**Journal of Neonatal Surgery**
*ISSN(Online): 2226-0439*
*Vol. 14, Issue 15s (2025)*
https://www.jneonatalsurg.com

OPEN ACCESS

# Evolution of Machine Learning to Predict Prognostic Outcomes in Patients Hospitalized with Congestive Heart Failure Using Random Forest Modelling Technique

**Dr. K. Sundravadivelu[1]**

[1]Assistant Professor, Department of Computer Science, School of Information Technology, Madurai Kamaraj University, Madurai, Tamil Nadu, India.

Email ID: svadivelu2021@gmail.com

Cite this paper as: Dr. K. Sundravadivelu, (2025) Evolution of Machine Learning to Predict Prognostic Outcomes in Patients Hospitalized with Congestive Heart Failure Using Random Forest Modelling Technique. *Journal of Neonatal Surgery*, 14 (15s), 1983-1993

**ABSTRACT**

One of the applications for which data mining tools are achieving success is disease diagnosis. Using Single Machine Learning Technique in the diagnosis of heart disease has been comprehensively investigated showing acceptable levels of accuracy Human heartbeat dynamics has been demonstrated to provide promising markers of Congestive Heart Failure (CHF). The main objective of this paper is to develop an Intelligent System using data analytics modelling technique and machine learning, namely, random forest SVC super vector classifier. Thus we propose to develop an application which can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate etc.. Heart failure (HF) is a complex clinical syndrome resulting from structural or functional impairment of ventricular filling or ejection of blood. Affecting over 64 million people globally, HF represents a significant burden on healthcare systems. This paper reviews current diagnostic strategies, classifications, treatment modalities, and emerging research trends in heart failure management, including pharmacological innovations and the use of artificial intelligence (AI) in predicting patient outcomes. The machine learning algorithm classifier has proven to be the most accurate and reliable algorithm and hence used in the proposed system.

***Keywords:*** *CHF, SVC, ECG, Naive Bayes, Random Forest, Artificial Intelligence, etc.,*

## 1. INTRODUCTION

Consider the self-driving Google car; cyber fraud detection, online recommendation engines like Facebook friend suggestions; Netflix showing you movies and shows you might like, and Amazon's "more items to consider" and "get yourself a little something" are all examples of applied machine learning. This will help you better understand the uses of machine learning [4].

[20] People in this fast-paced world want to live a very luxurious life, so they work like machines to earn a lot of money and have a comfortable life. Because of this, they forget to take care of themselves, and their eating habits change as a result. They become more stressed, have high blood pressure and high sugar levels at a young age, don't get enough sleep, eat what they get, and even don't care about the quality of the food. If they get sick, they take their own medicine because the heart is the most important organ in the human body, and if that organ is damaged, other important parts of the body are also affected [16]. Therefore it is very important for people to go for a heart disease diagnosis [6]. Over the past two decades, nonlinear dynamics of human cardiovascular oscillations has been recognized [13].

Heart failure is a major public health issue characterized by the heart's inability to pump sufficient blood to meet the body's metabolic demands [18]. [15] Despite advances in treatment, HF continues to be associated with high morbidity, mortality, and hospital readmission rates [21]. Heart failure is broadly divided into three categories: heart failure with reduced ejection fraction (HFrEF), heart failure with preserved ejection fraction (HFpEF), and heart failure with mid-range ejection fraction (HFmrEF). Each category has its own patho physiological characteristics and treatment options. All these examples echo the vital role machine learning has begun to take in today's data-rich world [3]. Machines can aid in filtering useful pieces of information that help in major advancements, and we are already seeing how this technology is being implemented in a wide variety of industries [17]. The process flow depicted here represents how machine learning works. With the constant evolution of the field, there has been a subsequent rise in the uses, demands, and importance of machine learning. All these are by products of applying machine learning to analyze huge volumes of data [15]. Trial and error has always characterized data analysis, but when data sets are large and diverse, this method becomes impossible. Because it offers clever alternatives to analyzing huge amounts of data, machine learning is touted as the solution to all of this chaos [7]. [11] By developing fast and efficient algorithms and data-driven models for real-time processing of data, machine learning is able to produce accurate results and analysis [12]. Machine learning tasks are classified into several broad categories. In supervised learning, an algorithm creates a mathematical model of a set of data that includes both the desired outputs and the inputs [19]. Supervised Machine Learning: Supervised learning is used in the majority of practical machine learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm for acquiring the mapping function Y = f(X) from the input to the output [8].

[1] The objective is to get so close to the mapping function that you can predict the output variables (Y) for new input data (x). 4 Techniques of Supervised Machine Learning algorithms include linear and logistic regression, multi-class classification, Decision Trees and support vector machines. The data used to train the algorithm must already be labelled with the correct answers for supervised learning [2]. For example, a classification algorithm will learn to identify animals after being trained on a dataset of images that are properly labelled with the species of the animal and some identifying characteristics. Supervised learning problems can be further grouped into Regression and Classification problems [5]. The development of a concise model capable of predicting the value of the dependent attribute based on the attribute variables is the objective of both problems [7]. The difference between the two tasks is the fact that the dependent attribute is numerical for regression and categorical for classification. A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”. Many different models can be used the simplest is the linear regression [14]. It tries to fit data with the best hyper-plane which goes through the points.

## 2. RELATED WORK

• Based on the previous research, it is safer to conclude that the KNN tends to produce a higher accuracy rate while comparing with other machine learning models. KNN has few significant drawbacks; the higher accuracy rate comes with the cost of a long training period.

• There hasn't been a previous study that shows which machine learning method is better at accurately determining the best treatment for heart disease patients. Practical use of healthcare database systems and knowledge discovery is difficult in heart disease diagnosis.

• The heartbeat parameter in the ECG signal is noticed, and mean heart rate, standard deviation, and frequency domains (e.g., LF and HF powers) are derived. It is difficult to select the "correct" K value, and it is computationally inefficient. However, this algorithm has the advantages of being memory-based, intuitive, and adaptable to a variety of proximity calculations. In order to find heart disease patients suitable treatments, we are utilizing machine learning methods. Apply single techniques to the heart disease diagnosis Kaggle dataset to apply machine learning algorithms to establish baseline accuracy in the diagnosis of heart disease patients. Classification model recognizes the characteristics of patients with heart disease. Algorithm - Random Forest Classifier, Gussian Naïve bayes, It shows the probability of each input attribute for the predictable state in Figure 1: Architecture of proposed model.

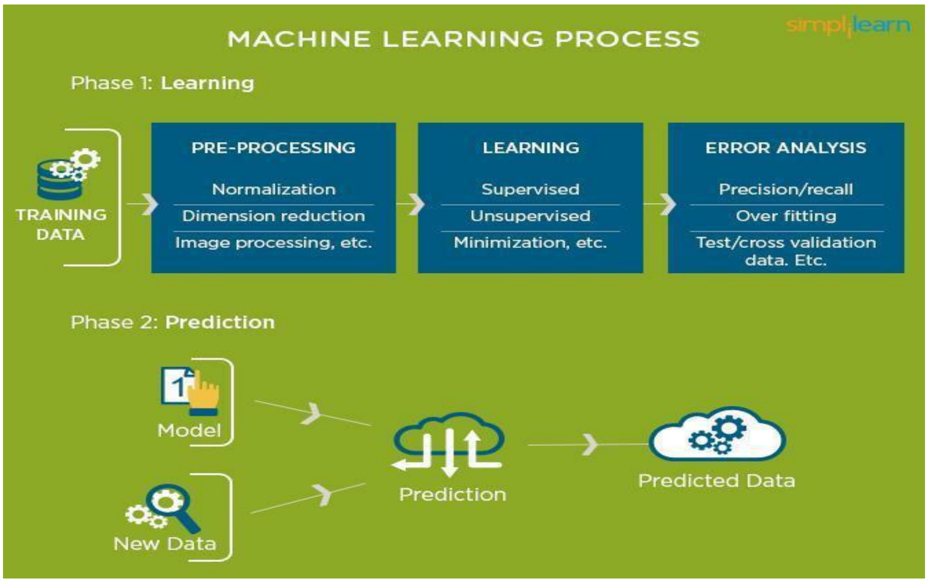


**Figure 1: Architecture of Proposed model**

## 3. EXPERIMENTAL RESULTS:

**3.1 Data Collection:** The dataset has 12 attributes:

Age: age in years, Sex: sex (1 = male; 0 = female), CP: chest pain type (Value 0: typical angina; Value 1: atypical angina;

Value 2: non-anginal pain; Value 3: asymptomatic) , trestbps: resting blood pressure in mm Hg on admission to the hospital, chol: serum cholesterol in mg/dl, fibs: fasting blood sugar > 120 mg/dl (1 = true; 0 = false), restecg: resting electrocardiographic results (Value 0: normal; Value 1: having ST-T wave abnormality; Value 2: probable or definite left ventricular hypertrophy), thalach: maximum heart rate achieved, exang: exercise-induced angina (1 = yes; 0 = no), oldpeak: ST depression induced by exercise relative to rest, slope: the slope of the peak exercise ST segment (Value 0: upsloping; Value 1: flat; Value 2: downsloping), target: heart disease (1 = no, 2 = yes).

**3.2 Data Pre-Processing**

Preparing raw data for use in a machine learning model is the goal of data pre-processing. It is the first and crucial step while creating a machine learning model.  When starting a machine learning project, clean and formatted data are not always available. Additionally, data must be formatted and cleaned before any operation can be performed on it [9]. To implement system performance with a total of TB of dataset, the Intel Core (TM) i7 with Programming with Python version 4 and Windows 11 operating systems, Flask with web application, and 8 GB of RAM capacity were utilized.

Real-world data generally contains noise, missing values, and maybe an unusable format, which cannot be directly used for machine learning models.  Data pre-processing is a required task for cleaning the data and making it suitable for a machine learning model, which also increases the accuracy and efficiency of a machine learning model.

**3.3 Random Forest Algorithm:**

Step 1: In the Random Forest Algorithm is to select a subset of data points and features for each decision tree in the Random Forest Model. Simply put, n random records and m features are taken from the data set having k number of records.

Step 2: For each sample, individual decision trees are created.

Step 3: An output will be produced by each decision tree.

Step 4: For classification and regression, respectively, the final output is taken into consideration using majority voting or averaging.Random forests offer the highest level of precision.

The random forest method can also handle large amounts of data with thousands of variables. When a class in the data occurs more frequently than other classes, it can automatically balance the set.

• Low generalization error, easy to implement, works with a wide range of classifiers, no parameters to adjust.  Special attention is needed to data, as this algorithm is sensitive to outliers.

• It aids in the reduction of bias and variance in the machine learning ensemble depicted in figure 2. They automate trading to generate profits at a frequency impossible for a human trader.  It helps in the conversion of weak learners into strong learners by combining N number of learners.

**3.4. Epidemiology**

Heart failure prevalence increases with age and is more common in men than women for HFrEF, while HFpEF is more prevalent among elderly women. In the United States alone, HF affects over 6 million adults, with projections estimating a 46% increase in prevalence by 2030. The direct and indirect costs of HF are estimated to exceed $30 billion annually. Heart Disease in human ratio of 55% yes and 45% no in figure 2.

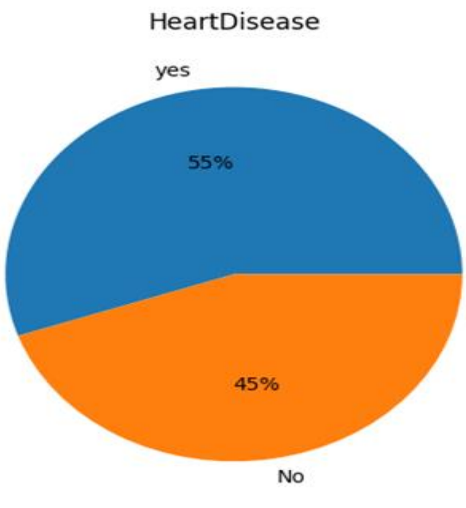


**Figure 2: Heart disease in human**

Chest pain is a common symptom among patients with heart failure (HF), but its prevalence varies depending on the type of chest pain considered in figure 3, Heart failure in frequency according to chest pain type. A study involving 5,786 patients with heart failure due to left ventricular systolic dysfunction found that:

### 3.5 Angina Chest Pain:

**No Angina in the Previous Week:** Angina was reported by 73% of patients with ischemic heart disease (IHD) and 84% of patients without IHD. Minimal Chest Pain: 79% of those with IHD and 82% without IHD experienced little to no chest pain at rest and during exertion. These findings suggest that while pain is commonly reported among HF patients, it is unlikely to be due to angina, even in those with underlying coronary heart disease. Additionally, a study comparing pain characteristics in ambulatory patients with HF and non-HF conditions revealed:

### 3.6 Chest Pain:

Compared to non-HF patients, approximately 33% of HF patients reported experiencing chest pain. This underscores the importance of distinguishing chest pain related to HF from other potential causes. It is essential to keep in mind that the severity of heart failure, the presence of ischemic heart disease, and individual pain thresholds can all have an impact on how HF patients experience chest pain. To accurately assess and treat chest pain in heart failure patients, it is necessary for healthcare professionals to conduct a comprehensive evaluation.

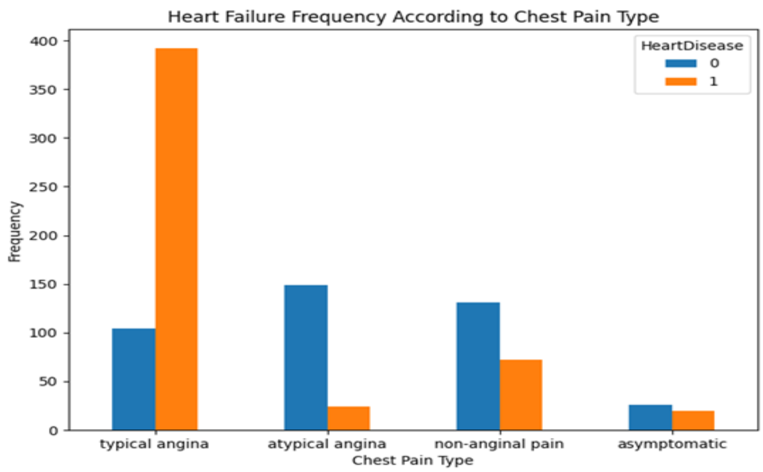


**Figure 3: Heart failure frequency according to chest pain type**

### 3.7 Prevalence and Incidence

- **Young to Middle-Aged Adults:** Men generally have a higher prevalence of heart failure, especially in their 60s and 70s. **Older Adults:** Women tend to develop heart failure at older ages, particularly after 80, due to factors like longer life expectancy and the increased prevalence of heart failure with preserved ejection fraction (HFpEF) in Figure 4: Heart failures according to gender.

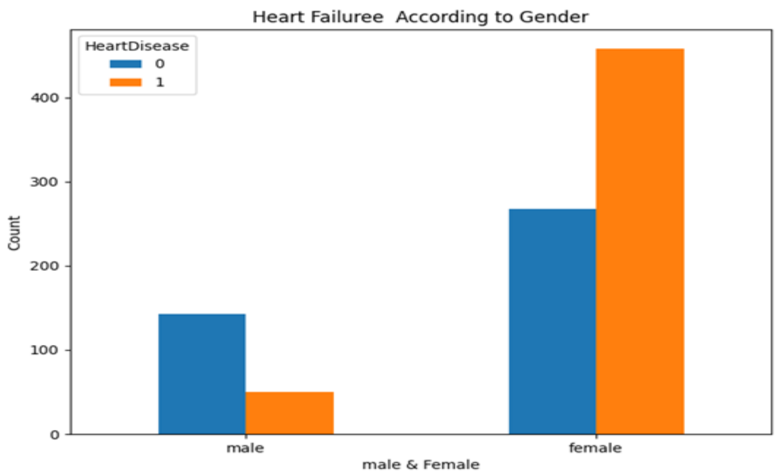


**Figure 4: Heart Failures according to Gender**

**Table 1: Gender Ratio**

| Type of Heart Failure | Male (%) | Female (%) |
|---|---|---|
| HFrEF | ~60-70% | ~30-40% |
| HFpEF | ~30-40% | ~60-70% |

**3.8 Gender Ratio:**

- In **HFrEF**: **Male-dominated**, about **60–70% male**. In **HFpEF**: **Female-dominated**, around **60–70% female**. **Women** tend to have **better survival rates** than men with heart failure in figure 4:.However, they may experience **worse quality of life** and **more symptoms** (like fatigue, breathlessness). **Men** have a **higher mortality risk**, especially with HFrEF in table 1.

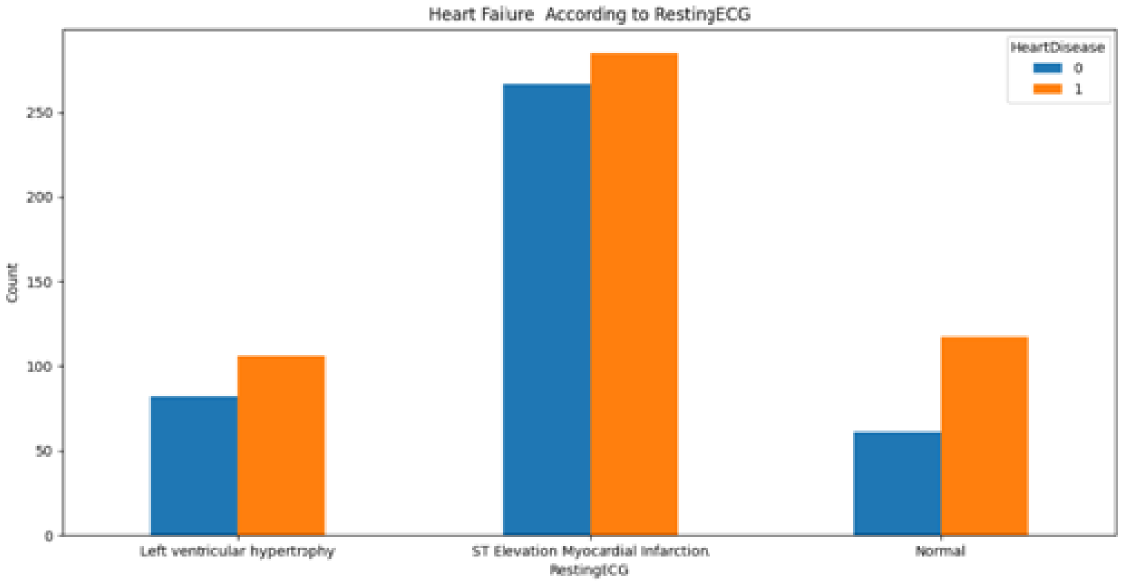


**Figure 5: Heart failure according to Resting ECG**

Resting electrocardiograms (ECGs) can reveal various patterns in patients with heart failure, including those with left ventricular hypertrophy (LVH), ST-elevation myocardial infarction (STEMI), or a normal ECG in Figure 5.

**3.8.1.** Left Ventricular Hypertrophy (LVH)

The left ventricle thickens in response to increased workload, which is frequently caused by hypertension or aortic stenosis, leading to LVH.

**Voltage Criteria:** An elevated QRS amplitude, such as a S wave in V1 and a R wave in V5 greater than 35 mm (Sokolow-Lyon criteria), can indicate LVH.

**Non-Voltage Criteria:** ST segment depression and T wave inversion in lateral leads (V5, V6, I, aVL), indicating a "strain" pattern. Additional findings include prominent U waves, a left axis deviation, and a prolonged R wave peak time in V5 or V6 (>50 ms).

**3.8.2.** ST-Elevation Myocardial Infarction (STEMI)

STEMI is characterized by acute myocardial injury and is typically identified by:

- **ST Segment Elevation**: Elevations in two or more adjacent leads in the ST segment are an indication of acute injury.
- **Reciprocal Changes:** ST segment depression and T wave inversion in leads opposite to the infarct.
- **Pathological Q Waves:** Developing hours to days after infarction, indicating myocardial necrosis. It's important to differentiate STEMI from other conditions that can cause ST elevation, such as LVH or pericarditis.

**3.8.3.** Normal ECG

A normal ECG in a patient with heart failure may suggest:

- **Early Stage Heart Failure**: Where structural or electrical changes have not yet manifested.
- **Non-Cardiac Etiologies**: Heart failure symptoms may be due to non-cardiac causes, such as pulmonary or renal conditions.
- **Normal Variant**: Some individuals may have normal ECGs despite underlying cardiac pathology.

**Exercise angina** refers to **chest pain or discomfort** that occurs during **physical exertion** (like walking uphill, climbing stairs, or during a treadmill stress test). It is usually caused by **insufficient blood flow to the heart muscle**, especially when the heart's demand for oxygen increases during exertion in figure 5.

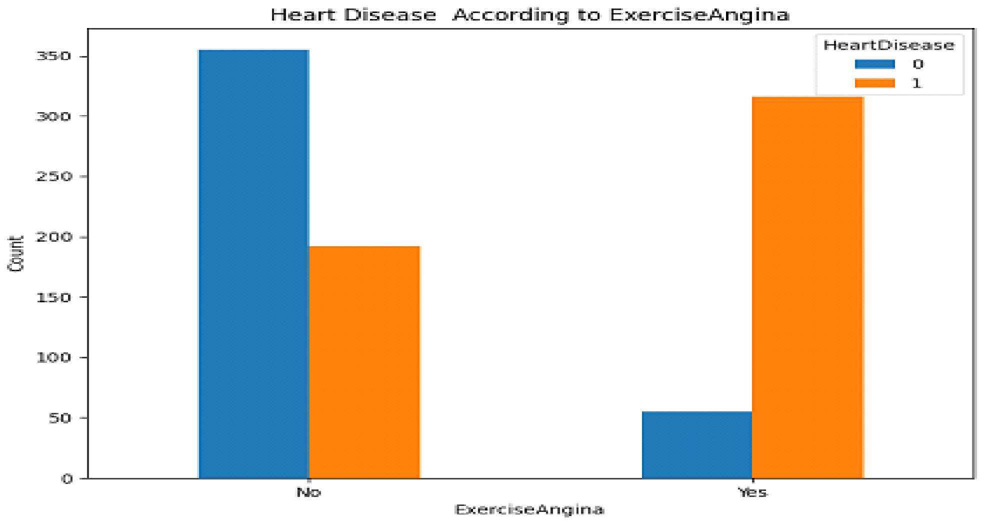


**Figure 6: Heart Disease according to ExerciseAngina**

**3.9 Significance in Diagnosing Heart Disease:**

**Positive Exercise Angina is strongly correlated with coronary artery disease (CAD).** Combined with abnormal ECG findings (e.g., ST-segment changes), it significantly increases the likelihood of a **clinically relevant blockage.** Patients with exercise angina are often referred by Coronary angiography, Cardiac CT or MRI, Further cardiology evaluation in figure 6.

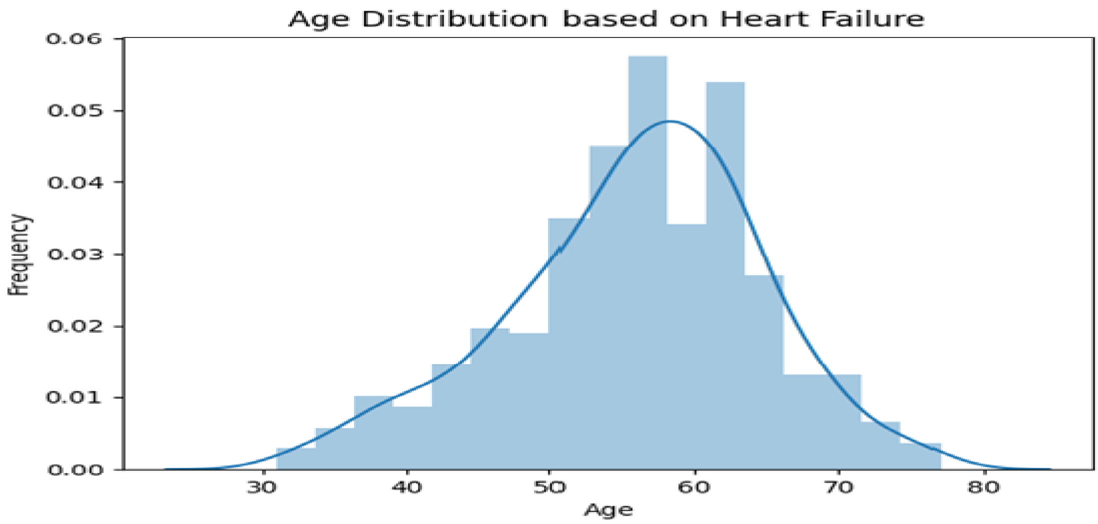


**Figure 7: Age distribution based on Heart Failure**

**Incidence Trends**

- **Framingham Heart Study**: Annual incidence of heart failure increases with age, from 0.4% in individuals aged 45–54 years to 4.0% in men aged 85–94 years.
- **AGES-Reykjavík Study**: Prevalence of heart failure in individuals aged ≥85 years was
- 21%, indicating a significant rise in older populations in Table 2.

**Table 2: Age distribution based on Heart Failure**

| **Age Group (Years)** | **Prevalence (%)** | **Notes** |
|---|---|---|
| 18 – 44 | 0.05 | Rare in this group; typically associated with congenital or genetic conditions. |
| 45 – 54 | 0.4 | Early onset may be linked to hypertension or lifestyle factors. |
| 55 – 64 | 9.9 | Increasing prevalence with age; early signs may be subtle. |
| 65 – 74 | 3.3 | Noticeable rise; often associated with hypertension and coronary artery disease |
| 75 – 84 | 9.0 | Significant increase; multiple comorbidities common. |
| ≥85 | 21 | Often due to cumulative effects of aging and Chronic conditions |

## 4. METRICS

We used the following evaluation metrics to assess the execution of the proposed Random Forest model:

### 4.1 Accuracy:

One of the performance measures is accuracy, which is simply the ratio of correctly predicted observations to total observations.

**Accuracy = TP+TN/TP+FP+FN+TN**

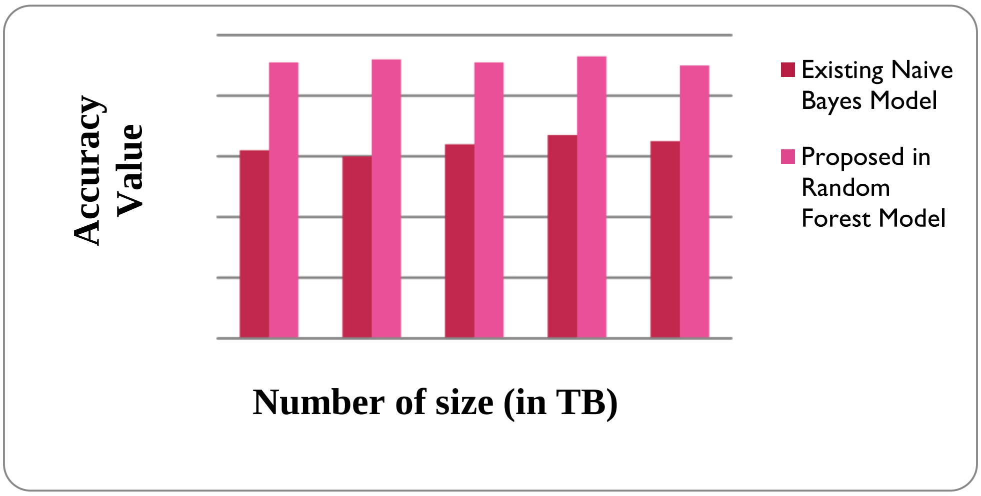


**Figure 8: Comparison of Existing Naive Bayes Model and Proposed Random Forest model**

Mathematically, accuracy is defined as:

It is evaluated by the percentage of items which are predicted exactly and to the absolute number of expectations presented. This evaluates images and their corresponding terms usage.

**Accuracy =TP +TN / TP +FP +FN +TN**

Figure 8 presents the accuracy of two models Naive Bayes and Random Forest Model. The exactness could be estimated by the proportion of genuine positive and genuine negative in the dataset. The Random Forest model achieves an average accuracy score of 91% whereas the Naive Bayes model provides an average accuracy score of 64% as illustrated in Figure 8. In Figure 8 X-axis denotes Number of documents and Y-axis represents the accuracy value. The performance results show Random Forest model produces greater accuracy than Naive Bayes model.

**4.2 Precision:**

For a classifier with a given class, precision is the ratio of true positives to false positives, which should be added together. The correctly predicted topic for the given document in the data source is True Positives (TP), which indicates that both the actual class and the predicted class have values of yes. False Positives (FP)—When actual class is no and predicted class is yes Figure. 9. Comparison of Existing Naive Bayes Model and Proposed Random Forest model.

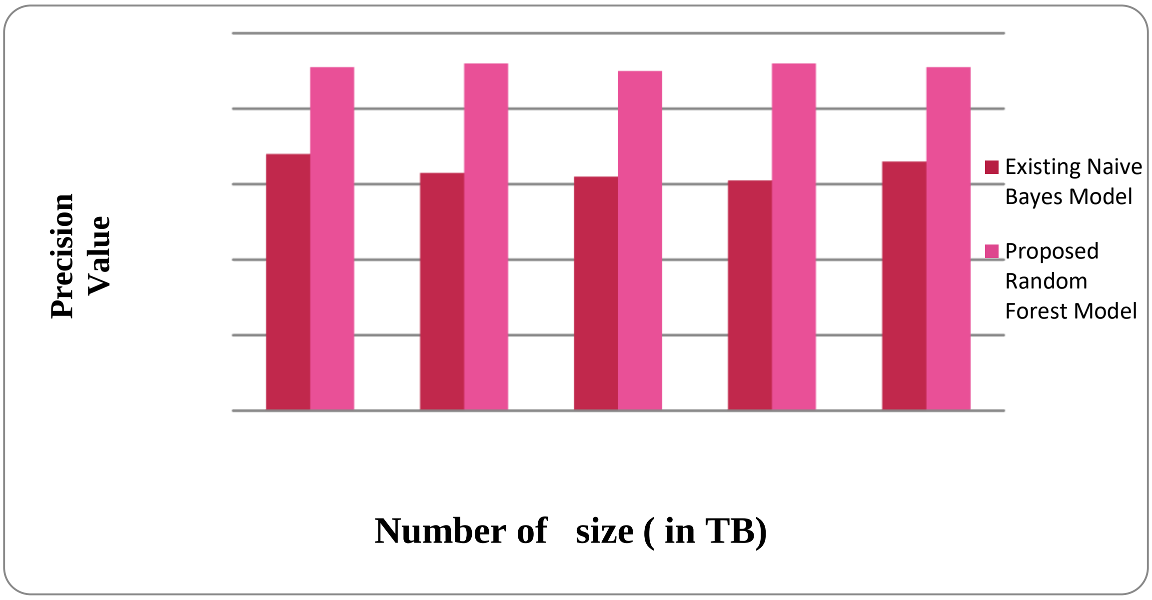


**Figure 9: Comparison of Existing Naive Bayes Model and Proposed Random Forest model**

Mathematically,

**Precision=TP/TP+FP**

This measurement is vital when the expense of misleading up-sides is high (e.g., predicting an irrelevant object as an object of interest in image classification). The genuine positive proportion of the model is evaluated by the positive prediction which resides in the positive class. In Figure 9, The Accidental Forest model achieved a precision score of 91%, whereas the Naive Bayes model derived a precision score of 64%. This indicates that the developed framework defeated the performance of the compared Naive Bayes framework. In order to produce accurate results, the precision value is determined with respect to the data of different sizes which range from 1TB to 5TB.

**4.3 Recall:**

The ratio of True Positive Spam Positive words identified from the data set by Clustering, which serves as the base data source for the LIME classification technique, is used to calculate recall.

Mathematically, recall is defined as:

**Recall=TP/TP+FN**

When the cost of false negatives is high, recall is crucial. The measure used to indicate the minimum number of derived terms in the corpus is recall. The term TP is used to represents positive documents whereas the term FP represents the false documents and FN is used to indicate the failed documents. The recall rate could be measured in terms of the percentage of relevant cases that are retrieved. In Figure 10 the number of size is indicated in X-axis and the recall score is shown in the

Y-axis.

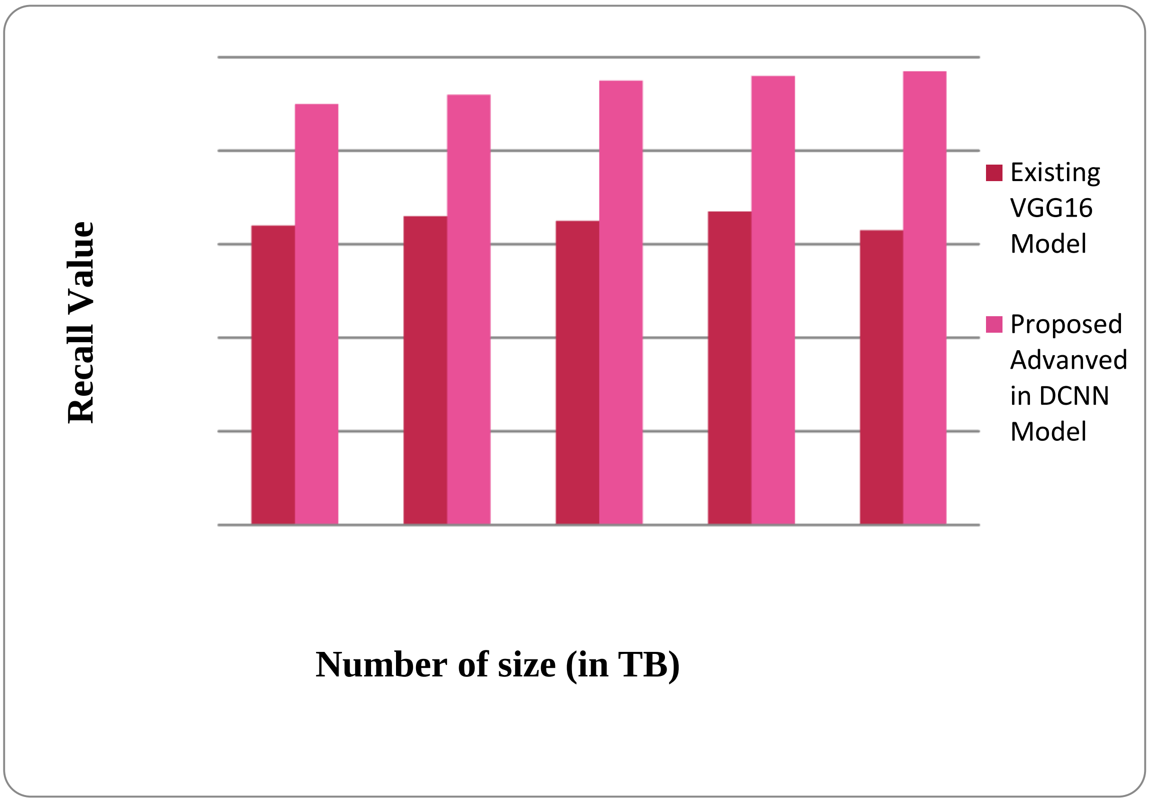


**Figure 10: Comparison of Existing Naive Bayes Model and Proposed Random Forest model**

The existing Naive Bayes model has gotten a Review score of 63%. The proposed Random Forest model has obtained a Recall score of 93% which is a best result than the baseline framework.

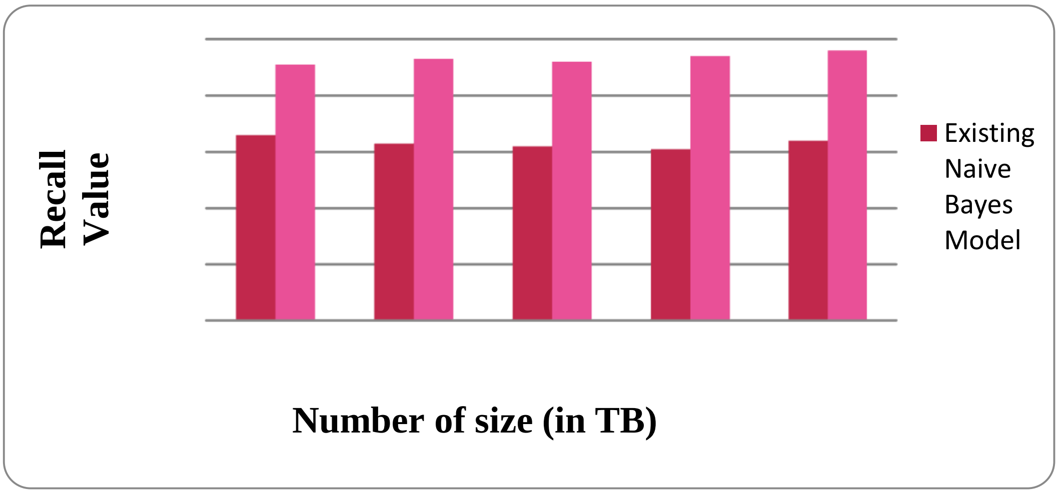


**Figure 10: Comparison of Existing Naive Bayes Model and Proposed Random Forest model**

The existing Naive Bayes model has gotten a Review score of 63%. The proposed Random Forest model has obtained a Recall score of 93% which is a best result than the baseline framework.

**4.4 F1 Measure:**

The weighted average of Precision and Recall determines the F1 score. F1 Score = 2*(Recall * Precision) / (Recall + Precision), Based on this Graph and is plotted as below in Figure 11. Comparison of Existing Naive Bayes Model and Proposed Random Forest model.

**F1-Score=2 × (Precision × Recall) / (Precision + Recall)**

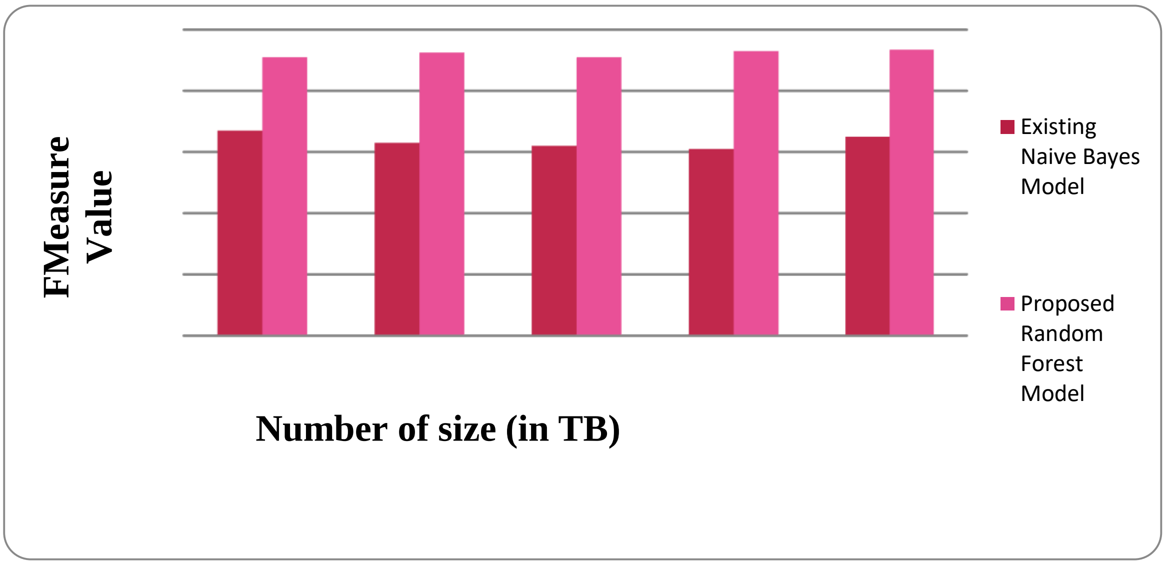


**Figure 11: Comparison of Existing Naive Bayes Model and Proposed Random Forest model**

In the case of the comparative Existing Naive bayes model, the F-Measure score is 63%, while the proposed Random forest model has a higher F-Measure score of 92% which indicates a significant improvement. The data includes in the measuring of F- Measure value has of varying sizes ranging from 1 TB to 5 TB. The result shows better improvement over the existing model. The output of the F-measure shows our proposed model has better performance than the comparative model.

## 5. CONCLUSION

Predicting the presence or absence of heart diseases, locomotor disorders, and other diseases can all benefit from machine learning. Such information, if predicted well in advance, can provide important insights to doctors who can then adapt their diagnosis and treatment on a per-patient basis. A classifier in machine learning is an algorithm that automatically orders or categorizes data into one or more of a set of "classes." One of the most common examples is a heart disease classifier that scans emails to filter them by class label: The patient is not likely to have heart disease, or the patient is likely to have heart disease. With proper data processing, the paper analyzes the heart disease patient dataset. Then, 2 models were trained and tested with maximum scores as follows: Naïve Bayes Accuracy: 0.636; Random Forest Accuracy: 0.914. In the future, look into using more types of data beyond what's currently used, like genetic information or data from wearable devices (like fitness trackers). This could give a more complete picture of a patient's health.

## REFERENCES


[1] Aggarwal, A., & Toshniwal, D. (2019). Frequent pattern mining on time and location aware air quality data. IEEE Access, 7, 98921-98933.

[2] Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., ... & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. Review of Research in Education, 44(1), 130-160.

[3] Ghojogh, B., Samad, M. N., Mashhadi, S. A., Kapoor, T., Ali, W., Karray, F., & Crowley, M. (2019). "Feature selection and feature extraction in pattern analysis: A literature review", arXiv preprint arXiv:1905-02845.

[4] Gurcan, F., &Cagiltay, N. E. (2019) "Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling", IEEE Access, 7, 82541-82552.

[5] Jankowski, M. (2018, June), "Boost multi-class sLDA model for text classification". In International Conference on Artificial Intelligence and Soft Computing (pp. 633-644). Springer, Cham.

[6] Ko, A., & Gillani, S. (2020), "A research review and taxonomy development for decision support and business analytics using semantic text mining", International Journal of Information Technology & Decision Making, 19(01), 97-126.

[7] Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021), "Applications of text mining in services management: A systematic literature review", International Journal of Information Management Data Insights, 1(1), 100008.

[8] Li, Y., & Zhang, D. (2020, September), "Hadoop-Based University Ideological and Political Big Data Platform

Design and Behavior Pattern Mining", In 2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI) (pp. 47-51), IEEE.

[9] Min, F., Zhang, Z. H., Zhai, W. J., & Shen, R. P. (2020) "Frequent pattern discovery with tri-partition alphabets", Information Sciences, 507, 715-732.

[10] Saquicela, V., Baculima, F., Orellana, G., Piedra, N., Orellana, M., & Espinoza, M. (2018). Similarity Detection among Academic Contents through Semantic Technologies and Text Mining. In IWSW (pp. 1-12).

[11] Sundravadivelu, K, Senthilvel, P.G.,Duraimutharasan, N.,Esther T, H.R.,Kumar. K, R." Extensive Analysis of IoT Assisted Fake Currency Detection using Novel Learning Scheme", ICAISS 2023, pp. 1469–1477, 10.1109/ICAISS58487.2023.10250560.

[12] Sundravadivelu,K, Senthilvel, P.G.,Thirupurasundari, D.R.,Rajesh Kumar, K.,Palani, H.K., "Automated Drone Based Imaging Systems For Plant Health Monitoring Using Deep LearningTechniques",2023, Intelligent Computing and Control for Engineering and Business Systems, ICCEBS 2023, IEEE , 10.1109/ICCEBS58601.2023.10449190.

[13] Sundravadivelu, K., Thangaraj, M., Gnanambal, S. (2022). An extensive work on comparing sentiment patterns in twitter archives between two persons. International Journal of Health Sciences,6(S7), 5170- 5180. https://doi.org/10.53730/ijhs.v6nS7.13104.

[14] Sundravadivelu. K, M. Thangaraj, "Analyzing Educational Tweets using LDA Model", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), Volume 10, Issue 4, ISSN:2147-6799, PP. 100- 104, Dec. 2022.

[15] Sundravadivelu. K, M. Thangaraj, ―A Novel Approach for Discovering the Patterns by using PDBD Model in Big Data‖, Journal of Computer Science, (Science Publications), Volume 18 Issues 5, DOI: 10.3844 / jcssp.2022.382.395, ISSN: 1549-3636, pp.382-395, May 2022.

[16] Sundravadivelu. K, Suraj Rajesh Karpe, Harish V Mekali, Shital Nalgirkar, K. Abdul Rasak, Dr. V S Narayana Tinnalur, Information Theory and Coding: Techniques for Error Control and Data Compression, Journal of Electrical Systems 20-10s (2024): 5665-5674.

[17] Sundravadivelu. K, Thangaraj. M,"Mining effective patterns from text data - a survey", International Journal of Scientific and Technology Research, Volume 9, Issue 1, Pages 1930 – 1934, January 2020.

[18] Sung, S. F., Lin, C. Y., & Hu, Y. H. (2020). EMR-based phenotyping of ischemic stroke using supervised machine learning and text mining techniques. IEEE Journal of Biomedical and Health Informatics, 24(10), 2922-2931.

[19] Thangaraj.M, & Sundravadivelu.K, "Mining effective patterns from text data-a survey", International Journal of Scientific & Technology Research, ISSN-10: 2277-86161930 IJSTR, 2020.

[20] S. Shalev-Shwartz and S. Ben-David, "Understanding machine learning," From Theory to Algorithms, Cambridge University Press, Cambridge, UK, 2020.View at: Google Scholar

[21] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning," Data Mining, Inference, and Prediction, Springer, Cham, Switzerland, 2020.View at: Google Scholar