iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46900

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

Prediction of Cardiac Arrhythmia using Machine Learning

Kavyshree B[1], Rakesh M D[2]
[1,2] *Department of Electronics and Communication Engineering, JSS Science and Technology University, Mysore*

Abstract: ***The Heart is one of the most important organ responsible for sustaining Human life. The Normal functioning of it is very important but the irregular functioning of it will causes few problems which may be classified as different heart disease. Arrhythmia an Irregular Heart Beat, which is considered as one of the Cardio Vascular Disease. Electrocardiogram (ECG) is the most preferred tool used to capture Heart Beat. Without taking proper pre-cautionary measures this may lead to sudden death, blood clots, heart failure, stroke, etc.. Machine learning is the study of computer algorithms. In this work by adopting Machine learning algorithms such as Logistic Regression, Decision Tree, SVM[Support Vector Machine]are done to foresee the Cardiac Arrhythmia. The data-sets are collected from UCI Repository & processed using python programming .From all the three applied algorithms the SVM model showed the better results of 91.41\% in terms of accuracy for 80/20 combinations of Train and Test data sets. Therefore from this work SVM model is considered as best algorithm for the prediction of Cardiac Arrhythmia.***

Keywords: Machine learning, Arrhythmia, ECG Intervals, Logistic Regression, Decission Tree, SVM [Support Vector Machine]

I. INTRODUCTION

Heart diseases are one in several diseases that affects a huge population. Anxiety is also a main reason for many people's heart attack.This unwanted heart attack and sudden death can be prevented by initial detection and getting treated regularly of arrhythmia which reduces the heart attack in the society and also avoids the loss of life.ECG is the most broadly utilized diagnosing gadget or instrument for capacity of heart Which is being recorded when cathodes set on the body that produces examples of the electrical drive of the heart. ECG signals are of 'P' waves, 'QRS' waves,'T' waves. The connection between these P waves, QRS waves, T waves and RR interms of time term and shape are required for looking at a heart understanding .The main uses of the HRV signal are the identification and classification of cardiac arrhythmia[2].

Nowadays society is affected by different chronic diseases. Nearly quarter (24.8 percent) of deaths in India is due to Cardio vascular problems said by the Global Burden of Disease[16].The most effective and affordable diagnostic method for assessing patients data with cardiac arrhythmias is the electrocardiogram (ECG).It takes a long time to manually diagnose arrhythmia beats since the ECG is nonlinear and complex. Similarly,it is challenging to discern minute fluctuations in time-domain parameters such as amplitude, segments, and intervals with the naked eye [3]

Arrhythmia is a type of abnormalities in heart beat where heart pumps excessively quick or too moderate which results in heart sicknesses. AI systems can be connected to improve exactness of heart arrhythmia order from ECG signals.Classification of heart arrhythmia relies upon the setting of use, information investigation pre-requisite of the predetermined patient for choosing a proper strategy. The proposed method gives a productive framework that arrange ECG signal into healthy or unhealthy[4].

The key factors that are examined during a heartbeat's normal beat phase includes the ECG Signals durations, and relationships to one another. The alterations in these indicators point to a heart condition that could develop for any reason. Arrhythmia is the broad term for all irregular beat phases, and some arrhythmias can be quite dangerous for the patient[6].

Clinical and biological data are increasingly being made available in digital form. These information varies from very few data points that are available for a wide number of people (such as demographics, blood tests, medications used, etc.) to much richer data that are only available for a limited number of patients.[9]

II. RELATED WORK

J P. Kelwade et.al, has used ANN as a classifier to foresee cardiac arrhythmia into five classes.The time series between 'R-R' interval is collected by adopting the ECG usage.The algorithm is implemented in MATLAB Version R2014a.[1]Mr.Santhana Krishnan.J ,Dr.Geetha.S et.al, they have classified into two data mining algorithms and were applied on the data-sets to foresee the occurences of having heart disease and were analyzed with classification model namely Naive Bayes Classifier and Decision tree

classification.[2]Nir Kalkstein, Yaron Kinar et.al,proposed that the Electrocardiogram (ECG) data easily can be transformed into digital format with approaches.[3]Pratiksha Shetgaonkar,Dr. Shailendra Aswale et.al has used Data Mining technique and gathered different types of heart diseases.In this paper they have considered Neural network, Naive bayes , Decision Tree for their study.[4]Aditya Methaila ,Prince Kansal has foreseen heart diseases using data mining approaches. The main approach used to foresee is KNN Algorithms,Decision Trees and Naive Bayes methods. This speaks about thirteen medical types.[5]Nasreen Sultana ,Yedukondalu Kamatham et.al,has used an efficient and most foreseen methods for relevant classification of cardiac arrhythmia using ANN. ECG beats are classified adopting Multiclass SVM classifier demonstrates most accuracy compared to other classifiers which makes the system more effective [6].

III. METHEDOLOGY

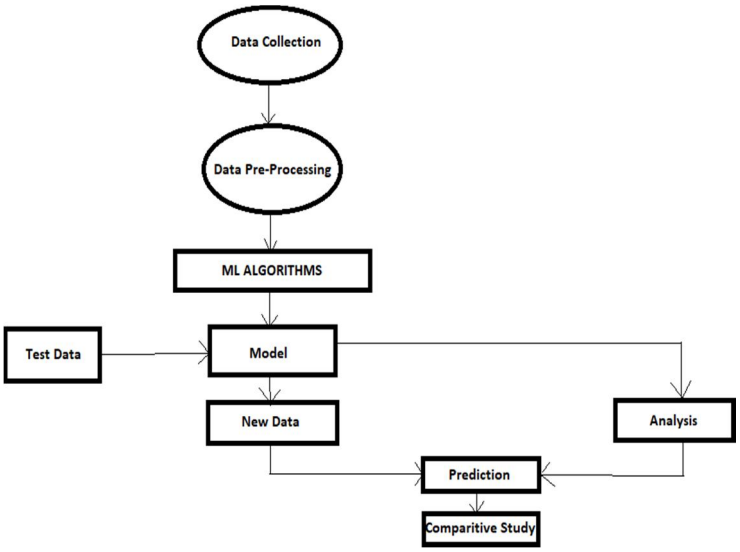


Fig 1 : Block Diagram of Proposed Work

The Figure 1 depicts the Block diagram of proposed system. The UCI is the repository where the data sets are collected from. Then the pre-processing of data is done. The pre-processed data is then divided into Train & Test data sets and given to model .The data obtained is analyzed and predicted. This work deals with different ML algorithms for the prediction of cardiac arrhythmia. Hence, these algorithms are differentiated on the basis of their efficiency and accuracy. So that the most accurate machine learning classifier can be found for arrhythmia prediction Study on Regular and Irregular ECG Signals

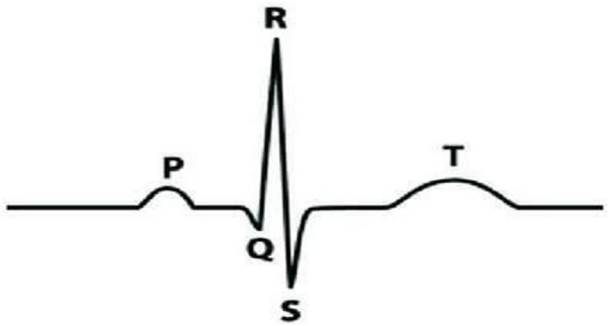


Fig. 2: Regular ECG Signal

The Figure 2 shows the Regular ECG signal which consists of Three parts that is P,QRS ,T. If these waves are obtained in a synchronous manner then it is said to be as a Normal or Regular ECG Signal ,or else it is considered as a Irregular Signal which is called by the name Arrhythmia an Irregular Hear Beat.

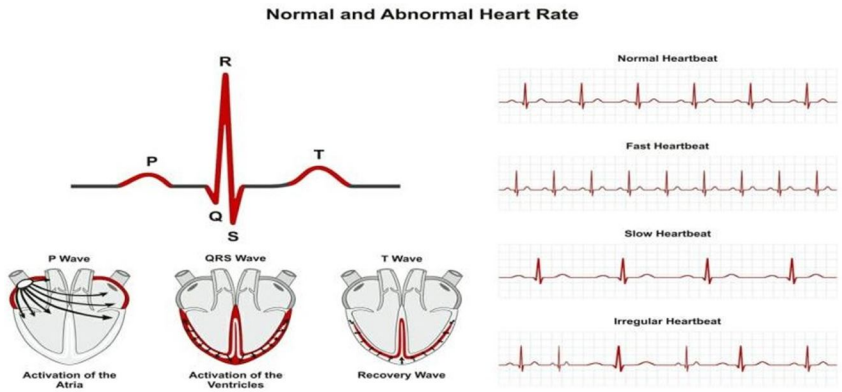


Fig. 3: Regular and Irregular Heart Rate Representation

The Figure 3 depicts the Regular and Irregular Heart Rate representation, where 'P' waves are caused due atrial depolarization that is when two atria contracting.The 'QRS' waves Complex is formed due to ventricular contraction .The 'T' waves are formed when ventricals relax.

A. Steps for ECG Interpretation

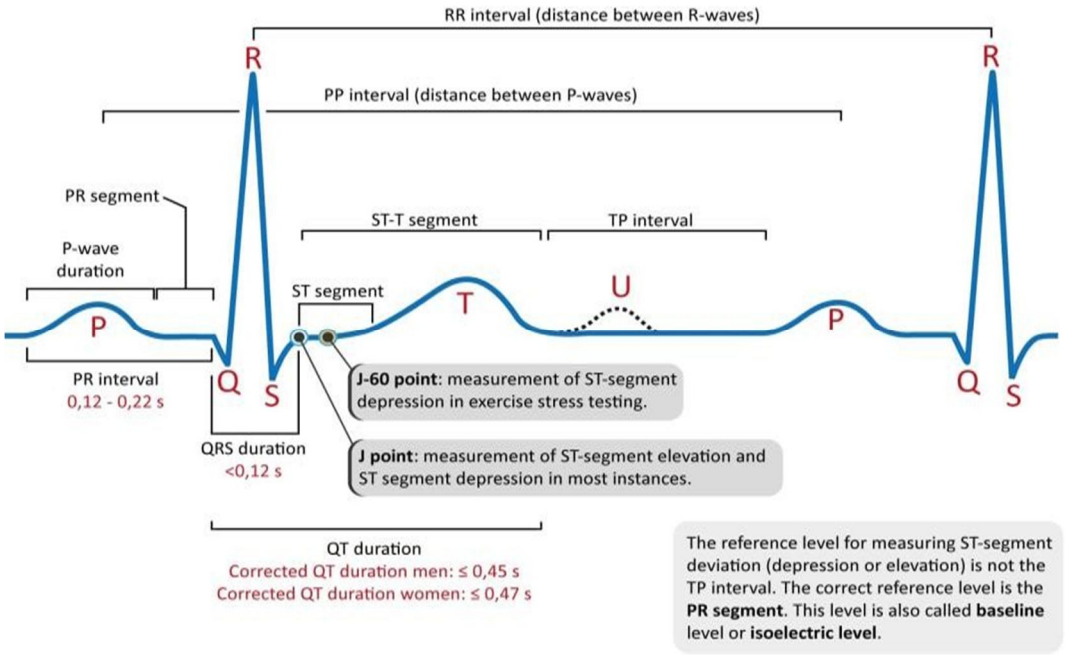


Fig. 4: Representation of ECG Signals and its Intervals between wave

1) *To Identify and Examine 'P' Waves:* Interval of 'P' wave is between 0.00 to 0.12 secs.
2) *To Compute the 'PR' Interval:* The beginning of 'P' to beginning of 'QRS' interval should be 0.12 to 0.20 secs.
3) *To compute 'QRS' Complex:* The range of 'QRS' interval is 0.06 t 0.11 secs .
4) *To identify the Rhythm 'R-R':* This is to measure Regular or Irregular ECG Signal. If 'R' to 'R' is having same distance then it is considered as Regular ECG Signal if not it is Irregular ECG Signal . The Interval between 'R-R' is 0.06 to 0.11 secs .
5) *To identify QT interval :* This is to measure 'QT ' interval in range 0.33 to 0.43 secs [17]

B. Conduction System of the Heart

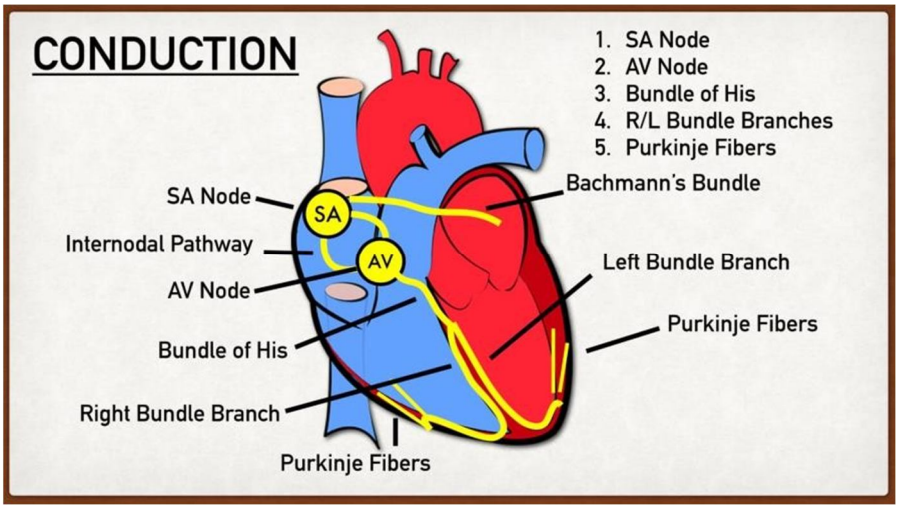


Fig. 5: Conduction system of the Heart

The Figure 5 represents the conduction system of Heart , where its consists of five main parts namely SA node , AV node , Bundle of His, Right and Left Bundle of His, Purkinje Fibers .

1) *SA Node:* It is a Siano atrial node where it is present in Right atrium of the heart , also know as Natural pacemaker of the Heart where beginning of the conduction takes place .It produces 60-100 BPM .
2) *AV Node:* It is atrial ventricular node , present in the boarder of Right atrium and Right ventrical .Also know as Gatekeeper of the Heart .If the SA node is blocked for some reason then AV node activates and gives 40-60 BPM .
3) *Bundle of His:* Form the AV node it goes to the single structure through the ventricles called Bundle of His.
4) *Right & Left Bundle of His:* It is Byfercated into two types namely Right and Left Bundle of His on either side of Ventricals, which produces 20-40 BPM.
5) *Purkinje Fibers:* The Right and Left Bundle of His is extended to the Apex of the Heart that is named as Purkinje Fibres , which produces 20-40 BPM [7].

C. Algorithms used for Prediction of cardiac Arrhythmia

The algorithm for machine learning is an approach by which the system of AI capabilities performs the processes, normally by foreseeing the values as output from already provided data as input. The important actions of algorithms of machine learning are regression and classification of data

D. Logistic Regression

It is a machine learning algorithm that uses the supervised learning method. It is a statistical technique for creating machine learning models with dichotomous dependent variables. It foresees a binary value's output. In addition to providing the exact values of 0 and 1, it also provides the probabilistic values that fall between 0 and 1. The output can be either Yes or No, 0 or 1, true or false[6]

It is as much alike to the Linear Regression except the usage. The usage of Linear Regression is for solving Regression problems, but the custom of using Logistic regression is for answering the classification problems . It is used for Binary classification problems ,to foresee a binary outcome such as yes/no based on prior observations of data set[6].

The straight line equation can be written as: $y=b0+b1x1+b2x2+b3x3+.....+bnxn$ (1)

In this y is between 0 and 1 only, so for this by dividing the above equation by (1-y): y/y-1 0 for y=0 ; and infinity for y=1 (2)

But the is between -[infinity] to +[infinity], then by taking algorithm of the equation it will becomes: $Log\ (y/y-1) = b0 +b1x1+b2x2+b3x3+.....+bnxn$---------- (3)

The equation 3 gives is the final equation of Logistic Regression.

The main advantage of this algorithm is that it is simple to implement and train a model .It provides a measure of how relevant a predictor is, and its positives or negatives. Their is no issues of scaling the input features,& also which does not require tuning.

E. Decision Tree

It is a Supervised machine learning technique and tree-structured classifier that can be used for two things namely for classification and Regression problems. But,preferably used for answering Classification problems. The internal node shows the features of data sets as it is a tree structured, The decisions are made in branches and the outcome represents each leaf node.Based on certain features itgives yes/no , True/false or 0/1 outputs [2] .

\par This model analyses in three different nodes namely root node ,interior node, leaf node .The decision node role is to make decisions, interior node is to make the condition of dependent variables which is handled by this node ,where decisions outputs are in Leaf nodes and do not contain any further branches.It is called a decision tree[2].

To find the root node:

Information Gain = Class Entropy - Entropy Attributes To find Class Entropy:

Pi+Ni = - P/P+N log P/(P+N) – N/P+N log N/P+N ----------(4)

Where , P is possibility of yes and N is possibility of no .

The main advantage of this is to know as it follows the similar process as that of a human take decisions in day to day life .This is very useful for answering decision making problems.

F. Support Vector Machine [SVM]

It is categorised as a supervised machine learning algorithm that takes both classification and regression issues into account.But it is mostly used for classification problems.By plotting it in n-dimensional space on each data item .Where 'n' is a several features you have. By performing the classifications and for locating the hyper-plane that evovle into classes. The hyperplane's dimension is influenced by the number of features. The SVM model supports few features as shown below:

Support Vectors, Hyperplane and Margin [8].

IV. IMPLEMENTATION AND RESULTS

A. Plan of Execution

1) Using the UCI Machine Learning repository which comprises of a data set containing arrhythmia data .

2) The collected datasets are pre-processed and analyzed using machine learning library .

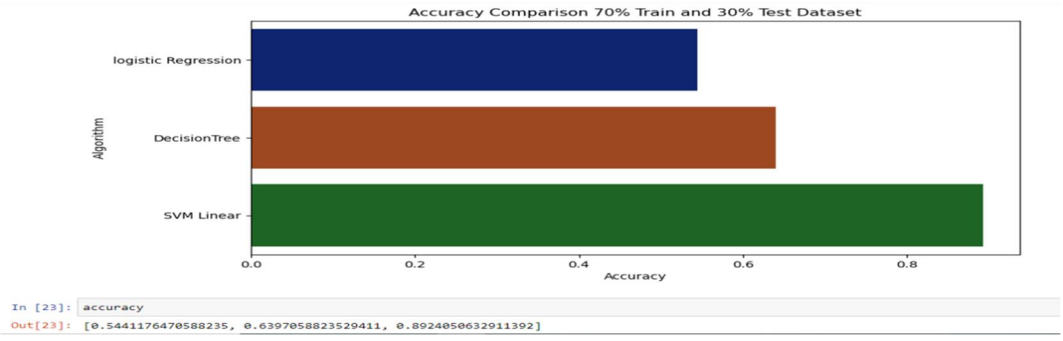
3) The pre-processed datasets are spitted into training and testing and passed to the machine learning algorithm .

4) The trained datasets are compared with test result with help of algorithm and results are shown in Percentage with bar graph.

5) The results are compared with the applied algorithms and the algorithm showing the best results is considered .

As per the above plan of execution the data sets are taken from the standard repository ,then based on the requirement certain features like Age, height , sex , weight , and the ECG signal durations as like P,PR,QRS,R-R,QT etc are considered and these data are pre-processed .The pre-processed data is divided into Train and Test data sets. In our work we have considered two combinations like one is 70/30 and the other is 80/20 as Train and Test data sets . After applying these two combinations of data sets into the algorithms the one which shows the accurate results is considered as the best model for the prediction of cardiac arrhythmia.

Fig. 6: Comparison graph



The above Figure 6 shows the comparison graph of all three models [Logistic Regression, Decision Tree SVM] by considering 70% and 30% as Train and Test data set . In the figure it shows Logistic Regression as 54.41%, Decision Tree as 63.97%, SVM as 89.24% , where SVM shows better results than other two models.

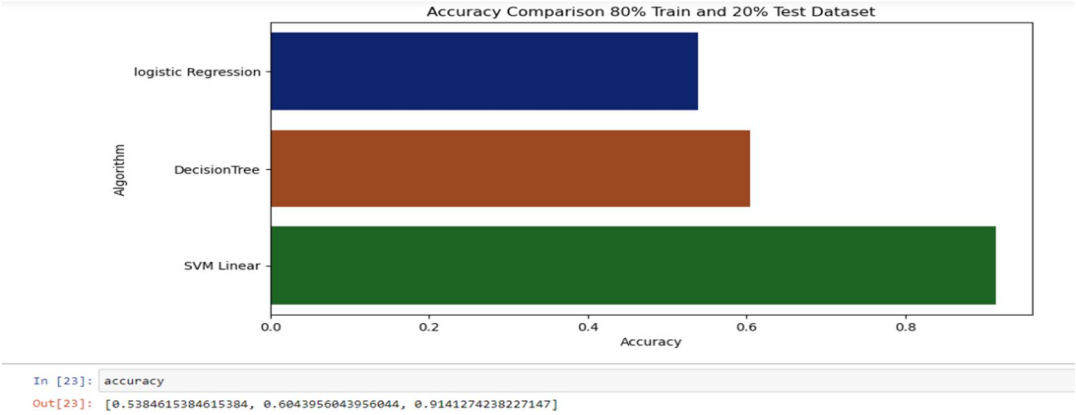


Fig. 7: Comparison graph

The above Fig 7 shows the comparison graph of all three models [Logistic Regression, Decision Tree , SVM] by considering 80% and 20% as Train and Test data set . In the figure it shows Logistic Regression as 53.84%, Decision Tree as 60.43% , SVM as 91.41% ,where SVM shows better results than other two models.

A. Comparative Study of Applied Algorithms

Algorithms Used	Prediction Based on	
	70% Training & 30% Testing data sets	80% Training & 20% Testing data sets
Logistic Regression	54.41%	53.84%
Decision Tree	63.97%	60.43%
Support Vector Machine [SVM]	89.24%	91.41%

Fig. 7: Comparison table of three algorithm

The below Fig 8 shows the comparison table of all three algorithms for two different combinations of data sets percentage such as 70/30 and 80/20 as Train and Test data. By this we can understand that 80/20 combination of Training and Testing data sets is showing best results for SVM model as 91.41%

B. GUI Interface

The GUI Interface has been designed for checking the prediction of Healthy or unhealthy Heart beat.

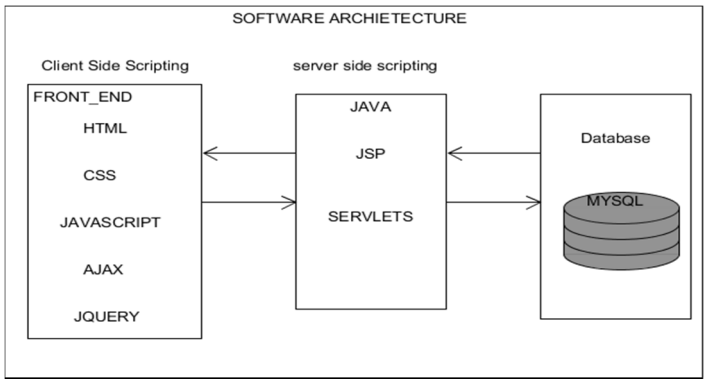


Fig. 9: Software Architecture

The Figure 9 shows the Software Architecture of the design.It has three groups namely Data set,Server Side and Client Side.For the visibility and accessing for client side certain web application needs to be built.This is built using a framework called Flaskframework.

Blow are the steps to be followed :

1) Require installed python.
2) In file system, create a folder for project, such as arrhythmia prediction.
3) Open the created project folder in VS Code.
4) In VS Code, press Ctrl+Shift+P and then select the installed python interpreter.
5) Open a new terminal in VS Code.
6) Update pip using:\textbf{python -m pip install --upgrade pip}.
7) Install Flask using command:\textbf{python -m pip install flask}.
8) Using html for rendring web page and css for styling purposess.along side Java script.
9) To run the application type: python app.py in the terminal from the project folder.
10) Open the browser and goto the default flask application would run the application at 127.0.0.1:5000 IP and port address.

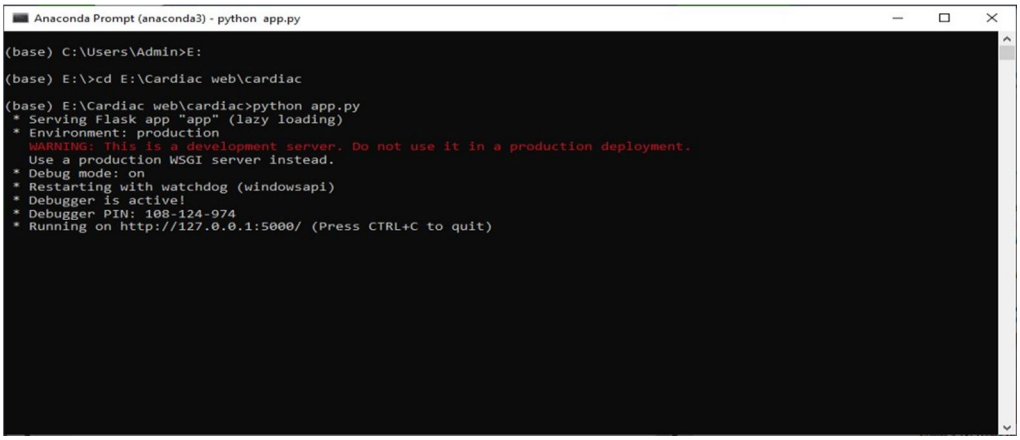


Fig. 10: Running web appplication using cmd

The above Figure 10 shows how to Run web appplication using cmd[command prompt].The following are few steps to be followed

a) Select the folder E first as shown above. C:\Users\Admin>E:
b) Under folder cardiac web select E:\>cd E:\Cardiac web\cardiac
c) Under that folder select python E:\Cardiac web\cardiac>python app.py
d) After that the obtained ip address can be opened in the URL. Web Page designed for the system

Below are the web pages designed for the system which includes Home Page, Value entry page and Results Page which shows Healthy or unhealthy Heart Beat.

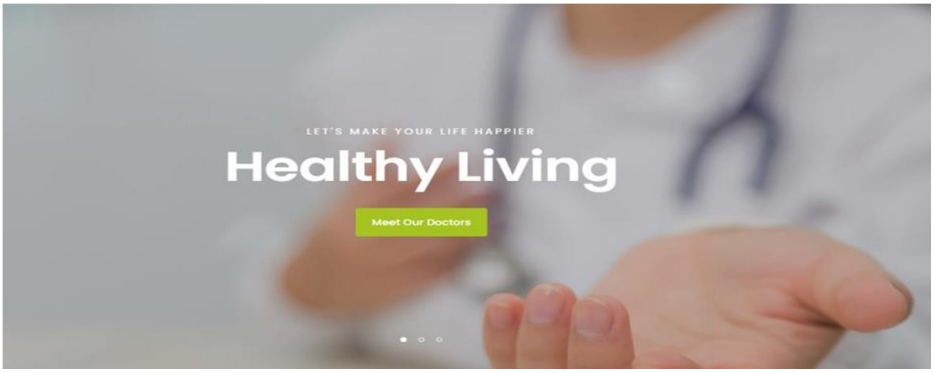


Fig 11 Home Page

The above Fig 11 shows the Home page of the designed system

Fig 12 Value Entry Page

The above Fig 12 shows the value entry page of the outlined System

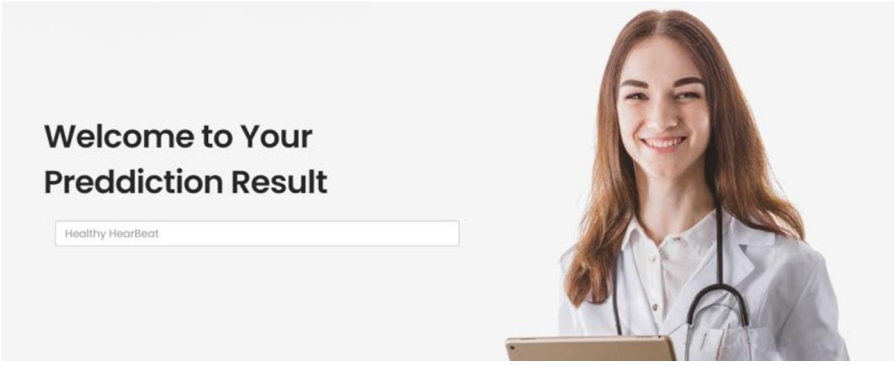


Fig 13 Results Page

The above Figure 13 shows the Results Page of the designed system.As per the above discussion in Results section the SVM[Support Vector Machine] model showed the best results for 80/20 combination of Train and Test data sets.The website was built for the SVM model to showcase whether the person is having an Healthy Heart Beat or not.

V. CONCLUSION

In this proposed work all the objectives were achieved. A website for Prediction of an Healthy or Unhealthy Heart Beat was developed.As discussed in the implementation and results section the implementation of three machine learning algorithms has been carried out.

The three machine learning algorithms which was appealed on to the data sets to foresee the probability of having cardiac Arrhythmia of a patient were examined with models namely Logistic Regression, Decision Tree and SVM.All three algorithms are put in to the same data sets for the sake of examining the best model in terms of efficiency and accuracy.

Firstly all the algorithms were applied to 70/30 Combination of Train and Test data. From that the obtained results are Logistic regression as 54.41 % ,Decision Tree as 63.97% and SVM as 89.24%.

Next the combination of data sets were changed to 80/20 as Train and Test data. From this the obtained accuracy for the applied algorithms was shown the better results than 70/30 combination the details are as follows Logistic Regression model showed with an accuracy of 53.84%, Decision Tree has predicted shows an accuracy level of 60.43% and the Support vector machine [SVM] model has shown with the accuracy level of 91.41%.

Therefore as per the above implementation results and discussion it can be seen that when the data sets are applied for both combinations of 70/30 and 80/20 of Train and Test data the obtained accuracy results for 80/20 combination of data set SVM Algorithm is showing 91.41% of accuracy. It can be concluded that the results of SVM algorithm which is showing the accurate results is considered as best algorithm for the Prediction of Cardiac Arrhythmia.

A. Future Scope

1) This Web application can be further deployed into Android/IOS and make them available to mobile devices so that it can be used by all the users.

2) Few more algorithms can be considered for the comparative study.

3) This can be further enhanced to keep the track of health records of patients data which helps the doctors for study.

4) By employing cloud computing to store the data and run more demanding applications, we may further enhance the system's performance in terms of operating speed and memory capacity.

REFERENCES

[1] Kelwade, J., & Salankar, S. (2015). Prediction of cardiac arrhythmia using artificial neural network. International Journal of Computer Applications, 115(20).

[2] J., S. K., & S., G. (2019). Prediction of heart disease using machine learning algorithms. 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 1–5. https : / / doi . org / 10 . 1109 / ICIICT1 . 20198741465.

[3] Kalkstein, N., Kinar, Y., Na'aman, M., Neumark, N., & Akiva, P. (2011). Using machine learning to detect problems in ecg data collection. 2011 Computing in Cardiology, 437–440.

[4] Shetgaonkar, P., & Aswale, S. (2021). Heart disease prediction using data mining techniques.

[5] Methaila, A., Kansal, P., Arya, H., & Kumar, P. (2014). Early heart disease prediction using data mining techniques. Computer Science Information Technology, 4, 53–59.https://doi.org/10.5121/csit.2014.4807.

[6] Sultana, N., Kamatham, Y., & Kinnara, B. (2016). Performance analysis of artificial neural networks for cardiac arrhythmia detection. 2016 IEEE 6th International Conference on Advanced Computing (IACC), 421–425. https://doi.org/10.1109/IACC.2016.84.

[7] Ozbay, Y., & Karlik, B. (2001). A recognition of ecg arrhytihemias using artificial neural networks. 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2, 1680–1683vol.2.https://doi.org/10.1109/IEMBS.2001.1020538.

[8] Rohan, M. D., & Patil, A. (2012). Layered approach for ecg beat classification utilizing neural network. Bioinformatics, 2(6), 1495–1500

[9] Desai, U., Martis, R. J., Nayak, C. G., K., S., & Seshikala, G. (2015). Machine intelligent diagnosis of ecg for arrhythmia classification using dwt, ica and svm techniques. 2015 Annual IEEE India Conference (INDICON), 1–4. https://doi.org/10.1109/INDICON.2015.7443220

[10] Shimpi, P., Shah, S., Shroff, M., & Godbole, A. (2017). A machine learning approach for the classification of cardiac arrhythmia. 2017 International Conference on Computing Methodologies and Communication (ICCMC), 603–607. https://doi.org/10.1109/ICCMC.2017.828253745

[11] Soman, T., & Bobbie, P. (2004). Classification of arrhythmia using machine learning techniques. WSEAS Transactions on Computers, 4

[12] Gandhi, M., & Singh, S. N. (2015). Predictions in heart disease using techniques of data mining. 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 520–525. https://doi.org/10.1109/ABLAZE.2015.7154917

[13] Suzuki, Y. (1995). Self-organizing qrs-wave recognition in ecg using neural networks. IEEE Transactions on Neural Networks, 6(6), 1469–1477. https://doi.org/10.1109/72.471381

[14] Udupa, J. K., & Murthy, I. S. N. (1980). Syntactic approach to ecg rhythm analysis.IEEE Transactions on Biomedical Engineering, BME-27(7), 370–375. https://doi.org/10.1109/TBME.1980.326650

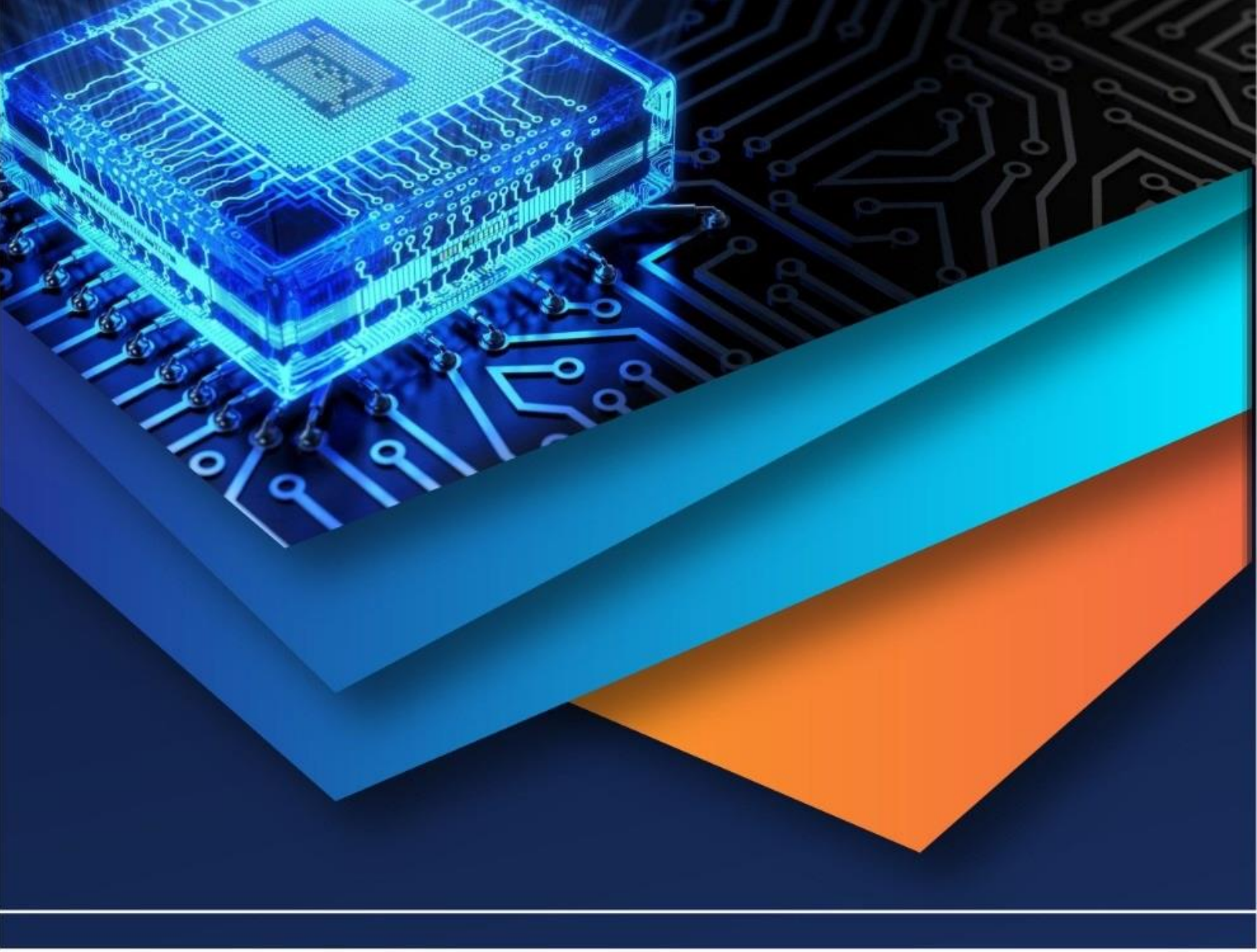
[15] Yu, R., Gao, Y., Duan, X., Zhu, T., Wang, Z., & Jiao, B. (2018). Qrs detection and measurement method of ecg paper based on convolutional neural networks. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 4636–4639. https://doi.org/10.1109/EMBC.2018.851313246.

[16] https://www.downtoearth.org.in/blog/health/india-s-burden-of-heart-diseases-study-says-elderly-women-more-at-risk-74993

[17] https://ecgwaves.com/topic/ecg-normal-p-wave-qrs-complex-st-segment-t-wave-j-point/

[18] https://www.mayoclinic.org/diseases-conditions/heart-arrhythmia/symptoms-causes/syc-20350668

[19] https://www.ibm.com/cloud/learn/machine-learning#:~:text=Machine%20learning%20is%20a%20branch,learn%2C%20gradually%20improving%20its%20accuracy.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY