SURVIVAL ANALYSIS OF HEART FAILURE PATIENTS USING MACHINE LEARNING

BY

Al-Mamun
ID: 191-15-2576

AND

Riad Shalahin Leon
ID: 191-15-2613

This Report Presented in Partial Fulfilment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. S.M. Aminul Haque

Associate professor & Associate Head

Department of CSE

Daffodil International University

Co-Supervised By

Md. Mahfujur Rahman

Lecturer (Senior Scale)

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project titled "**Survival Analysis of Heart Failure Patients Using Machine Learning**" submitted by Name: **Al-Mamun** ID:**191-15-2576** and Name: **Riad Shalehin Leon** ID: **191-15-2613** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 25-01-23.

BOARD OF EXAMINERS

Chairman

____

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

____

Dr. Md. Atiqur Rahman
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Shayla Sharmin
____

Shayla Sharmin 25.1.23
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

25-01-23
____

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S.M. Aminul Haque, Associate professor & Associate Head, Department of CSE**, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree.

Supervised by

____

Dr. S.M. Aminul Haque

Associate professor & Associate Head

Department of CSE

Daffodil International University

Co-Supervised by

____

Md. Mahfujur Rahman

Lecturer (Senior Scale)

Department of CSE

Daffodil International University

Submitted by

AL-MaMUN

____

Al-Mamun

ID: 191-15-2576

Department of CSE

Daffodil International University

____

Riad Shalahin Leon

ID: 191-15-2613

Department of CSE

Daffodil International University

© Daffodil International University

ACKNOWLEDGEMENT

First, we want to express our gratitude to powerful Allah for providing us with the chance to complete our senior project in a fortunate manner.

We are very appreciative of our Project Supervisor **Dr. S.M. Aminul Haque (Associate professor & Associate Head) Department of Computer Science and Engineering (Daffodil International University)** and Md. Mahfujur Rahman Lecturer (Senior Scale) Department of CSE Daffodil International University. Because of our supervisor's expertise, enthusiasm, amiable demeanor, and creative suggestions in the "**Artificial Intelligence**" field, the project was completed.

He has a terrific attitude, shows a lot of patience, and his nice demeanor gives us greater assurance that we can complete this job successfully.

We also wish to recognize a few of our Daffodil International University students who have provided us with vital information for finishing this project.

I won't let the opportunity pass to impress my folks with the interesting details of our project. Our esteemed faculty members' extensive knowledge, insightful suggestions, and contributions during the project's development are also acknowledged.

ABSTRACT

The global burden of death from heart attacks has increased dramatically in the modern era. South Asians are more likely than those in other parts of the world to get a heart attack at a young age. Being able to accurately and rapidly forecast the stage of a heart attack patient requires extensive expertise as well as a deep level of understanding. The medical industry has access to a great quantity of data that may be utilized to make informed judgments thanks to all the concealed information. We will be able to predict heart attack patients' states or stages rapidly with good judgment and a few excellent data mining methods like logistic regression and decision trees. Support vector machine (SVM), random forest classifier, decision tree, logistic regression, KNN, and Gaussian Naive Bayes are the six algorithms we employed in our system (GaussianNB). The accuracy of our final model, which applies the SVM method, is 92%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

A reduction in blood supply to the heart muscle is the main factor in heart attack fatalities. Having a blood clot results in the heart's lack of oxygen. Heart attacks come in three different flavors. NSTEMI, SEMI, and unstable angina are all types of myocardial infarctions that don't involve ST segments (coronary spasm).

Heart failure disease is currently a significant worry for Bangladesh. Hospitalizations are most frequently caused by heart illness (National Health News). The top cause of death worldwide is now heart disease, overtaking cancer. More people die from heart attacks each year than from any other illness. 17.9 million people worldwide passed away from cardiovascular disease in 2019. 32 percent of all deaths were attributable to that. "Heart failure and stock account for 85% of their deaths." According to WHO's country profiles for 2018, non-contagious illnesses including angina pectoris are Bangladesh's leading cause of death.

The World Health Organization claims that heart disease-related mortality has been rapidly increasing in Bangladesh during the last several years. Heart failure must be detected with precision. It must be completed accurately. Using standard procedures, doctors determine the patient's condition. Due to this, medical costs are surprisingly expensive. Therefore, computerized software that can precisely anticipate a cardiac patient's state would be beneficial.

1.2 Motivation

Doctors must act quickly to assess the patient's status after a heart attack has occurred. Doctors may not always be accessible locally around-the-clock. However, heart attacks can happen at any time. If every hospital had a system like the "Survival Analysis of Heart Failure Patients Using Machine Learning" It quickly generated a result on the patient's condition after receiving information. It will therefore be quite advantageous to continue with system acknowledgment. Using this technology, anyone can check their heart condition at any time. Additionally, people can become more conscious of their heart condition.

1.3 The purpose of the research

Today, heart attacks are a serious problem that requires cautious management. It is a very hazardous illness because it can cause an early death at any time without showing any previous symptoms. Heart failure has been chosen as the focus of our investigation. Our focus of exploration is to determine the number of fatal heart attacks. We aim to lessen the number of fatalities. Our method will enable the medical center to forecast patients' heart failure conditions. We anticipate that it will aid in folks better comprehending the scope of a heart attack.

1.4 Questions About Research

1. Does that system accurately predict a patient's probability of survival?
2. Does it use a machine learning system to identify the stage of survival?

Many diseases can affect humans. Because people can take action to prevent diseases, every sickness can be avoided. But a lot of others are also asleep. Nobody wants to lead a miserable life when they are always filled with fear due to this illness. Our approach can forecast the survival status with a high degree of accuracy.

We created a prediction system that will show the prognosis for each patient. Heart failure ranks rather well among the most common diseases. Many people died of heart attacks, including our relatives. Our approach uses a machine learning algorithm to categorize the survival status. Worldwide, heart failure claims the lives of many men and women. In Bangladesh, there isn't a reliable or widely used approach for immediately identifying heart failure situations. To create and execute a reliable system for heart failure survival analysis, we decided to utilize this issue as the main research topic for our project.

1.5 Research Objectives

- To predict the patient heart condition immediately.
- To reach timely appropriate treatment for patient.
- Interpret the actual state of the patient.
- Using different types of machine learning algorithms to predict patient situation.
- Finding out the best algorithm for predicting patient Circumstance.

1.6 Expected Result/Output

The approach for Heart Failure Patient Survival Analysis will help in generating findings that are suitable for the given data. We can rapidly comprehend the patient's state or status by using this technique. Our algorithm used 70% of its training data to make predictions that were more accurate. The training dataset determines how accurate the findings are when they are produced. The system will generate more accurate data if our data collection is rich, but it also depends on how large the data set is. Whether the data set will support our prediction. We used a variety of tactics to improve our accuracy and produce the desired results.

1.7 Report layout

Chapter 1: A overview of the study's justification, objectives, and anticipated outcomes were presented in Chapter 1.

Chapter 2: The introduction, pertinent writings, a synopsis of the study, and issues are all shown in the "Background" part of Chapter 2.

Chapter 3: Chapter 3 of the book will address research methodology.

Chapter 4: Discussion of the analytical outcomes will take place in Chapter 4.

Chapter 5: The overview and conclusion are included in Chapter 5.

CHAPTER 2

Background Study

2.1 Introduction

We'll go into great detail regarding heart failure disease in this section. A summary of the study, certain works that are relevant to us, and the difficulties facing our system. We spoke about a few initiatives that are comparable to our project in the section on related works. In this part, we also highlighted our project's approaches and the many difficulties we encountered while working on them.Heart failure is a serious medical condition in which the heart is unable to pump enough blood to meet the body's needs. It is a chronic and progressive condition that affects millions of people worldwide.

There are several risk factors associated with heart failure, including hypertension, diabetes, obesity, and smoking. Additionally, certain medical conditions, such as myocardial infarction (heart attack) and valvular heart disease, can also lead to heart failure.

There are several different types of heart failure, including systolic heart failure, diastolic heart failure, and heart failure with preserved ejection fraction.

Systolic heart failure occurs when the heart muscle is unable to contract effectively, leading to decreased blood flow. Diastolic heart failure occurs when the heart muscle is unable to relax effectively, leading to increased blood pressure and decreased blood flow. Heart failure with preserved ejection fraction occurs when the heart muscle is able to contract effectively but unable to relax effectively.

Treatment for heart failure includes lifestyle changes, such as maintaining a healthy diet and exercise regimen, and taking medications, such as ACE inhibitors and beta blockers. In severe cases, a heart transplant or a mechanical circulatory support device may be necessary.

Survival analysis is a statistical method used to study the time until an event of interest occurs, such as death. In the context of heart failure, survival analysis can be used to study the time until death due to heart failure, or the time until a patient experiences a significant worsening of their condition. This information is important for understanding the natural history of the disease and for developing and evaluating interventions to improve outcomes.

Overall, heart failure is a complex and multifactorial condition that requires a comprehensive approach to management and treatment. Survival analysis can provide valuable insights into the natural history of the disease and can be used to improve patient outcomes through the development of effective interventions.

2.2 Heart failure

Within that part, we'll cover a few subjects linked to heart failure.Heart failure is a serious medical condition in which the heart is unable to pump enough blood to meet the body's needs. It is a chronic and progressive condition that affects millions of people worldwide. The symptoms of heart failure include shortness of breath, fatigue, and swelling in the legs and ankles.

There are several risk factors associated with heart failure, including hypertension, diabetes, obesity, and smoking. Additionally, certain medical conditions, such as myocardial infarction (heart attack) and valvular heart disease, can also lead to heart failure.

2.3 Definition

The inability of the heart to properly pump blood throughout the body is known as heart failure. Cardiac failure often happens when the heart has become stiff or feeble. Congestive heart failure is the term used to describe the ailment, however, it is no longer often used. Heart failure may not usually indicate that the heart has ceased beating entirely.

2.4 Causes and Risk Factors

Heart attacks may have many different causes. There are several things that cause heart disease. Stress is one factor that contributes to heart attacks, and being in a stressful situation might increase your risk. A heart attack may also be caused by a bad diet that is low in nutrients and by not getting enough exercise.

➔One that is very high in fat.

➔Deficits in Nutrition.

➔Exercising lake.

➔Smoking.

➔Usage of drugs.

➔Stress.

➔Infection.

The condition is at risk from a number of circumstances.

Which are given below:

➔Breathing issues during physical activity or when resting down.

➔Sickness.

➔The legs, feet, and other extremities have a lump.

➔Heartbeat that is rapid or erratic

➔Reduced capacity to engage in physical activity.

➔Putting on weight due to accumulation of fluid.

➔Disadvantage a lowering or dulling of one's awareness

➔Chest discomfort may occur if a heart attack is the cause of heart failure.

➔Linked to both smoking and eating an unhealthy diet.

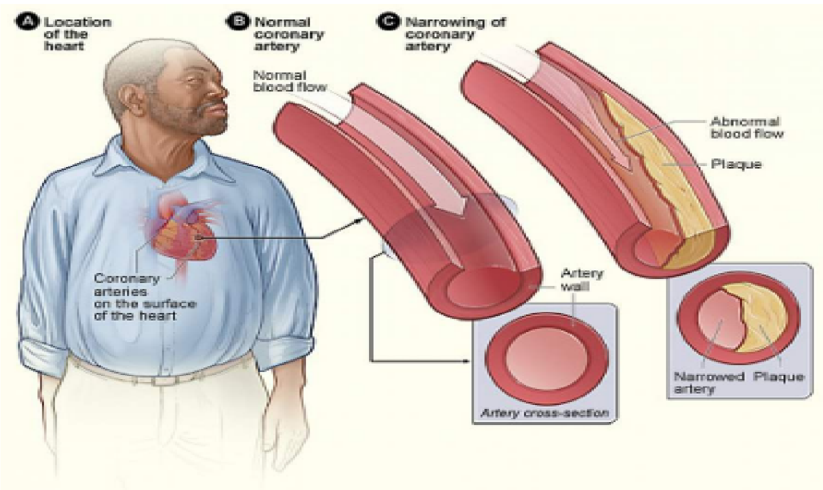

Fig 3.1: Blood vessel plaque[11]

2.5 Similar Work’s

- **Davide et al:** Machine learning can predict the survival of people with heart failure based on serum creatinine and ejection fraction alone. They divided the dataset into a training set made up of 239 randomly chosen patients for 80% of the dataset and a test set made up of the remaining 60 patients for 20%.[1]
- **Tanvir et al:** Based on their thesis (96)32% of patients died from CHD, and they focused on the survival study of heart failure patients.[2]
- **Asha et al:** The Naive Bayes, Decision List, and KNN algorithms were used to diagnose heart disease patients, with the Naive Bayes algorithm having the greatest accuracy at 52.33 percent.[3].
- **Giolo et al:** Continued to work on a study of heart failure patients' survival, According to their thesis, they provide a summary of the data showing that (37%) of the patients passed away at the conclusion of the follow-up period. Of these, males made up 62.8 percent of the population while women made up percent.[4]
- **Asif et al:** Focused on predicting patient survival in heart failure using machine learning approaches, The Faisalabad Institute of Cardiology initially provided the dataset, which includes the medical files of 299 people who have heart failure. Finally, they achieved an accuracy of 76.25% in their investigation.[5]
- **Ketut et al:** Conducted research using KNN, Naive Bayes, and a straightforward CART algorithm. KNN produced the best results, with an accuracy rate of 81.85%. [6].

2.6 The Problem's Scope

Several algorithms were included in our system, and we always paid attention to accuracy. However, as we anticipated greater accuracy, our algorithms presented us with challenges.

While some algorithms gave us more accuracy, others had downsides or took a lot of time. As a result, we concentrated on developing more accurate algorithms in a short amount of time. While looking for the optimal algorithm, there are a few issues that come up,

- We were inputting different kinds of data, yet a certain algorithm was still giving us similarly accurate findings.
- Data implementation, algorithms, and other areas all had issues.

2.7 Challenges

Primarily The finest difficulties for obtaining predictions are with data recruiting. Because we can't forecast anything without data, data collecting is crucial. We attempted to get in touch with the Bangladesh Heart Failure Foundation, but they weren't responsive, so we ended up gathering our information via kaggle.com. The following phase is a data processing and cloning, after which null or empty values were decreased and the data was made clean. With regard to value, all features are used with the aid of feature scaling. For more accurate results, we have always worked to lower the number of incorrect values.

There are additional difficulties,

- The best algorithm to use.
- Complete the shareholder study.
- Working with our data.
- Excel file data creation for input.
- The algorithms must make good use of the dataset.

CHAPTER 3

Research Methodology

3.1 Introduction

Our system collects data from Kaggle. We looked for a special method and made precise forecasts in our project inquiry. 12 characteristics attributes and 299 total data points were utilized. Then we had to repair a number of inaccurate feature values that we had found. In order to provide precise forecasts, we have already finished the procedures of feature scaling and data cleansing. We employed datasets for both training and testing as well as methods such as the Support Vector Machine (SVM), Random Forest Classifier, KNN, Decision Tree, Logistic Regression, and Gaussian Naive Bayes (GaussianNB). Our model, which is based on a working process, was created using these techniques.

3.2 Data Collection Process

Through Kaggle, we acquired this data collection. Making our analysis superior. A number of research publications were also examined; several of them focused on cardiac attacks. Our study examines the survivability of heart attacks that occur early specifically. Heart disease is on the rise right now. We aim to create a system that will help our people live longer by reducing heart disease deaths.

3.3 The Dataset's Reference Value

Table 3.1: Attribute range

Attribute	**Value**
high Blood Pressure	<120/80
anemia	<200mg/dL
creatinine Phosphokinase	<120
diabetes	120-200ms
ejection fraction	170bpm
Platelets	True
serum creatinine	Absence
serum Sodium	1.3-1.5
sex	Male/Female
smoking	Yes/NO
time	< .5
age	>40

Our goal in creating table 3.1 of the dataset was to provide a collection of reference values for 12 dataset attributes. High_Blood_Pressure, Anemia, creatinine phosphokinase, diabetes, ejection fraction, platelets, smoking, serum creatinine, serum sodium, sex, time, and age are some of the features of the dataset displayed in table 3.1

3.4 Statistical Analysis

We have 299 datasets in our dataset with 12 characteristics and a target attribute. The target property DEATH EVENT has a significant data difference ratio since 200 data are survived events and 99 are not serviced. We had to consider how to taste it. We need to make an effort to determine how many incorrect and null values there are. Thankfully, we did not get that type of value. Features related to age are the most useful features. The only available feature data types are float and int. How often a patient visits the doctor is another highly useful component of our program. 80% of the data are used to train and 20% to test. The steps for data collection and processing using a flowchart are shown below.

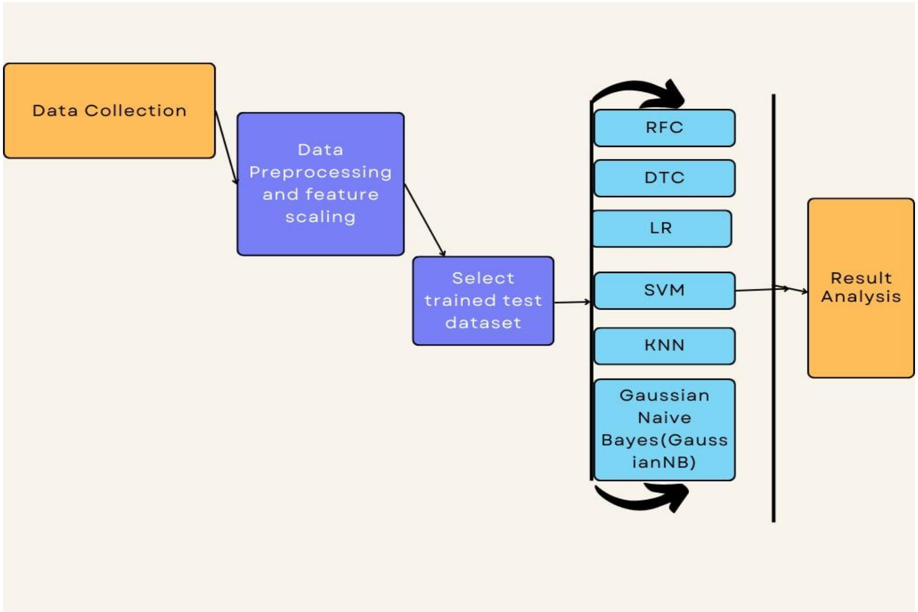

Fig 3.2: Work Flow.

3.5 Research Topic and Equipment

More people are becoming aware of ML every day. PCs may utilize machine learning algorithms to learn knowledge via quantitative methods. Finding all pertinent information, for instance, and utilizing machine learning to produce a conclusion or predictive learning without the need for code, is the most crucial phase. Then, without altering the general structures, a comparable computation may be conducted comparing datasets from different locations. There are several machine learning computations, some of which we have included in our system. The algorithm's specifics are displayed below:

3.5.1 Logistic Regression

The supervised machine learning includes logistic regression. The principal problem these addresses are classification. This has a dual effect for providing a free factor arrangement. A discrete variable is a target. Understanding machine learning logistics The system to induce backslide. The method is matched portrayal difficulties. Uses of portrayal include calculated backsliding. Appreciate direct backsliding in certain loads or coefficients for a fraction or yield for information for joining directly. One straight backslide differs from another by showing a produce regard that matches grade (0 OR 1) rather than number regard.

The condition of the model key reversal is: $y = \frac{e^{(b0 + b1*x)}}{(1 + e^{(b0 + b1*x)})}$ …………… (1)

When b0 is the skew or intercept term, b1 is the coefficient for the single input value, and y is the anticipated output (x). You must learn the associated b coefficients (constant real values) for each column in your input data from your training set.

3.5.2 Decision Tree

For machine learning forecasting, the Decision Tree Classifier is the key computation. One of the more record-breaking techniques for traditional decision tree calculations is arbitrary branching in day varieties. The decision tree is also known as a CART, or classification and regression tree. Because it generates reliable outcomes as rapidly as possible, the decision tree technique is the computer's greatest strength. The decision tree differentiates between a wide range of computations, like CART, C 4.5, ID3, H48, and CHH. One of these algorithms, the J48, is gaining popularity. J48 employs a pruning strategy to build a tree. The recursive procedure continues until the required results are obtained. This delivers precision and flexibility. This formula is used in the equations that follow.

E = ∑i=K Pi*log2*Pi……………..(2)

From 2, K represents the number of target traits, And The symbols Pi and I stand for the total number of instances and the count of existing code, respectively. The CART counting technique is commonly known as "DecisionTrees," but on two stages such as R, it is known as something else.

3.5.3 Support VectorMachine(SVM)

SVM accurately labels a detached hyperplane as a point is an important classifier. The algorithm results in a perfect plane that neatly arranges extra reference points. This hyperplane is a perfectly positioned two-dimensional surface with two classes, where each class represents half of the dimensions. If there were just two variables, this would represent a two-dimensional space. One way to characterize the space of variables is with the help of a line or hyperplane. When using SVM, class 0 or 1 is the most effective method for narrowing in on the relevant information variable space. Two-dimensional representations of the line exist, and we can reasonably assume that it will cleanly split most of the data points we're working with. Instances of this might be:

$$B0 + (B1 * X1) + (B2 * X2) = 0 \quad (3)$$

With X1 and X2 as inputs, the learning process determines the coefficients (B1 and B2) that set the slope of the line and the intercept (B0).

3.5.4 K-nearest neighbor classifier(KNN)

Using a similarity metric to classify new cases, The KNN algorithm provides a simple way to keep track of all the data. The KNN algorithm is also known as

1) Using case-based logic

2) The k-nearest neighbor

3) Definitions based on examples

4) Instance-based instruction

5) memory-based inference

6) Lazy education.

In most cases, where all the attributes are continuous, nearest neighbor classification is utilized. Illustrates a straightforward K-nearest neighbor technique.

Step 1: Determine which K training cases are the most similar to the unknown instans.
Step 2: Select the classification that applies to the majority of these K cases.

It's important to note that the k-NN algorithm doesn't have a model, it only stores the training data, so it's a lazy learning algorithm. The classification decision is made at the time of the prediction, based on the k closest neighbors.

Also, the choice of k value is crucial for the performance of the algorithm. A small k value will make the model more sensitive to noise and outliers, while a large k value will make the model less sensitive to noise and outliers. Therefore, the k value should be chosen carefully and can be done using cross-validation techniques.

3.5.5 Gaussian Naive Bayes

The Gaussian Naive Bayes algorithm uses Bayes' Theorem to calculate the probability of a class label given the feature values. Bayes' Theorem states that:

P(A | B) = P(B | A) * P(A) / P(B)

Where P(A | B) is the conditional probability of event A given that event B has occurred, P(B | A) is the likelihood of event B given that event A has occurred, P(A) is the prior probability of event A, and P(B) is the prior probability of event B.

In the case of Gaussian Naive Bayes, the class label is event A, and the feature values are event B. The likelihood of the feature values given the class label is modeled as a Gaussian distribution, which is also called a Normal Distribution.

So, the formula for Gaussian Naive Bayes algorithm is:

P(y|x) = P(y) * (1/(sqrt(2pisigma^2))^(exp(-1/2 * (x-mu)^2/sigma^2))

Where:

P(y|x) is the conditional probability of class label y given the feature values x.

P(y) is the prior probability of class label y.

mu and sigma are the mean and standard deviation of the feature values for class labely.

x is the feature value.

pi is the mathematical constant pi.

The class label y with the highest probability is chosen as the prediction.

It's important to note that this is a simplified explanation of the Gaussian Naive Bayes algorithm, in a real-world scenario the algorithm would be applied to multiple features, and the formula will be slightly different, but the concept remains the same.

3.6 Algorithm Choice

In this part, we utilize a variation method to gain improved accuracy. Here is a demonstration of the efficacy of all of the algorithms we've built so far:

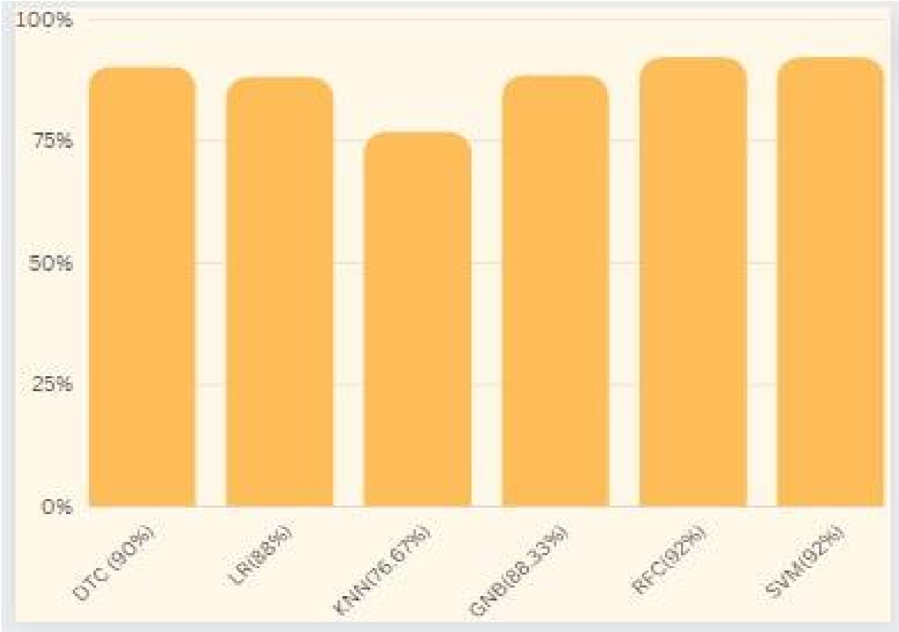

Fig 3.3: Using Algorithm’s

3.7 Proposed Algorithm

We use the Logistic Regression technique in our declared approach. We work to develop a model that estimates the survival analysis of patients with heart failure. Our suggested improvement method is logistic regression in practice. The most important aspect of a system is its implementation. The setting for creating our model is Jupyter Notebook.

Our model was created using a variety of libraries. In our approach -
Step-1:Our dataset is chosen, which has Heart Failure Clinical Records.

Step-2:Who can live and who can't is classified in the dataset.

Step-3:To load the dataset, type in the data.

Step-4:Using ML algorithm python.

Step-5:Various machine learning algorithms are compared for their levels of accuracy.

Step-6:The Logistic Regression Algorithm ensures the greatest precision.

Step-7:Calculate the performance of the model.

LG uses the data to determine which individuals have a chance of survival and which do not.In the beginning, we utilized our LG Algorithm with raw data. Seventy-five percent was my average score in the class. The data was then processed by me. By using Normalization, I scaled everything from 0 to 1. We made sure there weren't any blank or empty spots in the dataset. Fortunately, I don't have that problem. After that, I do a principal component analysis in an effort to rectify the aforementioned data discrepancy (Principal component analysis). This time, our algorithm will be used. We found that Linear Regression could predict outcomes with 83% precision. DecisionTreeClassifier was then employed, and a respectable 90% accuracy was achieved. So I tried Logistic Regression, and it gave me an accuracy of 88%.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

The outcomes and rationale of the experiment will be discussed below. In this part, we will evaluate and contrast the efficacy and efficiency of several classifiers. Here is the prognostic data we utilized from 299 people on kaggle.com to make our prediction. Several charts are produced in our data collection depending on the statistical data we collect using various algorithms. To ascertain the rate of our system in our data set, we examined three different ways. The findings will be shown in tables and a graph.

4.2 Experimental Results

Our dataset contains 299 data points for the survival analysis of patients with heart failure, and we obtained 90% accuracy. We use actual data for our analysis, which significantly improves the accuracy of our results. Our data are normalized to have a range of 0 to 1. Our accuracy is higher after normalization than it was before. Additionally, we employ voting classifiers, which enable us to combine several algorithms into a single model. Designers likewise utilized SMOTE (synthetic minority over-sampling) to help us zero down on even more data-driven insights. We again matched our dataset into the algorithm after incorporating this, however, we made no significant advancements in how this technique was applied. After all of this, our logistic regression technique produces results that are superior to all others. Furthermore, we employ the true negative, true positive, false positive, and false negative prediction rates. A confusion matrix is a useful tool for explaining a data categorization problem. Four terms are included in the confusion matrix's table.

- **True Positive (TP):** We calculated the outcome as the number of people who could actually live.
- **True Negative (TN):** According to our calculations, people who we predicted couldn't have survived probably.
- **False Positive (FP):** Our presumption was that those who could survive couldn't.

- **False Negative (FN):** At first, we didn't think we had a chance, but eventually, we did.

Precision: The ratio of true positives to overall positive predictions is used to determine precision.

$$\textbf{Precision} = \frac{TP}{TP+FP} = \frac{45}{45+1} = \frac{45}{46} = 0.98$$

recall: Recall is measured as a percentage of retrieved relevant examples to all relevant events. The great majority of useful results may be retrieved using algorithms with good recall.

$$\textbf{Recall} = \frac{TP}{TP+FN} = \frac{45}{45+4} = \frac{45}{49} = 0.92$$

F Measure: When calculating test accuracy, the f-score takes into account both precision and recall. The integration of precision and memory serves as a powerful tool.

$$\textbf{F Score} = \frac{2*precision*recal}{precision+recall} = \frac{2*0.98*0.92}{0.98+0.92} = \frac{1.8}{1.9} = 0.95$$

Accuracy: How closely a measured value matches an expected value is what we call its accuracy.

$$\textbf{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{45+10}{45+10+1+4} = \frac{55}{60} = 0.92$$

True Positive Rate: The percentage of times our proposed technique mistakenly determines that a stroke is not occurring while one really is is known as the "True positive rate." In order to calculate the real success rate, use the following formula:

$$\textbf{True Positive Rate} = \frac{TP}{TN+FP} = \frac{45}{10+1} = \frac{45}{11} = 4.090$$

Specificity: The concept of "specificity" describes the manner in which our recommended method accurately foretells a brain stroke when one is in fact, occurring. The specificity of the following equation has to be determined:

$$\textbf{Specificity} = \frac{TN}{TN+FP} = \frac{10}{10+1} = 0.91$$

4.3 The Outcome of Experiments

Table 4.1:Experimental Results

precision	0.98
recall	0.92
f-Score	0.95
Accuracy	0.92
true Positive Rate	4.090
specificity	0.91

The so-called confusion matrix may help us decide on a more sophisticated arrangement metrics for our classifier, including accuracy, recall, specificity, and sensitivity. Table 4.1 provides information about them.

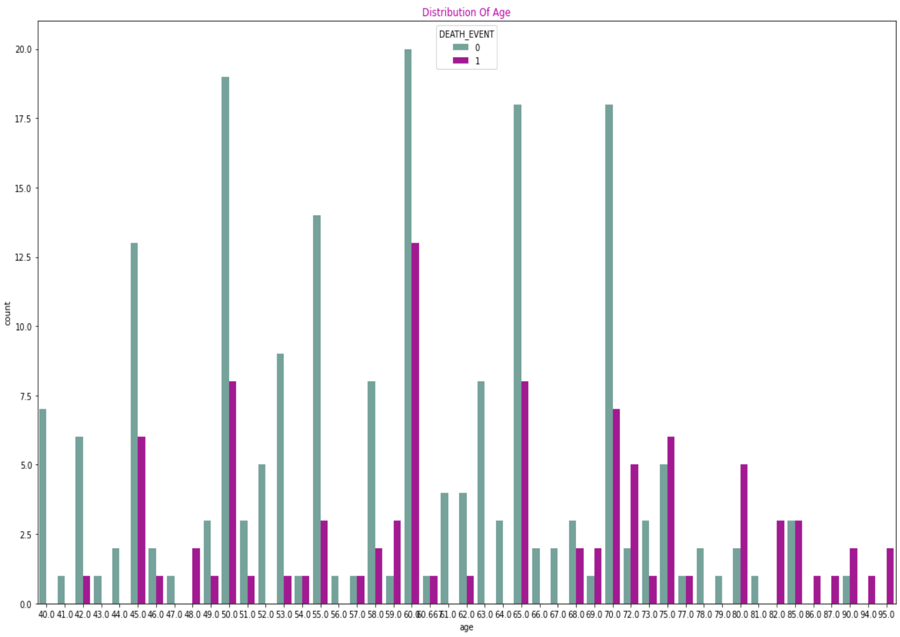

Fig4.1: Death Event

There is a clear correlation between age and passing away, as seen by this drunken chitchat. How many persons in a certain age group are expected to live or die is predictable based on their age.

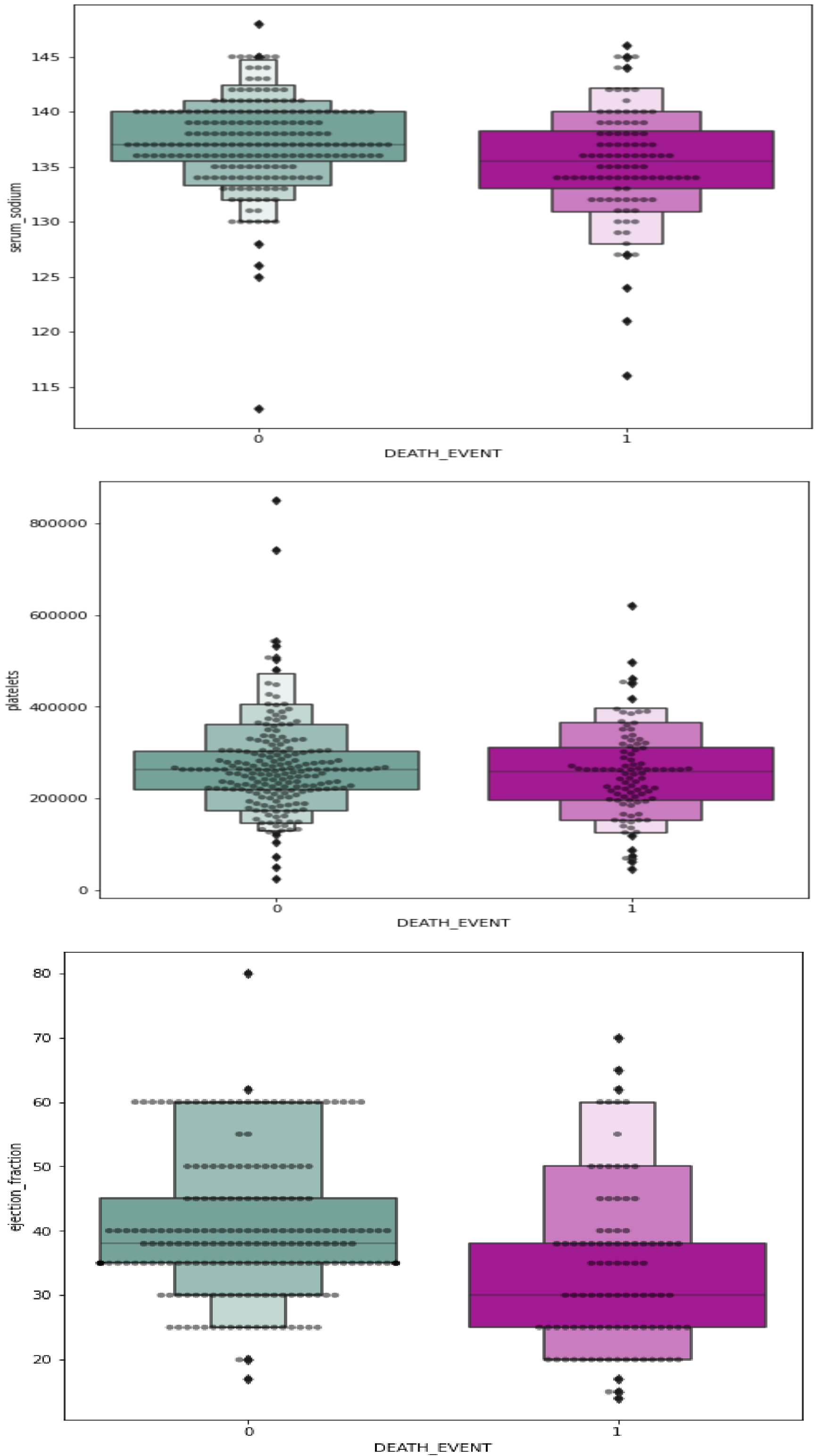

Fig 4.2: Features

The impact of serum sodium, ejection, and fraction features on the target characteristic may be seen in the cluster plot and graphs below. Those with a serum sodium level below 135 have a better chance of survival. At 135–140, it's better for human survival. Increased survival is shown in platelets with a rating of 20k-35k. More lives are spared when the ejection fraction is below 30.

Table 4.2: Confusion matrix

	No-Event	Event
No-Event	True negative 10	False positive 1
Event	False negative 4	True positive 45

The confusion matrices between genuine and false positives and negatives are shown in table 4.2.

Table 4.3: Confusion matrix result

	precision	recall	f1-score	support
0	0.92	0.98	0.95	46
1	0.91	0.71	0.80	14
accuracy			0.92	60
macro avg	0.91	0.85	0.87	60
weighted avg	0.92	0.92	0.91	60

A confusion matrix is a table that is often used to describe the performance of a classification algorithm, such as a machine learning model. It is used to visualize the performance of the model in terms of true positive, true negative, false positive, and false negative predictions.

After applying the algorithm, we have achieved accuracy:

Table 4.4: Accuracy

algorithm	accuracy
SVM	92%
random forest classifier	90%
decision Tree	89%
logistic Regression	88%
Gaussian Naive Bayes(GNB)	88%
K-Nearest Neighbour	77%

The Support Vector Machine (SVM) technique we use in our final model provides an accuracy of **92%**.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Heart failure is a significant health issue that affects millions of people worldwide. It can have a significant impact on individuals and their families, as well as on society as a whole. The condition can lead to increased healthcare costs, reduced quality of life, and a decreased ability to work and participate in society. It is important for individuals to be aware of the risks of heart failure and to take steps to reduce their risk, such as maintaining a healthy diet and exercise regimen and managing any underlying medical conditions. Additionally, it is important for healthcare providers and policymakers to work towards improving access to care and developing effective treatments for heart failure. It can also lead to hospitalization, chronic disability, and even death. Individuals with heart failure often require long-term care and management, which can place a significant burden on healthcare systems and caregivers. Additionally, heart failure can lead to decreased productivity and increased absenteeism, which can have an economic impact on both individuals and society as a whole. Therefore, it is important for individuals to be aware of the risks of heart failure and to take steps to reduce their risk, as well as for healthcare providers and policymakers to work towards improving access to care and developing effective treatments for heart failure.

5.2 Impact on Environment

Heart failure does not have a direct impact on the environment. However, the treatment and management of heart failure can have an indirect impact on the environment. For example, the use of certain medications, such as ACE inhibitors, can lead to increased levels of nitrogen and phosphorus in the environment, which can contribute to water pollution and the growth of harmful algae. The use of medical equipment, such as pacemakers and defibrillators, can also contribute to electronic waste. Additionally, the transportation required for individuals with heart failure to receive medical treatment can contribute to air pollution. It's important to note that these impacts are not specific to heart failure, but to the healthcare system in general.

Also, it's worth mentioning that a healthy lifestyle and environment can reduce the risk of heart failure, by reducing the risk factors such as high blood pressure, obesity, and diabetes. Therefore, promoting a healthy environment and reducing air pollution, water pollution and promoting physical activity can have a positive impact on reducing the risk of heart failure.

5.3 Ethical Aspects

Survival analysis is a statistical method used to study the time until an event of interest occurs, such as death. In the context of heart failure, survival analysis can be used to study the time until death due to heart failure, or the time until a patient experiences a significant worsening of their condition. There are several ethical considerations to keep in mind when conducting survival analysis in the context of heart failure. One important consideration is patient privacy and confidentiality. Researchers must ensure that patient data is protected and kept confidential, and that patient consent is obtained before using their data in a study. Additionally, researchers must consider the potential for bias in their study design and data analysis, as well as the generalizability of their findings to the larger population. Another ethical aspect to consider is the patient's autonomy and informed consent for treatment options. Heart failure patients have a complex decision-making process with their physicians when it comes to the treatment options. For example, the decision of whether to undergo a heart transplant or a mechanical circulatory support device. The physicians should ensure that the patients are fully informed of the risks, benefits, and alternatives of the treatment options, and that they are making an informed decision based on their values and preferences. Another ethical consideration is the allocation of resources, particularly in cases where the treatment options are costly. Health care providers should ensure that the resources are allocated in a fair and equitable way, based on the patient's clinical need and not on their ability to pay. Finally, it's worth mentioning that survival analysis data can be used to support the development of policies and guidelines for the management of heart failure patients. However, it's important to consider the ethical implications of the use of this data in policy-making, and to ensure that the policies and guidelines promote the best interests of the patients.

5.4 Sustainability Plan

A sustainability plan for a survival analysis of heart failure would involve several key components to ensure the long-term success and impact of the study.

Data management: A robust data management plan should be in place to ensure the accuracy, integrity, and security of the data collected. This includes regular backups and version control, as well as protocols for data access and sharing.

Study design and analysis: The study design and analysis should be robust and well-powered to ensure the validity and generalizability of the findings. This includes appropriate sample size calculations and statistical methods, as well as a clear and transparent reporting of the results.

Stakeholder engagement: Engaging with stakeholders, such as patients, caregivers, and healthcare providers, can help to ensure that the study is relevant and responsive to their needs and concerns. This can also help to build trust and support for the study.

Impact and dissemination: The study should have a clear plan for how the findings will be disseminated and used to improve the care and outcomes for patients with heart failure. This includes publishing the findings in peer-reviewed journals, presenting the results at conferences, and working with healthcare providers and policymakers to implement the findings in practice.

Long-term sustainability: The study should be designed in a way that allows for ongoing follow-up and data collection, to ensure the long-term sustainability of the research. This can include ongoing monitoring of the outcomes of the study, as well as the opportunity to update the study design and analysis as needed.

Ethical considerations: The study should be designed and conducted in an ethical manner, ensuring patient privacy and autonomy, informed consent, and fair and equitable resource allocation.

All of these components should be integrated into a comprehensive sustainability plan to ensure the long-term success and impact of the study on the management and outcomes of heart failure patients.

CHAPTER 6

Conclusion And Future Work

5.1 Summary of the Study

Here, we show how to utilize logistic regression here on collected data to draw conclusions. We compared predictions using an 80% model train and 20% testing sets and improved the overall predictive performance in comparison to previous methods.

5.2 Conclusion

'survival status' could be terrifying. It's wonderful that medical care and procedures are getting better all the time so that people who suffer from heart attacks can still have fulfilling lives. It's crucial to identify the medical condition that causes heart attack death. The ratio of heart attack deaths can be decreased if the general public can determine their health condition with ease. In this study, we covered logistic regression algorithms. We have estimates based on patient characteristics that enable patients and the general public to determine whether a patient's body has recovered from a heart attack. Incorporating our effort A physician may easily check on a patient's status. Helpful for a doctor to make a quick decision. Our method aids in precise and appropriate decision-making for heart disease patients. treatment. People might use it to lessen the number of fatal heart attack cases.

5.3 Future Research Directions

We got our information from kaggle.com, as we said before. Our current accuracy is 92% and we are always striving to increase that number. Without receiving every feather from the user, our output would be faultless, but we also work on that. We'll strive to gather data from other cardiac hospitals in the future. Additionally, look for additional data with the effective organization. Additionally, we'll work to make it simpler for all types of individuals to access.

REFERENCES

[1] Effect of heart failure on the outcome of COVID-19 — A meta-analysis and systematic review Year published Accepted 3 July 2020

[2] Rajkumar, A., & Reena, G. S. (2010). Diagnosis of heart disease using data mining algorithm. Global journal of computer science and technology, 10(10), 38-43.

[3] HEART FAILURE PREDICTION USING MACHINE LEARNING TECHNIQU Year published Accepted 29 Jan 2021.

[4] Wisaeng, K. (2014). Predict the diagnosis of heart disease using feature selection and the k-nearest neighbor algorithm. Applied Mathematical Sciences, 8(83), 4103-4113.

[5] Enriko, I. K. A., Suryanegara, M., & Gunawan, D. (2016). Heart disease prediction system using k-Nearest neighbor algorithm with simplified patient's health parameters. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 8(12), 59-65.

[6] https:// core.ac.uk/download/pdf /2292 79242. pdf >> last accessed on:: 23, October 2021 07:00 pm

[7] Donnan GA, Fisher M, Macleod M, Davis SM (May 2008). "Stroke". Lancet. 371 (9624):1612–23.

[8] Rymarczyk, T., Kozłowski, E., Kłosowski, G., & Niderla, K. (2019). Logistic regression for machine learning in process tomography. Sensors, 19(15), 3400.

[9] Loh, W. Y. (2011). Classification and regression trees. Wiley interdisciplinary reviews: data mining and knowledge discovery, 1(1), 14-23.

[10] Provost, F. (2000, July). Machine learning from imbalanced data sets 101. In Proceedings of the AAAI'2000 workshop on imbalanced data sets (Vol. 68, No. 2000, pp. 1-3). AAAI Press.

[11] https://www.nih.gov/news-events/news-releases/hiv-infected-men-increased-risk-heart-disease-large-study-finds

paper report

ORIGINALITY REPORT

18%	17%	6%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
3	Submitted to European University Student Paper	1%
4	www.oecd-ilibrary.org Internet Source	1%
5	Submitted to Leeds Beckett University Student Paper	1%
6	www.igi-global.com Internet Source	<1%
7	Hancock, Mark J., Christopher G. Maher, Lucíola da Cunha Menezes Costa, and Christopher M. Williams. "A guide to survival analysis for manual therapy clinicians and researchers", Manual Therapy, 2014. Publication	<1%
8	www.drfuhrman.com Internet Source	

<1%

9	desklib.com Internet Source	<1%
10	Leondes, Cornelius T. "NON-PARAMETRIC PIXEL APPEARANCE PROBABILITY MODEL USING GRID QUANTIZATION FOR LOCAL IMAGE INFORMATION REPRESENTATION", Medical Imaging Systems Technology, 2005. Publication	<1%
11	Pushpa Singh, Narendra Singh, Krishna Kant Singh, Akansha Singh. "Diagnosing of disease using machine learning", Elsevier BV, 2021 Publication	<1%
12	peerj.com Internet Source	<1%
13	investingnews.com Internet Source	<1%
14	www.leadgatepharmacy-leadgate-northconsett.co.uk Internet Source	<1%
15	www.medicalnewstoday.com Internet Source	<1%
16	Submitted to Nottingham Trent University Student Paper	<1%

17	tel.archives-ouvertes.fr Internet Source	<1 %
18	morioh.com Internet Source	<1 %
19	www.researchgate.net Internet Source	<1 %
20	"Proceedings of International Joint Conference on Computational Intelligence", Springer Science and Business Media LLC, 2020 Publication	<1 %
21	Submitted to Jacksonville University Student Paper	<1 %
22	www.benkazie.com Internet Source	<1 %
23	www.narayanahealth.org Internet Source	<1 %
24	core.ac.uk Internet Source	<1 %
25	www.coursehero.com Internet Source	<1 %
26	Andreea-Clara Pricopi, Alin Razvan Paraschiv, Diana Besliu-Ionescu, Anca-Nicoleta Marginean. "Predicting the Geoeffectiveness of CMEs Using Machine Learning", The Astrophysical Journal, 2022	<1 %

Publication

27	Bei Hui, Yue Wu. "A New Anytime Classifier Basing on AAPE: Anytime Averaged Probabilistic with Weight Estimator (AAPWE)", 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008 Publication	<1%
28	chaos3.fkp.physik.tu-darmstadt.de Internet Source	<1%
29	link.springer.com Internet Source	<1%
30	Yu Guo, Xiaoxiao Yang, Yinuo Wang, Fei Wang, Badong Chen. "Online robust echo state broad learning system", Neurocomputing, 2021 Publication	<1%
31	docs.oracle.com Internet Source	<1%
32	eco.nahrainuniv.edu.iq Internet Source	<1%
33	ks3-cn-beijing.ksyun.com Internet Source	<1%
34	scholarsjunction.msstate.edu Internet Source	<1%

ujcontent.uj.ac.za

35	Internet Source	<1 %
36	Giovanna Sannino, Giuseppe De Pietro, Ivanoe De Falco. "Automatic Extraction of Interpretable Knowledge to Predict the Survival of Patients with Heart Failure", 2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2021 Publication	<1 %
37	"Theory: Probabilistic Classifiers", Computational Imaging and Vision, 2005 Publication	<1 %
38	Nitish Biswas, Khandaker Mohammad Mohi Uddin, Sarreha Tasmin Rikta, Samrat Kumar Dey. "A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach", Healthcare Analytics, 2022 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

paper report

PAGE 1

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Proper Noun If this word is a proper noun, you need to capitalize it.

PAGE 2

Sentence Cap. Remember to capitalize the first word of each sentence.

Proper Noun If this word is a proper noun, you need to capitalize it.

PAGE 3

Article Error You may need to use an article before this word. Consider using the article **the**.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Article Error You may need to use an article before this word.

P/V You have used the passive voice in this sentence. Depending upon what you wish to emphasize in the sentence, you may want to revise it using the active voice.

Article Error You may need to use an article before this word.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Proper Noun If this word is a proper noun, you need to capitalize it.

PAGE 4

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Proper Noun If this word is a proper noun, you need to capitalize it.

PAGE 5

Missing "," You may need to place a comma after this word.

Proofread This part of the sentence contains a grammatical error or misspelled word that makes your meaning unclear.

PAGE 6

Missing "," You may need to place a comma after this word.

Dup. You have typed two **identical words** in a row. You may need to delete one of them.

Proper Noun If this word is a proper noun, you need to capitalize it.

Missing "," You may need to place a comma after this word.

Possessive This word may be a plural noun and may not need an apostrophe.

PAGE 7

Confused You have used **Choice** in this sentence. You may need to use **chose** instead.

PAGE 8

Possessive This word may be a plural noun and may not need an apostrophe.

PAGE 9

PAGE 10

Article Error You may need to remove this article.

Garbled Grammatical or spelling errors make the meaning of this sentence unclear. Proofread the sentence to correct the mistakes.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Confused You have used **WHO's** in this sentence. You may need to use **whose** instead.

Missing "," You may need to place a comma after this word.

Prep. You may be using the wrong preposition.

PAGE 11

Article Error You may need to use an article before this word.

Article Error You may need to use an article before this word.

Article Error You may need to use an article before this word.

PAGE 12

Article Error You may need to remove this article.

Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

Confused You have used **A** in this sentence. You may need to use **an** instead.

PAGE 13

Article Error You may need to use an article before this word. Consider using the article **a**.

Article Error You may need to use an article before this word.

PAGE 14

Article Error You may need to use an article before this word. Consider using the article **a**.

Missing "," You may need to place a comma after this word.

PAGE 15

Article Error You may need to use an article before this word.

Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

Article Error You may need to use an article before this word.

PAGE 16

Article Error You may need to remove this article.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Prep. You may be using the wrong preposition.

Article Error You may need to use an article before this word.

Run-on This sentence may be a run-on sentence. Proofread it to see if it contains too many independent clauses or contains independent clauses that have been combined without conjunctions or punctuation. Look at the "Writer's Handbook" for advice about correcting run-on sentences.

PAGE 17

PAGE 18

Missing "," You may need to place a comma after this word.

PAGE 19

Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Missing "," You may need to place a comma after this word.

PAGE 20

Missing "?" Remember to use a question mark at the end of a question.

Article Error You may need to use an article before this word. Consider using the article **the**.

Article Error You may need to use an article before this word.

Article Error You may need to use an article before this word.

Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.

Article Error You may need to use an article before this word.

Word Error Did you type "**the**" instead of "**they**," or have you left out a word?

Prep. You may be using the wrong preposition.

Missing "," You may need to place a comma after this word.

Proper Noun If this word is a proper noun, you need to capitalize it.

Article Error You may need to use an article before this word.

Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.

Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.

Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

Article Error You may need to use an article before this word. Consider using the article **the**.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Run-on This sentence may be a run-on sentence. Proofread it to see if it contains too many independent clauses or contains independent clauses that have been combined without conjunctions or punctuation. Look at the "Writer's Handbook" for advice about correcting run-on sentences.

Missing "," You may need to place a comma after this word.

Article Error You may need to remove this article.

Confused You have used **A** in this sentence. You may need to use **an** instead.

Confused You have used **A** in this sentence. You may need to use **an** instead.

Article Error You may need to remove this article.

Run-on This sentence may be a run-on sentence. Proofread it to see if it contains too many independent clauses or contains independent clauses that have been combined without conjunctions or punctuation. Look at the "Writer's Handbook" for advice about correcting run-on sentences.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sentence Cap. Remember to capitalize the first word of each sentence.

Run-on This sentence may be a run-on sentence. Proofread it to see if it contains too many independent clauses or contains independent clauses that have been combined without conjunctions or punctuation. Look at the "Writer's Handbook" for advice about correcting run-on sentences.

PAGE 24

Possessive This word may be a plural noun and may not need an apostrophe.

Missing "," You may need to place a comma after this word.

Verb This verb may be incorrect. Proofread the sentence to make sure you have used the correct form of the verb.

Article Error You may need to remove this article.

Article Error You may need to use an article before this word.

PAGE 25

P/V You have used the passive voice in this sentence. Depending upon what you wish to emphasize in the sentence, you may want to revise it using the active voice.

Article Error You may need to use an article before this word.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

PAGE 26

Verb This verb may be incorrect. Proofread the sentence to make sure you have used the correct form of the verb.

Missing "," You may need to place a comma after this word.

PAGE 27

Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

Dup. You have typed two **identical words** in a row. You may need to delete one of them.

PAGE 28

Missing "?" Remember to use a question mark at the end of a question.

PAGE 29

PAGE 30

Missing "," You may need to place a comma after this word.

Article Error You may need to use an article before this word.

Missing "," You may need to place a comma after this word.

PAGE 31

Article Error You may need to use an article before this word.

PAGE 32

Missing "," You may need to place a comma after this word.

Missing "," You may need to place a comma after this word.

Article Error You may need to remove this article.

PAGE 33

Article Error You may need to remove this article.

Article Error You may need to remove this article.

PAGE 34

Article Error You may need to remove this article.

Prep. You may be using the wrong preposition.

Missing "," You may need to place a comma after this word.

Article Error You may need to remove this article.

PAGE 35

PAGE 36
