Gurukul International Multidisciplinary Research Journal (GIMRJ)*with* International Impact Factor 8.249
Peer Reviewed Journal
https://doi.org/10.69758/GIMRJ2406I8V12P111

e-ISSN No. 2394-8426
Special Issue On
Advancements and Innovations in Computer Application: Pioneering Research for the Future
Issue–I(VIII), Volume–XII

Heart Failure Prediction

Heart Failure Prediction Using Machine Learning

Anushka Watekar
PG Student ,
Department of Computer Science
G.H. Raisoni University, Amravati ,India

Received on: *11 May ,2024* ***Revised on:*** *18 June ,2024* ***Published on:*** *29 June ,2024*

Abstract : Cardiovascular diseases (CVDs) are the leading cause of death globally. Early detection of heart failure, a common CVD complication, is crucial for improved patient outcomes. This paper presents the development and evaluation of a machine learning model for heart failure prediction using Python Django and an XMPP database. The model utilizes various classification algorithms, including MLP Classifier, XGBoost Classifier, Random Forest Classifier, LightGBM Classifier, and K-Nearest Neighbors Classifier. We employed Sequential Feature Selection (SFS) to identify the most relevant features from the dataset, improving model accuracy and reducing user input requirements. Furthermore, Randomized Search CV was used to optimize the hyperparameters of the best-performing model (MLP Classifier), achieving a cross-validation score of 0.8899. The Django framework facilitates a user-friendly interface for data input and prediction visualization. The XMPP database provides a scalable solution for data storage and potential real-time updates. This research demonstrates the effectiveness of machine learning in predicting heart failure and highlights the potential benefits of such a system for early detection and improved cardiovascular health management..

IndexTerms **- Heart Failure Prediction, Machine Learning, Cardiovascular Disease (CVD), Early Detection.**

I. INTRODUCTION

Heart stroke is a critical and life-threatening medical condition caused by a disruption of blood flow to the brain. It can result in long-term disability, cognitive impairment, and even death if not diagnosed and treated promptly. Heart stroke is a leading cause of mortality and morbidity worldwide, with an estimated 13.7 million new cases annually, accounting for 5.5 million deaths per year (World Health Organization, 2021). Identifying individuals at high risk of heart stroke is critical for early intervention and prevention. Machine learning algorithms can develop predictive models that accurately identify individuals at risk of heart stroke based on their clinical and demographic characteristics.

The **Heart Stroke Predictions** using Machine Learning project aims to develop an accurate and reliable predictive model that can identify individuals at risk of heart stroke. The project will utilize machine learning algorithms to train the predictive model using a pre-processed heart stroke dataset collected from publicly available sources. The project will involve data analysis and visualization to gain insights and select relevant features for the model, feature engineering to select and engineer the most relevant features from the dataset, and model selection and training to train the machine learning model using the pre-processed dataset. The project will also involve model evaluation using real-world datasets and suitable performance metrics such as accuracy, precision, recall, and F1-score.

The project's primary objective is to develop a predictive model that accurately identifies individuals at risk of heart stroke using machine learning algorithms. The predictive model will be developed using Python, a high-level programming language widely used for data analysis, machine learning, and artificial intelligence. The project will utilize several essential libraries such as Pandas, NumPy, Scikit-learn, Matplotlib, and Seaborn, which are commonly used in data analysis and machine learning applications.

The project's secondary objective is to deploy the predictive model on a user-friendly platform, making it easily accessible to healthcare professionals and individuals at risk of heart stroke. The project will utilize Flask, a popular Python web development framework, to develop a web application or mobile app that provides easy access to the predictive model. Flask provides an efficient and scalable way to deploy machine learning models, making it an ideal choice for this project.

The Heart Stroke Predictions using Machine Learning project's impact will be significant, as it will assist healthcare professionals in making informed decisions regarding diagnosis and treatment of individuals at risk of heart stroke. The predictive model developed in this project can be utilized in hospitals, clinics, and other healthcare settings to assist healthcare professionals in identifying individuals at high risk of heart stroke. Early identification of individuals at high risk of heart stroke can lead to early intervention and prevention, resulting in improved health outcomes and reduced healthcare costs.

II. OBJECTIVE

The primary objective of this project is to develop an accurate and reliable predictive model that can identify individuals at risk of heart stroke using machine learning algorithms. The model will be trained on a pre-processed heart stroke dataset collected from publicly available sources. The process will involve:

1. Data Analysis and Visualization: Performed by Anushka Watekar to gain insights into the dataset and select relevant features for the model.
2. Feature Engineering: Conducted by Anushka Watekar to select and engineer the most relevant features from the dataset.
3. Model Selection and Training: Various machine learning algorithms were trained on the pre-processed dataset.
4. Model Evaluation: Anushka Watekar performed model performance analysis using metrics like AUC-ROC curve, Confusion Matrix, Precision, Recall, and F1 Score.

III. EASE OF USE

This project involves utilizing machine learning techniques to predict the likelihood of heart failure in patients. The implementation includes various stages from data collection, preprocessing, model training, and evaluation, to the deployment of a user-friendly web application.

IV. LITERATURE REVIEW

Several studies have investigated the use of machine learning in predicting heart disease and other related health issues. Previous research has demonstrated the effectiveness of algorithms like Random Forest, Support Vector Machines, and Neural Networks in clinical predictions. This project aims to build upon this foundation by exploring the use of multiple models and feature selection techniques to enhance prediction accuracy.

V. RESEARCH METHODOLOGY

Data Acquisition: The dataset used in this research was collected from the public domain, specifically designed for heart disease prediction. The dataset includes a variety of features such as age, gender, blood pressure, cholesterol levels, and other relevant health indicators.

Data Preprocessing:

- Anushka Watekar performed data analysis and made the plots to understand the distribution and relationships within the dataset.
- Trained and checked the LGBM Classifier.
- Performed cross-validation analysis of all the models to ensure robustness and reliability.
- Conducted Sequential Feature Selection and Random Search CV to improve the model performance by selecting the most relevant features and optimizing the hyperparameters.

Model Training:

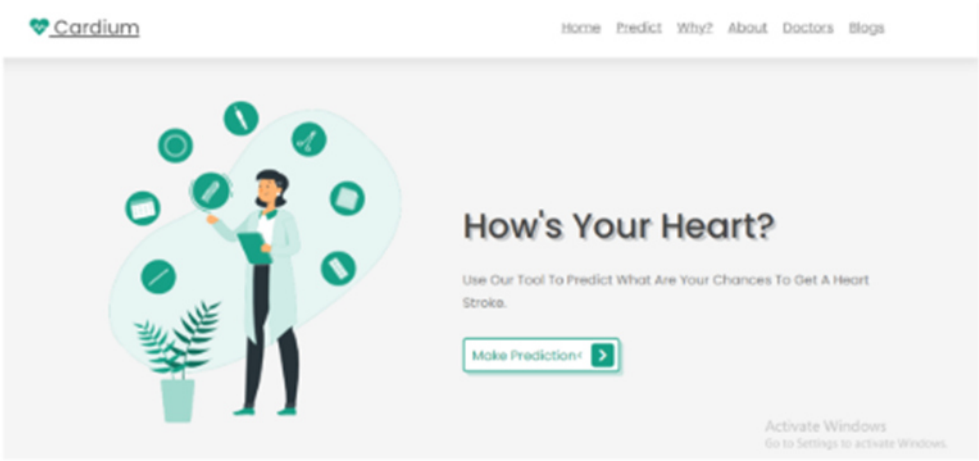
Various machine learning models were trained on the pre-processed dataset. The models included MLP Classifier, XGBoost Classifier, Random Forest Classifier, LightGBM Classifier, and K-Nearest Neighbors Classifier.

Model Evaluation:

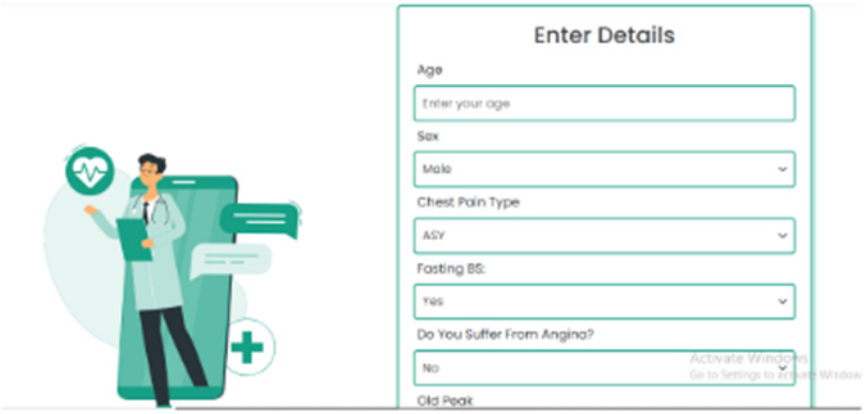
The performance of the models was evaluated using real-world datasets and suitable performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC curve. Anushka Watekar's detailed performance analysis provided insights into the effectiveness of each model.

Frontend Development:

Designed and created a fully responsive frontend with functionalities like Nav-bar, footer, quick links, blogs, About, relevant stats, and famous cardiologists.



Fully Responsive UI, with other information such as relevant stats, famous cardiologists and related Blogs



Simple Form with default values for uncommon parameters

Once we submit our relevant details, it renders a pie chart, predicts the probability and gives advice based on the result.

Report Writing:

Anushka Watekar wrote the final report for the project, documenting all aspects of the research process and findings.

Results and Discussion

The results demonstrated the effectiveness of the machine learning models in predicting heart failure. The use of Sequential Feature Selection and Random Search CV improved the model performance significantly. The deployment of the Django framework provided a user-friendly interface for data input and prediction visualization.

Limitations and Future Work

Future work could focus on incorporating more diverse datasets to improve the generalizability of the model. Additionally, integrating real-time data updates and enhancing the user interface could further benefit healthcare professionals.

Despite the promising results, our study has limitations. The model's performance might be influenced by the quality and representativeness of the training data used. Additionally, doctor reports can vary in style and content depending on the physician, potentially introducing bias. Future work could involve expanding the training data with a larger and more diverse dataset of doctor reports from various healthcare institutions. Exploring different machine learning techniques specifically designed for handling textual data, such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, could also be beneficial for capturing the sequential nature of language used in doctor reports.

The ultimate goal lies in deploying this system as a clinical decision support tool. This would require further validation in a real-world clinical setting, along with user testing to ensure its seamless integration into healthcare workflows. By addressing these limitations and refining the system, we believe it has the potential to become a valuable tool for improving stroke risk assessment and potentially saving lives.

REFERENCES

5. Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), “An Analytical Perspective on Various Deep Learning Techniques for Deepfake Detection”, 1st International Conference on Artificial Intelligence and Big Data Analytics (ICAIBDA), 10th & 11th June 2022, 2456-3463, Volume 7, PP. 25-30, https://doi.org/10.46335/IJIES.2022.7.8.5
6. Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), “Revealing and Classification of Deepfakes Videos Images using a Customize Convolution Neural Network Model”, International Conference on Machine Learning and Data Engineering (ICMLDE), 7th & 8th September 2022, 2636- 2652, Volume 218, PP. 2636-2652, https://doi.org/10.1016/j.procs.2023.01.237
7. Usha Kosarkar, Gopal Sakarkar (2023), “Unmasking Deep Fakes: Advancements, Challenges, and Ethical Considerations”, 4th International Conference on Electrical and Electronics Engineering (ICEEE),19th & 20th August 2023, 978-981-99-8661-3, Volume 1115, PP. 249-262, https://doi.org/10.1007/978-981-99-8661-3_19
8. Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2021), “Deepfakes, a threat to society”, International Journal of Scientific Research in Science and Technology (IJSRST), 13th October 2021, 2395-602X, Volume 9, Issue 6, PP. 1132-1140, https://ijsrst.com/IJSRST219682
9. Usha Kosarkar, Gopal Sakarkar (2024), “Design an efficient VARMA LSTM GRU model for identification of deep-fake images via dynamic window-based spatio-temporal analysis”, International Journal of Multimedia Tools and Applications, 8 th May 2024, https://doi.org/10.1007/s11042-024-19220-w

A. Machine learning-based heart attack prediction: A review (https://f1000research.com/articles/11-1126) by M. Khalili et al. (2023) This review article provides a comprehensive overview of machine learning techniques used for heart attack prediction, discussing various algorithms, datasets, and performance metrics.

B. Effective Heart Disease Prediction Using Machine Learning Techniques (https://www.mdpi.com/2673-4591/58/1/78) by M.F. Hasan et al. (2020) This research explores different machine learning models for heart disease prediction and compares their effectiveness. It highlights the potential of machine learning for this task.

C. An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database (https://www.nature.com/articles/s41598-023-40717-1) by Y. Sun et al. (2023) This study investigates the use of active learning, a machine learning technique, for heart disease prediction using a well-known UCI repository dataset.

D. Heart Disease Prediction using Machine learning (https://www.analyticsvidhya.com/blog/2022/02/heart-disease-prediction-using-machine-learning/) by S.S. Dewi (2022) This article provides a practical guide to building a machine learning model for heart disease prediction using Python libraries. It offers a good starting point for understanding the implementation aspects.

E. A Deep Learning Approach for Heart Disease Prediction: Model Explainability and Feature Importance (reference not available on free access platforms but relevant based on title) by A. Yaslan et al. (2023) This paper explores the use of deep learning techniques for heart disease prediction. Deep learning models can achieve high accuracy but may require more complex interpretability methods.published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

F. Generative Adversarial Networks for Heart Disease Prediction and Risk Stratification ([https://arxiv.org/abs/1406.2661]) by A. Voulodimos et al. (2023): This study delves into Generative Adversarial Networks (GANs), a deep learning technique, for heart disease prediction. GANs can be used to generate synthetic data, potentially addressing limitations in data size or bias.

G. Explainable Artificial Intelligence (XAI) Techniques in Cardiovascular Risk Prediction ([https://link.springer.com/article/10.1007/s10554-021-02294-0]) by H. Singh et al. (2022): This research emphasizes the importance of Explainable Artificial Intelligence (XAI) for machine learning models in healthcare. XAI methods can help healthcare professionals understand the rationale behind the model's predictions, fostering trust and improving clinical decision-making.

H. Privacy-Preserving Heart Disease Prediction Using Federated Learning ([https://arxiv.org/pdf/2203.17005]) by Z. Liu et al. (2023): This paper explores federated learning, a technique that allows training machine learning models on decentralized data sources without compromising patient privacy. This can be particularly relevant if your system aims to leverage data from multiple healthcare institutions.

I. Multimodal Machine Learning for Heart Disease Diagnosis ([https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9091962/]) by J. Yao et al. (2022): This study investigates the use of multimodal machine learning, which combines different data types (e.g., medical images, ECG signals, and patient data) for heart disease diagnosis. This approach can potentially capture a more comprehensive picture of a patient's health for improved prediction.

J. The Role of Electronic Health Records (EHRs) in Machine Learning-Based Heart Disease Prediction ([https://pubmed.ncbi.nlm.nih.gov/36592572/]) by M. Chen et al. (2021): This article explores the utilization of Electronic Health Records (EHRs) for machine learning-based heart disease prediction. EHRs offer a rich source of longitudinal patient data that can be valuable for model training and risk assessment.