ECG Heartbeat Categorization Using Machine Learning & Deep Learning

Md Fahim Shahoriar Titu
Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
fahim.shahoriar@northsouth.edu

Hassan Mahmud Emon
Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
hassan.emon@northsouth.edu

Shahriar Jabin Abir
Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
shahariar.abir@northsouth.edu

Abdul Aziz Chowdhury
Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
abdul.chowdhury@northsuth.edu

Abstract: An electrocardiogram, or ECG, is a straightforward, non-invasive test used to identify abnormal conditions like irregular heartbeats. Notwithstanding the way that man-made reasoning and AI are used in a wide assortment of medical care related applications and datasets, various heartbeat classifiers utilizing profound and AI methods have been proposed lately. However, there are no widely explained public ECG datasets, and the datasets that are available for building and testing AI models frequently have very small sizes. A machine and deep transfer learning framework for classifying a small training dataset is presented in this paper. In order to predict the likelihood of developing a problem related to the heartbeat, a number of machine learning (ML) models were developed. Using a variety of physiological indicators and machine learning techniques like Logistic Regression (LR), Random Forest (RF) Classification, Support Vector Machine (SVM), and Voting Classifier, this study trains four alternative models for accurate prediction. With an accuracy of almost 98 percent, the Support Vector Machine method performed the best on this assignment. The open-access heartbeat categorization dataset was utilized in the development of the strategy. The fact that the models used in this study have a much higher percentage of accuracy than those used in previous studies indicates that these models are more reliable. The scheme can be found in the research's analysis, and numerous model comparisons have demonstrated their strength.

Keywords: Heartbeat problem; Disease; Machine Learning; Prediction; Support Vector Machine; Accuracy.

1 Introduction

Cardiologists and other medical professionals frequently use electrocardiograms (ECG) to monitor cardiac health. Similar to many other time-series data, the difficulty of identifying and classifying various waveforms and morphologies in the signal is the primary issue with manual ECG signal analysis. This is a task that takes a lot of time and is easy to make mistakes for a human. Due to the fact that cardiovascular diseases account for approximately one third of all deaths worldwide, accurate diagnosis is critical [1]. For example, a large number of individuals experience sporadic pulses which can be deadly at times. As a result, it is highly desirable to have a diagnosis of arrhythmias that is both accurate and affordable [2]. Numerous studies in the literature looked into using machine learning techniques to accurately detect anomalies in the signal in order to address the issues raised by manual ECG signal analysis [3, 4]. The majority of these strategies involve a phase of preprocessing to prepare the signal (such as running it through band-pass filters, for example). The handcrafted features, which are primarily statistical summaries of signal windows, are then extracted from these signals and utilized in subsequent classification analysis. Support Vector Machines, multi-layer perceptrons, decision trees, and other conventional machine learning approaches are utilized for ECG analysis's inference engine. 5], [6], [7]. The most common symptoms are: seizures, headaches, vision issues, mental shifts, vomiting, and other symptoms Other symptoms include difficulty speaking or moving, diminished sensations, and occasionally a loss of consciousness [4-5]. People frequently ignore these symptoms, which leads to an increase in their classification and worsens their condition. Gradually, the classification begins to change into heartbeat disease, which can even kill you. Therefore, a categorization of a heartbeat ought to be found early on. Numerous lives could be saved with the proper treatment of the disease. Adult heartbeat classifications are typically benign (non-diseases). However, malignant classifications are frequently observed in children. A CT (Computed Tomography) scan and an MRI (Magnetic Resonance Imaging) scan are the medical tests that are most frequently suggested for the purpose of determining the classifications of heartbeats. A CT examine is essentially a Xbeam of the head. Because they harm the cells that make up our heartbeats, X-rays are known to be harmful to our bodies. Therefore, MRI is the most secure diagnosis and better at detecting categorizations than the previous tests. A heartbeat categorization cell and a normal cell can also be distinguished more clearly using deep learning techniques on MR images. In addition, this may speed up the process of categorizing heartbeats and identifying them while maintaining higher precision. Basically, deep learning (DL) is a type of machine learning (ML) that uses AI to work with images. By distinguishing between images that a regular human cannot detect using standard methods, people can gain certain knowledge in this way. By combining a variety of layers of linear process parameters, the DL algorithms extract features. Data extraction becomes easier as we move deeper into the network because the result of each layer becomes the source for the next layer. A type of deep learning known as deep convolutional neural networks (CNNs) is designed to be easy to use and is frequently utilized in visual image analysis. It is based on the biological activities of the human heartbeat and has been used to order data in various arrays [6]. Over the years, a variety of machine learning algorithms have been used to identify heartbeat categorization. Research [7] was directed utilizing AI calculations, Multi-facet Perceptron (MLP) and Innocent Bayes to identify the heartbeat categorization from MR pictures, weighted-F1 scores were around 98.6% and 91.6% individually. In [8], the heartbeat classification was detected with a precision of 91% and 95%, respectively, using the C4.5 algorithm and MLP. With the help of a Deep Neural Network and a Machine Learning Algorithm, the researchers in the study [9] discovered an effective classification of heartbeats. They achieved incremental accuracies of 98.67%, 97.34%, and 94.24% by utilizing Softmax's Fully Connected layer, CNN with the Radial Basis Function (RBF) classifier, and the Decision Tree (DT) classifier. In [10], the authors looked at various approaches taken by other authors and discovered that MRI images were more effective at identifying categorizations of heartbeats. The creators in [11] zeroed in on convolutional brain organizations (CNN) for the location of heartbeat categorization MR pictures. Using the CNN algorithm, they were able to get a weighted-F1 score of around 94 percent. In [12], CNN and VGG-16 were used to identify categorizations of heartbeats with 91.6 percent and 91.9 percent, respectively. The K-NN classification algorithm for categorizing heartbeats, which had a weighted-F1 score of 86%, was the subject of the study in [13]. Using UTSU's method, MATLAB was utilized to detect images for heartbeat categorization detection in [14]. The level of accuracy was 95 percent. CNN was utilized for distinguishing heartbeat categorizations in [15], and had a general precision of 97.87%. Deep transfer learning is used in this study to demonstrate a method for categorizing heartbeats from MR images. The most common diagnosis at hand is an MRI, taking into account all of the effects of a CT scan. This paper will probably give a productive framework that recognizes heartbeat categorizations at a beginning phase with minimal expense and high effectiveness, which will decrease passings and experiencing because of heartbeat categorisations. The deep learning (DL) process is crucial to our ability to distinguish between cells that are affected by heartbeat categorisation and cells that are normal heartbeat cells. Furthermore, deep convolutional neural networks (CNN) are extremely useful for the study because the differences between the images are so small that traditional methods are unable to detect them. Most studies had an accuracy rate of 90–95 percent, which was considered excellent. However, the unique aspect of our study is that we employed a combination of well-known machine learning techniques to get the greatest results. With 97, 98, 97, and 96% F1-scores, the most effective algorithms were Random Forest (RF), Support Vector Machine (SVM), Voting Classifier (VC), and Logistic Regression (LR). The accuracy percentage of the models employed in this study is much higher than that of earlier studies, indicating that the models utilized in this study are more reliable. They've been proved to hold up in a variety of model comparisons, and the system might be derived from the study's findings.

As previously stated, the primary value of our study is that we utilized a publicly accessible dataset to test several machine learning models. Most researchers utilized a substantial model to predict heatbeat problem in earlier studies. However, we compared the findings to earlier studies and employed four distinct models. The next section discusses all of the findings and comparisons in detail. The following is the remainder of the article: Section 2 describes the procedure and experimental technique, whereas Section 3 examines the results. Section 4 has the conclusion.

2 Procedure and Experimental Methodology

Dataset description, block diagram, flow diagram, and evaluation matrices are included in this part, as well as the study's procedure and methodology.

2.1 Proposed System

The block diagram of the proposed system is shown in Figure 1.

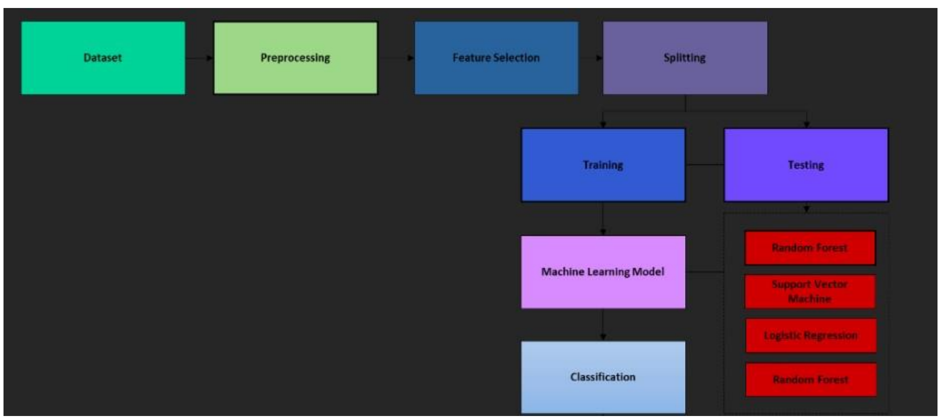

Figure 1: Proposed system architecture

After the data have been analyzed, they can be used to build models. A preprocessed dataset and machine learning algorithms are required for model creation. LR, SVM Classification, RF Classification, and a Voting Classifier are some of the methods used. Using the accuracy metrics Accuracy Score, Precision Score, Recall Score, and F1 Score, four alternative models are compared. The following subsections go over each one of the block diagram's components.

2.2 Dataset

The MIT-BIH Arrhythmia Dataset and The PTB Analytic ECG Data set, two notable heartbeat grouping datasets, contain the two assortments of heartbeat signals in this dataset [17]. A profound brain organization can be prepared involving the huge number of tests in the two assortments. This dataset was used to investigate heartbeat classification and observe some transfer learning capabilities by utilizing architectures of deep neural networks. The shapes of the heartbeats on the electrocardiogram (ECG) match the signals in both cases of normal heartbeat and cases of various arrhythmias and myocardial infarction. These signals have been preprocessed so that each segment can be compared to a heartbeat.

	0	1	2	3	4	5	6
count	87554.000000	87554.000000	87554.000000	87554.000000	87554.000000	87554.000000	87554.000000
mean	0.890360	0.758160	0.423972	0.219104	0.201127	0.210399	0.205808
std	0.240909	0.221813	0.227305	0.206878	0.177058	0.171909	0.178481
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.921922	0.682486	0.250969	0.048458	0.082329	0.088416	0.073333
50%	0.991342	0.826013	0.429472	0.166000	0.147878	0.158798	0.145324
75%	1.000000	0.910506	0.578767	0.341727	0.258993	0.287628	0.298237
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Figure 2: Statistical Description of the Dataset

On the other hand, the fact that some of the values in the dataset are missing indicates that some questions have been left blank due to privacy concerns raised by several patients. So, this study only looks at things that are easy to get to and have a less than 15% missing value for detecting heartbeat problems.

2.3 Preprocessing

The current heartbeat problem dataset will have many imbalances and missing values fixed by this step. Due to privacy concerns, many patients did not respond to questions, resulting in this lack of data. As a result, between 12 and 14 percent of all attributes lack data. The sample mean and mode were used as a data imputation strategy to fill in the values that were missing from the impacted characteristics. In addition, the dataset was subjected to the synthetic minority oversampling method (SMOTE) following the correction of missing data in order to correct the dataset's imbalance and generate new instances of the under sampled target class. After preprocessing, the data for the target column are shown in Figure 3.

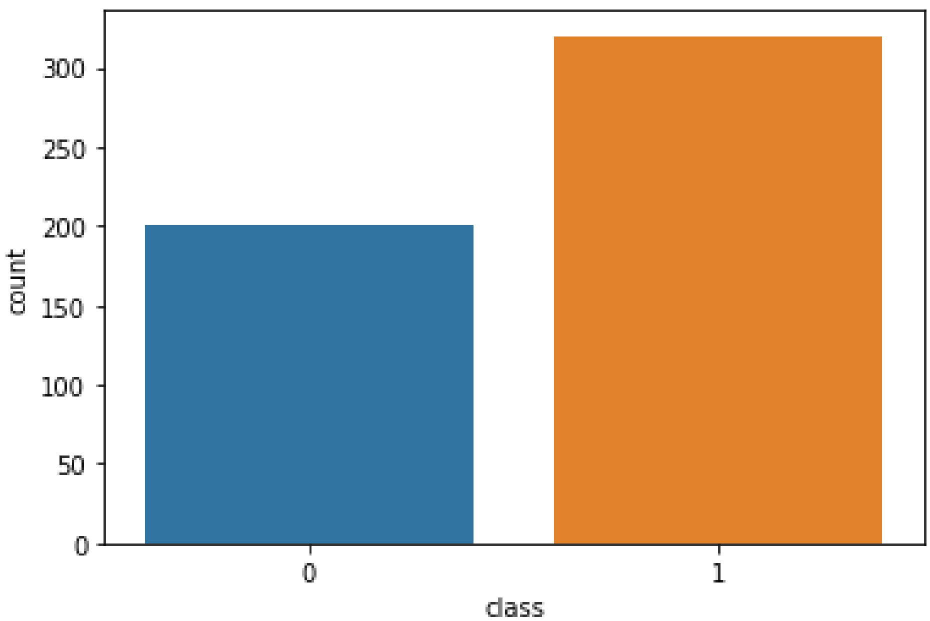

Figure 3: Target Column Data

Following data preparation and control of the unbalanced dataset, the model's design is the next step. In order to make this work more accurate and efficient, the data are divided into training and testing segments in an 80/20 ratio. After the model splits, a variety of classification algorithms are used to train it. This work employs Logistic Regression, Support Vector Machine, Random Forest, and Voting Classifier as classification techniques.

2.4 Algorithms

This part of the process uses algorithms like Random Forest Decision Tree (RF), Support Vector Machine (SVM), Logistic Regression (LR), and Voting Classifier (VC) to classify the data. A confusion matrix is used to figure out how well the algorithms did in terms of accuracy, precision, and recall.

2.4.1 Random Forest

Relapse and characterization related issues can be settled with the assistance of the irregular backwoods' calculation, a managed learning strategy. Due to its adaptability and ease of use, it is one of the machine learning algorithms that is used the most. Overfitting, which could be a major issue with such a complex algorithm, is reduced and accuracy is improved through the use of randomization. These methods construct decision trees from randomly selected data samples and derive predictions from each tree. After that, they vote on the option that is best for them. It is utilized, among other everyday tasks, in feature selectors, recommender systems, and image classifiers. Misrepresentation location, advance application order, and sickness forecast are just a few of its legitimate uses. The Boruta algorithm, which is used to find important data points, is based on it. Figure 4 depicts a block diagram of random forest categorization.

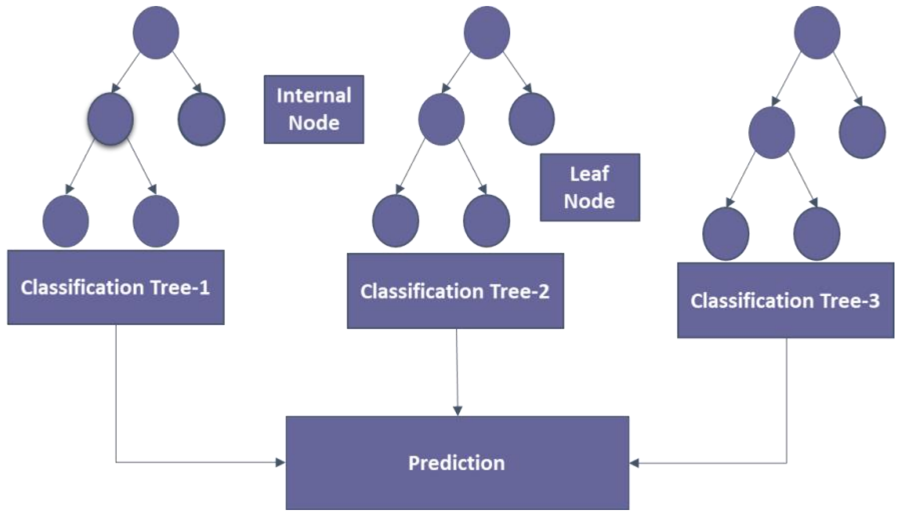

Figure 4: Block Diagram of Random Forest Classifier

The majority of non-linear classifiers fall short of the performance of the random forest method. Additionally, the fact that it draws its conclusion from a number of decision trees makes this approach quite resilient. Since it averages all predictions, the random forest classifier prevents overfitting by eliminating biases and resolving the overfitting issue. Missing values are not a problem for random forests. They could address this problem by either calculating the proximity-weighted average of the missing data or substituting median values for continuous variables. By comparing the importance of each characteristic, you can select the most crucial ones for your classifier.

2.4.2 Support Vector Machine

In N-dimensional space, where n is the number of features in a dataset, an SVM is a discriminative classifier that creates hyperplanes. Future data inputs can be distinguished with the help of these hyperplanes. The boundaries that aid in the classification of data points are called hyperplanes, decision boundaries, or decision planes. The side of the hyperplane where a new data point lands may belong to more than one class. The number of features assigned to a dataset is inversely proportional to the hyperplane's dimension. The hyperplane might be a straight line if the dataset only has two features. The hyperplane is a two-dimensional plane when there are three distinct characteristics in a dataset. The data points closest to the hyperplane that have an effect on its location are called support vectors. The term "Support Vector Machine Algorithm" comes from the fact that these vectors have an effect on where the hyperplane is positioned. Figure 5 depicts a block diagram of SVM categorization.

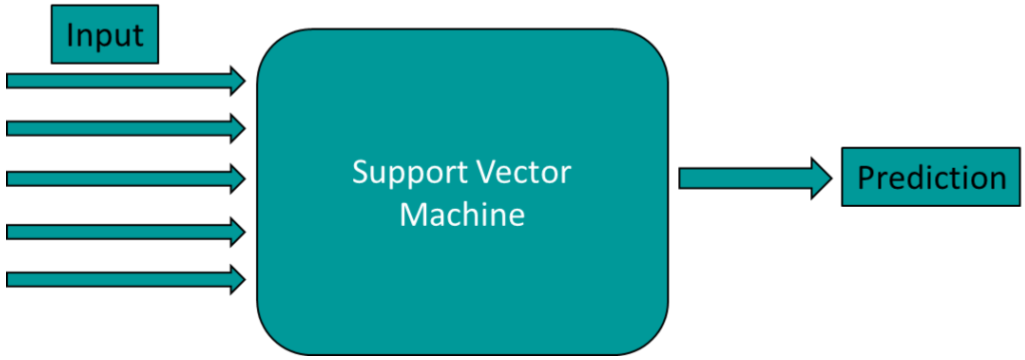

Figure 5: Basic structure of Support Vector Machine

Regression and classification problems can all be solved with SVM algorithms. It works well for getting non-linear data out of linear data. SVMs should not be used with large datasets. The task of selecting the best kernel for SVM is challenging. SVM also performs poorly when the number of features in each data set is less than the number of training data samples. The level of classification accuracy that the SVM algorithm must achieve is indicated by the C parameter. In a nutshell, the C parameter tells your model how much to penalize for each misclassified point on a particular curve. The model has an easier time correctly classifying all of its training examples when its C is low. The model is able to select a greater number of samples as support vectors thanks to its high C, which aids in the classification of all of its training examples.

2.4.3Voting Classifier

A voting classifier is a type of machine learning model that predicts an output (class) based on the class with the greatest likelihood of being chosen as the output and trains on an ensemble of multiple models. It forecasts the output class based on the class with the largest voting majority by simply combining the results of each classifier that are fed into the voting classifier. We create a single model that trains on these models and predicts output based on the overall majority of votes for each output class rather than creating discrete models for each output class and determining their accuracy. The block diagram for the voting classifier model is shown in Figure 6.

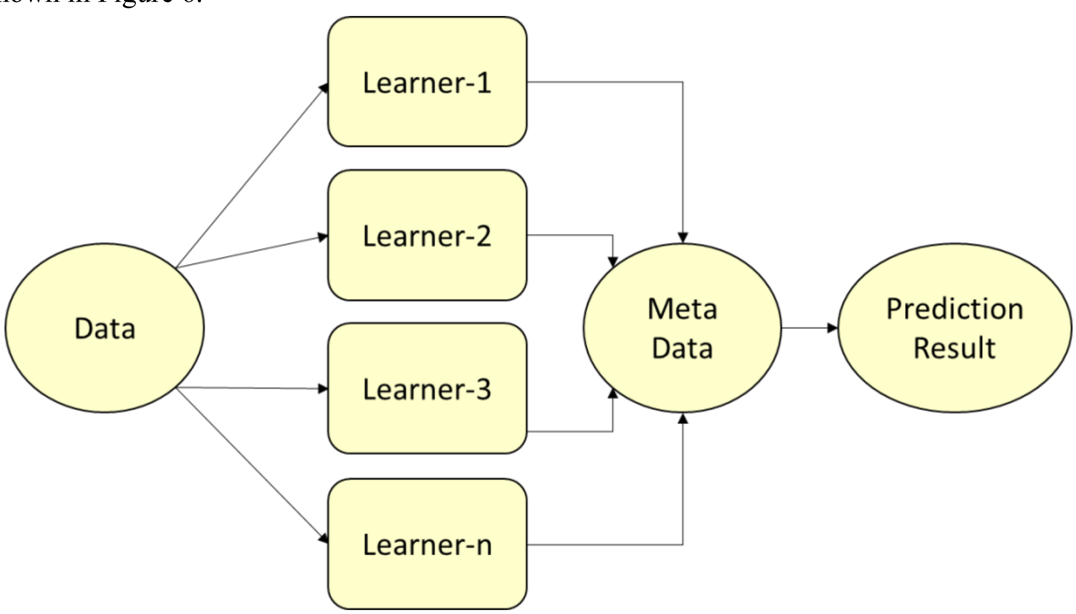

Figure 6: Block Diagram of Voting Classifier

The mechanism we will use to compare alternative training models is summarized by voting. There are two ways to vote:

- Soft Voting: For each model, the expected probability slopes are added and averaged in this step. The content of the category with the highest value is printed and declared the winner. Even though this seems like a good strategy, it should only be used if each subcategory is calibrated correctly. With the exception of the fact that the various models each contribute in a different way to the final output vector, this is similar to determining the weighted average of a group of integers.

- Hard Voting: The mode value of the final output is determined by combining the outputs of the various categorization models in this step. This method is comparable to finding the arithmetic mean of a set of integers because it ignores the specific probability values associated with each model. Only the output of each model is taken into consideration when evaluating it.

2.4.4 Logistic Regression

A discriminative model based on binary classification is the logistic regression classifier. The logistic sigmoid function that was utilized in the prediction model is referred to as "logistic regression." Additionally, the sigmoid function is frequently utilized as an activation function in neural networks. The Logistic Regression model's logic flowchart can be seen in Figure 7. In the context of supervised learning, LR is one of the machine learning algorithms that is utilized the most frequently. It is a modeling technique that uses a set of independent variables to forecast a categorical response variable.

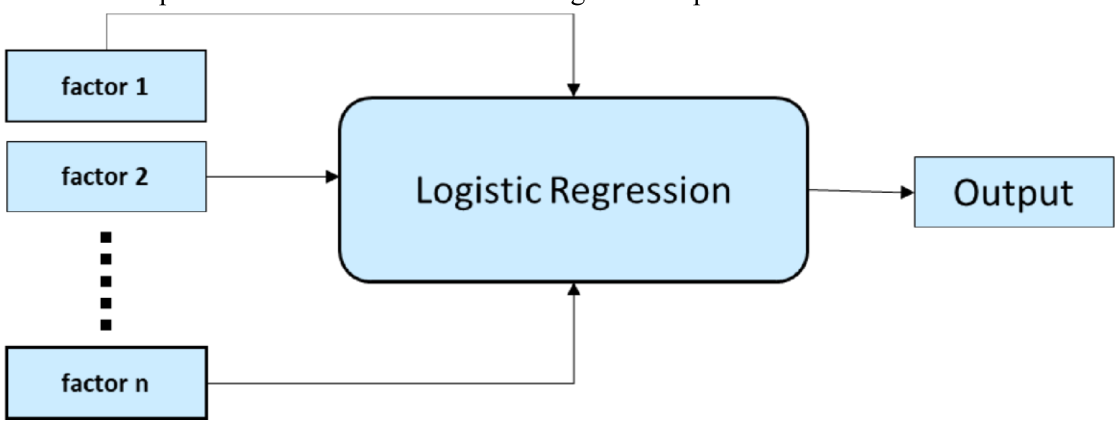

Figure 7: Block Diagram of Logistic Regression Classifier

When predicting the outcome of a categorical dependent variable, logistic regression is utilized. The output format must therefore be discrete or categorical. It could be true or false, yes or no, 0 or 1, true or false, or any other number between 0 and 1, but only probabilities between 0 and 1 are specified. Linear regression and logistic regression share a lot of similarities. Linear Regression is used to solve regression issues, whereas Logistic Regression is used to solve classification issues. We use a logistic function with an "S" shape that predicts two maximum values instead of a regression line.

2.5 Evaluation Matrix

Figure 8 depicts the evaluation matrix, also known as the confusion matrix. A metric for evaluating the efficiency of machine learning algorithms is the confusion matrix. Each model's efficacy was evaluated with the help of the confusion matrix. The confusion matrix depicts the frequency with which our models correctly predict and the frequency with which they underestimate. False positives and false negatives have been ascribed to values that were incorrectly predicted, whilst true positives and true negatives have been given to values that were correctly predicted. After grouping all the predicted values into a matrix, the accuracy, precision-recall trade-off, and AUC of the model were used to see how well it worked.

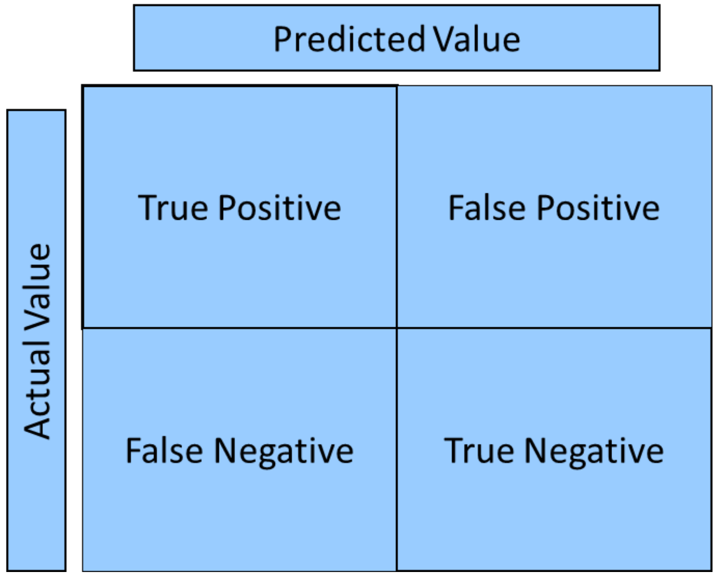

Figure 8: Block diagram of the confusion matrix

The classifier's errors and the kinds of errors they make are detailed in the confusion matrix. While making predictions, it demonstrates how disorganized and confusing a categorization model is. The limitations imposed by relying solely on categorization accuracy can be circumvented with the assistance of this feature. It is used when one class dominates over others and the categorization problem is extremely unbalanced. Recall, precision, specificity, accuracy, and the AUC-ROC Curve can all be calculated using the confusion matrix in an extremely efficient manner.

3 Result Analysis

The experiments' objective is to determine the optimal classification system for heatbeat problem diagnosis. The performance of random forest, logistic regression, voting classifier, and support vector machines is examined and compared in terms of accuracy, precision, and recall using a heatbeat problem risk factor dataset and classification in machine learning.

3.1 Data Visualization

The graphical representation of a frequency distribution with categorized continuous classes is known as a data visualization. Figure 9 shows the visualization of the heartbeat problem dataset.

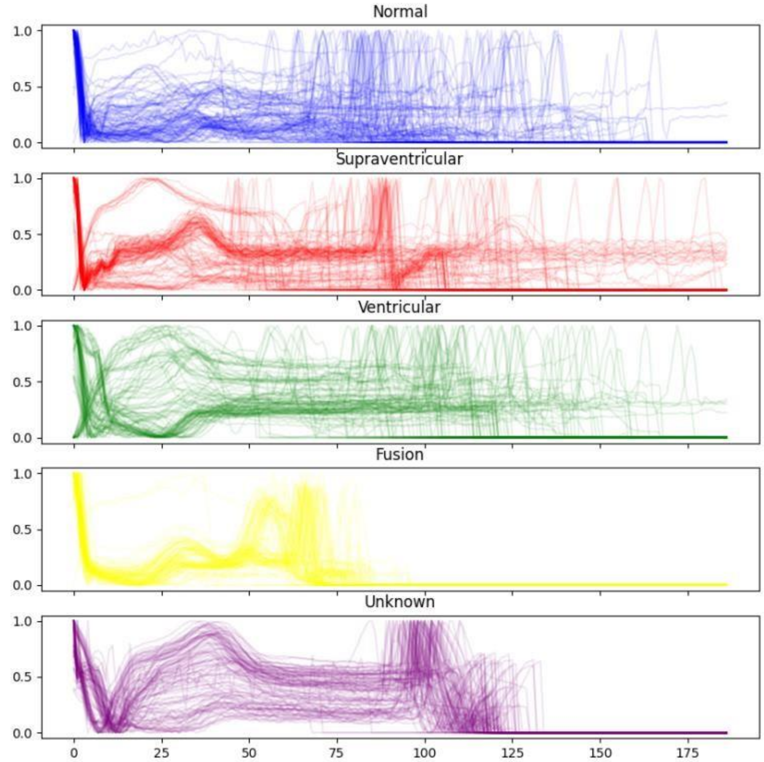

Figure 9: Visualization of some features of the dataset

It is known as an area diagram, and its definition states that it is a collection of rectangles with areas that correspond to frequencies in the associated classes and bases and intervals between class borders. In such representations, each rectangle is contiguous because the base spans the spaces between class borders. The matching frequencies of similar classes determine the heights of rectangles, whereas the matching frequency densities of dissimilar classes determine the heights of rectangles.

3.2 Evaluation of The Model

3.3.1 Random Forest (RF)

The classification result for the RF model is shown in Figure 10.

```
[ ]  print(classification_report(y_test,y_pred_rf))

              precision    recall  f1-score   support

           0       0.99      0.95      0.97       179
           1       0.94      0.99      0.97       143

    accuracy                           0.97       322
   macro avg       0.97      0.97      0.97       322
weighted avg       0.97      0.97      0.97       322
```

Figure 10: Classification result of Random Forest

The overall F1 score, in this case, is 97 percent. Individual F1 scores range from 97 percent for healthy people to 97 percent for people who have had heartbeat problem. Precision for 0 is 99 percent and for 1 is 94 percent. Prior to fine-tuning, the model had an accuracy of 93 percent. Figure 11 displays the Random Forest model's estimate.

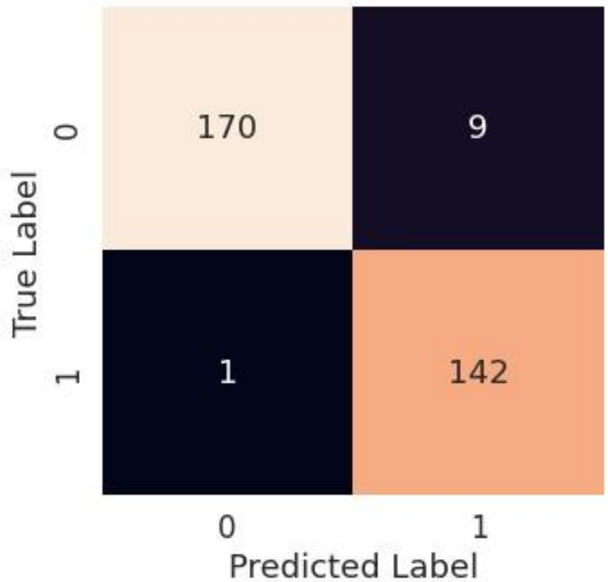

Figure 11: Confusion Matrix of Random Forest

The confusion matrix displays the projected result as well as the estimated performance of the model. There were 312 correct predictions and 10 incorrect ones.

3.2.2 Support Vector Machine

Figure 12 shows the SVM Classifier classification report.

```
[ ] print(classification_report(y_test,y_pred_svm))

              precision    recall  f1-score   support

           0       0.99      0.97      0.98       179
           1       0.96      0.99      0.97       143

    accuracy                           0.98       322
   macro avg       0.97      0.98      0.97       322
weighted avg       0.98      0.98      0.98       322
```

Figure 12: Classification Report of SVM

In this case, the final F1 score is 98 percent of the total. According to the study, the F1 score of a healthy individual is 98 percent, but the score of a person who has had heatbeat problem is 97 percent. Figure 13 also depicts the precision and recall of the test results. An SVM model that has been fine-tuned has also been incorporated. However, even after fine-tuning, the accuracy remained unchanged.

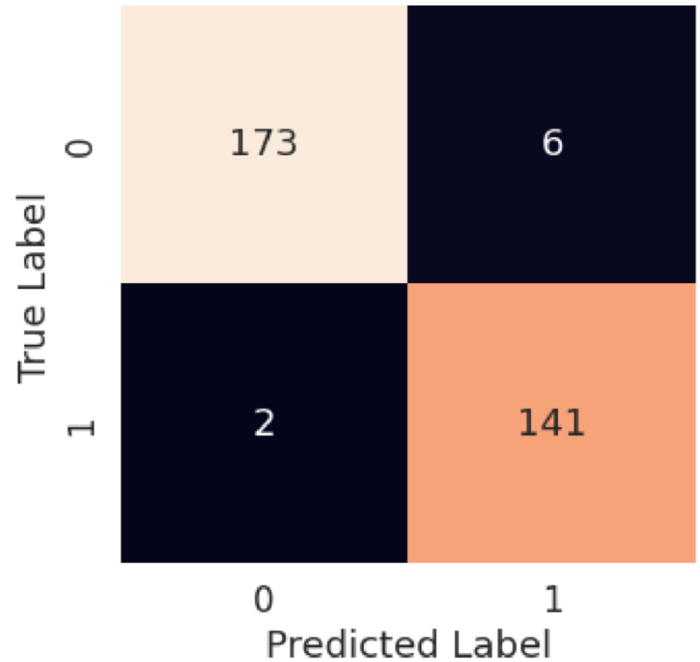

Figure 13: Confusion Matrix of SVM classifier

The forecast made by the SVM model is seen in Figure 14. There were 314 correct predictions and just eight incorrect predictions.

3.2.3 Voting Classifier

Figure 14 shows the categorization report that the Voting Classifier makes for use in voting.

```
y_pred_VC = VC.predict(X_test_s)
print(classification_report(y_test, y_pred_VC))

              precision    recall  f1-score   support

           0       0.99      0.96      0.97       179
           1       0.95      0.99      0.97       143

    accuracy                           0.97       322
   macro avg       0.97      0.97      0.97       322
weighted avg       0.97      0.97      0.97       322
```

Figure 14: Classification Report of Voting Classifier

In this case, the final F1 score is 97 percent of the total. According to the study, the F1 score of a healthy individual is 97 percent, but the score of a person who has had heatbeat problem is 97 percent. Figure 15 also depicts the precision and recall of the test results. A voting classifier model that has been fine-tuned has also been incorporated. However, just like SVM, even after fine-tuning, the accuracy remained unchanged.

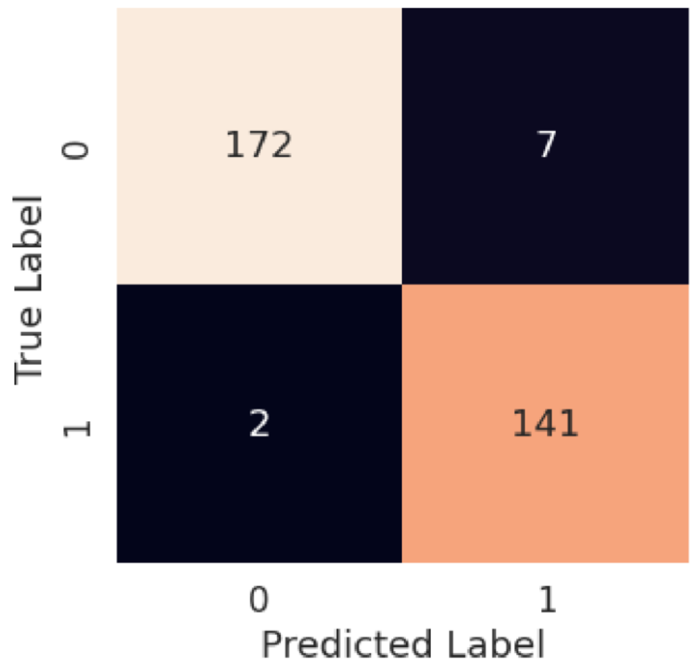

Figure 15: Confusion Matrix of Voting Classifier

The forecast made by the Voting classifier model is seen in Figure 16. There were 313 correct predictions and just 9 incorrect predictions.

3.2.4 Logistic Regression Classifier

Figure 17 shows the categorization report that the Logistic Regression classifier.

```
[ ]  print(classification_report(y_test,y_pred_lr))

              precision    recall  f1-score   support

           0       0.98      0.96      0.97       179
           1       0.95      0.97      0.96       143

    accuracy                           0.96       322
   macro avg       0.96      0.96      0.96       322
weighted avg       0.96      0.96      0.96       322
```

Figure 16: Classification Report of Logistic Regression Classifier

In this case, the final F1 score is 96 percent of the total. According to the study, the F1 score of a healthy individual is 97 percent, but the score of a person who has had heatbeat problem is 96 percent. Figure 17 also depicts the precision and recall of the test results. A Logistic Regression classifier model that has been fine-tuned has also been incorporated. However, just like SVM and Voting classifier, even after fine-tuning, the accuracy remained unchanged.

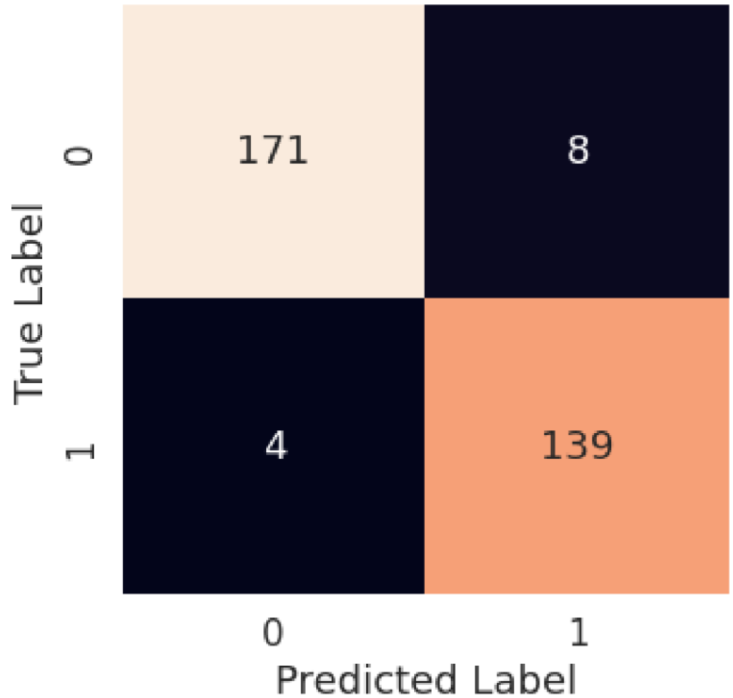

Figure 17: Confusion Matrix of Logistic Regression Classifier

The forecast made by the Logistic Regression classifier model is seen in Figure 17. There were 310 correct predictions and just 12 incorrect predictions.

3.2.5 CNN

The accuracy of the CNN model is shown in Figure 18. This model has the 99% accuracy in classifying objects.

```
[ ]  print(classification_report(test_y, predictions))

              precision    recall  f1-score   support

           0       0.99      1.00      0.99       915
           1       0.99      0.94      0.97       235

    accuracy                           0.99      1150
   macro avg       0.99      0.97      0.98      1150
weighted avg       0.99      0.99      0.99      1150
```

Figure 18: CNN Classifier Accuracy

In this example, the overall accuracy is 99 percent. Individual f1-score is 99 and 97 percent for 0 and 1, respectively. Figure 19 depicts the CNN model's final prediction number.

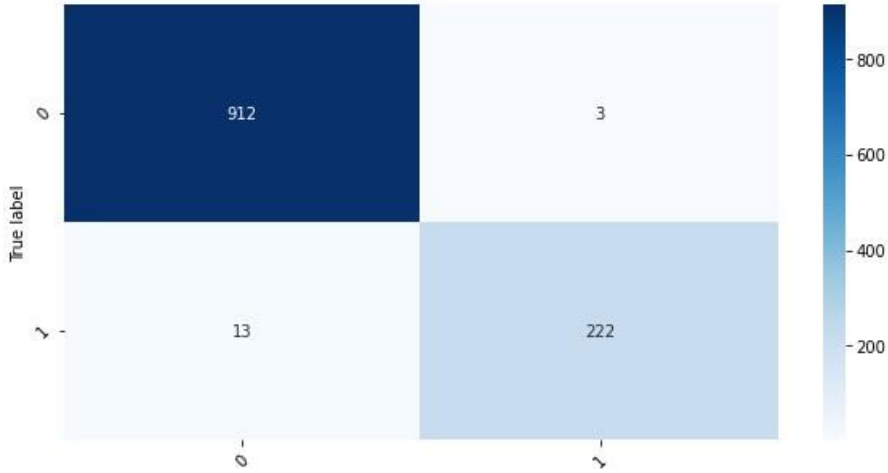

Figure 19: CNN Confusion Matrix

The confusion matrix illustrates the expected result as well as the estimated performance of the model. For a total of 99 percent accuracy, there were 1134 right predictions and 16 wrong forecasts

3.3 Model Comparison

The graphic clearly illustrates that the CNN model is the most effective model among the numerous models contained in the framework. Additionally, it has greater precision, better memory, and accuracy, in addition to having a higher F1 score.

Table 1: Algorithm Performance Comparison

This Paper (Model Name)	Accuracy (%)	Reference Paper (Model Name)	Accuracy (%)
Random Forest	99	Ref [10] Random Forest	80.18
Support Vector Machine	**94**	Ref [9] Support Vector Machine	90.0
Voting Classifier	97	Ref [11] Artificial Neural Network	54.0
Logistic Regression	96	Ref [15] Logistic Regression	74.56
CNN	99	Ref [18] ANN	83
Decision Tree	99	Ref [10] Decision Tree	80.18

All of the algorithms have an acceptable level of accuracy, as shown in Table 1, but the DL method is the preferred option due to its higher level of accuracy than the others. The accuracy of this study's CNN approach was 99 percent, whereas the authors of [9] achieved only 90 percent with their algorithm. The Random Forest method was used in this study to achieve 99 percent accuracy, while the same algorithm was used by the authors of [10] to achieve 80.18 percent accuracy. A voting classifier was used in this study to achieve 97 percent accuracy, while artificial neural networks were used in the Artificial Neural Network ref [11] to achieve 54 percent accuracy. On the other hand, this article achieved a 96% accuracy rate with a Logistic Regression classifier, while reference [15] achieved a 74% accuracy rate with the same approach.

4 Conclusion

In the field of machine learning-based illness prediction, medical data are essential. Data detection and diagnostics applications rely on machine learning to learn from previous data and make predictions based on that knowledge. The review shows that the CNN method and random forest could be used to better predict seizures caused by epilepsy. These findings could be added to the current research in the area of heatbeat problem prediction utilizing emerging machine learning technologies. The investigation revealed that their individual accuracy was almost 100% and 95%. The fact that the precision percentage of the models used in this request is significantly higher than that of the models used in previous studies suggests that the models used in this study are more trustworthy than those used in previous studies. When crossendorsement assessments are used in the assumption for epileptic seizures, the CNN system beats various cycles. Future research could build on this work by developing a web application that uses these calculations and a larger dataset than the one used in this review. This will help improve outcomes as well as the accuracy and efficacy with which medical professionals can anticipate epilepsy issues. This will be beneficial to the system's appearance as well as the structure's unwavering quality. The goal is to get people to get treatment for epilepsy early and make certain changes to their daily lives. On the other hand, the fact that we utilized a variety of well-known machine learning algorithms to achieve the best results makes our work distinctive. With F1 ratings of 97, 98, 97, and 96 percent, Random Forest (RF), Support Vector Machine (SVM), Voting Classifier (VC), and Logistic Regression (LR) were the most successful algorithms (LR). The fact that the models used in this investigation have a much higher percentage of reliability than the models used in previous studies indicates that these models are more reliable than their predecessors. They were found to be reliable in a number of model comparisons. The study's findings could serve as the basis for the system. Utilizing a larger dataset and machine learning algorithms like AdaBoost and Bagging to enhance the framework models are potential future directions for this study. This increase in dependability will have a positive impact on the framework's presentation as well as its dependability. In exchange for providing some basic information, the machine-learning architecture may be able to assist the general public in determining the likelihood of an adult patient with a heartbeat problem developing. It would, in an ideal world, aid patients in receiving prompt treatment for heartbeat issues and reestablishing their lives following a tragedy.

Data Availability Statement: The data utilized to support these research findings is accessible online at:

https://www.kaggle.com/datasets/shayanfazeli/heartbeat **Funding Statement:**

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. H. O. R. Health, W. H. O. C. Diseases, and H. Promotion, Comprehensive Heartbeat problem Control: A Guide to Essential Practice, World Health Organization, Geneva, Switzerland, 2006.

[2] Heartbeat problem: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow up N Colombo and others Annals of Oncology, 2012. Volume 23, Supplement 7

[3] I. C. Scarinci et al., "Heartbeat problem Prevention: New Tools and Old Barriers," Disease, vol. 116, no. 11, pp. 2531–2542, 2011.

[4] John Doorbar, "Molecular biology of human papillomavirus infection and heartbeat problem", Clinical Science,May 01, 2006,110(5)525-541;DOI: 10.1042/CS20050369.

[5] D. Saslow et al., "American disease society, American society for colposcopy and heartbeat problem pathology, and American society for clinical pathology screening guidelines for the prevention and early detection of heartbeat problem," CA-Disease J. Clin., vol. 62, pp. 147–172, 2012.

[6] Igor Kononenko,"Machine learning for medical diagnosis: history, state of the art and perspective", Artificial Intelligence in Medicine Volume 23, Issue 1, August 2001, Pages 89-109.

[7] NasserH. Sweilama , A.A.Tharwatb , N.K.Abdel Moniemc , "Support vector machine for diagnosis disease disease: A comparative study", Egyptian Informatics Journal,Volume 11, Issue 2, December 2010, Pages 81-92.

[8] Y. Weng, C. Wu, Q. Jiang, W. Guo and C. Wang, "Application of support vector machines in medical data," 2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), Beijing,2016,pp.200-204.

[9] W. Wu and H. Zhou, "Data-Driven Diagnosis of Heartbeat problem With Support Vector MachineBased Approaches," in IEEE Access, vol. 5, pp. 25189-25195, 2017.

[10] Y. E. Kurniawati, A. E. Permanasari and S. Fauziati, "Comparative study on data mining classification methods for heartbeat problem prediction using pap smear results," 2016 1st International Conference on Biomedical Engineering (IBIOMED), Yogyakarta, 2016, pp. 1-5.

[11] Priyanka K Malli , Dr. Suvarna Nandyal, “Machine learning Technique for detection of Heatbeat problem using k-NN and Artificial Neural Network”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), JulyAugust 2017. [12] R. Vidya1* and G. M. Nasira2, “Prediction of Heartbeat problem using Hybrid Induction Technique:

A Solution for Human Hereditary Disease Patterns”, Indian Journal of Science and Technology, August 2016.

[13] D. Kashyap et al., "Heartbeat problem detection and classification using Independent Level sets and multi SVMs," 2016 39th International Conference on Telecommunications and Signal Processing (TSP), Vienna, 2016, pp. 523-528.

[14] E. Njoroge, S. R. Alty, M. R. Gani and M. Alkatib, "Classification of Heartbeat problem Cells using FTIR Data," 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, 2006, pp. 5338-5341.

[15] J. Hyeon, H. J. Choi, K. N. Lee and B. D. Lee, "Automating Papanicolaou Test Using Deep Convolutional Activation Feature," 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 382-385.

[16] K. Teeyapan, N. Theera-Umpon and S. Auephanwiriyakul, "Application of support vector based methods for heatbeat problem cell classification," 2015 IEEE International Conference on Control System,Computing and Engineering (ICCSCE), George Town, 2015, pp. 514-519.

[17] Heatbeat problem (Risk Factors) Data Set. Available on: https://www.kaggle.com/datasets/shayanfazeli/heartbeat

[18] Auria, Laura and Moro, R. A., Support Vector Machines (SVM) as a Technique for Solvency Analysis (August 1, 2008). DIW Berlin Discussion Paper No. 811. Available at SSRN: https://ssrn.com/abstract=1424949 or http://dx.doi.org/ 10.2139/ssrn.1424949