Novel HealthCare Framework for Real Time Detection of Cardiac Arrest with the Application of AI Using ANN

Vivekananda Journal of Research
January - June 2022, Vol. 12, Issue 1, 114-131
ISSN 2319-8702(Print)
ISSN 2456-7574(Online)
Peer Reviewed Refereed Journal

https://vips.edu/journal/

Nasmin Jiwani*
Ketan Gupta**

ABSTRACT

Cardiovascular illnesses are the leading cause of mortality worldwide, killing an estimated 27.9 million people each year, accounting for 31% of all fatalities worldwide. Cardiovascular disorders are a common cause of heart failure. It is distinguished by the heart's inability to provide an appropriate amount of blood to the body. All main bodily processes are affected when there is insufficient blood flow. Heart failure is a disease or set of symptoms that causes the heart to weaken. The Main Findings of the research study lies on the three major factors First, the timing of the patient's follow-up appointment for the condition is critical, since early detection of a cardiovascular problem and treatment decreases the likelihood of a death. It has an inverse relationship. The second most significant aspect is the ejection fraction. It is to be anticipated given that it is essentially the efficiency of the heart. At last, the patient's age is the third most linked characteristic. Clearly, as one age, the heart's function deteriorates. The Data is Modeled Using Machine learning with ANN and Accuracy about 80% obtained shows the Framework is fairly useful for the Detection of Cardiac Arrest. Further the Accuracy can be increased to 90-95% by using deep learning Models

Keyword: *cardiovascular, heart, blood, symptoms*

INTRODUCTION

The term "heart disease" refers to a broad variety of diseases that affect the heart. Worldwide, cardiovascular illnesses now account for 27.9 million deaths per year, according

---

* Research Scientist, Department of IT, University of the Cumberlands, USA. Email: nasminjiwani@gmail.com
** Research Scientist, Department of IT, University of the Cumberlands, USA. Email: ketan1722@gmail.com

to World Health Organization data. The increased risk of heart disease can be attributed to a variety of unhealthy behaviors, such as high cholesterol, obesity, a rise in triglyceride levels, and hypertension. The American Heart Association mentions some symptoms, such as difficulty sleeping, an increase and drop in heart rate, swollen legs, and in rare cases, rapid weight gain of 1-2 kg per day.

All of these symptoms are similar to those of various diseases, such as those that affect the elderly, making a precise diagnosis difficult to come by. This, in turn, increases the risk of death in the near future.

However, with the passage of time, an increasing amount of research data as well as hospital patient records are becoming available. There are several free resources available for gaining access to patient information and conducting research so that various computer technologies may be utilized to diagnose patients correctly and discover the disease early enough to save their lives. Machine learning and artificial intelligence are increasingly well-known for their importance in the medical business.

When diagnosing the illness, several machine learning and deep learning models come in handy. Machine learning algorithms make it simple to do a comprehensive genetic data analysis. Models may be taught to anticipate knowledge pandemics, and medical records can be converted and studied more extensively to improve predictions.

Different machine learning algorithms have been used to classify and predict heart disease diagnoses in numerous researches. As a result of using a machine learning technique known as CART (Classification and Regression), doctors may now utilize an automated classifier to determine whether patients are at high and low risk of developing congestive heart failure.

Rahhal et al. Propose an electrocardiogram (ECG) technique for increasing performance by utilizing deep neural networks to choose and then use the best information. Guidi et al. then provide a clinical decision support system for early detection of heart failure. They compared a variety of machine learning and deep learning models, particularly neural networks, such as support vector machines, random forests, and CART algorithms. There was a difference in accuracy between random forest and CART (87.6 percent) and all other methods employed in the categorization.

Zhang et al. identified the NYHA HF class with 93.37 percent accuracy using a natural language processing technique combined with a rule-based approach. Researchers Parthiban and Srivatsa utilised a 94.60 percent accuracy rate of SVM methods to diagnose diabetes and subsequently forecast heart disease in patients. The features collected were common such as the patient’s blood sugar level, age at diagnosis, and systolic and diastolic blood pressure data.

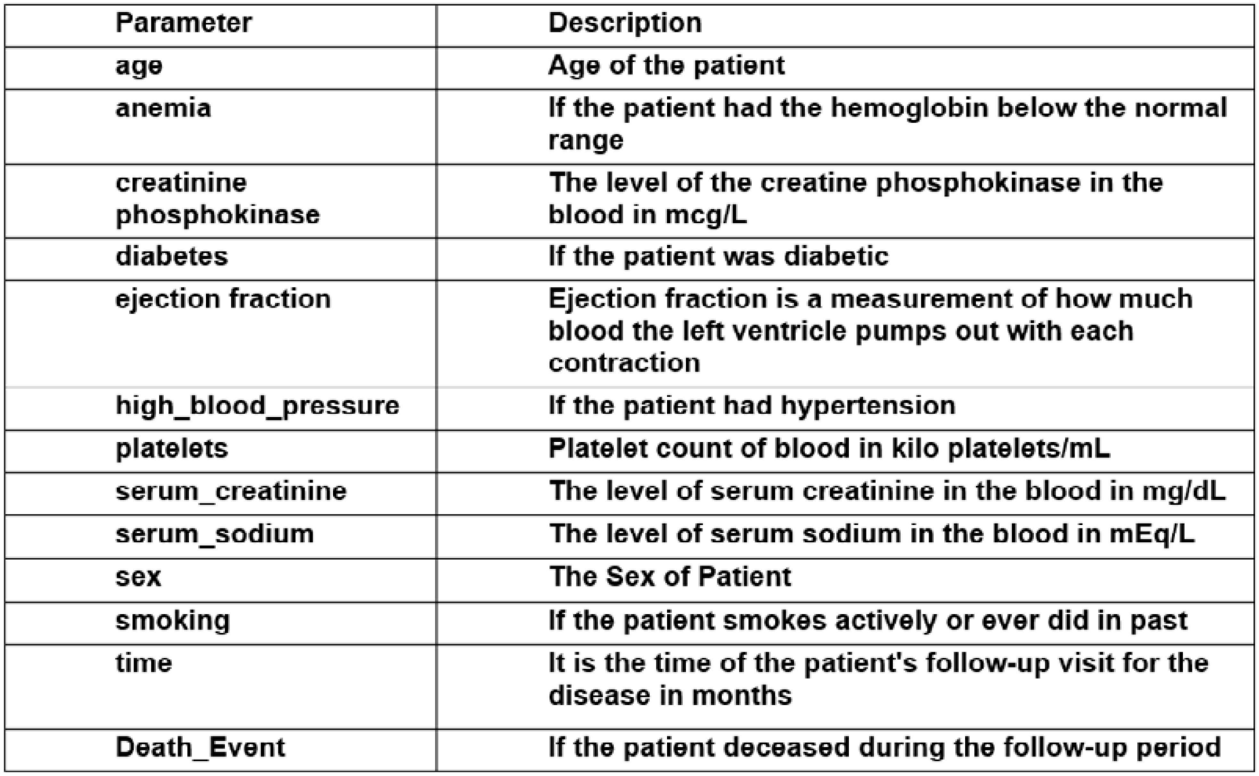
Data dimensionality is a typical difficulty in machine learning; the datasets we use include a lot of data and sometimes we can’t see it even in 3D, which is also known as the curse of dimensionality. As a result, processing this data necessitates a large amount of memory, and data might expand rapidly, leading to overfitting. This dataset may be made more efficient by using weighted features to reduce redundancy. This results in a reduction in processing time.

It is possible to reduce the dataset’s dimensionality by employing feature engineering and feature selection approaches to exclude material that isn’t of critical relevance.

METHODOLOGY

The data set used for the analysis and prediction is shown in the Table below. It consists of Various parameters like age, Anemia, with their description as shown in the Table 1

Table 1 Description of Dataset



Parameter	Description
age	Age of the patient
anemia	If the patient had the hemoglobin below the normal range
creatinine phosphokinase	The level of the creatine phosphokinase in the blood in mcg/L
diabetes	If the patient was diabetic
ejection fraction	Ejection fraction is a measurement of how much blood the left ventricle pumps out with each contraction
high_blood_pressure	If the patient had hypertension
platelets	Platelet count of blood in kilo platelets/mL
serum_creatinine	The level of serum creatinine in the blood in mg/dL
serum_sodium	The level of serum sodium in the blood in mEq/L
sex	The Sex of Patient
smoking	If the patient smokes actively or ever did in past
time	It is the time of the patient's follow-up visit for the disease in months
Death_Event	If the patient deceased during the follow-up period

DATA PREPARATION AND ANALYSIS

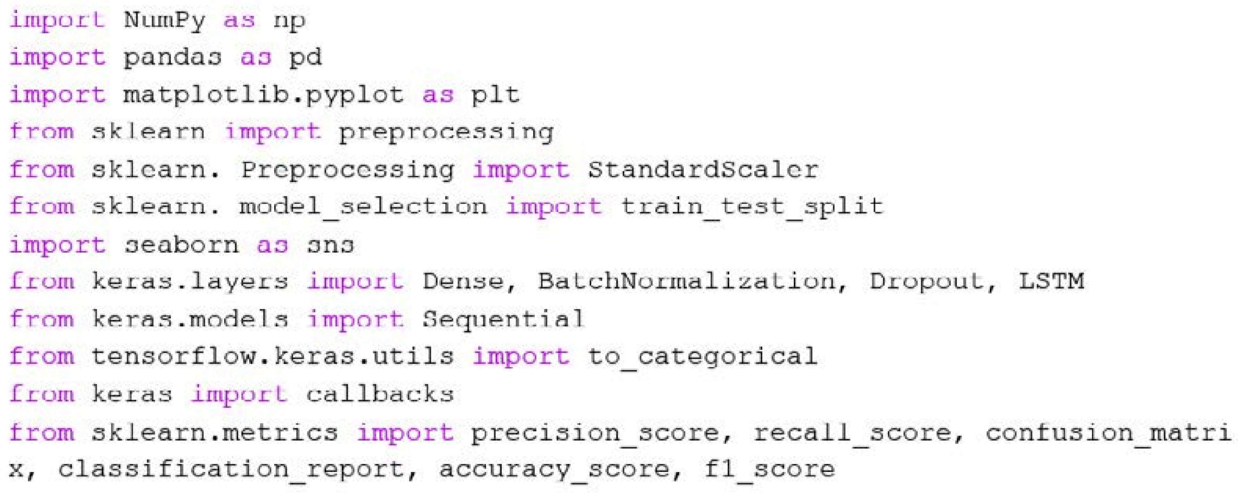
There are no null values in the dataset. However, numerous outliers must be handled correctly, since the dataset is not adequately distributed. There were two techniques utilized. One without outliers and a feature selection procedure, as well as immediately applying the data to machine learning algorithms, had unimpressive results. However, after utilizing the normal distribution of the dataset to overcome the over fitting problem and then employing Isolation Forest to find outliers, the results obtained are extremely encouraging. Various charting approaches were employed to assess the skewness of the data, locate outliers, and examine the data distribution.

LIBRARIES AND PACKAGES

A computer's ability to learn from diverse types of data is called machine learning, and it is a branch of computer science. According to Arthur Samuel, "Machine learning is the field of research that provides computers the ability to learn without being explicitly taught." They're commonly employed to deal with a wide range of issues in one's life.

People used to do Machine Learning jobs by manually coding all the algorithms and mathematical and statistical formulas in the past. It took a long time, and it was a painstaking procedure. The various Machine learning packages used in this research are given below

Figure 1 Various Libraries used



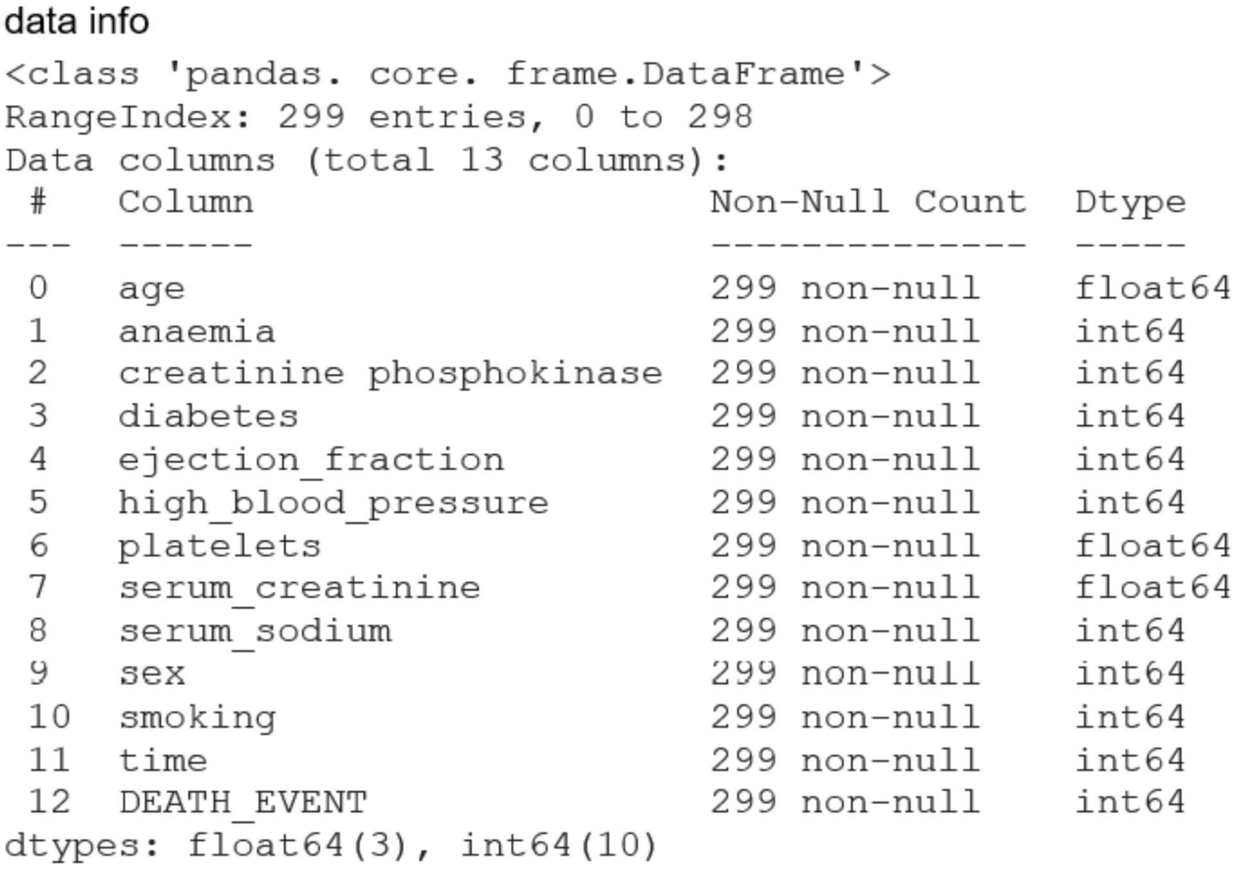
```
import NumPy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn. Preprocessing import StandardScaler
from sklearn. model_selection import train_test_split
import seaborn as sns
from keras.layers import Dense, BatchNormalization, Dropout, LSTM
from keras.models import Sequential
from tensorflow.keras.utils import to_categorical
from keras import callbacks
from sklearn.metrics import precision_score, recall_score, confusion_matri
x, classification_report, accuracy_score, f1_score
```

DATA INFORMATION

DataInfo is a machine learning function which is used to collect the gist of the dataset its count its Datatype . The information obtained after executing data.info is shown below

Data info

Table 2 data information



data info

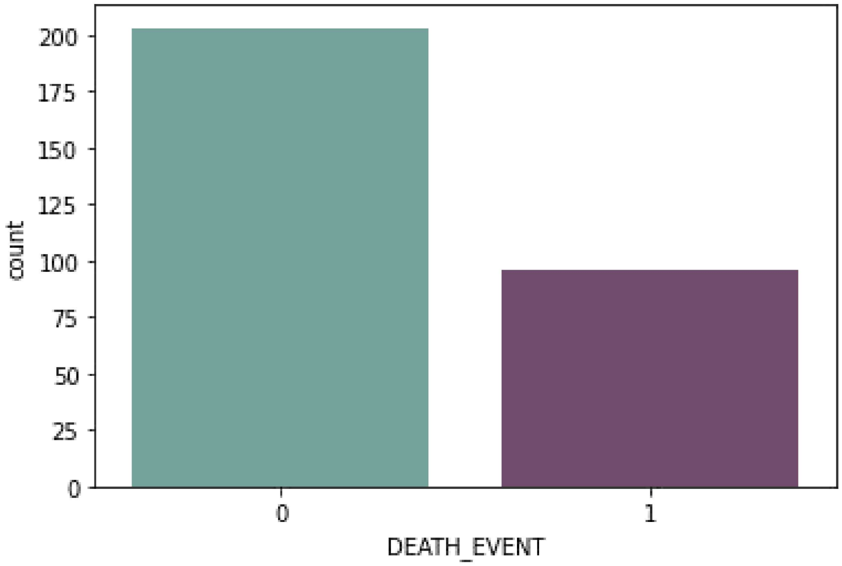
```
<class 'pandas. core. frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 13 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   age                       299 non-null    float64
 1   anaemia                   299 non-null    int64
 2   creatinine phosphokinase  299 non-null    int64
 3   diabetes                  299 non-null    int64
 4   ejection_fraction         299 non-null    int64
 5   high_blood_pressure       299 non-null    int64
 6   platelets                 299 non-null    float64
 7   serum_creatinine          299 non-null    float64
 8   serum_sodium              299 non-null    int64
 9   sex                       299 non-null    int64
 10  smoking                   299 non-null    int64
 11  time                      299 non-null    int64
 12  DEATH_EVENT               299 non-null    int64
dtypes: float64(3), int64(10)
```

CHECKING OF IMBALANCE OF DATA

A dataset with an uneven distribution of classes is inherently unbalanced. A dataset, on the other hand, is considered to be unbalanced when the number of samples of a problem class differs significantly, or even wildly. To check this the following command is executed and the result obtained. The result shows the data is balanced and ready for further processing

```
cols= ["#6daa9f","#774571"]
sns. count plot (x= data["DEATH_EVENT"], palette= cols)
```

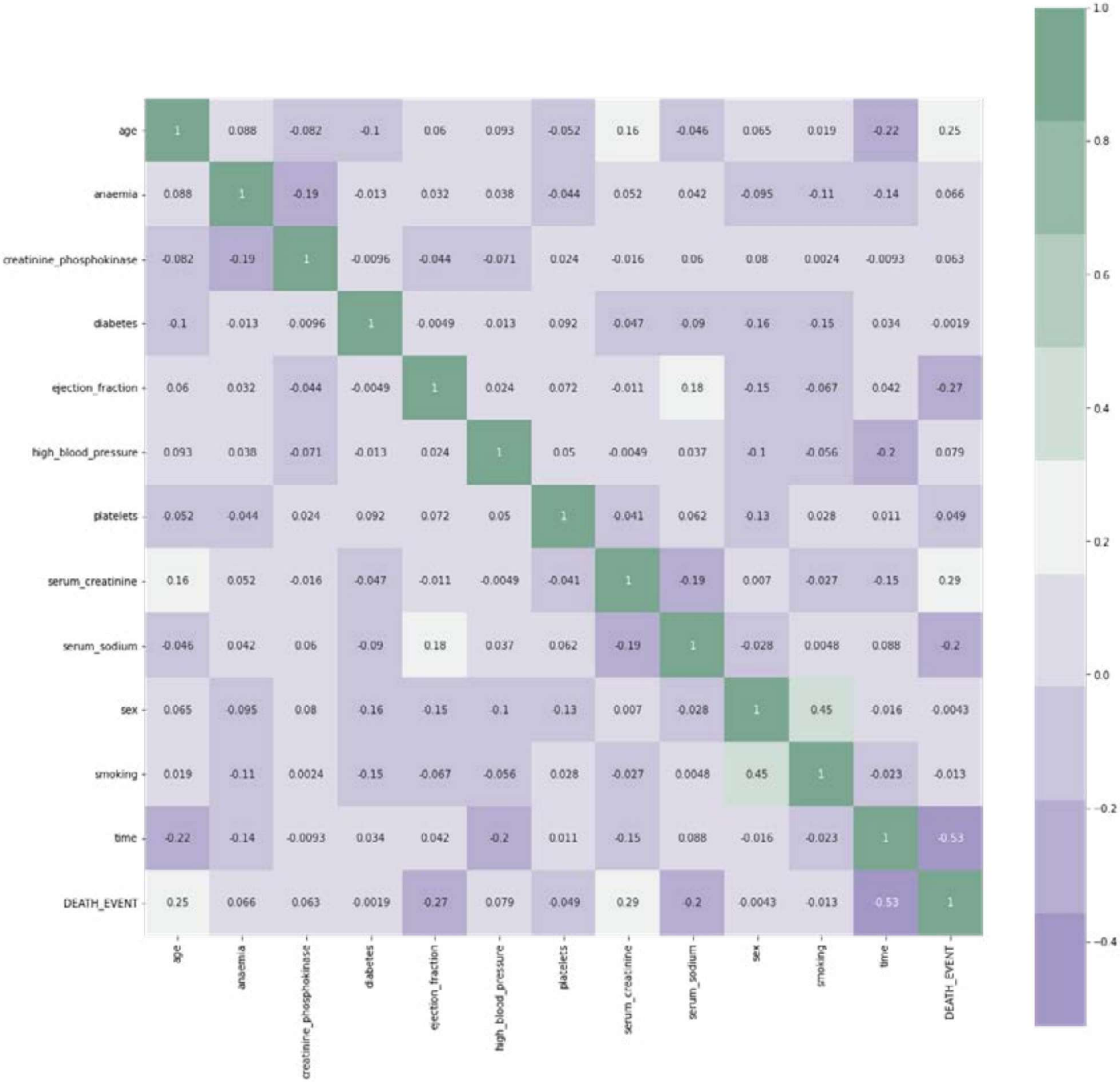
Figure 2 Figure to show the balancing of data



HEAT MAP

Data visualization approach that uses a heat map (or heatmap) to illustrate the magnitude of a phenomena in two dimensions as a colour map. The change in colour might be via hue or intensity, providing the reader clear visual signals about how the phenomena is grouped or fluctuates across space. The practice of evaluating and analyzing heat map data to get insights about user engagement on a page is known as heat map analysis. This data analysis can result in better site designs that have lower bounce rates, fewer drop-offs, more pageviews, and higher conversion rates. The heat map between all the parameters are plotted and shown in the Fig. Below . The colour and value of heat map tell the correlation between the parameters it is found that most of the values are positive and fairly correlated or depending upon each other. Platelets and diabatic are the two major factor which have high dependency to check the probability of cardiac arrest or not.

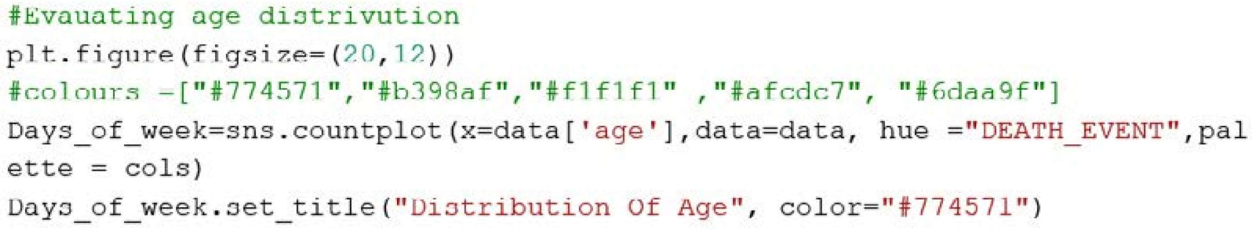
Figure 3 Heat Map for the given data



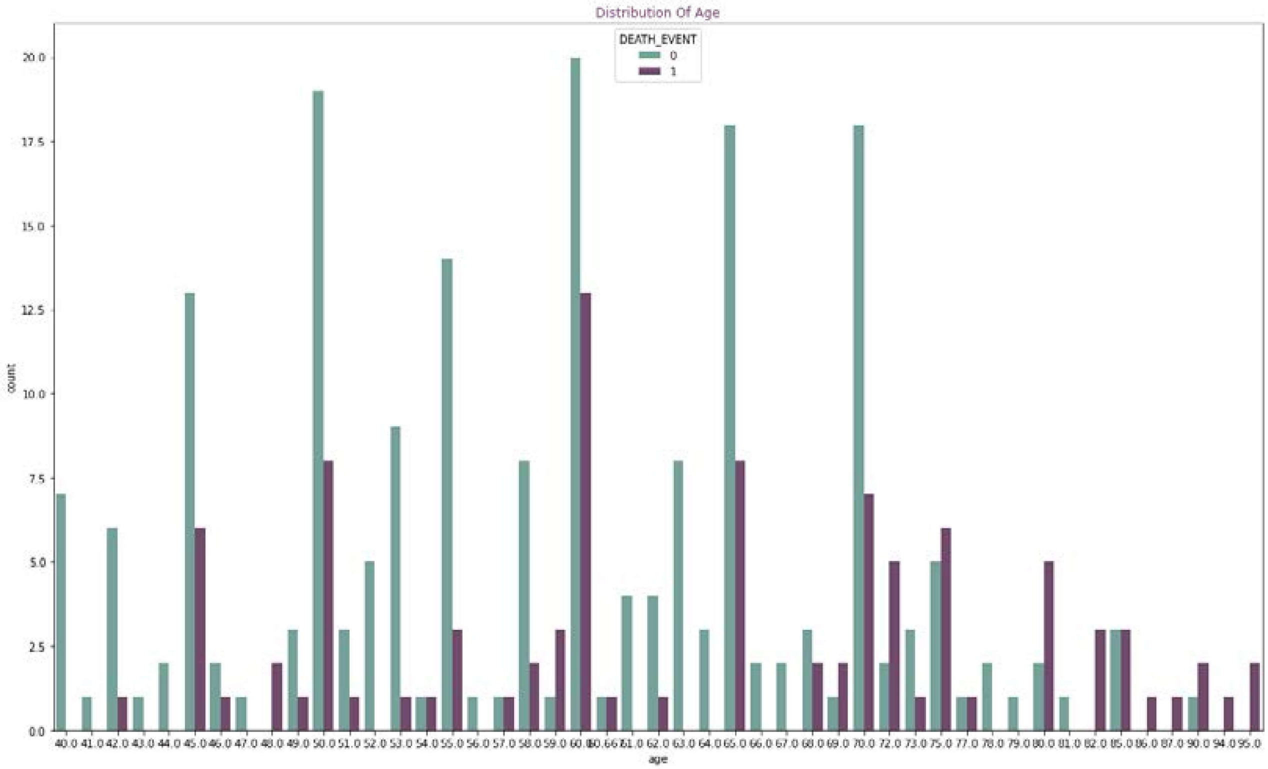
	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
age	1	0.088	-0.082	-0.1	0.06	0.093	-0.052	0.16	-0.046	0.065	0.019	-0.22	0.25
anaemia	0.088	1	-0.19	-0.013	0.032	0.038	-0.044	0.052	0.042	-0.095	-0.11	-0.14	0.066
creatinine_phosphokinase	-0.082	-0.19	1	-0.0096	-0.044	-0.071	0.024	-0.016	0.06	0.08	0.0024	-0.0093	0.063
diabetes	-0.1	-0.013	-0.0096	1	-0.0049	-0.013	0.092	-0.047	-0.09	-0.16	-0.15	0.034	-0.0019
ejection_fraction	0.06	0.032	-0.044	-0.0049	1	0.024	0.072	-0.011	0.18	-0.15	-0.067	0.042	-0.27
high_blood_pressure	0.093	0.038	-0.071	-0.013	0.024	1	0.05	-0.0049	0.037	-0.1	-0.056	-0.2	0.079
platelets	-0.052	-0.044	0.024	0.092	0.072	0.05	1	-0.041	0.062	-0.13	0.028	0.011	-0.049
serum_creatinine	0.16	0.052	-0.016	-0.047	-0.011	-0.0049	-0.041	1	-0.19	0.007	-0.027	-0.15	0.29
serum_sodium	-0.046	0.042	0.06	-0.09	0.18	0.037	0.062	-0.19	1	-0.028	0.0048	0.088	-0.2
sex	0.065	-0.095	0.08	-0.16	-0.15	-0.1	-0.13	0.007	-0.028	1	0.45	-0.016	-0.0043
smoking	0.019	-0.11	0.0024	-0.15	-0.067	-0.056	0.028	-0.027	0.0048	0.45	1	-0.023	-0.013
time	-0.22	-0.14	-0.0093	0.034	0.042	-0.2	0.011	-0.15	0.088	-0.016	-0.023	1	-0.53
DEATH_EVENT	0.25	0.066	0.063	0.0019	-0.27	0.079	-0.049	0.29	-0.2	-0.0043	-0.013	-0.53	1

COUNTPLOT

In order to show connections between category data and at least one numerical variable, bar graphs are commonly used seaborn. In a count plot, the dependent variable (the number of data points) is the independent variable (the data points). The count plot for age and death event is drawn to analyze which age group is highly effected by cardiac arrest . It is observed from the graph the cardiac arrest is most profound in age group of 60 and it is very less in the children having age less than 10



```
#Evauating age distrivution
plt.figure(figsize=(20,12))
#colours =["#774571","#b398af","#f1f1f1" ,"#afcdc7", "#6daa9f"]
Days_of_week=sns.countplot(x=data['age'],data=data, hue ="DEATH_EVENT",pal
ette = cols)
Days_of_week.set_title("Distribution Of Age", color="#774571")
```



KDE

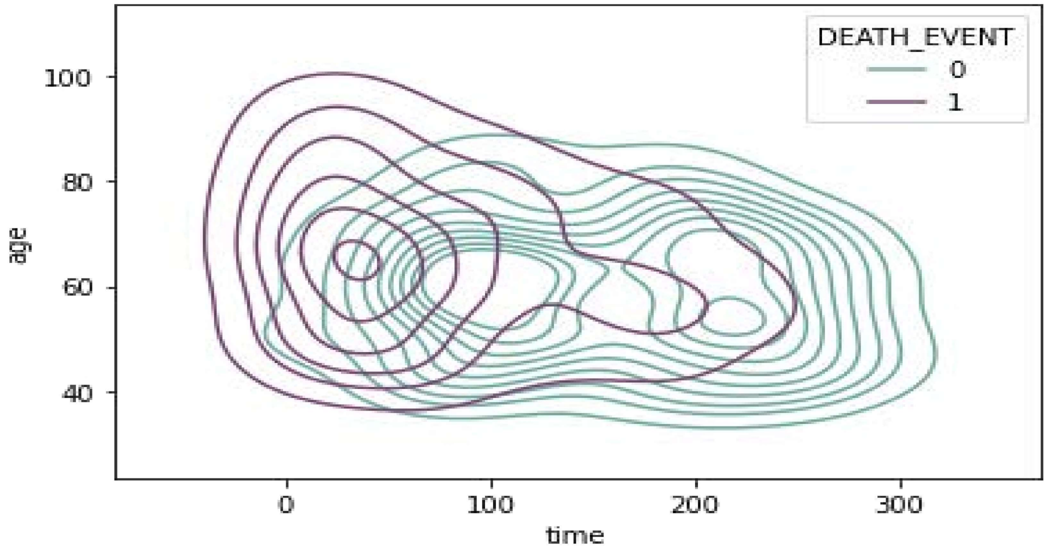
A kernel density estimate (KDE) plot, similar to a histogram, is a method for displaying the distribution of observations in a dataset. KDE uses a continuous probability density curve in one or more dimensions to describe the data.

In comparison to a histogram, KDE can generate a plot that is less crowded and more interpretable, especially when showing several distributions. However, it has the potential to produce distortions if the underlying distribution is limited or not smooth. The quality of the depiction, like that of a histogram, is determined by the use of suitable smoothing settings.

The Kenal Density Estiate graph between age and death event is plot using following command and the visualization is represented in Fig.

```
sns.kdeplot(x=data["time"], y=data["age"], hue =data["DEATH_EVENT"], palet
te=cols)
```

Figure 4 Kernel density estimate Plot

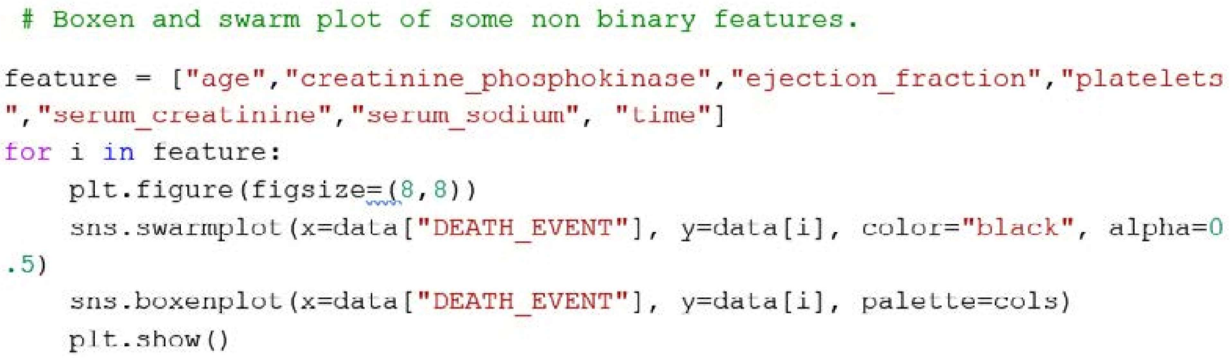


SWARM BOX PLOT

This function is identical to stripplot(), except the points are modified such that they do not overlap (just along the categorical axis). This provides a more accurate depiction of the value distribution, but it does not scale well to large numbers of data. This plot type is frequently referred to as a “beeswarm.”

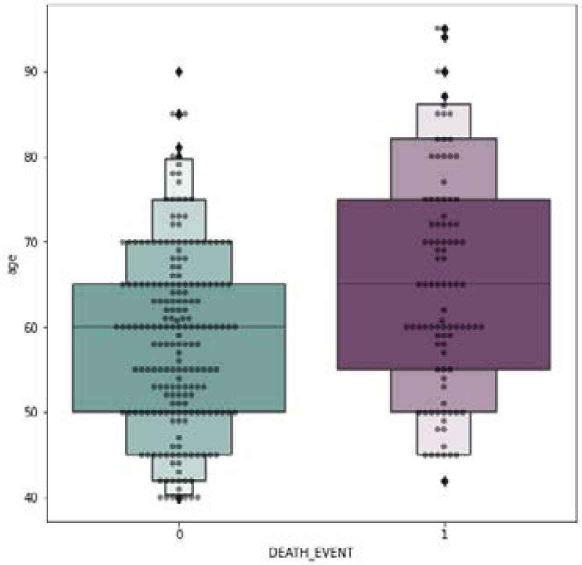
A swarm plot may be created on its own, but it also works well in conjunction with a box or violin plot when you want to display all observations as well as some representation of the underlying distribution.

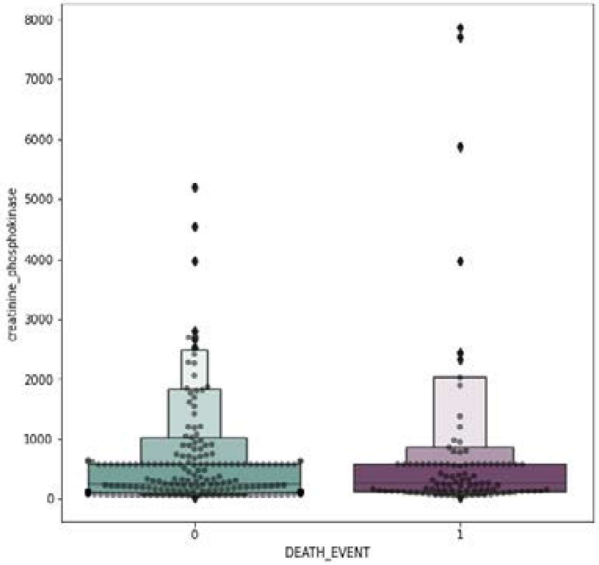
This type of plot was initially known as a “letter value” plot because it displays a high number of quantiles described as “letter values.” It’s comparable to a box plot in that it plots a nonparametric representation of a distribution with all characteristics corresponding to real data. It gives additional information about the shape of the distribution, particularly in the tails, by plotting more quantiles. The Swarm and box plot of Death event is ploted w.rt to the features like age, cp, ef, platelets ,sc,ss and time the command used to plot is given below , and the plots are shown in the Fig.

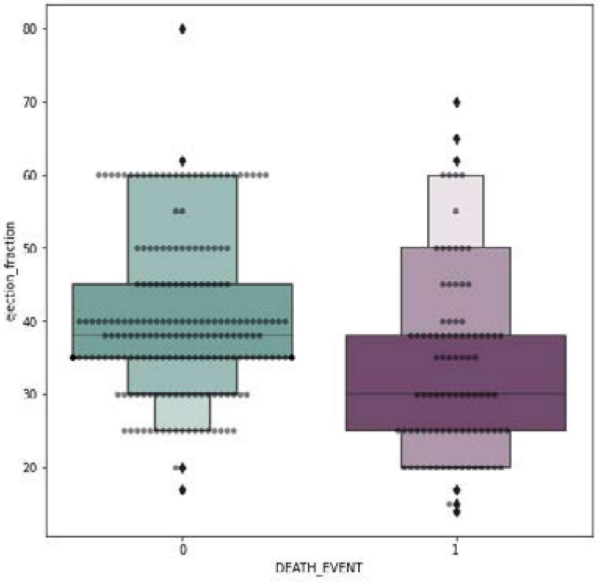


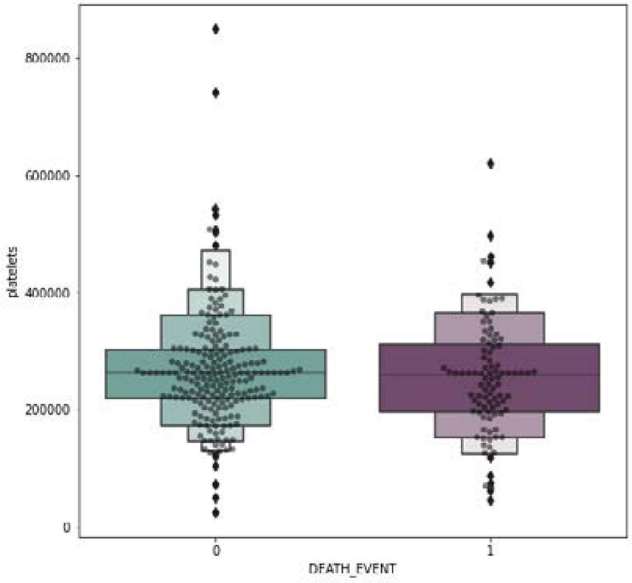
```
# Boxen and swarm plot of some non binary features.

feature = ["age","creatinine_phosphokinase","ejection_fraction","platelets
","serum_creatinine","serum_sodium", "time"]
for i in feature:
    plt.figure(figsize=(8,8))
    sns.swarmplot(x=data["DEATH_EVENT"], y=data[i], color="black", alpha=0
.5)
    sns.boxenplot(x=data["DEATH_EVENT"], y=data[i], palette=cols)
    plt.show()
```









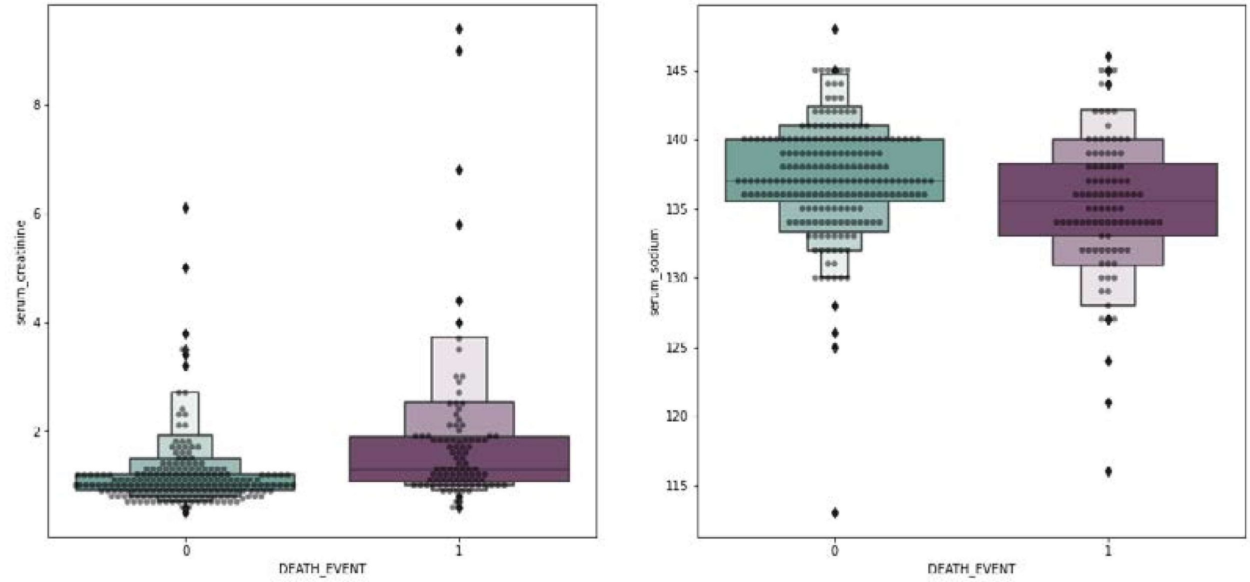


Figure 5 Swarm box plot of various parameters with Death_event

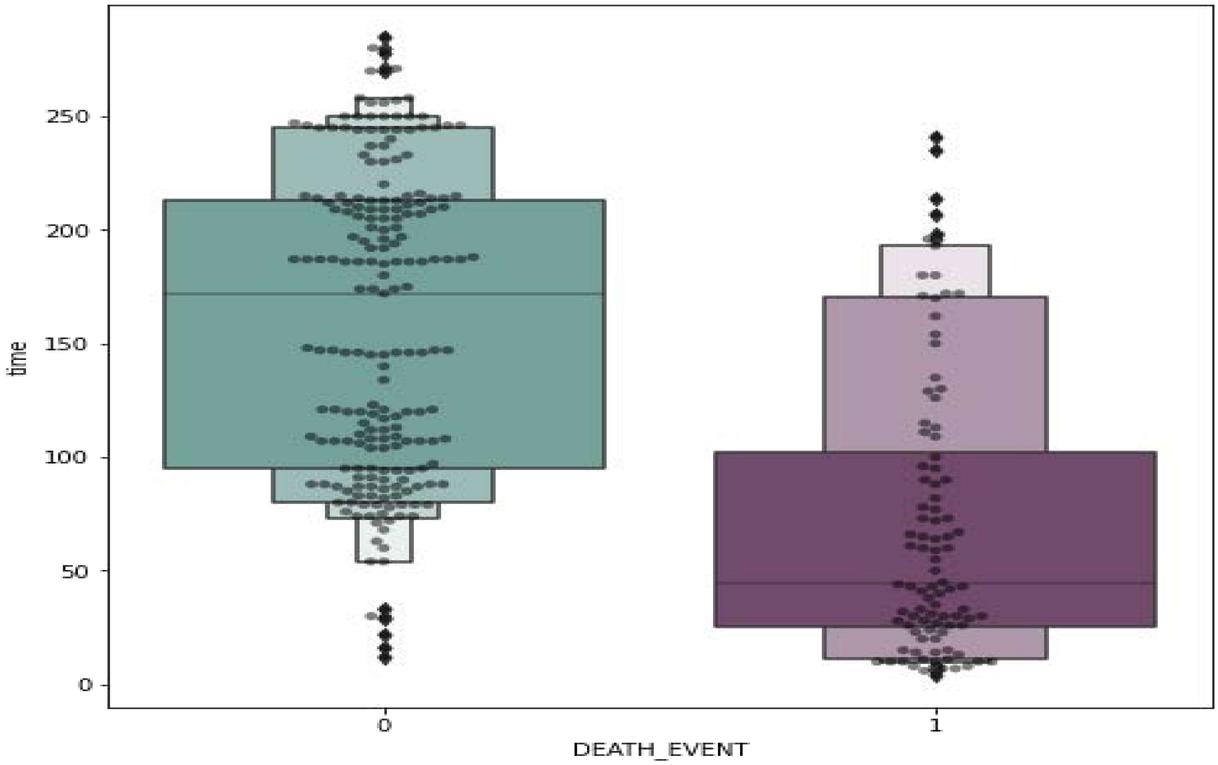
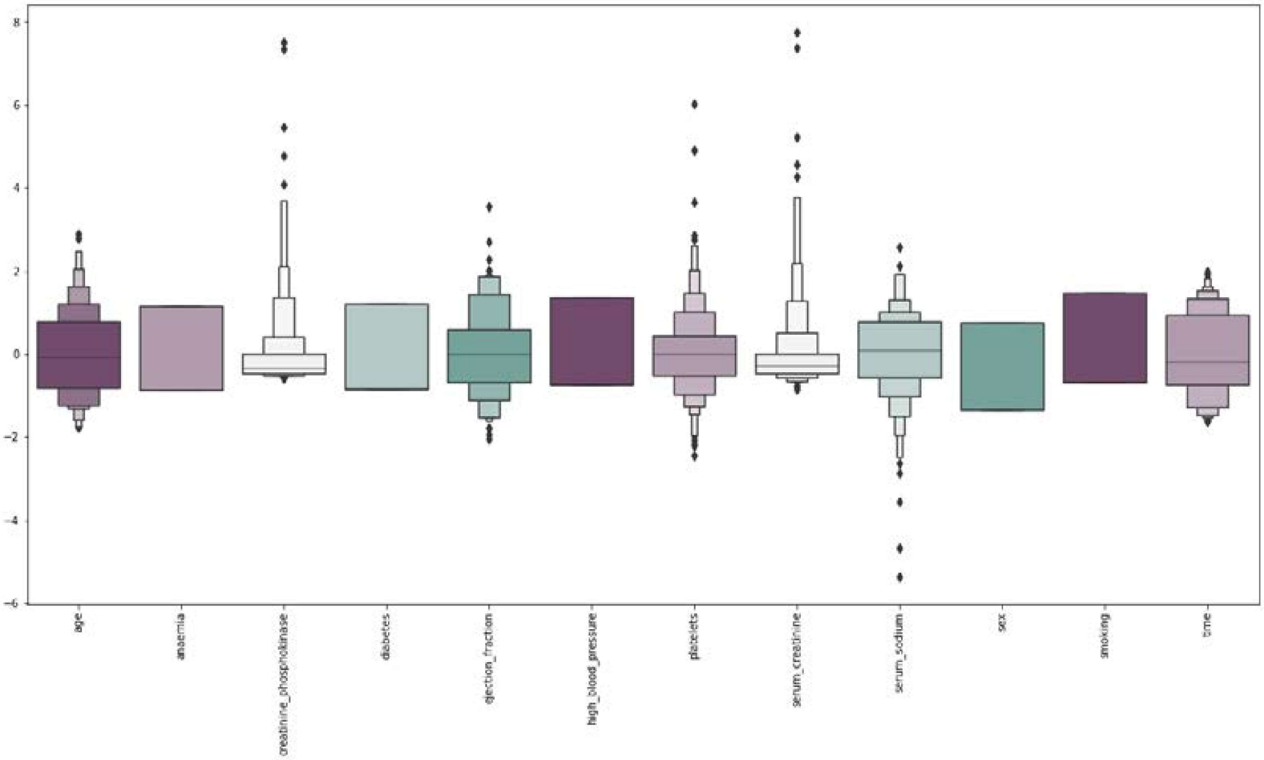


Figure 6 Combined Box Plot with parameters



DESCRIBE COMMAND

The describe() function is used to calculate statistical statistics from the numerical values of a Series or DataFrame, such as percentile, mean, and standard deviation. It examines both numeric and object series, as well as DataFrame column sets containing mixed data types. Various statistical results are described in the given table and the 25 % and 75% value indicate lower and upper quartile va;ues while 50% indicate the value of median along with min and max values for each parameters,

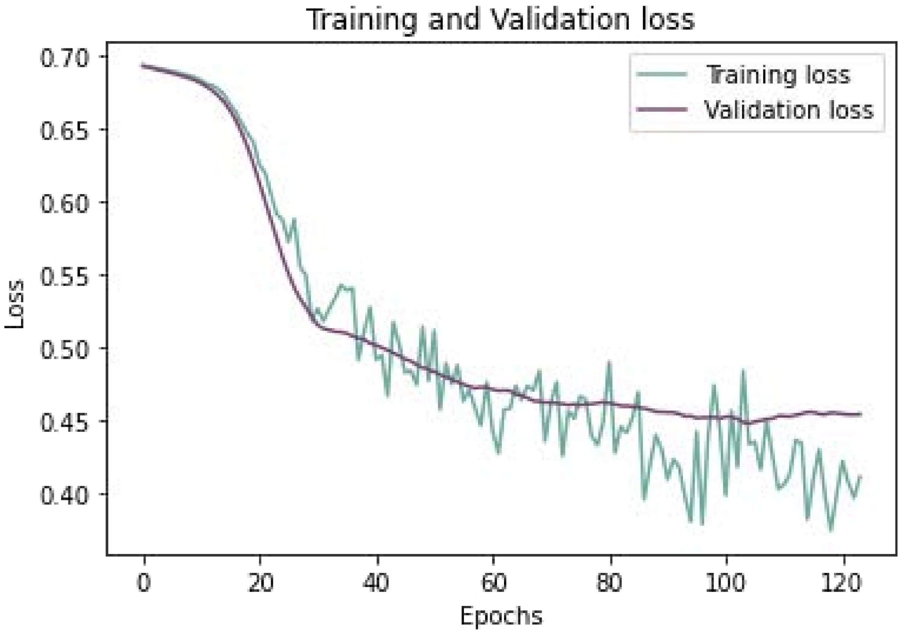
Table 3 Statistical analysis of the various parameters

	count	mean	std	min	25%	50%	75%	max
age	299.0	60.833893	11.894809	40.0	51.0	60.0	70.0	95.0
anaemia	299.0	0.431438	0.496107	0.0	0.0	0.0	1.0	1.0
creatinine_phosphokinase	299.0	581.839465	970.287881	23.0	116.5	250.0	582.0	7861.0
diabetes	299.0	0.418060	0.494067	0.0	0.0	0.0	1.0	1.0
ejection_fraction	299.0	38.083612	11.834841	14.0	30.0	38.0	45.0	80.0
high_blood_pressure	299.0	0.351171	0.478136	0.0	0.0	0.0	1.0	1.0
platelets	299.0	263358.029264	97804.236869	25100.0	212500.0	262000.0	303500.0	850000.0
serum_creatinine	299.0	1.393880	1.034510	0.5	0.9	1.1	1.4	9.4
serum_sodium	299.0	136.625418	4.412477	113.0	134.0	137.0	140.0	148.0
sex	299.0	0.648829	0.478136	0.0	0.0	1.0	1.0	1.0
smoking	299.0	0.321070	0.467670	0.0	0.0	0.0	1.0	1.0
time	299.0	130.260870	77.614208	4.0	73.0	115.0	203.0	285.0
DEATH_EVENT	299.0	0.321070	0.467670	0.0	0.0	0.0	1.0	1.0

TRAINING AND VALIDATION LOSS

Training loss + validation loss over time is one of the most commonly utilised metric combinations. The training loss measures how well the model fits the training data, whereas the validation loss measures how well the model fits new data.

Figure 7 Training and Validation Loss

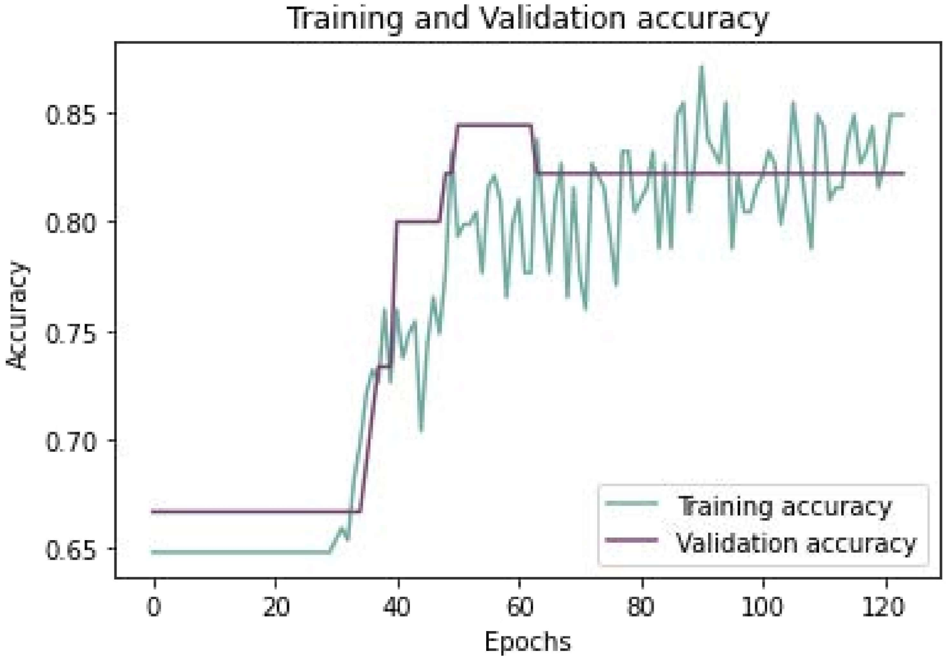


TRAINING AND VALIDATION ACCURACY

As a result, the test (or testing) accuracy typically refers to the validation (or “testing”) accuracy, which you compute on a data set that was not used for training but was used (during the training process) to validate (or “test”) the generalisation capacity of your model or to “early stop.”

With the help of the training data, we build the model and evaluate its results on both the training and validation sets (evaluation metric is accuracy). The accuracy of the training is 85%, but the accuracy of the validation is just 83%.

Figure 8 Training and Validation Accuracy



CONFUSION MATRIX

A confusion matrix is a table that shows how well a classification model (or “classifier”) performs on a set of test data for which the real values are known. The Confusion Matrix for the dataset is shown in Fig. the True Positve Value is 0.65 while false positive is 0.11 and true negative value is 0.0093 and false negative is 0.11 these value indicates the model is best fit . The Accuracy score with precision recall and f1- score is shown in Table

Figure 9 Confusion Matrix

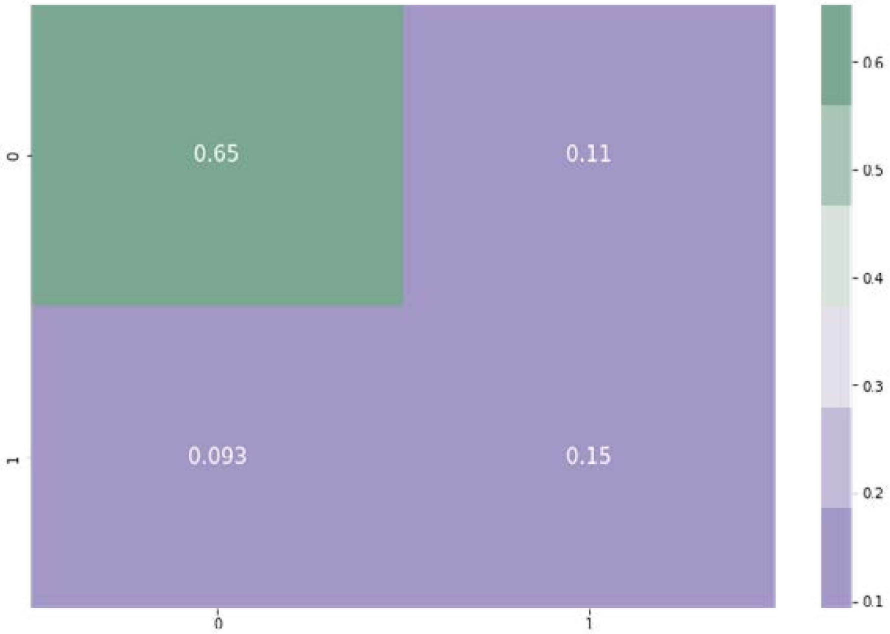
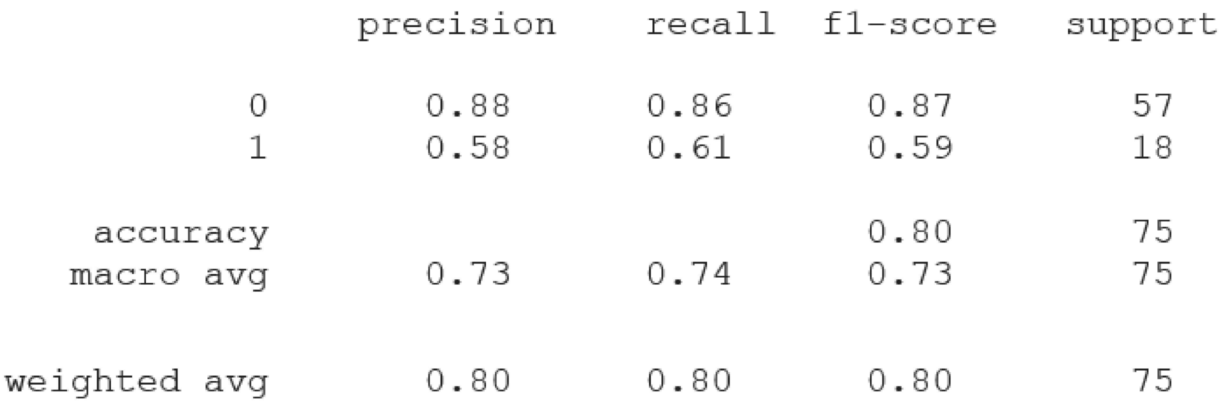


Table 4 Accuracy table with precision recall and fi-score



```
              precision    recall  f1-score   support

           0       0.88      0.86      0.87        57
           1       0.58      0.61      0.59        18

    accuracy                           0.80        75
   macro avg       0.73      0.74      0.73        75

weighted avg       0.80      0.80      0.80        75
```

CONCLUSION AND FUTURE SCOPE

In this research paper a novel Framework for Real Time Detection of Cardiac Arrest with the application of AI using ANN is proposed. It is discovered that machine learning is much better alternative for cardiac arrest detection as it is fast and reliable. With the application of ANN, it is found that the validation accuracy is 83% and the training accuracy is 85% which is a fairly good contribution. Further the accuracy may be increased by increasing the data and using deep learning model. Various results using ANN like confusion matrix, precision, specificity, sensitivity, and F1 score are calculate and presented in the tabular form. The results present in the paper are fair and this research study will be very helpful for the researchers working in the same field

REFERENCES

World Health Organization, *Cardiovascular Diseases*, WHO, Geneva, Switzerland, 2020, American Heart Association, *Classes of Heart Failure*, American Heart Association, Chicago, IL, USA, 2020, American Heart Association, *Heart Failure*, American Heart Association, Chicago, IL, USA, 2020, Shalev-Shwartz S. and S. Ben-David. (2020). Understanding Machine Learning. *From Theory To Algorithms*, Cambridge University Press, Cambridge, UK,.

Hastie, T., R. Tibshirani, and J. Friedman.(2020). The Elements Of Statistical Learning, *Data Mining, Inference, and Prediction*, Springer, Cham, Switzerland,

Marsland, S. , (2020) "Machine learning," *An Algorithmic Perspective*, CRC Press, Boca Raton, FL, USA, 2020.

Melillo, P., N. De Luca, M. Bracale, and L. Pecchia. (2013). Classification Tree For Risk Assessment In Patients Suffering From Congestive Heart Failure Via Long-Term Heart Rate Variability.*IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 727–733,

Rahhal, M. M. A., Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager.(2016) Deep Learning Approach For Active Classification Of Electrocardiogram Signals. *Information Sciences*, vol. 345, pp. 340–354,

Guidi, G., M. C. Pettenati, P. Melillo, and E. Iadanza. (2014). A Machine Learning System To Improve Heart Failure Patient Assistance. *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1750–1756..

Liu H., and H. Motoda, (1998). *Feature Extraction, Construction and Selection*. Springer, Cham, Switzerland,

Dun, B. E. Wang, and S. Majumder. (2016). Heart Disease Diagnosis On Medical Data Using Ensemble Learning.

Singh, R. S., B. S. Saini, and R. K. Sunkaria. (2018). Detection Of Coronary Artery Disease By Reduced Features And Extreme Learning Machine. *Medicine and Pharmacy Reports*, vol. 91, no. 2, pp. 166–175,

Yaghouby, F , A. Ayatollahi, and R. Soleimani. (2009) "Classification Of Cardiac Abnormalities Using Reduced Features Of Heart Rate Variability Signal. *World Applied Sciences Journal*, vol. 6, no. 11, pp. 1547–1554,.

Asl, B. M., S. K. Setarehdan, and M. Mohebbi.(2008). "Support Vector Machine-Based Arrhythmia Classification Using Reduced Features Of Heart Rate Variability Signal. *Artificial Intelligence in Medicine*, vol. 44, no. 1, pp. 51–64Guyon, I., S. Gunn, M. Nikravesh, and L. Zadeh(2008). *Feature Extraction: Foundations and Applications*. Springer, Cham, Switzerland..

Rajagopal R., and V. Ranganathan. (2017). Evaluation Of Effect Of Unsupervised Dimensionality Reduction Techniques On Automated Arrhythmia Classification. *Biomedical Signal Processing and Control*, vol. 34, pp. 1–8.Zhang, D., L. Zou, X. Zhou, and F. He. (2018). Integrating Feature Selection And Feature Extraction Methods With Deep Learning To Predict Clinical Outcome Of Breast Cancer. *IEEE Access*, vol. 6, pp. 28936–28944

Negi, S., Y. Kumar, and V. M. Mishra. (2016). Feature Extraction And Classification For EMG Signals Using Linear Discriminant Analysis, *Proceedings of the 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall)*, IEEE, Bareilly, India, September 2016.

Zhang, R., S. Ma, L. Shanahan, J. Munroe, S. Horn, and S. Speedie. (2017). Automatic Methods To Extract New York Heart Association Classification From Clinical Notes. *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1296–1299, IEEE, Kansas City, MO, USA, November

Parthiban G., and S. K. Srivatsa. (2012).Applying Machine Learning Methods In Diagnosing Heart Disease For Diabetic Patients. *International Journal of Applied Information Systems*, vol. 3, no. 7, pp. 25–30, 2012.

Keogh E., and A. Mueen. (2017). Curse Of Dimensionality. in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. I. Webb, Eds., Springer, Cham, Switzerland,Wettschereck D., and T. G. Dietterich. (1995). An Experimental Comparison Of The Nearest-Neighbor And Nearest-Hyperrectangle Algorithms. *Machine Learning*, vol. 19, no. 1, pp. 5–27

Wettschereck, D., D. W. Aha, and T. Mohri. (1997) A Review And Empirical Evaluation Of Feature Weighting Methods For A Class Of Lazy Learning Algorithms. *Lazy Learning*, vol.

11, no. 1/5, pp. 273–314,Yang M.-S., and Y. Nataliani. (2018). A Feature-Reduction Fuzzy Clustering Algorithm Based On Feature-Weighted Entropy. *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 817–835,

Chen, R., N. Sun, X. Chen, M. Yang, and Q. Wu. (2018). “Supervised Feature Selection With A Stratified Feature Weighting Method. *IEEE Access*, vol. 6, pp. 15087–15098Imani M. and H. Ghassemian.(2015). Feature Extraction Using Weighted Training Samples. *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1387–1391