CARDIAC RISK PREDICTION

[1]K. Lavanya, [2]Y. Praveesha, [3]M. Pratap, [4]G. Pavan, [5]G. Prasanna Priya

[1,2,3,4]Student, [5]Assistant Professor
Department of CSE,
Vignan's Institute of Information Technology, AP, India.

***Abstract*: Sudden cardiac arrest is one of the many cardiovascular disorders that is on the rise. There are a variety of factors that affect our hearts, and changes in our everyday routines may also contribute to cardiovascular disease. If health measures are not taken, a person's sudden cardiac arrest might result in death. We will use machine learning to forecast the likelihood of a person developing a cardiac risk. Using a machine learning technique, this module calculates the chances of being diagnosed with this condition based on several variables.**

***Index Terms*: Cardiac Risk, Support Vector Machine (SVM), Decision Tree, Random Forest, Bagging.**

I. INTRODUCTION

Sudden cardiac arrest is a global health issue that is becoming more prevalent. However, if proper measures are taken to resurrect a person, the chances of survival are minimal because technological advancements have not yet reached their full potential. Cardiac arrest can arise without symptoms, making it difficult to know the cause of death. It is particularly noticeable when a person is sleeping or not moving.

Sudden cardiac arrest happens when the heart beat or rhythm is irregular, which is known as arrhythmia [1]. Cardiac arrest is characterized by a rapid stoppage of breathing and loss of consciousness. Cardiac arrest and heart attack are not the same condition, however they both fall under the category of cardiovascular illnesses. When blood flow to any chamber of the heart is blocked, a heart attack results [2]. Sudden cardiac arrest can occur as a result of a heart attack.

Various symptoms such as chest tightness, weakness, and shortness of breath can develop during cardiac arrest. Cardiac arrest is irregular occurrence that occurs without notice. Previous heart attacks and coronary artery disease are both risk factors for cardiac arrest. Other indicators include a previous cardiac arrest episode, a family history of cardiac arrest, and episodes of fainting for unspecified reasons [3].

In biomedical research, machine learning has proven to be an invaluable tool for solving complex challenges. It is simple to locate key features and develop ways to use medical signals and increase the accuracy and functionality of prediction models to solve medical problems using modern algorithm technology [4]. Traditional approaches have clinical limits, as well as poor performance and a high proportion of false alarms. Random forest, Decision tree, Support vector machines, Linear regressions, and Naive bayes are the examples of predictive classification algorithms. We also use ensemble approaches, which use the base models to identify the most optimal and effective predictive model. That would have been a major accomplishment in healthcare industry if we could diagnose cardiac risk using these specific algorithms.

II. LITERATURE REVIEW

To predict cardiac risk, R Karthikeyan et al used support vector machines, random forests, decision trees, logistic regression, and artificial neural networks. Artificial neural networks have the maximum obtained accuracy of 84 %. Because of the abnormality of the dataset utilized, the algorithm appears to be recognized inadequate [5].

L. Murukesan, M. Murugappan, M. Iqbal, and Dr. K. Saravanan implemented HRV signals (heart rate variability) to test the electrophysiological indication used to detect heart conditions. To detect optimal features of cardiac arrest using a tree bag classifier, a sequential feature selection technique was applied. The generated optimal features are given to Support Vector Machine and Probabilistic Neural Network [6].

III. PROPOSED SYSTEM

A. Parameter Used

The parameters utilized to implement the algorithms are Sex, Age, and Chest Pain Type (Angina, Atypical angina, Non anginal, Asymptomatic), Resting blood pressure (in mm Hg), cholesterol, fasting blood sugar, ECG results, maximum heart rate, old peak, slope (up sloping, flat, down sloping), number of major vessels (arteries, arterioles, veins, ventricular, capillaries), thallium stress level (normal blood flow, abnormal blood flow, low blood flow, no thallium), exercise-induced angina The algorithm employed is a Random Forest, Decision Tree, Support Vector Machine with Bagging method.

B. Decision Tree

A decision tree is used to tackle classification and regression problems in supervised learning. The judgments made are based on the characteristics in the dataset. The leaf node symbolizes the decision's conclusion, whereas the nodes represent the characteristics and decisions. The bigger node serves as the data collecting indication. It is simple to understand the structure and decision using a decision tree. A decision tree has a high variance, which means that minor differences in input value can result in large changes in output [7]. The model grows more fit as the tree becomes highly complicated and deep. The accuracy obtained is 88%. The below figure displays the structure of Decision Tree [8].

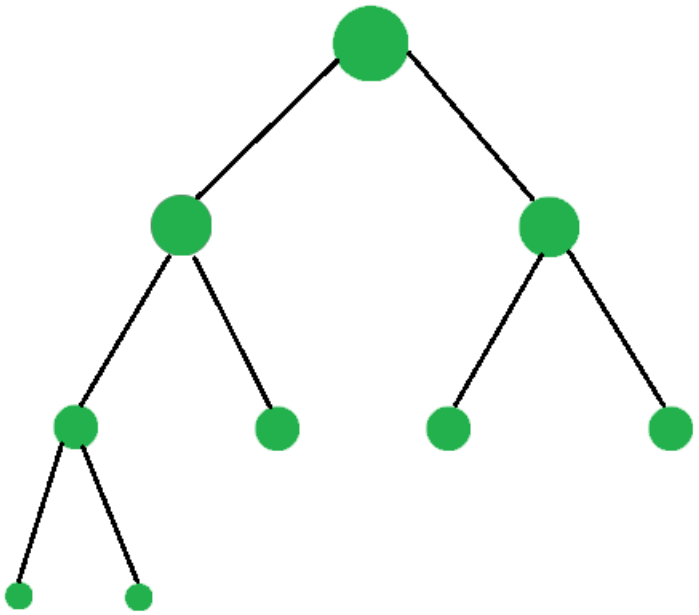


Figure 1 Decision Tree Structure

C. Support Vector Machine

A support vector machine (Classifier) is a model for classification and regression. When there is a huge amount of data, a support vector machine is the preferred algorithm. Different kernels are often used in SVM to perform classification. [9] Kernels play a critical part in SVM's ability to tackle complex problems. Kernel also aids with the problem of overfitting. It splits classes using hyperplane. The amount of features gathered determines the size of the hyperplane. The fundamental benefit of SVM is that it provides significant accuracy while requiring less computational effort. SVM is a bioinformatics software tool. Where new data may be efficiently added and can easily used for future research. The accuracy obtained using support vector machine is 63%. The structure below is SVM [10].

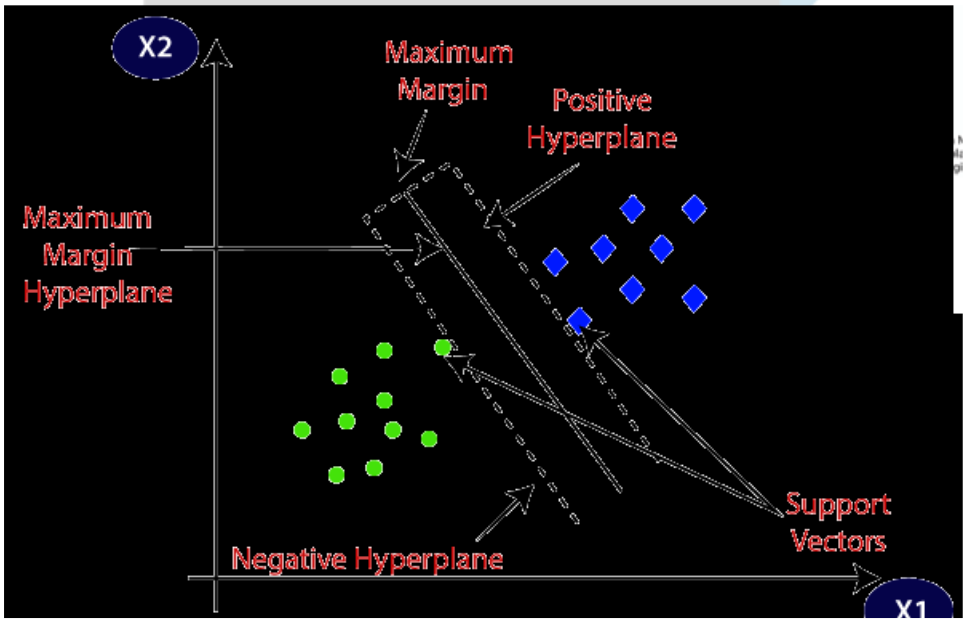


Figure 2 Support Vector Machine

D. Random Forest

The Random Forest model is an ensemble learning model. Random Forest is a classifier that does use a number of decision trees on different subgroups of a given dataset and averages them to enhance the predicted accuracy of that dataset. The greater the number of trees used, the higher the predictive output and the lower the overfitting. It operates effectively even with bigger datasets

and missing values. Using a higher number of trees may result in delayed and poor prediction [11]. This method is quicker during training and slower when predicting. This problem is solved by bagging with an accuracy of 89 %.

E. Bagging Method

[12] Bagging is an ensemble model that integrates many models to enhance productive accuracy. The disadvantage of using above algorithms is that they have a large variance, which leads to overfitting. The bagging approach reduces the unstable that occurs and increases the accuracy with correct prediction. Each base classifier's training set is independent of the others. Bagging lowers overfitting by averaging or voting; nevertheless, this increases bias, which is offset by the decrease in variance. The highest accuracy produced is random forest with 89%. Below structure shows Bagging [13].

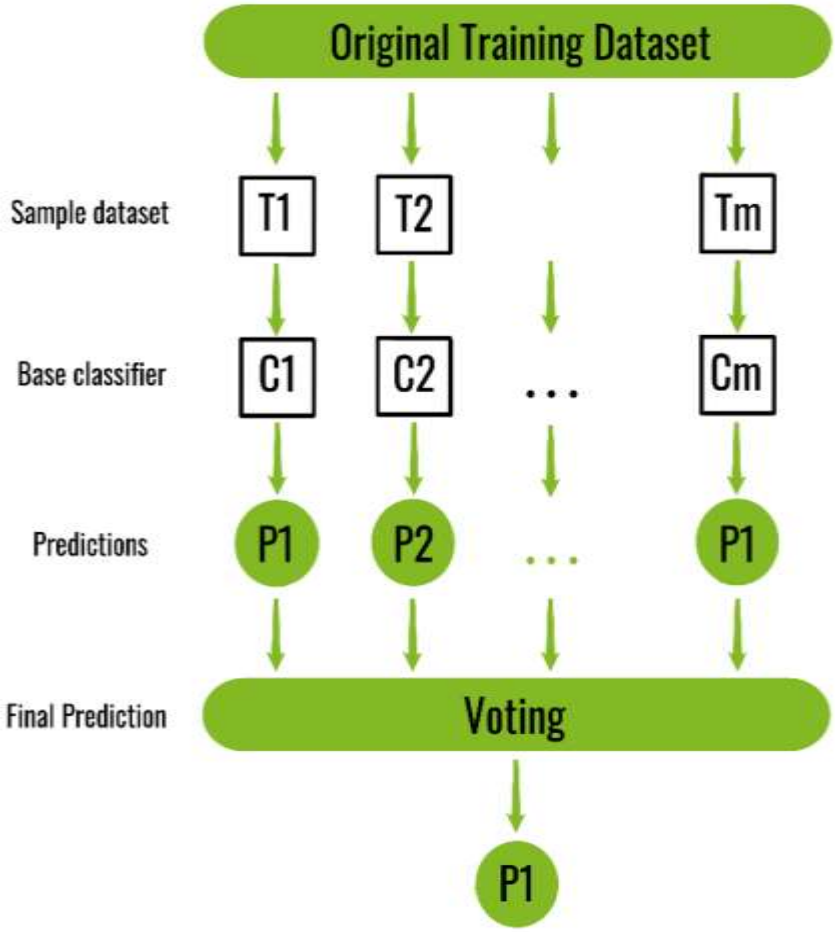
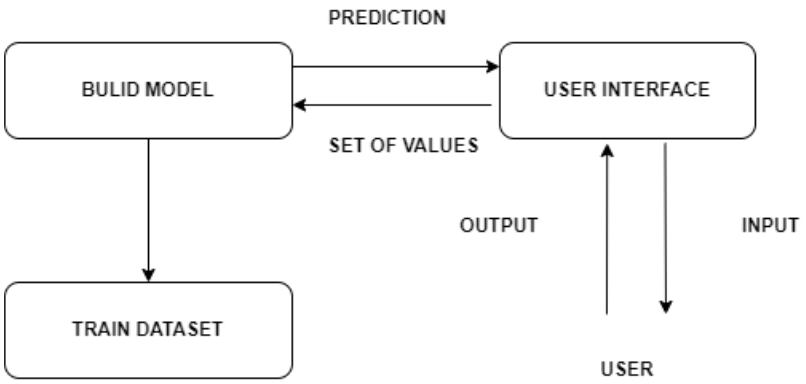


Figure 3 Bagging Technique

F. Flow Chart

The flowchart of the system is user entered into the website, the website displays the required data that need to be entered for the output. The user enters the input data which is set values that are predicted using the model that is built. Based on the given set values the model predicts the output, if the patient has the chances of getting cardiac arrest or not. As shown on Figure 7.



IV. RESULTS AND DISCUSSION

Various methods, such as Decision Tree, Support Vector Machine, and Random Forest with Bagging approach, are used. Figure 4 depicts the accuracy of Support Vector Machine, Figure 5 shows accuracy of Random Forest, and Figure 6 represents the accuracy of decision tree. The forecasting is done by adding new values to the model, and the output is presented in Figure 7 Output prediction. The maximum accuracy is attained by using a Random Forest with a Bagging technique for predicting new values.

```
In [14]: from sklearn.svm import SVC
         bag_model = BaggingClassifier(
             base_estimator=SVC(),

             n_estimators=100,
             max_samples=0.8,
             oob_score=True,
             random_state=0
         )
         bag_model.fit(X_train, y_train)
         bag_model.oob_score_

Out[14]: 0.6343612334801763

In [15]: bag_model.score(X_test, y_test)

Out[15]: 0.6578947368421053
```

Figure 4 Support Vector Machine (63%)

```
In [12]: from sklearn.ensemble import RandomForestClassifier
         bag_model = BaggingClassifier(
             base_estimator=RandomForestClassifier(n_estimators=100,criterion='gini'),

             n_estimators=100,
             max_samples=0.8,
             oob_score=True,
             random_state=0
         )
         bag_model.fit(X_train, y_train)
         bag_model.oob_score_

Out[12]: 0.788546255506608

In [13]: bag_model.score(X_test, y_test)

Out[13]: 0.8947368421052632
```

Figure 5 Random Forest (89%)

```
In [8]: bag_model = BaggingClassifier(
            base_estimator=DecisionTreeClassifier(),

            n_estimators=100,
            max_samples=0.8,
            oob_score=True,
            random_state=0
        )
        bag_model.fit(X_train, y_train)
        bag_model.oob_score_

Out[8]: 0.7841409691629956

In [9]: bag_model.score(X_test, y_test)

Out[9]: 0.881578947368421
```

Figure 6 Decision Tree (88%)

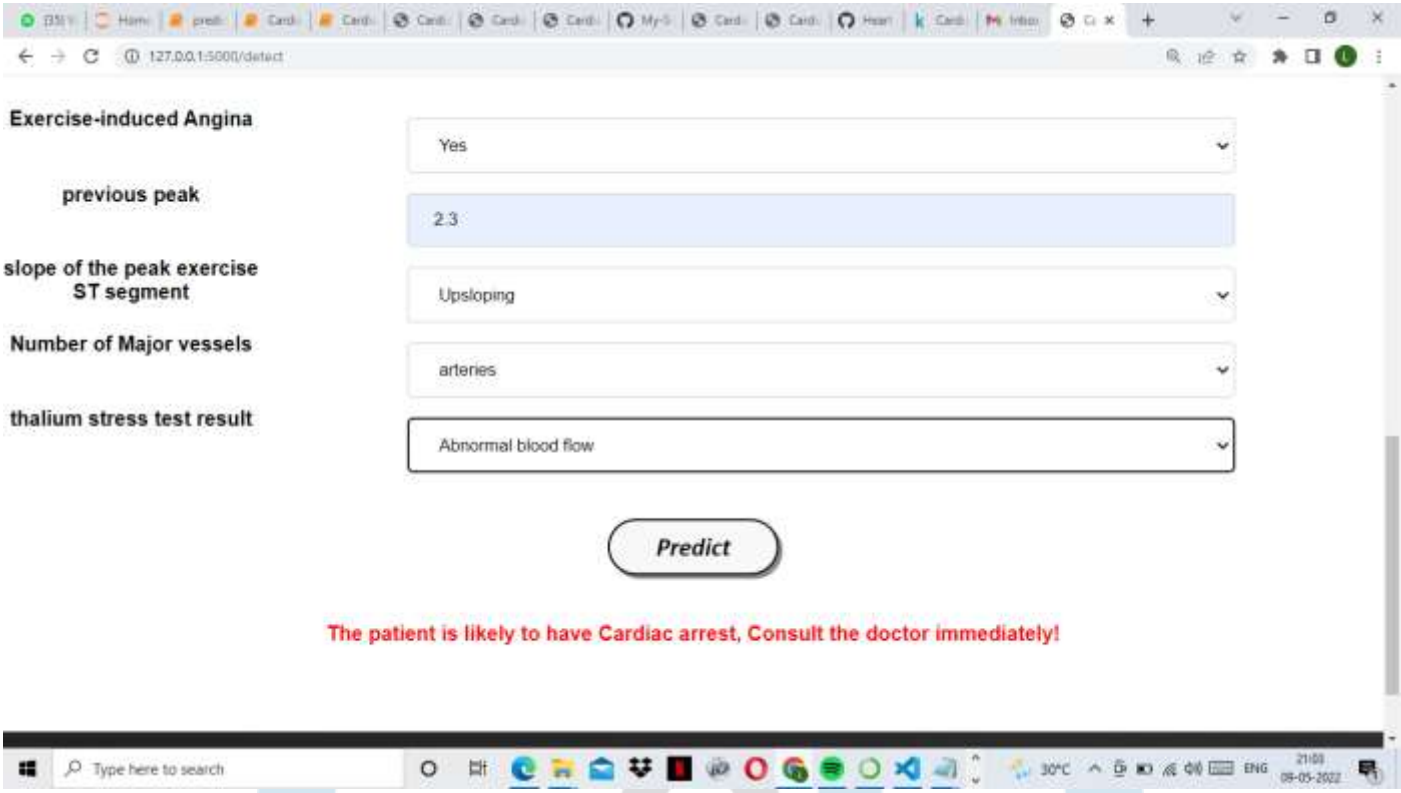


Figure 7 Prediction Output

V. CONCLUSION

This article helps us comprehend the factors taken in to consideration and how the modifications made when detecting the disease. Random Forest has the greatest prediction accuracy when the following algorithms, Support Vector Machine, Random Forest, Decision Tree, and Bagging Technique, are applied to the dataset. The Random Forest algorithm is used in connection with the Bagging method to generate low variance and faster predictions in real time with an accuracy of 89%. It also helpful in avoiding overfitting when introducing fresh data into the model. This algorithm can forecast a person's probability of experiencing cardiac risk.

REFERENCES

[1]https://www.mayoclinic.org/diseases-conditions/sudden-cardiac-arrest/symptoms-causes/syc20350634#:~:text=The%20usual%20cause%20of%20sudden,and%20rhythm%20of%20your%20heartbeat.
[2] Priyanshi Mittal, Sushil Bansal published in International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 11 Issue 01, January-2022
[3] https://my.clevelandclinic.org/health/diseases/17522-sudden-cardiac-death-sudden-cardiacarrest
[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7943901/
[5] R Karthikeyan *et al* 2021 *J. Phys.: Conf. Ser.* 1964062076 published in IOP publishing doi:10.1088/1742-6596/1964/6/062076
[6] L. Murukesan, M. Murugappan, M. Iqbal, and Dr. K. Saravanan Article *in* Journal of Medical Imaging and Health Informatics · August 2014doi:10.1166/jmihi.2014.1287
[7] https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm
[8] Utsav Chauhan,Vipul Chauhan, Amit Kumar, Vikas Kumar, Sumit Tiwary, published in 2019 2nd International Conference on Intelligent Computing ,Instrumentation and Control Technologies(ICICICT)
[9]https://techvidvan.com/tutorials/svm-kernelfunctions/#:~:text=A%20kernel%20is%20a%20function,number%20of%20dimensions%20using%20kernels.
[10] https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machinealgorithm.png
[11] https://www.javatpoint.com/machine-learning-random-forest-algorithm
[12]https://www.simplilearn.com/tutorials/machine-learning-tutorial/bagging-in-machinelearning#:~:text=Bagging%2C%20also%20known%20as%20Bootstrap,variance%20of%20a%20prediction%20model.
[13] https://media.geeksforgeeks.org/wp-content/uploads/20190515171714/cb7feb7cd065-4da7-bea6-5f6017038059.png