Heart Failure Prediction by Using Machine Learning

Kunde Greeshma

School of CS & AI

SR University

Warangal, India

2203A51052@sru.edu.in

Dhulipalla Hasini

School of CS & AI

SR University

Warangal, India

2203A51044@sru.edu.in

Venisetty Varsha

School of CS & AI

SR University

Warangal, India 2203A51132@sru.edu.in

**ABSTRACT**— Heart failure remains a major and growing public health issue leading to higher rates of illness and death worldwide early detection of heart failure could allow for timely treatment improving patient outcomes this study aimed to develop and test a machine learning model to predict heart failure early by analyzing various patient factors methods we used data from a group of patients each with a range of health information such as age blood pressure medical history and other health markers the data was split into two groups 70 for training the model and 30 for testing its accuracy we explored different machine learning methods like random forest gradient boosting machines and deep neural networks to see which provided the most accurate predictions the main focus was on predicting whether heart failure would occur within a set period conclusion our findings showed that machine learning models especially gradient boosting machines can accurately predict heart failure early these models could work alongside traditional diagnostic tools to help healthcare providers identify patients who could benefit from earlier treatment future research should focus on testing these models in different patient groups and integrating them into real-world clinical practices.

Keywords— Linear Regression, Support Vector Machine, Random Forest Regression, Logistic Regression, KNN.

# **Introduction:**

Heart failure stands as a primary contributor to illness and death and health-related expenditures globally being an increasingly pressing global health issue heart failure leads to social and economic burdens by requiring significant hospital admissions and long-term care this progressive disease progressively deprives sufferers of effective pumping action by the heart subsequent blood flow to the bodily organs and results in a variety of clinical symptoms including fatigue.

Heart failure has a varying prospect for prognosis of the patients they may continue to survive with the disease without suffering a significantly decline in their well-being but others deteriorate rapidly this variability in the course of patients emphasizes that the prediction models developed for an early possibility are far from being accurate and timely earlier prediction would alert doctors to identify at-risk patients thus preventing adverse consequences through appropriate early interventions involving tailor-made treatment measures resulting in better prognosis previously forecasting heart failure outcomes assessment was fundamentally based on clinical judgment and conventional risk factors such as coronary artery disease hypertension and diabetes among others such methods have inherent shortcomings owing to the exponential growth of digital health records and innovations in data science the healthcare sector is shifting toward harnessing the strength of machine learning and artificial intelligence to refine with demonstrating dependable accuracy across numerous conditions including heart failure.

# **RELATED WORK**

Ways like support vector machines support vector machine arbitrary timbers and combined styles similar as grade boosting algorithms have all proven extremely promising in terms of perfecting the delicacy of prognostications likewise deep literacy ways were particularly advanced similar as artificial neural networks and recurrent neural networks which have been used to capture complex patterns in clinical and temporal data similar as longitudinal health records feature creation was also espoused by using top element analysis principal component analysis to drop the number of features and identify the most influential predictors further perfecting model performance despite these advancements challenges remain for illustration the failure of quality- labeled data as well as lower robustness when varied case populations or healthcare settings suddenly change these studies give a starting point toward developing more accurate interpretable and scalable models for predicting the future of heart failure.

1. **METHODOLOGY:**

Improved data quality to achieve more effective design data cleaning random forest regression it is an important metric used in this project classifiers are tested based on several criteria such as accuracy precision decision tree regression logistic regression and rfr this will lead to the growth of e-commerce business and also volumes of products with improved website usability navigate to the product one wants to explore quickly this can be made possible through the technique called random forest regression in this case many different decision trees worked together for analyzing data by dividing them into subsets and generating accurate predictions finds patterns in each piece and combining predictions from all the trees by calculating their average improves the overall accuracy of the result.

## **Dataset and Preprocessing**

Data transformation is useful as it helps to clean and prepare the data to make it easier for machine learning models to make proper predictions the main practice is in collecting raw images of rice grains that could be very variable with regard to lighting conditions orientation and clarity of detail undergo resizing the images followed by normalization to guarantee all images have the same dimensions and scale and pixel values image enhancement techniques including contrast enhancing quality and minimizing costs unwanted interference are also applied to enhance image quality and reduce undesired variations color spaces are adjusted are adjusted to improve the process of collecting elements and histogram equalization is used when the illumination is not uniform techniques that aid in o reduce over fitting techniques such as rotating reversing cropping and other forms of data modification are normally used that increase the diversity of the dataset to which a model is being trained on in addition image segmentation can be applied to separate rice grains from interfering background noise generally speaking data preprocessing is basically essential for classifying rice as this ensures the use of clean well-processed input data that improves the feature collection and classification works towards improving the accuracy and stability of arranging rice grains in the machine learning model.

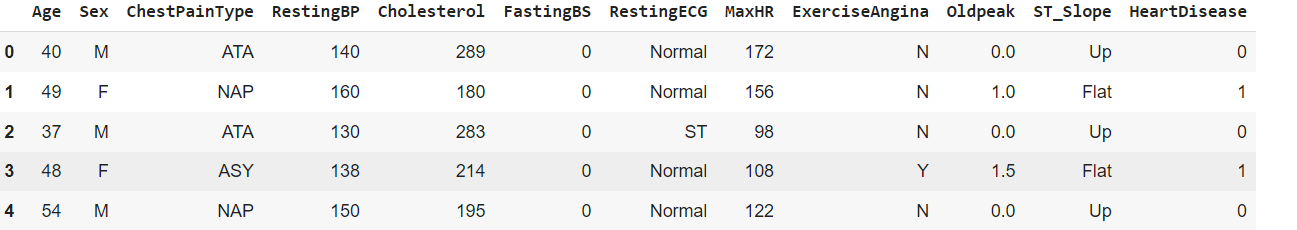


Figure 1. Dataset of Heart failure Prediction

## **B. Implementation**

Heart failure prediction is nearly related to prophetic analysis of heart failure issues using machine literacy ways for the first step choose a rich dataset that will contain features like age gender blood pressure cholesterol position heart rate ejection bit and medical history every preprocessing step ensures to handle missing values remove outliers scale numerical features and render categorical variables so that the data is ready for analysis the training of the model on preprocessed data involves algorithms like logistic retrogression random forest and support vector machines to establish patterns between features and threat of heart failure ensemble styles similar as grade boosting can be used for better delicacy by combining multiple weak learners the model is trained on the training dataset and estimated using criteria similar as accuracy precision recall and f1- score over the test set cross-validation ensures that it generalizes veritably well across unseen data once its estimated the model stylish stationed for predicting the furture purposes point significance analysis is carried out to interpret the model as far as which factors most significantly contribute to pitfalls concerning heart failure perpetration featuring data- driven perceptivity this perpetration offers data- driven perceptivity with practical operations for healthcare professionals and stakeholders in the medical field.

**a. Support Vector Machine:**

A Support Vector Machine (SVM) is a technique in machine learning that, based on analyzing data, can classify whether a person is at risk of heart failure or not. The basic idea behind this is very simple: SVM views the data (like age, blood pressure, heart rate, and other health indicators) and attempts to find the boundary dividing people into two categories-those at risk and those not. Steps to follow:

1. Input Data: SVM uses the data related to the patient's health, like some medical test results and lifestyle factors.

2. Training the Model: The SVM model learns from past cases (data where we already know if the patient developed heart failure or not) to figure out the best way to separate the two groups.

3. Determine the Boundary: SVM constructs a "decision boundary" in which the optimal dividing line or curve that best splits two sets is determined. It aims to make the boundary clear and as accurate as possible.

4. Prediction: When data is introduced into the model of a new patient, then SVM looks across where this data falls on the side of this boundary and further predicts that whether this patient is at risk or not for heart failure.

The thing that makes SVM powerful is the ability to handle complex data and find boundaries even when the groups are not clearly separable. For instance, if data points overlap, SVM employs a technique called the "kernel trick," where the data is mapped into a higher dimension in which the groups become easier to separate.

SVM, in fact, would help doctors better predict heart failure by identifying high-risk patients with early identification, even when the risk factors are complex or subtle. This then allows for timely interventions, better management of health, and improved outcomes for the patients.

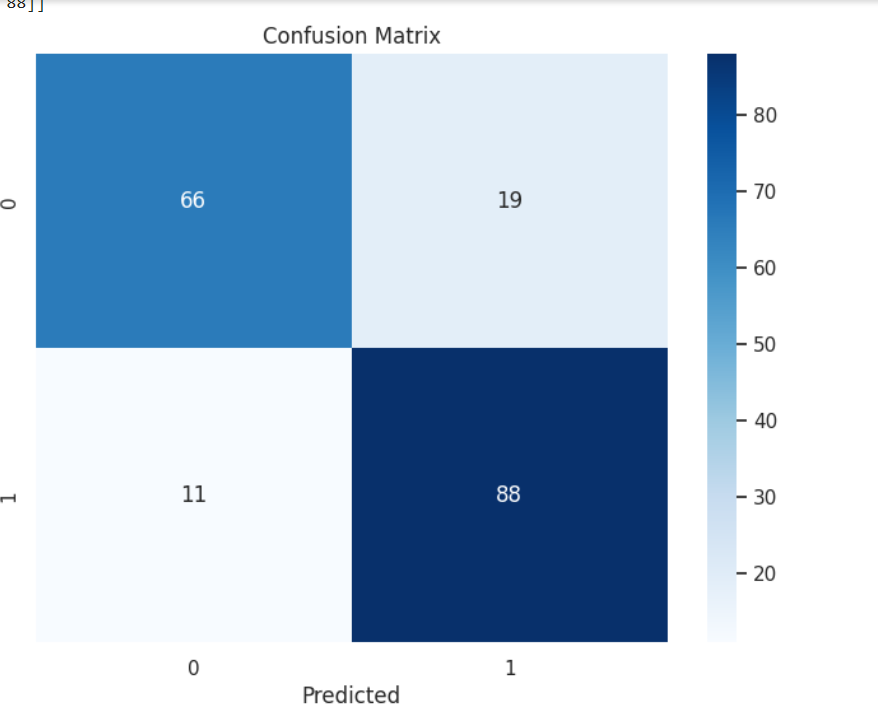


Figure 2: Confusion Matrix of Support Vector Machine

**b. Random Forest:**

Random Forest is a powerful machine learning algorithm that plays a key role in heart failure prediction. It works by creating multiple decision trees from the data and combining their results to make accurate predictions. Each decision tree acts like an expert that examines different factors, such as age, heart rate, blood pressure, and medical history, to determine whether a person is at risk of heart failure.

The strength of Random Forest lies in its ability to handle complex data and avoid over fitting, which happens when a model performs well on training data but poorly on new, unseen data. By averaging the predictions from multiple trees, Random Forest produces reliable results and reduces errors.

In heart failure prediction, this algorithm can process large datasets from medical records or wearable devices to identify patterns that indicate risk. For example, if several patients with similar symptoms and medical histories developed heart failure, Random Forest could learn from these cases and predict risk for new patients with similar conditions.

Overall, Random Forest is widely used because it is easy to use, interprets important features contributing to predictions, and delivers accurate results. This makes it a valuable tool for doctors to assess heart failure risk and recommend timely treatments or preventive measures.

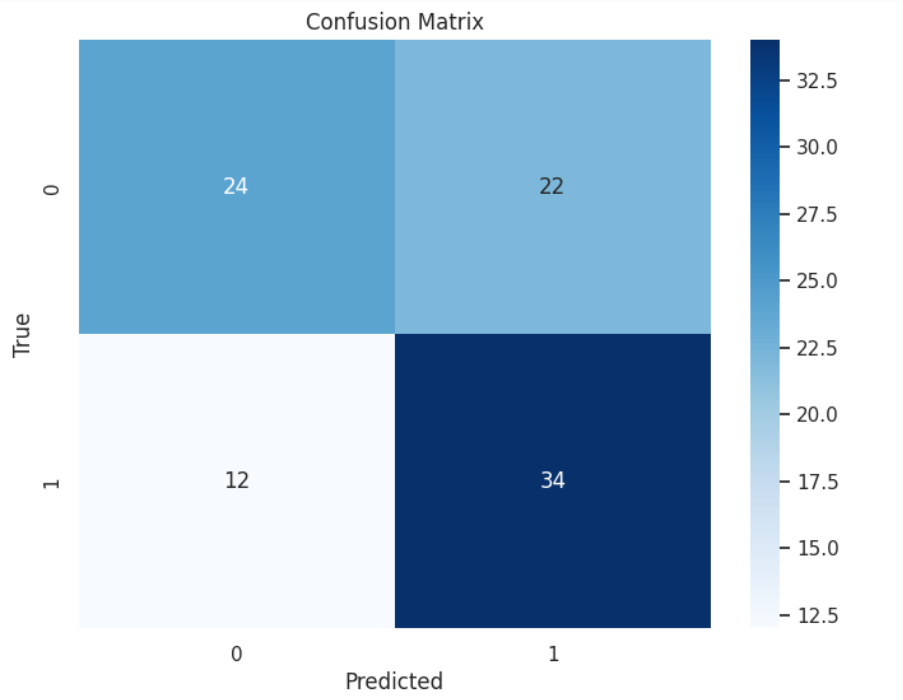


Figure 3: Confusion Matrix of Random Forest Algorithm

1. **IMPACT:**

Heart failure vaticination has a significant impact on individual health healthcare systems and society as a whole prophetic tools play a pivotal part in early opinion enabling timely interventions that can drastically ameliorate patient issues by relating individualities at threat healthcare providers can apply preventative measures reducing the liability of exigency hospitalizations beforehand vaticination also makes possible personalized curatives acclimatized to a cases specific requirements which maximizes the success of curatives and generally makes life much easier from a health care point of view heart failure vaticination cuts up massive cost- towards forestallment preventative care costs much lower than taking care of advanced heart failure and hospitals can optimize their resource use by targeting high- threat cases this approach does limit the number of tests and interventions applied to low- threat individualities therefore optimizing healthcare delivery it also contributes to public health strategies similar as targeted interventions in vulnerable populations and education juggernauts that call for healthy changes through better diet and increased exercise technological innovation is another area significantly affected in terms of heart failure vaticination artificial intelligence and machine literacy have developed advanced algorithms that number indeed more accurate threat estimation wearable and iot- enabled bias give nonstop health monitoring bringing prophetic perceptivity closer these developments are nt only perfecting findings but also making individualities visionary in their case operation for health enhancement vaticination of heart failure has psychosocial goods communicating to cases and families that they can have certainty over the health pitfalls and anxiety reduced when theyre uncertain when well- informed a case is suitable to take control and make better health choices and take preventative measures with confidence prophetic perceptivity explain studies about biology genetics and life factors behind heart failure leading to new medicine development and remedy eventually heart failure vaticination shifts focus from reactive to visionary care thus the end results of saving lives reducing healthcare burdens and fuelling medical invention and technological imagination represent a significant corner in advancing patient care as well as addressing one of the leading causes of morbidity and mortality worldwide.

1. **RESULT AND DISCUSSIONS:**

This dataset is structured to support the analysis and prediction of heart health conditions, particularly focusing on the risk of heart failure. The **Age** column represents the patient's age in years, while **Sex** is encoded as binary (e.g., 0 for female, 1 for male). **Chest Pain Type** categorizes the type of chest pain experienced by the patient, which is a key indicator of cardiac health. **Resting BP** (resting blood pressure) and **Cholesterol** provide critical cardiovascular metrics. **Fasting BS** (fasting blood sugar) is another binary variable that indicates whether fasting blood sugar is above 120 mg/dL, an important factor for assessing diabetes and its impact on heart health. **Resting ECG** records the electrocardiogram results at rest, which is crucial for detecting heart abnormalities.

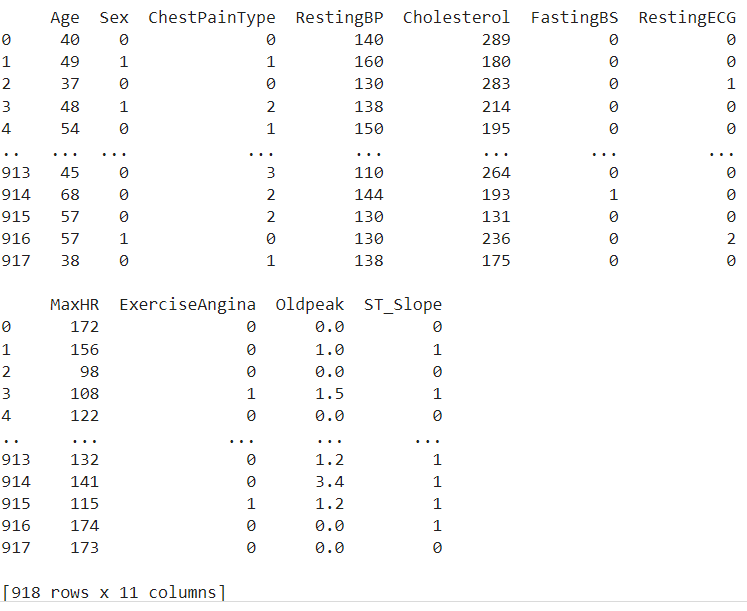
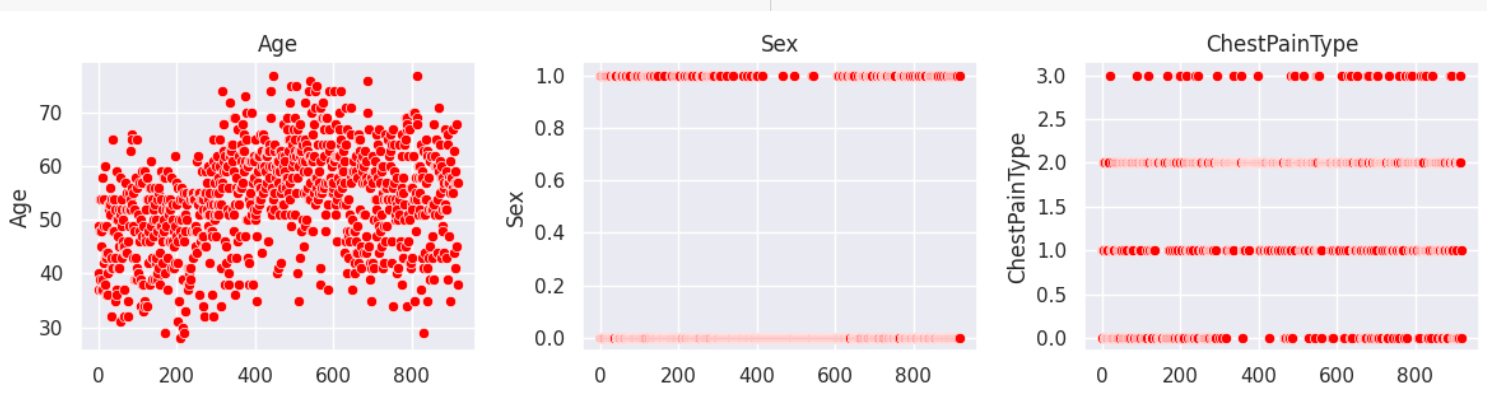
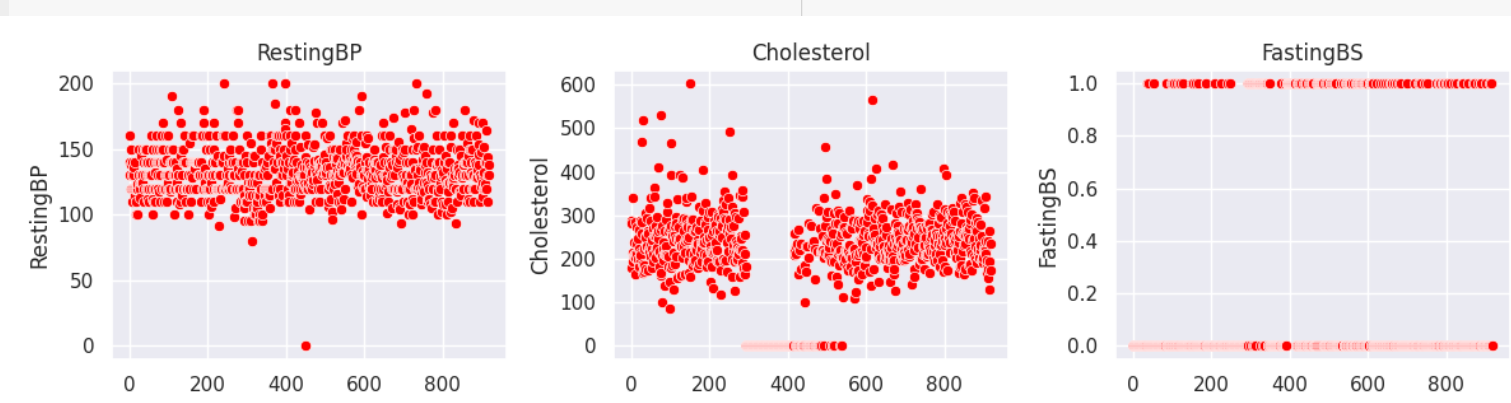
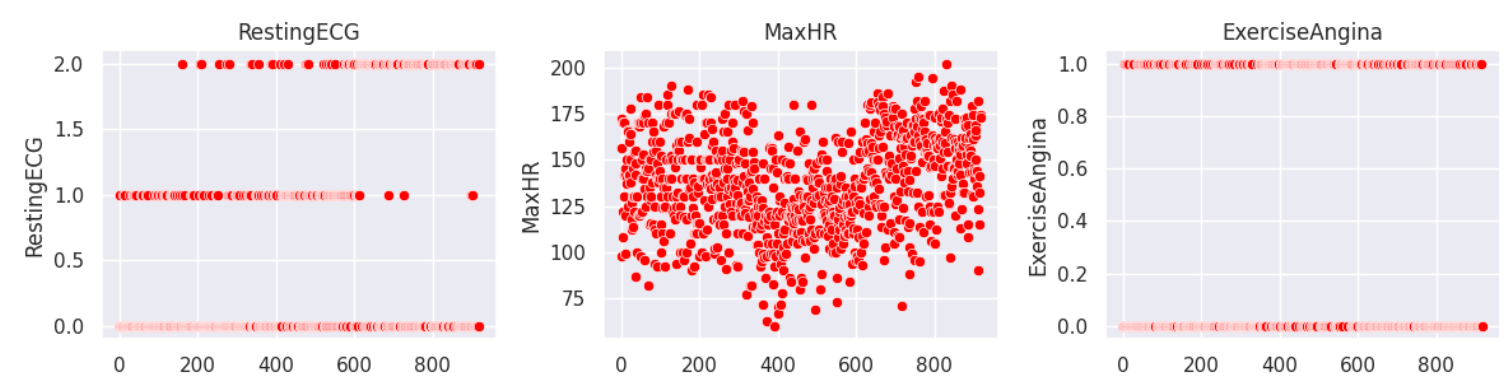


Figure 4: Content of Heart Disease Prediction

The dataset also includes **MaxHR** (maximum heart rate achieved during exercise) and **ExerciseAngina**, a binary feature that indicates if exercise-induced angina (chest pain) was reported. **Oldpeak** measures ST depression induced by exercise relative to rest, which helps in diagnosing ischemia. Finally, **ST\_Slope** represents the slope of the peak exercise ST segment, an essential marker for identifying heart stress. With a total of 918 entries, this dataset offers a comprehensive basis for machine learning applications in heart failure prediction and cardiovascular risk assessment. Proper preprocessing and feature analysis will be necessary to handle the categorical variables and scale numerical ones, ensuring accurate and interpretable model results.







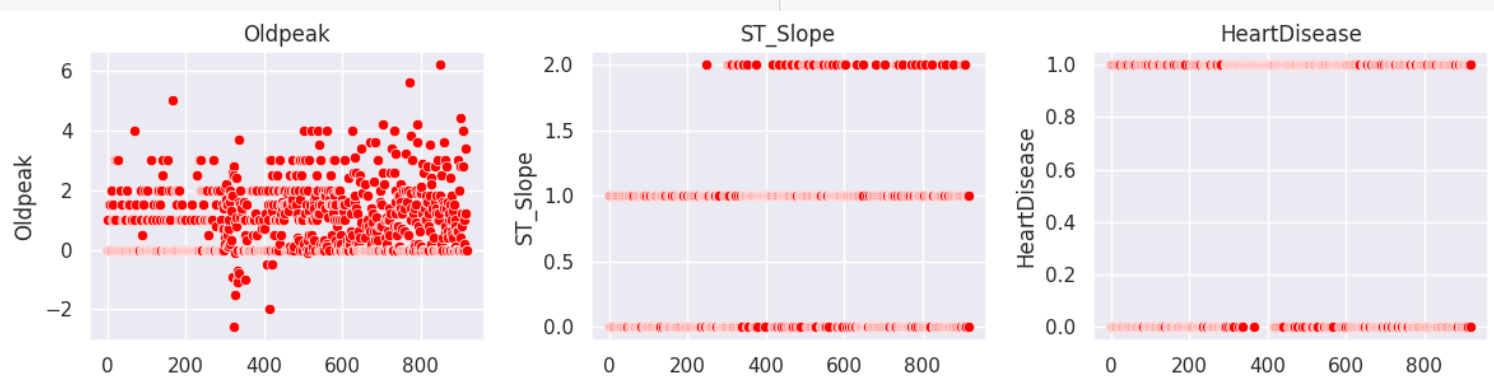


Figure 5: Graphs of each attribute

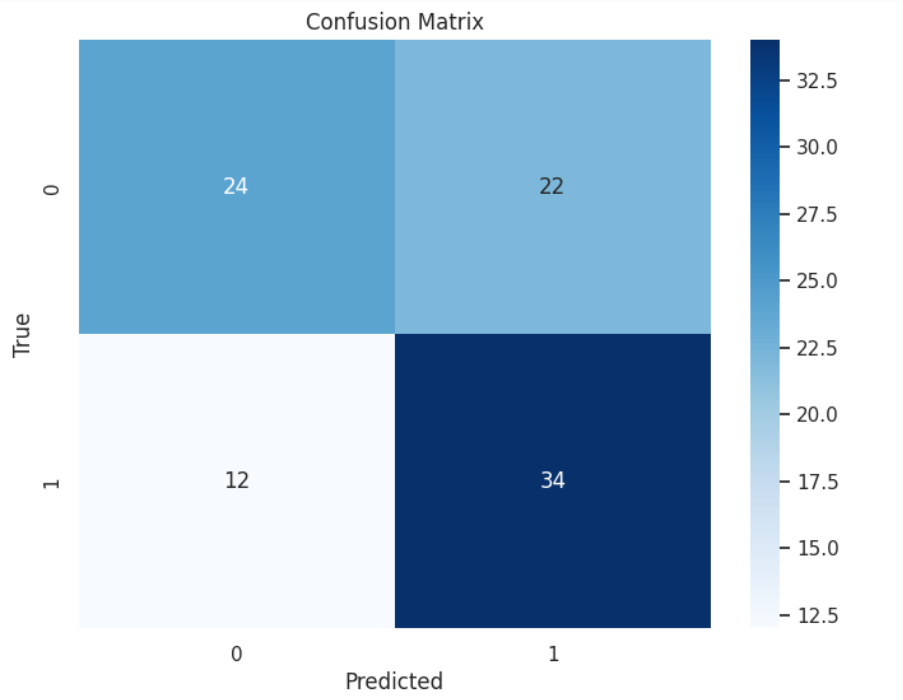


Figure 6: Confusion Matrix of KNN

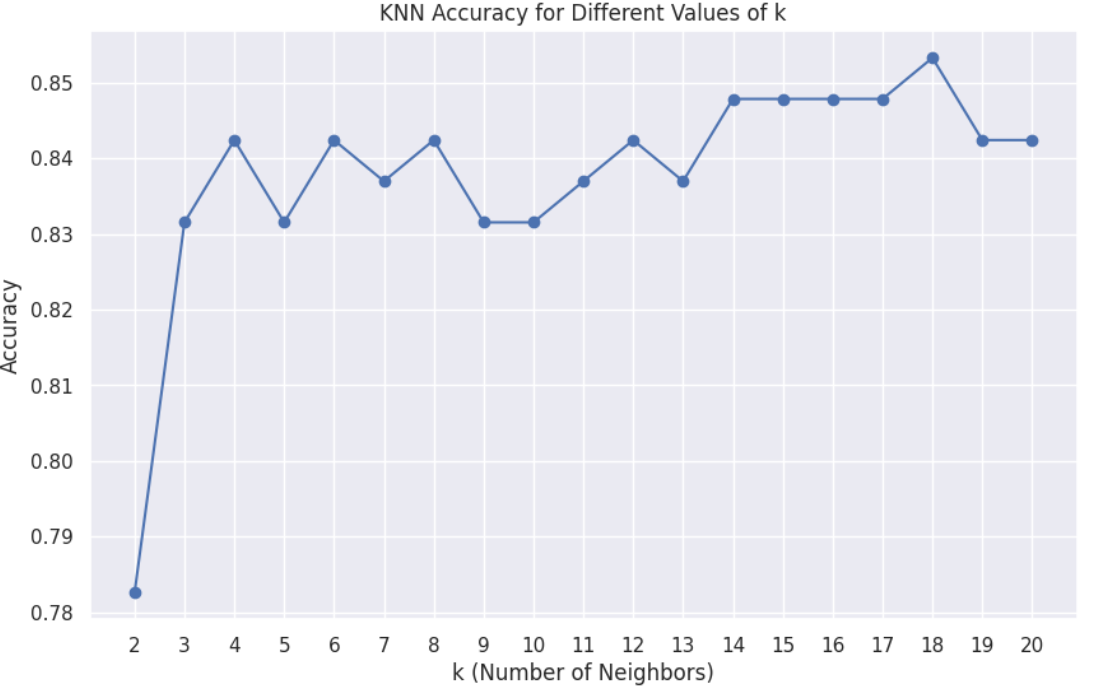


Figure 7: KNN Accuracy for different k values

The graph illustrates the accuracy of the K-Nearest Neighbors (KNN) algorithm for different values of kkk (the number of neighbors considered). The x-axis represents the values of kkk, ranging from 2 to 20, while the y-axis shows the corresponding accuracy scores. Initially, the accuracy is relatively low at k=2k = 2k=2, but it increases significantly and stabilizes as kkk rises. The highest accuracy, approximately 85%, is observed around k=18k = 18k=18, after which a slight drop is noticeable.

This pattern highlights the importance of selecting an optimal kkk value to balance bias and variance. A smaller kkk may lead to overfitting (high variance), while a larger kkk could result in underfitting (high bias). Based on this graph, k=18k = 18k=18 or values close to it appear to be the most suitable for achieving high accuracy with this dataset.

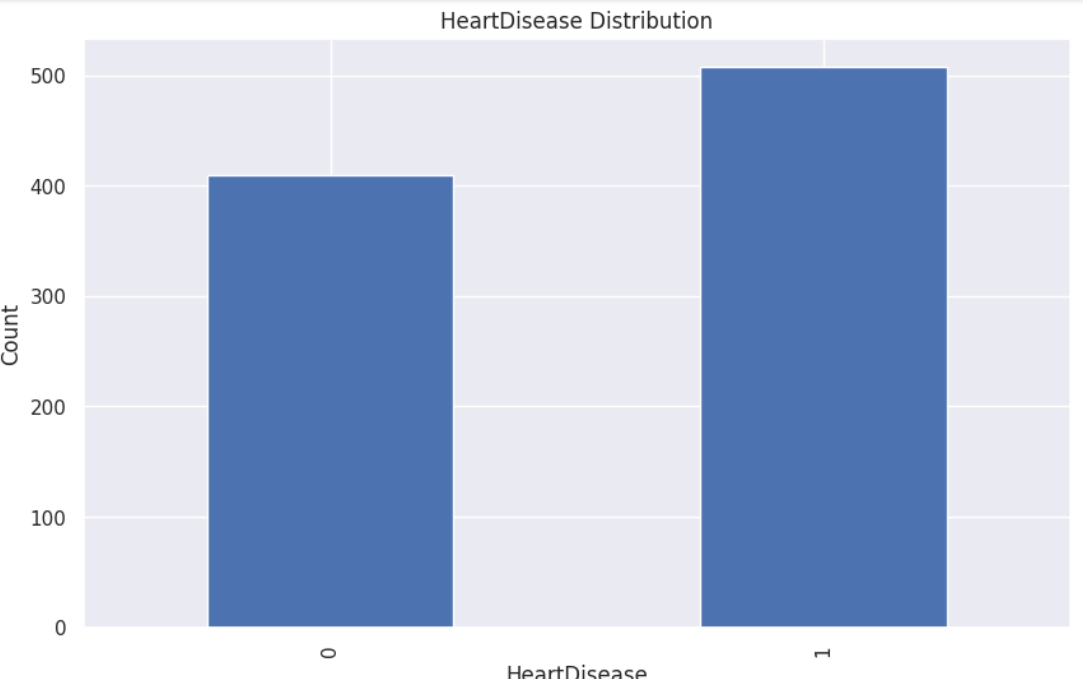


Figure 8: Heart Disease Distribution

The bar chart represents the distribution of heart disease in a dataset. The x-axis shows the presence or absence of heart disease, where "0" represents the absence and "1" represents the presence of heart disease. The y-axis shows the count of individuals in each category.

**Table 1.** Classification report of Heart failure Prediction

|  |  |  |
| --- | --- | --- |
| S. No | Machine Learning | Accuracy |
| 1 | Logistic regression | 0.875 |
| 2 | SVM | 0.84 |
| 3 | Random forest regression | 0.64 |
| 4 | Perceptron | 0.63 |

This table summarizes the performance of four different machine learning algorithms: Logistic Regression, SVM, Random Forest Regression, and Perceptron. The accuracy of each algorithm is shown, with Logistic Regression achieving the highest accuracy at 0.875, followed by SVM with 0.84. Random Forest Regression and Perceptron show lower accuracies, at 0.64 and 0.63 respectively. This data suggests that Logistic Regression and SVM are the most effective algorithms in this context.

1. **FUTURE SCOPE:**

The future will likely hold great prospects for heart failure prediction because of technology and medical research breakthroughs which are more particular, simple, and accessible predictive tools. Artificial intelligence and machine learning is believed to revolutionize the role by analyzing gigantic datasets - electronic health records, genetic information, and real-time data from wearable devices. Such technologies will allow for the development of extremely accurate predictive models that may otherwise remain too subtle to be observed in other methods. Such tools, as they evolve, will become much better at predicting the inception and progression of heart failure so that interventions can be achieved earlier on and better results attained.

The introduction of wearable and IOT devices in integrating continuous monitoring of heart failure would certainly empower patients and healthcare providers with good predictability of heart failure. Continuous monitoring by smart watches, fitness trackers, or even implantable devices will collect vital signs such as heart rate, blood pressure, and oxygen levels in real-time. Combined with AI algorithms, these will provide immediate risk appraisals with alerts for the patient and clinicians about potential issues before they appear as symptoms. It will give the individual capabilities to be prepared ahead of time and allows healthcare systems the ability to manage patients who are at risk most immediately.

In addition, genomics and personalized medicine advances will also be used in research to predict heart failure. The dearth of understanding about the basic genetic causes contributing to heart failure risk will improve when genetic sequencing becomes cheaper and widely accessible. Based on this new information, personalized predictive models will be designed for an individual's unique genetic makeup and lifestyle, enhancing the precision of assessments of risk and permitting focused prevention strategies, including specific lifestyle suggestions and pharmacological interventions.

Another integration of heart failure prediction with telemedicine and remote care is expected to increase access, especially from patients seen in rural or underserved areas. Such predictive insights can be easily transmitted to healthcare providers via health platforms, thus allowing for timely consultations and management plans without the need to visit in person too often. This will reduce healthcare disparities and improve outcomes among diverse populations.

Prediction of heart failure will likely be a foundation of preventative cardiology in the future. The reduction of the heart failure burden all over the world will, therefore, result in significantly improved patient quality of life and optimal health-care systems worldwide with the integration of technological innovation, personalized approaches, and improved accessibility.

# **Conclusion:**

Machine literacy algorithms are computational frameworks by which computers understand trends and make vaticinations or judgments without clearly being programmed these algorithms form the base of ultramodern artificial intelligence and have been decreasingly developed for wide operations in areas similar as image and voice understanding content manipulation discovery independent buses etc machine literacy can be a supervised or unsupervised if you have lower quantum of data and easily labelled data for training conclude for supervised literacy self-organising literacy would generally give better performance and results for large data sets if you have a huge data set fluently available go for deep literacy ways you also have learned underpinning literacy and deep underpinning learning you now know what neural networks are their operations and limitations.

##### **References:**

1. Soni J, Ansari U, Sharma D and Soni S 2011 Predictive data mining for medical diagnosis: an overview of heart disease prediction *International Journal of Computer Applications*
2. Rjeily, G. Badr, A. H. A. Hassani and E. Andres, "Predicting heart failure class using a sequence prediction algorithm," 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME), Beirut, Lebanon, 2017, pp. 1-4, doi: 10.1109/ICABME.2017.8167546.
3. Bashir S, Qamar U and Javed M Y An ensemble-based decision support framework for intelligent heart disease diagnosis *International Conference on Information Society (i-Society 2014)* (IEEE) 259-64
4. R. Rani, "Optimized Heart Failure Prediction using Support Vector Machine Algorithms," 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, pp. 1265-1268, doi: 10.1109/ICOSEC61587.2024.10722131.
5. V. Jain, "Heart Failure Prediction Using Machine Learning Algorithms with Cross Validations," 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 2023, pp. 1420-1423, doi: 10.1109/ICCPCT58313.2023.10245865.
6. B. Gnaneswar and M. R. E. Jebarani, "A review on prediction and diagnosis of heart failure," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-3, doi: 10.1109/ICIIECS.2017.8276033.
7. C. De Silva and P. Kumarawadu, "Performance Analysis of Machine Learning Classification Algorithms in the Case of Heart Failure Prediction," 2022 International Wireless Communications and Mobile Computing (IWCMC), Dubrovnik, Croatia, 2022, pp. 1160-1165, doi: 10.1109/IWCMC55113.2022.9824214.