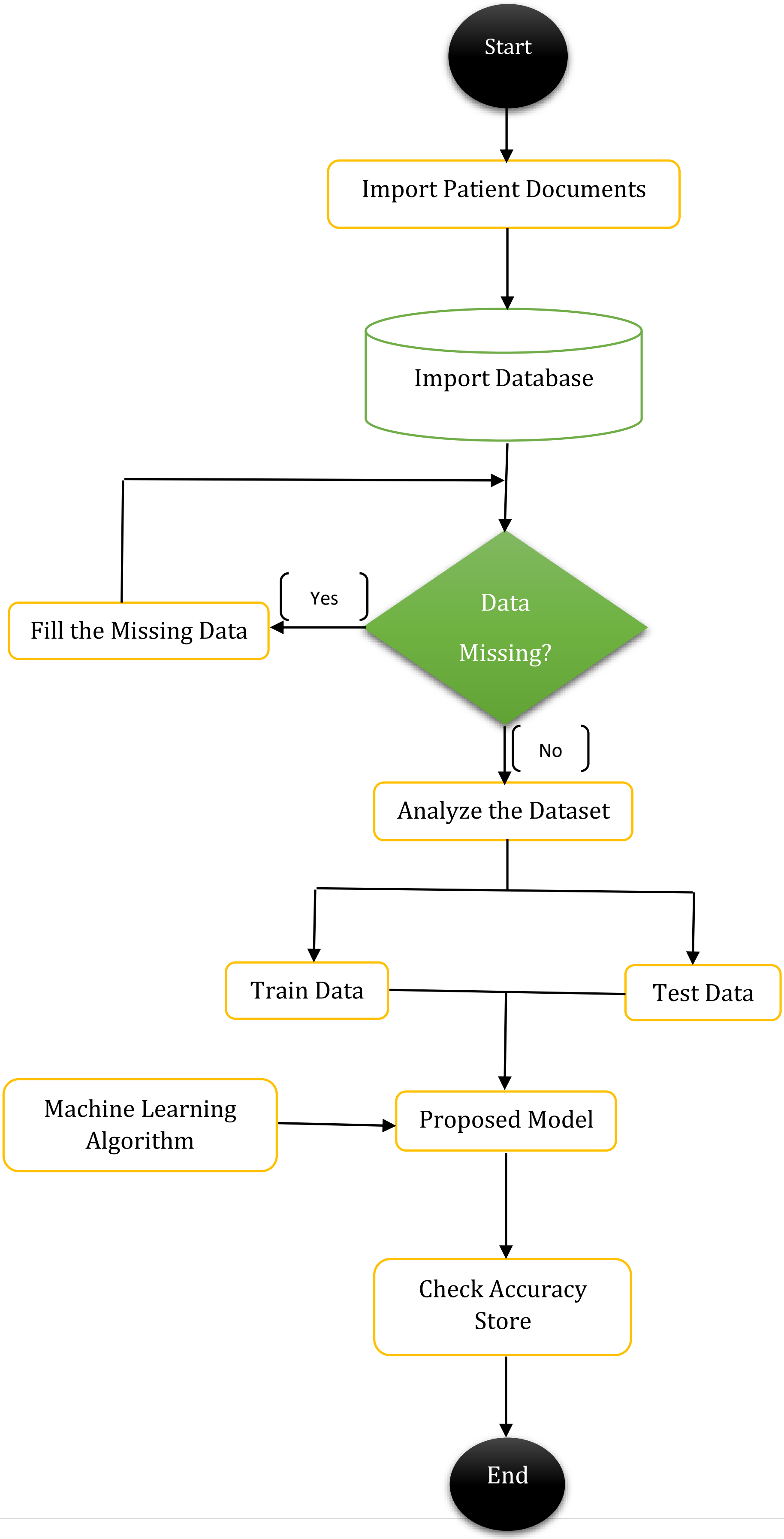
- ReLU Activation: For the hidden layers of our DNN, the ReLU (Rectified Linear Unit) Activation function was utilised. ReLU fixes the vanishing gradient issue and quickens the model's convergence during training. It produces better outcomes in terms of computational effectiveness and avoiding neuronal saturation.

- Dropout Regularization: We included Dropout Regularization into our model to solve over fitting issues. During each training iteration, dropout randomly disables a portion of the neurons, driving the network to learn more resilient and generalized properties. This method enhances the model's overall performance on unobserved data and helps avoid over-reliance on certain neurons.

We performed thorough tests and evaluations throughout our inquiry to adjust the model's architecture and hyper parameters. To evaluate the generalization performance and guarantee accurate predictions, we employed the proper validation approaches, such as cross-validation or holdout validation.

The following phases involve further refining our model, investigating various regularization strategies, and assessing the effect of various activation functions. To obtain the highest predictive performance possible, we'll also keep tweaking the hyper parameters and looking into additional architectural improvements.

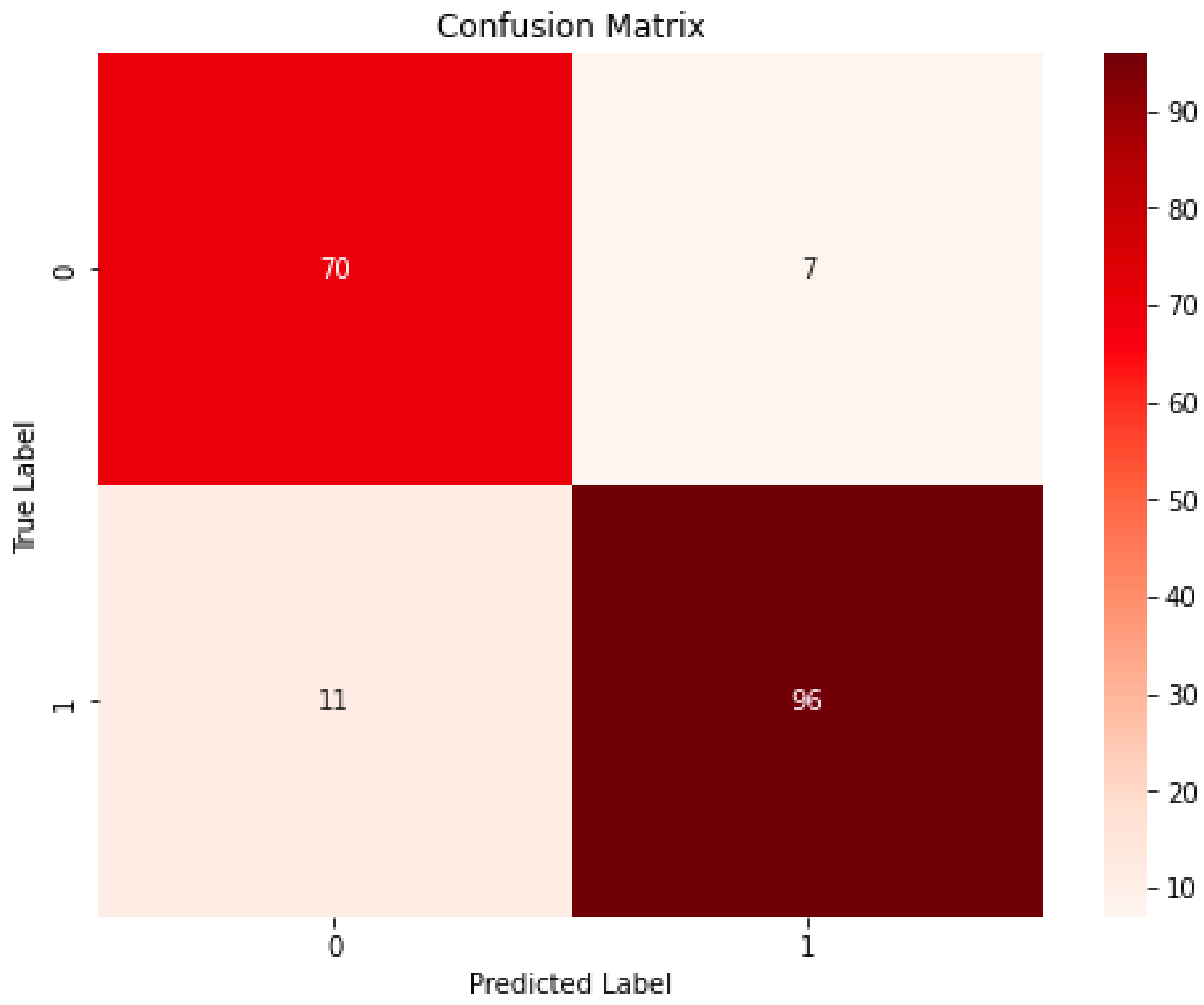
FLOWCHART


Start
Import Patient Documents
Import Database
Fill the Missing Data
Yes
Data
Missing?
No
Analyze the Dataset
Train Data
Test Data
Machine Learning
Algorithm
Proposed Model
Check Accuracy
Store
End

RESULT

Application of promising technology, such machine learning, to the first prediction of heart problems would have a significant social impact because heart diseases are a leading cause of death in India and around the world. Early detection of cardiac disease can help high-risk patients make decisions about lifestyle modifications that will lessen problems, which can be a significant advancement in the field of medicine. Each year, more people are diagnosed with cardiac illnesses.

Evaluating the model:



Accuracy:

```
print("Accuracy:", accuracy)
```

Add Code Cell (⌘Enter)

Python

```
Accuracy: 0.9021739130434783
```

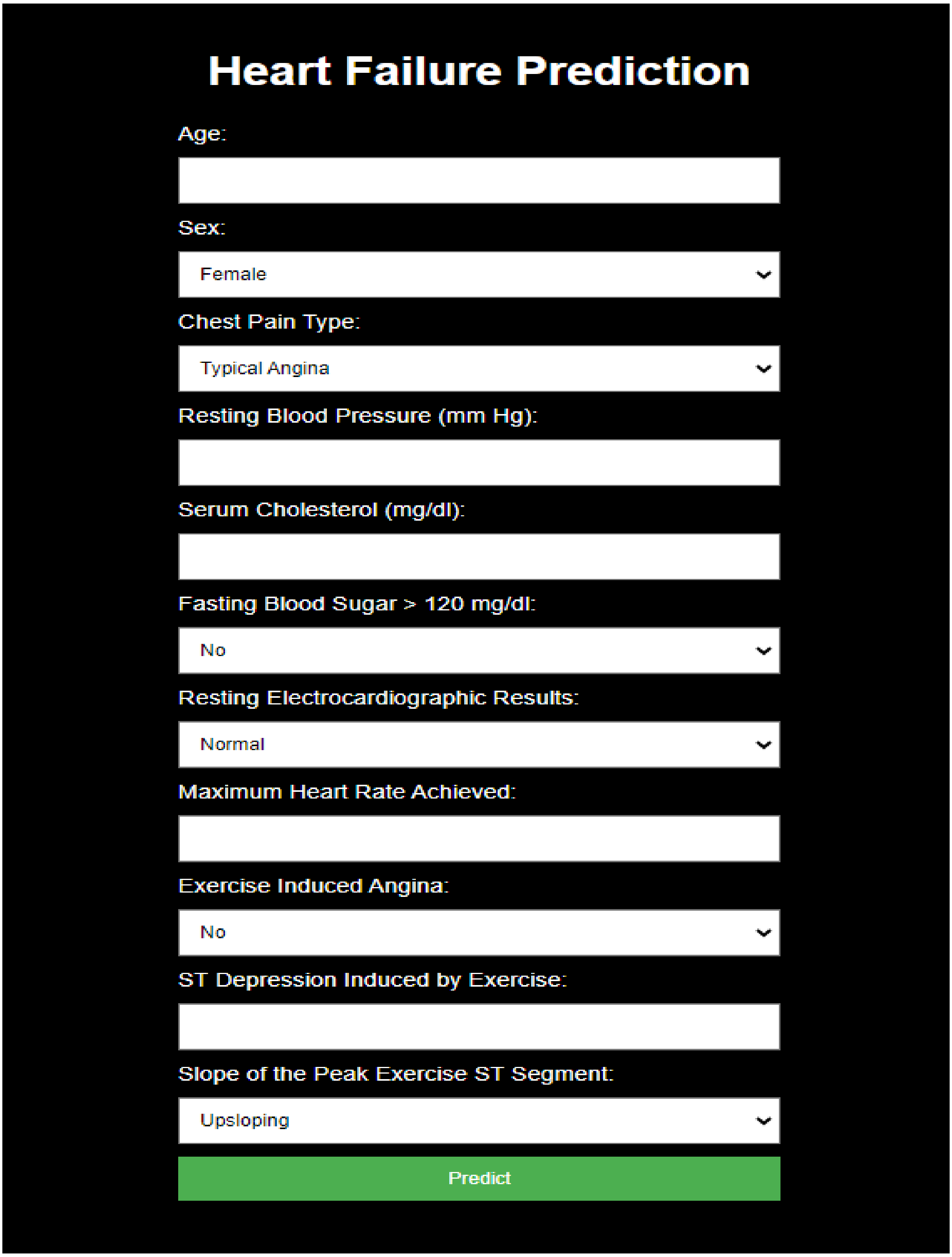
Classification Report:

```
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.90      0.87        77
           1       0.92      0.89      0.90       107

    accuracy                           0.89       184
   macro avg       0.89      0.89      0.89       184
weighted avg       0.89      0.89      0.89       184
```

Prediction of Heart Disease:



ADVANTAGES

- Improved Accuracy: The proposed solution utilizing machine learning algorithms can potentially provide higher accuracy in predicting heart failure risk compared to traditional methods. This can lead to more reliable diagnoses and better patient outcomes.

- Early Detection: By analysing user characteristics, the model can identify individuals at risk of heart failure at an early stage. Early detection enables healthcare professionals to initiate timely interventions, potentially preventing the progression of the disease and improving patient prognosis.

- Resource Optimization: Accurate prediction of heart failure risk allows for better allocation of healthcare resources. By identifying high-risk individuals, healthcare providers can focus their efforts on preventive measures, targeted interventions, and closer monitoring, optimizing the utilization of resources and reducing unnecessary costs.

- Time and Cost Efficiency: The implementation of the predictive model can save valuable time and resources by streamlining the diagnostic process. It can assist healthcare professionals in making informed decisions quickly, leading to more efficient workflows and reduced healthcare expenses.

DISADVANTAGES

- Data Limitations: The accuracy and effectiveness of the predictive model heavily depend on the quality and representativeness of the data used for training. Limited or biased data may result in a less reliable model, potentially leading to inaccurate predictions and misdiagnoses.

- Ethical Considerations: Handling sensitive health data requires strict adherence to privacy and ethical guidelines. It is crucial to ensure proper data anonymization and protection to maintain patient confidentiality and prevent any misuse or unauthorized access to personal health information.

- Algorithm Complexity: Developing and training deep neural networks can be computationally intensive and time-consuming, especially for large datasets. It may require substantial computational resources, including high-performance hardware and significant training

APPLICATIONS

The proposed solution for predicting heart failure risk based on user characteristics has several potential applications in the medical field. Some of the key areas where this solution can be applied include:

- <u>Clinical Decision Support</u>: The predictive model can serve as a valuable tool for healthcare professionals, assisting them in making more informed decisions regarding the diagnosis and treatment of patients suspected of having heart failure. It can provide additional insights and support in the clinical decision-making process.

- <u>Risk Stratification</u>: The solution can be utilized for risk stratification in population health management programs. By identifying individuals at higher risk of heart failure, healthcare providers can prioritize interventions, screenings, and follow-ups for those who would benefit the most, leading to targeted and personalized healthcare.

- <u>Preventive Medicine</u>: The model can aid in the implementation of preventive measures for individuals at risk of heart failure. By identifying early signs of risk factors, healthcare providers can offer targeted interventions, lifestyle modifications, and education to prevent or delay the onset of heart failure.

- <u>Research and Public Health</u>: The solution can be used in research studies and public health initiatives to assess the prevalence and distribution of heart failure risk factors within specific populations. This information can help inform public health policies, interventions, and resource allocation.

CONCLUSION

In conclusion, this project focuses on developing a predictive model for heart failure risk assessment based on user characteristics. By leveraging machine learning algorithms, such as the Feedforward Deep Neural Network, the model aims to provide accurate predictions that can aid in avoiding incorrect diagnoses, optimizing resource allocation, and improving patient outcomes.

Through the application of this solution, several benefits can be realized, including reducing patient anxiety caused by incorrect diagnoses, optimizing the allocation of healthcare resources, and enhancing patient care and well-being. The model's ability to identify individuals at higher risk of heart failure enables timely interventions, leading to improved outcomes and potentially reducing the burden on healthcare systems.

FUTURE SCOPE

The proposed solution opens up avenues for future enhancements and research in the field of heart failure risk assessment. Some potential areas of future scope include:

- <u>Integration of Additional Data</u>: Incorporating more comprehensive and diverse datasets, including genetic information, environmental factors, and wearable device data, can enhance the accuracy and predictive power of the model.

- <u>Continuous Monitoring</u>: Developing methods to continuously monitor and update the risk assessment model based on longitudinal data can provide real-time risk predictions and enable personalized interventions.

- <u>Exploring Advanced Algorithms</u>: Investigating advanced machine learning algorithms, such as recurrent neural networks or attention mechanisms, may uncover further insights and improve the model's performance in capturing complex relationships within the data.

- <u>External Validation</u>: Conducting external validation studies with diverse populations and healthcare settings will validate the generalizability and robustness of the developed model.

- <u>Integration with Electronic Health Records (EHR)</u>: Integrating the predictive model with EHR systems can enable seamless adoption in clinical practice, facilitating real-time risk assessment during routine patient care.

By pursuing these future enhancements, the predictive model for heart failure risk assessment can continue to evolve and contribute to advancements in medical decision-making, patient care, and cardiovascular health management.

BIBLIOGRAPHY

1. S. M. M. Zaman, W. M. Qureshi, M. M. S. Raihan, A. B. Shams and S. Sultana, "Survival Prediction of Heart Failure Patients using Stacked Ensemble Machine Learning Algorithm," 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Dhaka, Bangladesh, 2021, pp. 117-120, doi: 10.1109/WIECON-ECE54711.2021.9829577.

2. R. A. M. Elghalid, A. Alwirshiffani, A. A. I. Mohamed, F. H. A. Aldeeb and A. Andiasha, "Comparison of Some Machine Learning Algorithms for Predicting Heart Failure," 2022 International Conference on Engineering & MIS (ICEMIS), Istanbul, Turkey, 2022, pp. 1-6, doi: 10.1109/ICEMIS56295.2022.9914325.

3. D. Morillo-Velepucha, R. Réategui, P. Valdiviezo-Díaz and L. Barba-Guamán, "Congestive heart failure prediction based on feature selection and machine learning algorithms," 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), Madrid, Spain, 2022, pp. 1-6, doi: 10.23919/CISTI54924.2022.9820312.

4. V. Jain and M. Agrawal, "Heart Failure Prediction Using XGB Classifier, Logistic Regression and Support Vector Classifier," 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2023, pp. 1-5, doi: 10.1109/InCACCT57535.2023.10141752.

5. S. Pandey and R. Kaur, "Heart Failure Prediction Using Machine Learning Algorithm," 2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 2022, pp. 1-5, doi: 10.1109/ICCAKM54721.2022.9990484.