International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Prediction of Cardiac Disease

[1]Lisha Shree N, [2]Himaja H B, [3]Pushpa Latha K

[1]*Department of ISI, Presidency University, Bengaluru, Karnataka, *20201ISI0011*
[2]*Department of IST, Presidency University, Bengaluru, Karnataka, *20201IST0028*
[3]*Department of IST, Presidency University, Bengaluru, Karnataka, *20201IST9004*

A B S T R A C T

—The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. Heart disease is the Leading cause of death worldwide. With the rampant increase in the heart stroke rates at juvenile ages, we need to put a system in place to be able to detect the symptoms of a heart stroke at an early stage and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease. So we proposed a system with the help of machine learning techniques and algorithms like Logistic Regression, KNN, SVC, Random Forest, Decision Tree , XGB Classifier and Naive Bayes to predict Heart Disease based on different parameters entered by the user in the front end

1. INTRODUCTION

N People in the modern world struggle with extreme stress and anxiety as a result fissional typically define a healthy blood pressure range of 120/80 to 140/90 mmHg and a healthy pulse of their hectic schedules and regular assignments. Furthermore, there is growing concern about people who become addicted to long-term habits like smoking cigars or drinking gutka, which can result in a number of chronic illnesses like cancer, heart disease, liver issues, and kidney failure. For well-known physicians, treating and curing these chronic illnesses is a major challenge. IT specialists have taken notice of this new difficulty and are assisting in the early detection and treatment of such diseases. Every person is different from the next in terms of appearance, behavior, and blood pressure and pulse rate readings. Medical pro rate of 60 to 100 bpm. The health sector uses a range of machine learning methods and tools in the market today to forecast chronic illnesses. In spite of these efforts, scientists have found certain short comings and are looking for more precise predictive algorithms to identify chronic illnesses in people early on and potentially save lives. Thus, based on user input parameters at the front end, we propose a system that uses machine learning techniques and algorithms, such as XGB Classifier, KNN, SVC, Random Forest, Decision Tree, and Naive Bayes.

2. RESEARCH OF EXISTING METHODS

Despite significant progress, there are still areas in need of study and improvement in the field of machine learning-based heart disease prediction. Understanding these gaps is crucial to developing prediction models that are more reliable and accurate. Some unfulfilled gaps and existing methods in the field of heart disease prediction are as follows: 3.1. Limited Variability within the Sets: 1. Research Gap: Many of the datasets currently used for the prediction of heart disease lack diversity in terms of demographics, lifestyle, and geographic representation. 2. Existing methods: Researchers have used popular datasets, such as the Framingham Heart Study and Cleveland Heart Disease datasets. To increase the generalizability of the model, more diverse datasets from different papulation to enhance models. 3.2. Choosing Features and Their Significance: 1. Research Gap: It's still difficult to determine which features are most important for predicting heart disease. Comprehending the attributes that substantially contribute to precise forecasting is crucial for both clinical applicability and model interpretability. 2. Current Approaches: Recursive Feature Elimination (RFE), feature importance derived from tree-based models, and domain knowledge driven feature selection are a few of the feature selection strategies that have been used. But more reliable and automated techniques are required. 3. Unbalanced Collections: School of Computer Science Engineering \& Information Science, Presidency University. Page 6 of 37 4. Research Gap: Biased models can result from imbalanced datasets, where one class (such as the presence of heart disease) is noticeably underrepresented. 5. Current Methods: To address class imbalance, methods such as under sampling, oversampling, and the use of synthetic data (SMOTE) have been used. Nevertheless, more research is required to find the best strategy for imbalanced datasets related to heart disease. 3.3. Unbalanced Collections: 1. Research Gap: Biased models can result from imbalanced datasets, where one class (such as the presence of heart disease) is noticeably underrepresented. 2. Current Approaches: To address class imbalance, methods such as under sampling, oversampling, and the use of synthetic data (SMOTE) have been used. Nevertheless, more research is required to find the best strategy for imbalanced datasets related to heart disease. 3.4. Model Interpretability: 1. Research Gap: Interpretability issues prevent many machine learning models—especially complex ones like ensemble methods—from being widely used in clinical settings. 2. Current Approaches: Interpretability strategies have been investigated, including model-agnostic approaches, LIME, and SHAP values. Nonetheless, creating models that strike a balance between interpretability and complexity is a never-ending task. 3.5. Temporal Elements: 1. Research Gap: The onset of heart disease is a dynamic process that is impacted by time. Current models frequently ignore the data's temporal dimensions.

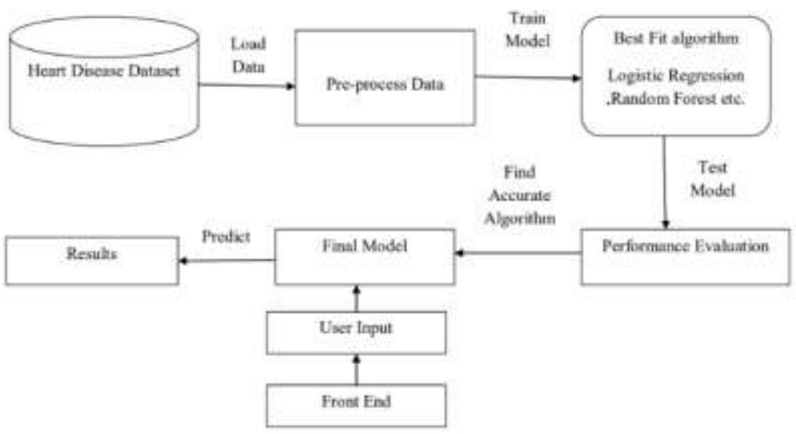
3. METHODOLOGY

A Heart Disease Dataset is taken. The dataset is loaded and preprocessed with various machine learning techniques. The preprocessed data is divided as training and testing data. The prediction model is built using machine learning algorithms like Logistic Regression, KNN, SVC, Random Forest Decision Tree, Naïve Bayes and XGB Classifier. The model is trained using training dataset and once the model has been trained successfully it has to be tested. The trained model is tested using testing dataset and accuracy is calculated. The algorithm which gives the best accuracy is taken as our final prediction model. The finalized model is converted into pickle model (binary format data) and saved. A Front End is developed with the help of Flask and HTML. Now user will enter the various parameters required to predict the heart disease in the front end. The collected parameters from the front end are given as input to our finalized algorithm to predict whether the person has the heart disease or not. Finally, the predicted output is displayed on the front end

4. OBJECTIVE

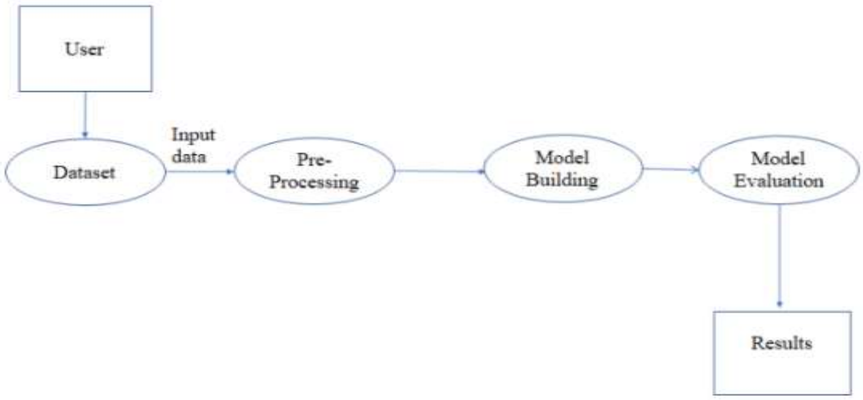
Addressing critical gaps in machine learning-based heart disease prediction is imperative for improving model reliability and accuracy. Notably, current datasets lack diversity in demographics and lifestyle, emphasizing the need for School of Computer Science Engineering \& Information Science, Presidency University. Page 9 of 37 more varied datasets to enhance model generalizability. Feature selection remains challenging, and while methods like Recursive Feature Elimination are utilized, there is a demand for more reliable automated techniques. Dealing with imbalanced datasets, a common issue, requires further investigation into the most effective strategies beyond current methods. Model interpretability is another gap, with a need for a better balance between complexity and interpretability, especially for complex models like ensembles. Additionally, incorporating the temporal dimension in models is essential, recognizing heart disease as a dynamic process impacted by time. Efforts should focus on developing models that consider the evolution of risk factors over time to capture the dynamic nature of the disease.

5. SYSTEM DESIGN & IMPLEMENTATION

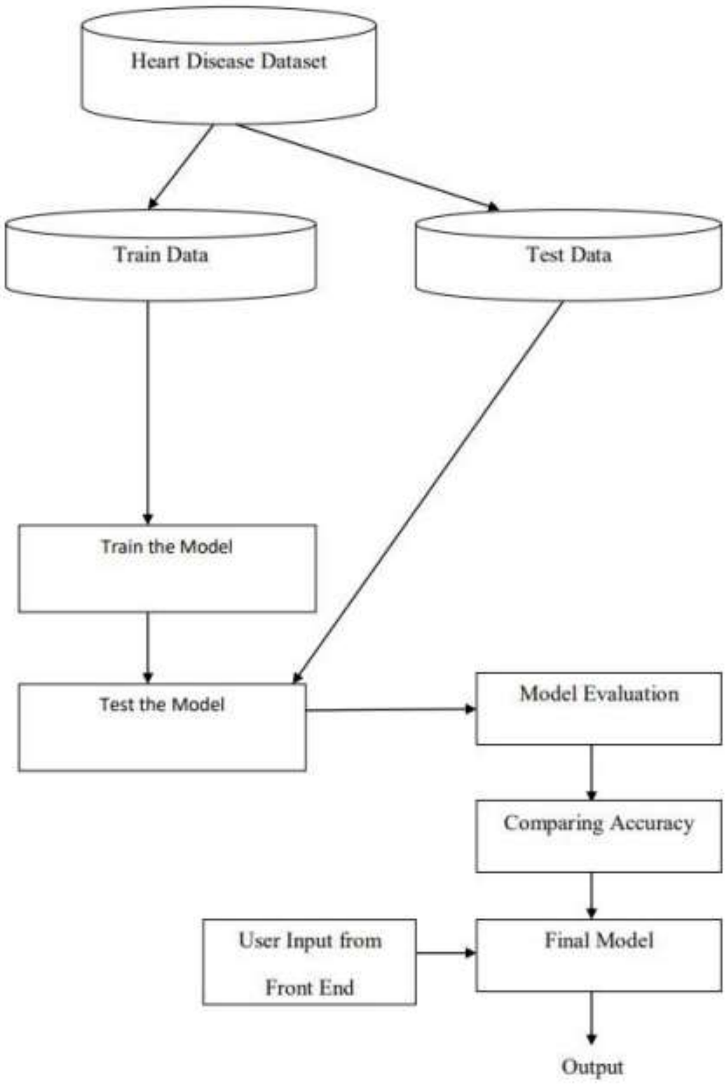


DATA FLOW DIAGRAM

In the heart disease prediction system, the model evaluation process is a critical step to ensure its reliability and accuracy. Initially, a diverse dataset encompassing demographics, lifestyle, and health history is acquired. This dataset is then split into training and testing subsets to facilitate model training and evaluation. The training phase involves utilizing algorithms like Random Forest to train the model on the training dataset, adjusting parameters for optimal performance. Subsequently, the model is tested on the reserved dataset to assess its ability to generalize to new, unseen data. During the evaluation, various metrics such as accuracy, precision, recall, and F1 score are calculated to quantify the model's performance. The comparison of these metrics aids in gauging the effectiveness of the model in predicting heart disease. Accuracy, representing the overall correctness of predictions, is a key metric, but other measures provide a more nuanced understanding, especially considering potential class imbalances in the dataset. The output of the model evaluation phase includes not only quantitative metrics but also qualitative insights into the model's strengths and potential areas for improvement. These insights are crucial for refining the model and enhancing its predictive capabilities. The iterative nature of this process ensures continuous improvement based on feedback from the testing phase. The user interface (UI) plays a pivotal role in this entire process, serving as the gateway for users to input their data securely. The UI should not only collect user inputs effectively but also present the model's predictions in a clear and understandable format. User feedback from the UI is invaluable for refining the model further, addressing any discrepancies, and making the system more user-friendly. This closed feedback loop between users and the system ensures ongoing optimization and adaptation to user needs.

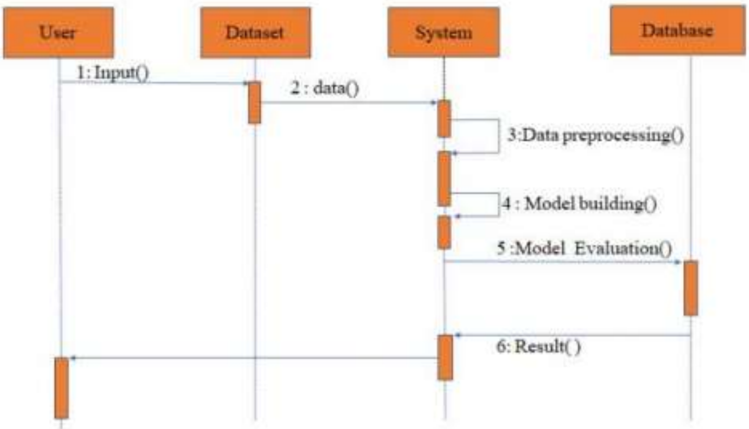


SYSTEM ARCHITECTURE

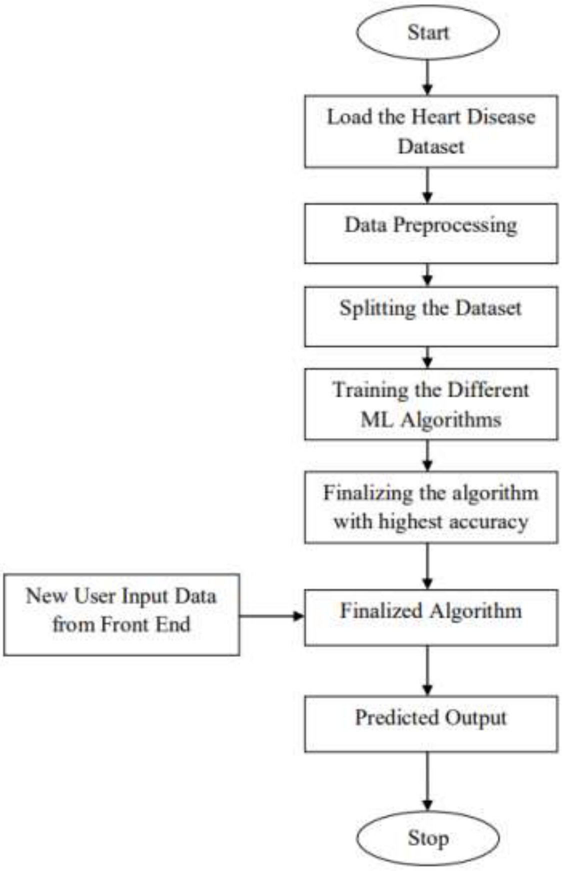


The user will provide the dataset to the system. The dataset is preprocessed in order to increase the accuracy of the model. The model is built using different algorithms. The model is evaluated and model with best accuracy is finalized. The finalized model will predict the results.

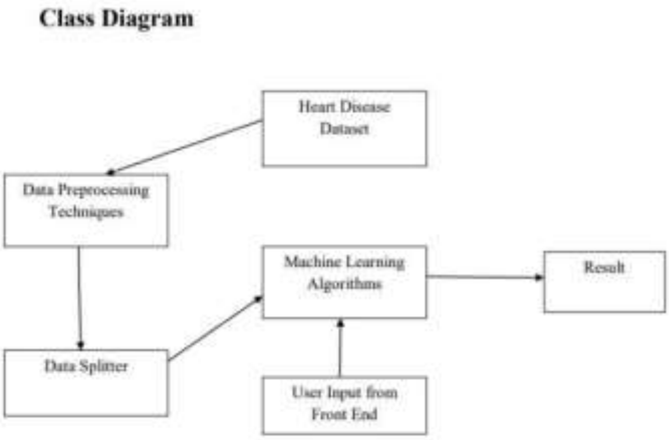
Sequence Diagram



The user will give dataset as input to the system. The system will store the dataset given by the user in its database. The system will do preprocessing of the data stored. The model is built using various ML algorithms and trained using preprocessed data. The model is evaluated and the algorithm with best accuracy is finalized. The finalized model will predict the results.



FLOWCHART



6. IMPLEMENTATION

Python: Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting language. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available as open source. Python is widely considered as the preferred language for teaching and learning ML (Machine Learning). Features: Easy to code, Free and Open Source, Object-Oriented Language, High-Level Language, Extensible feature, Python is Portable language, Python is Integrated language, Interpreted Language, Large Standard Library, Dynamically Typed Language.

Python IDLE: An IDE (or Integrated Development Environment) is a program dedicated to software development. IDEs integrate several tools specifically designed for software development. These tools usually include: An editor designed to handle code (with, for example, syntax highlighting and auto-completion) Build, execution, and debugging tools. Some form of source control. When you install Python, IDLE is also installed by default. Its major features include the Python shell window (interactive interpreter), auto-completion, syntax highlighting, smart indentation, and a basic integrated debugger. We used python version 3.6.8 HTML: HTML (Hypertext Markup Language) is the code that is used to structure a web page and its content. For example, content could be structured within a set of paragraphs, a list of bulleted points, or using images and data tables. HTML is a markup language that defines the structure of your content. HTML consists of a series of elements, which you use to enclose, or wrap, different parts of the content to make it appear a certain way, or act a certain way. The enclosing tags can make a word or image hyperlink to somewhere else, can italicize words, can make the font bigger or smaller, and so on. Anatomy of HTML Element: The main parts of our element are as follows: 1. The opening tag: This consists of the name of the element (in this case, p), wrapped in opening and closing angle brackets. This states where the element begins or starts to take effect — in this case where the paragraph begins. 2. The closing tag: This is the same as the opening tag, except that it includes a forward slash before the element name. This states where the element ends — in this case where the paragraph ends. Failing to add a closing tag is one of the standard beginner errors and can lead to strange results. 3. The content: This is the content of the element, which in this case, is just text. 4. The element: The opening tag, the closing tag, and the content together comprise the element. Elements can also have attributes that look like the following: Attributes contain extra information about the element that you don't want to appear in the actual content. Here, class is the attribute name and editor-note is the attribute value. The class attribute allows you to give the element a non-unique identifier that can be used to target it (and any other elements with the same class value) with style information and other things. An attribute should always have the following: 1. A space between it and the element name (or the previous attribute, if the element already has one or more attributes). 2. The attribute name followed by an equal sign. 3. The attribute value wrapped by opening and closing quotation marks.

7. OUTCOMES

The implementation of the heart disease prediction system has yielded several significant outcomes. Foremost among these is the successful creation of an accurate prediction model, achieved through the application of various machine learning algorithms, including Logistic Regression, Random Forest Classifier, KNN, SVC, Naïve Bayes, Decision Tree Classifier, and XGB Classifier. This ensures a comprehensive and robust model capable of providing reliable predictions for heart disease. Additionally, the system boasts a user-friendly interface, thanks to the incorporation of HTML. This enables users to seamlessly interact with the system, providing datasets and visualizing results with ease. The interface enhances the overall user experience, making the system accessible and intuitive. Moreover, the implementation leverages Python's capabilities for efficient data processing. With its high-level built-in data structures and dynamic semantics, Python proves to be instrumental in handling and processing the datasets effectively. This efficiency contributes to the system's overall performance and reliability. In conclusion, the outcomes of the heart disease prediction system implementation include the successful development of an accurate model, a user-friendly interface, and efficient data processing capabilities. These achievements collectively contribute to a system that is not only reliable and accurate but also accessible and user-centric in its design and functionality.

8. RESULTS AND DISCUSSION

Predicting and detection of heart disease has always been a critical and challenging task for healthcare practitioners. Hospitals and other clinics are offering expensive therapies and operations to treat heart diseases. So, predicting heart disease at the early stages will be useful to the people around the world so that they will take necessary actions before getting severe. So, we proposed a system with the help of machine learning techniques and algorithms like Logistic Regression, KNN, SVC, Random Forest, Decision Tree, XGB Classifier and Naïve Bayes to predict Cardiac Disease based on different parameters entered by the user in the front end.

9. CONCLUSION

In conclusion, our heart disease prediction system harnesses the power of various machine learning algorithms, achieving an impressive 91.80\% accuracy, with Random Forest standing out as a key contributor. The multi-algorithmic approach provides a comprehensive analysis of user-entered parameters. The user-friendly front end, developed with Flask, HTML, and pymysql, facilitates easy interaction, empowering users to proactively engage with their health data. This system holds immense potential for revolutionizing early diagnosis and intervention in heart disease cases, exemplifying the symbiosis of healthcare and machine learning. Ongoing refinement and exploration of additional features promise to further enhance the system's predictive capabilities, shaping the future of preventive healthcare practices globally.

10. REFERENCES

[1] P. Kola Sujatha and K. Mahalakshmi. Performance evaluation of supervised machine learning algorithms in prediction of heart disease. 2020 IEEE International Conference for Innovation in Technology (INOCON), pages 1– 7, 2020.

[2] Pahulpreet Singh Kohli and Shriya Arora. Application of machine learning in disease prediction. 2018 4th International Conference on Computing Communication and Automation (ICCCA), pages 1–4, 2018.

[3] Abderrahmane Ed-Daoudy and Khalil Maalmi. Real-time machine learning for early detection of heart disease using big data approach. pages 1–5, 04 2019.

[4] A. Lakshmanarao, A. Srisaila, and Srinivasa Tummala. Heart disease prediction using feature selection and ensemble learning techniques. pages 994–998, 02 2021.

[5] Alperen Erdoğan and Selda Guney. Heart disease prediction by using machine learning algorithms. 2020 28th Signal Processing and Communications Applications Conference (SIU), pages 1–4, 2020.

[6] Shaik Farzana and Duggineni Veeraiah. Dynamic heart disease prediction using multi-machine learning techniques. 2020 5th International Conference on Computing, Communication and Security (ICCCS), pages 1–5, 2020.