University of Ibadan
Journal of Science and Logics in ICT Research

Development of an Intelligent Model for Cardiac Arrest Prediction using Radio Frequency Identification (RFID) And Machine Learning Algorithm

[1]OGUNTUNDE T., [2]AYINDE O. A.

Computer Science Department, University of Ibadan
[1]toyin.oguntunde@gmail.com, *[2]* ayindeoyedamola@gmail.com

Abstract

Physiological instability or abnormal vital signs such as heart beat rate, respiratory rate or blood pressure suggests altered physiology. It is well recognized that abnormal physiology is associated with adverse clinical outcomes. The higher physiology deviates from normal, the higher the risk of mortality such as cardiac arrest. Most Patients showed evidence of physiological abnormality prior to the event of the arrest. This research work is intended to provide an improved model for the investigation and prediction of physiological characteristics of cardiac arrest via prediction using Radio Frequency Identification (RFID) and Machine Learning techniques. The model was developed using Bidirectional Long Short Term Memory with Conditional Random Field and a One Dimensional Convolutional Neural Network. The model takes as input a vector of the heart disease dataset and produces as output a classification of the prediction based on the analysed heart disease dataset. The accuracy of this research work is rated 98.3%. It was also discovered that the use of an autoencoder to actively learn the dataset played an important role in the outcome of the results

Keywords: Case reports, Abnormal vital signs, Radio frequency identification, Hidden states, Highest positive correlation

1. INTRODUCTION

In medicine, a case report is a detailed report of the symptoms, signs, diagnosis, treatment, and follow-up of an individual patient. Case reports may contain a demographic profile of the patient, but usually describe an unusual or novel occurrence. Some case reports also contain a literature review of other reported cases. They can be shared for medical, scientific or educational purposes.

Physiological instability or abnormal vital signs such as heart rate, respiratory rate or blood pressure suggests altered physiology. It is well recognized that abnormal physiology is associated with adverse clinical outcomes. The higher physiology deviates from normal, the higher the risk of mortality such as cardiac arrest [6]. This is a key finding which precipitated further research in to the recognition and management of physiological abnormalities in preventing cardiac arrest.

Healthcare is a high-risk industry. Urgent, unanticipated admission to critical care from acute care wards is an untoward occurrence which constitutes a serious adverse event. The prevention of the occurrence of adverse event, like cardiac arrest, requires the constant monitoring of patients physiological instability.

Physiological instability or abnormal vital signs such as heart rate, respiratory rate or blood pressure suggests altered physiology. Case reports are professional narratives that provide feedback on clinical practice guidelines and offer a framework for early signals of effectiveness, adverse events and cost.

This research seeks to prevent cardiac arrest by monitoring patient's physiological instability through their case report and signaling if there are risks of cardiac arrest in the future. This would be

OGUNTUNDE, T., AYINDE, O. A. (2021), Development of an Intelligent Model for Cardiac Arrest Prediction Using Radio Frequency Identification (RFID) and Machine Learning Algorithm, *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 6 No. 2, pp. 171-183

achieved by developing a model that uses Radio Frequency Identification (RFID) and machine learning algorithm to monitor and make predictions based on what is learned from studying the patient's physiological information via the case report.

2. Literature Review

Frank [4] worked on creating Decision tree quickly with clinical data of the physician or service. He suggested few data mining techniques which can help cardiologists in the predication survival of patients. The main drawback of the system was that the user needs to have knowledge of the techniques and they needed to collect sufficient data so as to be able to create a suitable model.

Boleslaw [3] operated on a novel experiential to check the aptitude of calculation of scarce kernel in SUPANOVA. The author used this technique on a standard boston housing market dataset for discovering heart diseases, measurement of heart activities and prediction of heart diseases were found 83.7% correct which were measured with the help of Support Vector Machine and kernel equivalent to it. A quality result is gained by spline kernel with the help of standard boston housing market database.

Kuyong [5] made use of a classification technique for removal of multi-parametric structures by accessing HRV and ECG signals. Kuyong used the FP growth algorithm as the foundation of this technique that is associative. A rule consistency degree was gained which allows a robust press on trimming designs in the method of producing designs.

Anurag *et. al.* [2] focused their efforts to predict the possibility of occurring of these quick and killing attacks of cardiac arrest using decision model based predictive analytics techniques, so that they can analyze and find some patterns that are common in the happenings of Sudden Cardiac Death. They also proposed a framework that can be implemented for emergency situations of people with such medical history and the raw datasets are further analyzed from scratch to predict the upcoming life threatening pain that might cause death.

Shashikant and Chetankumar [8] compared the performance of logistical regression, decision tree, and random forest model to predict cardiac arrest in smokers. They applied machine learning technique implemented on the dataset received from the data science research group MITU Skillogies Pune, India. To know if the patient has a chance of cardiac arrest or not, they developed three predictive models as 19 input feature of HRV indices and two output classes. These model evaluated based on their accuracy, precision, sensitivity, specificity, F1 score, and area under the curve (AUC). They claimed that the model of logistic regression achieved an accuracy of 88.50%, precision of 83.11%, the sensitivity of 91.79%, the specificity of 86.03%, F1 score of 0.87, and AUC of 0.88. They also claimed that the decision tree model had an accuracy of 92.59%, precision of 97.29%, the sensitivity of 90.11%, the specificity of 97.38%, F1 score of 0.93, and AUC of 0.94 and Finally, they claimed that random forest achieved an accuracy of 93.61%, precision of 94.59%, the sensitivity of 92.11%, the specificity of 95.03%, F1 score of 0.93 and AUC of 0.95.

Furthermore, Nachket *et. al.* [7] observed that the changing risk of and changing influence of various clinical, angiographic and electrophysiological parameters on subsequent cardiac arrest recurrence with time. Various medical and synthetic datasets such as ECG dataset from PhysioNet, Pima Indian Diabetes dataset from UCI Machine Learning Repository and gene expression dataset from GEO were used, which are unique as compared with related works. Various classifiers such as LogitBoost with simple regression function, random forest and multilayer perceptron were used for recurrence risk prediction. Collection of these classifiers together forms the ensemble classifiers. The classifiers were compared based on various measures like accuracy and precision. Based on the classification, risk scores are calculated using logistic regression with backward elimination.

Moreover, Ahmed [1] used data-driven techniques based on machine learning (ML) to improve the performance of risk predictions by agnostically discovering novel risk predictors and learning the complex interactions between them. They tested (1) whether ML techniques based on a state-of-the-art automated ML

framework (AutoPrognosis) could improve CVD risk prediction compared to traditional approaches, and (2) whether considering non-traditional variables could increase the accuracy of CVD risk predictions. Using data on 423,604 participants without CVD at baseline in UK Biobank, they developed a ML-based model for predicting CVD risk based on 473 available variables. They claim that their ML-based model was derived using AutoPrognosis, an algorithmic tool that automatically selects and tunes ensembles of ML modeling pipelines (comprising data imputation, feature processing, classification and calibration algorithms).

According to Cornegruta *et al.* in figure 1, described a **Bidirectional LSTM**, or **biLSTM**, as a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction. BiLSTMs effectively increase the amount of information available to the network, improving the context available to the algorithm (e.g. knowing what words immediately follow *and* precede a word in a sentence).

3. Methodology

The methodology is such that employs both radio frequency identification and machine learning algorithm to perform a classification of clinical information from a case-note. RFID technology has been in use for several decades to track and identify goods, assets and even living things. RFID technology is already being used in many applications, such as for toll booth automation and self-service retail processes. In library and office environments, RFID technology allows the tracking and fast retrieval of physical files, thereby increasing overall efficiency and productivity.

Figure 2 depicts the proposed model for an Intelligent Cardiac Arrest Prediction using radio Frequency Identification and Machine Learning algorithm. The RFID generated dataset were fed into the preprocessing unit. In the preprocessing unit, data are cleaned, reduced and feature selection was done. The cleaned data was splitted into training and test datasets. The training dataset was to train the model. The test data was run through the trained model to be able to predict and the results of prediction were displayed with visualizations.

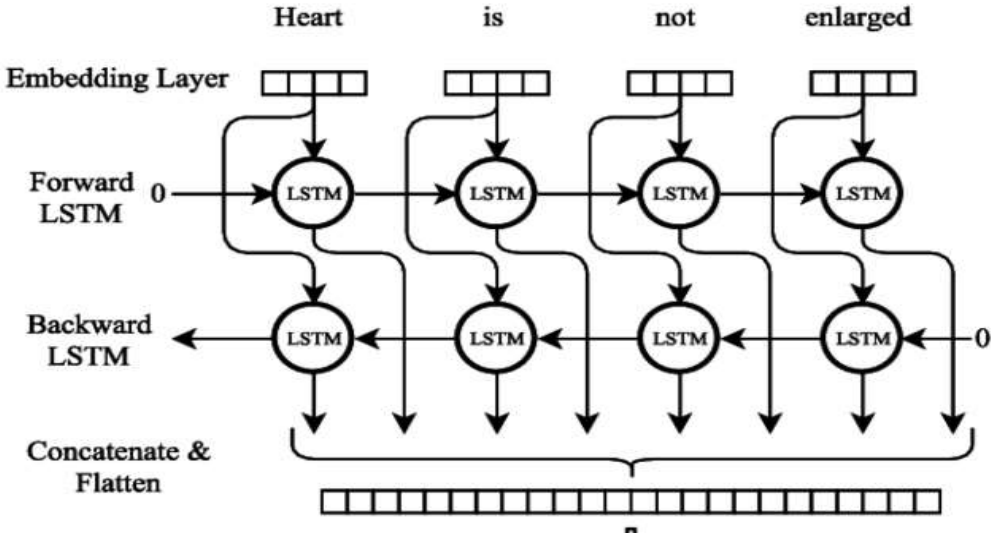

Figure 1: BiLSTM Model

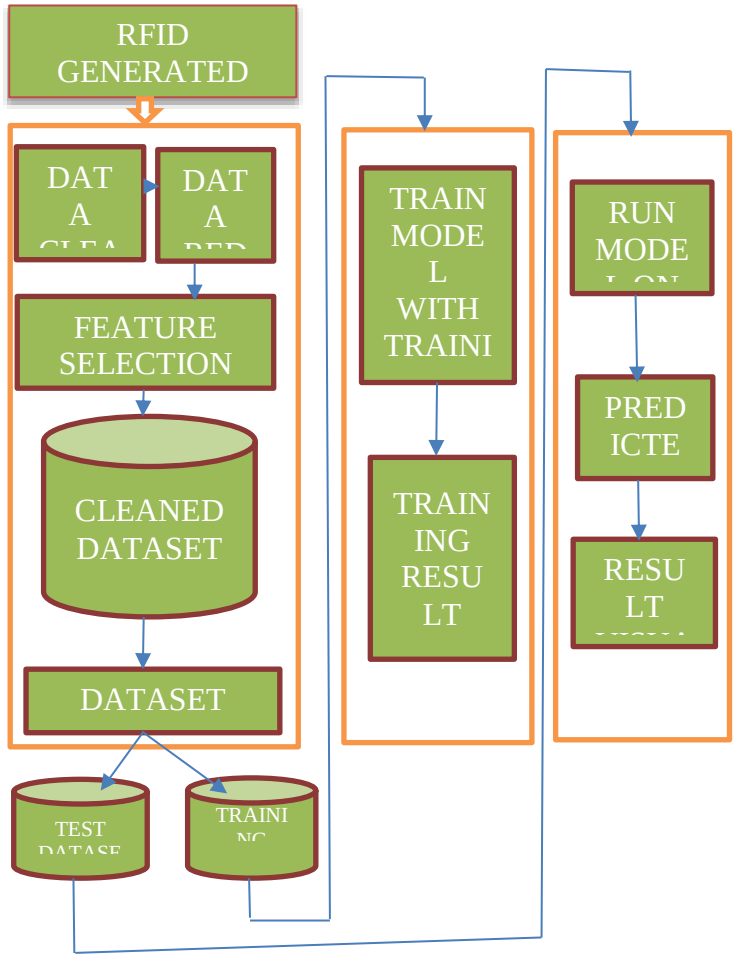

Figure 2: Model of an Intelligent Cardiac Arrest Prediction using Radio Frequency Identification and Machine Learning Algorithm

This next phase of the methodology involves the dataset used for the research work and how the dataset will be processed to make it ready for use in our analysis. There are two steps involved in this phase, the first being the setting up of an RFID chip to hold medical information about a patient. The idea is to make it available to the model that will be designed to access for analysis, the second phase involves the development and training of the model using the Cleveland Heart Disease (UCI Repository) dataset.

3.1 *The Cleveland Heart Disease (UCI Repository) Dataset*

The dataset from UCI repository consists of 303 individuals' data. There are 14 columns in the dataset, which are described below.

1. **Age:** displays the age of the individual.
2. **Sex:** displays the gender of the individual using the following format:1 = male, 0 = female
3. **Chest-pain type:** displays the type of chest-pain experienced by the individual using the following format: 1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptotic
4. **Resting Blood Pressure:** displays the resting blood pressure value of an individual in mmHg (unit)
5. **Serum Cholesterol:** displays the serum cholesterol in mg/dl (unit)
6. **Fasting Blood Sugar:** compares the fasting blood sugar value of an individual with 120mg/dl.
   If fasting blood sugar > 120mg/dl then: 1 (true) else: 0 (false)
7. **Resting ECG :** displays resting electrocardiographic results 0 = normal,1

= having ST-T wave abnormality, 2 = left ventricular hypertrophy

8. **Max heart rate achieved:** displays the max heart rate achieved by an individual.
9. **Exercise induced angina :** 1 = yes, 0 = no
10. **ST depression induced by exercise relative to rest:** displays the value which is an integer or float.
11. **Peak exercise ST segment :** 1 = upsloping, 2 = flat , 3 = downsloping
12. **Number of major vessels (0–3) coloured by fluoroscopy:** displays the value as integer or float.
13. **Thal: displays the thalassemia :**3 = normal , 6 = fixed defect, 7 = reversible defect
14. **Diagnosis of heart disease:** Displays whether the individual is suffering from heart disease or not: 0 = absence, 1, 2, 3, 4 = present.

These parameters were selected because, in the actual dataset, we had 76 features but for our study, we chose only the above 14 features.

3.2 *The Work Flow of the Proposed Model*

In figure 3, the proposed model's work flow is presented with all the components. Although the BiLSTM layer has proven powerful for handling temporal correlation, it contains too much redundancy for spatial data. To address this problem, an extension of BiLSTM was proposed which has convolutional structures in both the input-to-state and state-to-state transitions. By stacking multiple Convolutional layers and forming an encoding-forecasting structure with the use of the autoencoder, a network was built to model not only for the cardiovascular disease prediction problem but also for more general spatiotemporal sequence forecasting problem.

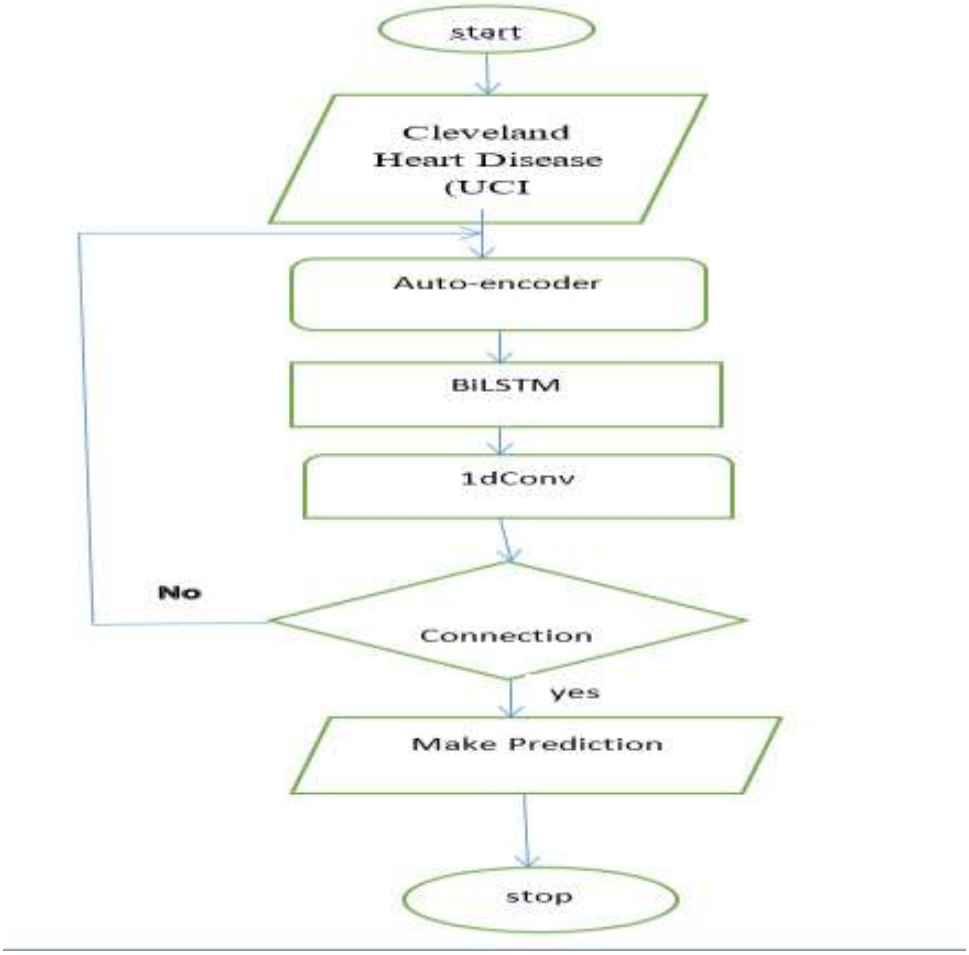

Figure 3: Work flow of the Model

In order to view the states as the hidden representations of moving objects, a BiLSTM and Impulse Detection Convolutional Neural Network (1dCNN) with a larger transitional kernel should be able to capture faster motions while one with a smaller kernel can capture slower motions. To ensure that the states have the same number of rows and columns as the inputs, padding is needed before applying the convolution operation. Here, padding of the hidden states on the boundary points can be viewed as using the state of the outside world for calculation. Usually, before the first input comes, all the states of the BiLSTM were initialized to zero which corresponds to "total ignorance" of the future. Similarly, if zero-padding were performed on the hidden states, the state of the outside world were are being set to zero and assume no prior knowledge about the outside.

This module is responsible for the extraction and aggregation of the cardiovascular disease predicted by the Bi-directional Long Short-Term Memory and Conditional Random Field (Bi-LSTM-CRF) and 1DConv algorithm. The predicted cardiovascular disease extractor extracts predicted cardiovascular disease made by the Bi-LSTM-CRF and Impulse Detection Convolutional Neural Network (1d-CNN) algorithm while the predicted cardiovascular disease aggregator aggregates the predicted precipitation extracted by the predicted cardiovascular disease extractor module. This way, the aggregated predicted cardiovascular disease that has been extracted can be presented in a coherent manner by the result presentation module. The collection of these three (3) modules ensures that the predictions made by the Bi-LSTM-CRF and 1DConv algorithm are presented in a human readable form.

3.3 Experimental Analysis

3.3.1 Data And Parameters

For training the BiLSTM and 1DConv classification model for Cardiac Arrest, Cleveland Heart Disease (UCI Repository) dataset was used. The dataset used for this study is publicly available. This is a sample of the dataset with statistical information in figure 4:

BiLSTM was employed with one dimensional convolutional neural network architecture to predict the presence of cardiovascular disease. The architecture is composed of two networks: an autoencoder network and a multilayer perceptron network. The autoencoder network is responsible for feature selection and as mentioned in chapter three, the autoencoder is a BiLSTM with a carry forward function for the feature treatment in the time series part of the dataset. A multilayer perceptron network is responsible for making classification and prediction.

The first element in the architecture is the autoencoder. An autoencoder is an unsupervised network that aims to extract non-linear features for a data in-put. Being more specific, an autoencoder is composed by three layers: the input layer, a hidden layer using the sigmoid activation function, and the output layer. The autoencoder is trained so that the output layer attempts to be as similar as possible to the input layer. This way, the hidden layer results in a non-linear compact representation of the input layer, achieved by the sigmoid activation function. The rationale behind this transformation is that, data will be more compact (i.e., less prone to over rating) and hopefully some interesting non-linear relationships that improve the explanation of the output variable would be discovered.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.0(
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.31
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.61
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.0(
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.0(
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.0(
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.0(
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.0(

Figure 4: sample dataset

The hidden layer of the autoencoder, the non-linear compact representation of the original input, is directly connected to a Multilayer perceptron i.e the BiLSTM directly connected to the convolutional neural network. This convolutional network is responsible for making predictions in the problem, by taking the new problem representation as an input.

4. Results

4.1 Heatmap Showing Feature Correlations

The Heatmap graph in figure 5 shows the density of the linear variation between each of the 14 features. The colour key at the right side of the graph shows the colour range value with cream having the highest positive correlation, the wine colour has the highest negative correlation. The correlation graph helps us to make some feature reduction thereby selecting 14 features.

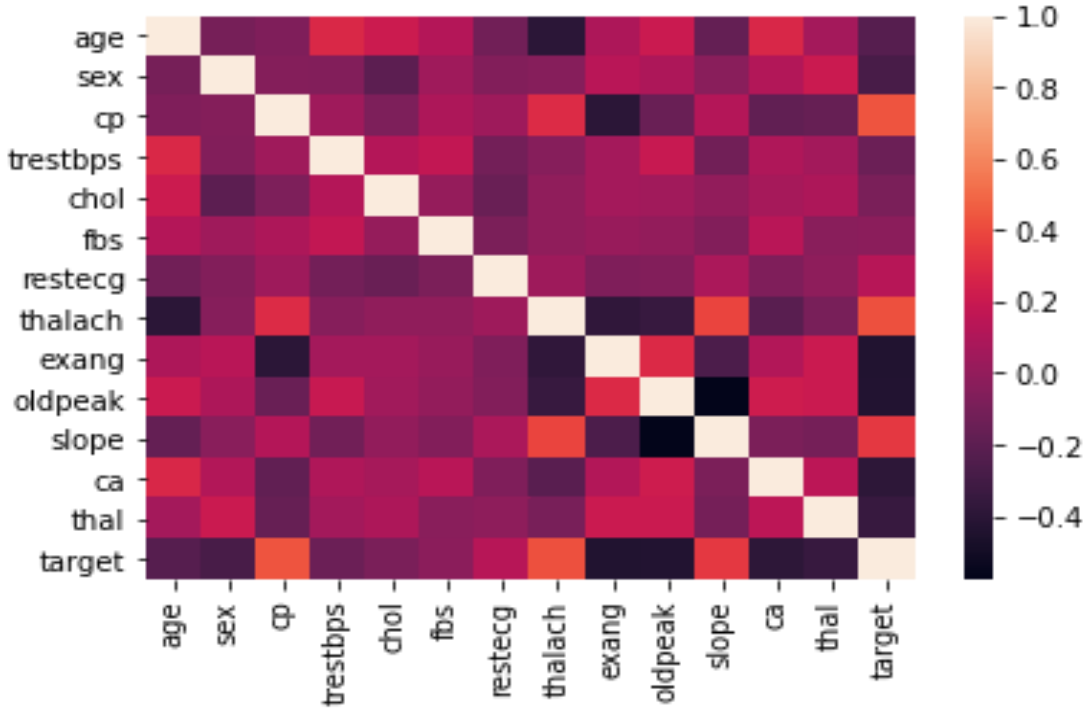

Figure 5: Heatmap Showing Feature Correlation

age:	age
sex:	1: male, 0: female
cp:	chest pain type, 0: typical angina, 1: atypical angina, 2: non-anginal pain, 3: asymptomatic
trestbps:	resting blood pressure
chol:	serum cholestoral in mg/dl
fbs:	fasting blood sugar > 120 mg/dl

restecg:	resting electrocardiographic result (values 0, 1, 2)
thalach:	maximum heart rate achieved
exang:	exercise induced angina
oldpeak:	oldpeak = ST depression induced by exercise relative to rest
slope:	the slope of the peak exercise ST seg-ment

ca:	number of major vessels (0-3) colored by flourosopy
thal:	thal: 0 = normal; 1 = fixed defect; 2 = reversable defect; 3 = defect
Checking correlations between Columns	
Target:	1.000000, exang 0.436757
Cp:	0.433798
Oldpeak:	0.430696
Thalach:	0.421741
ca:	0.391724
slope:	0.345877
thal:	0.344029
sex:	0.280937
age:	0.225439
trestbps:	0.144931
restecg:	0.137230
chol:	0.085239
fbs:	0.028046
Name: target	
dtype: float64	

This shows that most columns are moderately correlated with target, but 'fbs' is very weakly correlated.

4.2 Exploratory Data Analysis

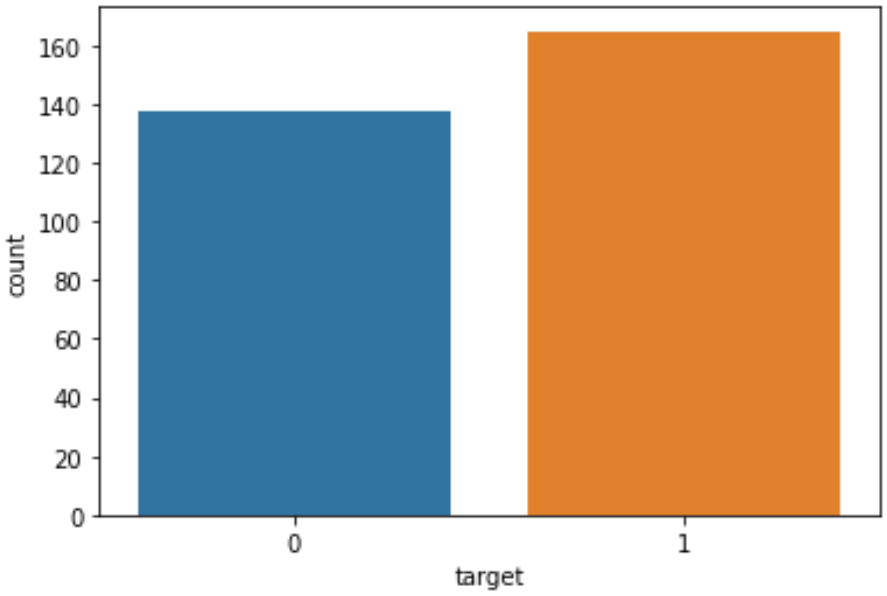

Figure 6: Target Variable

The Figure 6 represents the count on the y-axis and the Target on the x-axis. The count which is the total dataset used has a total of 303. The value 0 on the Target Axis represents Percentage of patience without heart problems which gives a value of 138 while the value 1 represents Percentage of patience with heart problems which gives a value of 165. Using the formula in this syntax: print("Percentage of patience without heart problems: " + str (round (target_temp[0] * 100/303,2)))

print("Percentage of patience with heart problems:" + str(round(target_temp[1] * 100/303,2)))

We have:
Percentage of patience without heart problems (0): 45.54
Percentage of patience with heart problems (1): 54.46

4.3 Analysing the 'Sex' feature

We notice that as expected, the ‘sex’ feature has two (2) unique features. Figure 7 represents the sex Variable of the dataset, value 0 on the sex axis represents the female Variable of the dataset while the value 1 represents the Male features of dataset. From the Graph we can see that female have high risk of heart problem as compared with the male. The female (0) has value 0.75 and the male (1) has value 0.4.

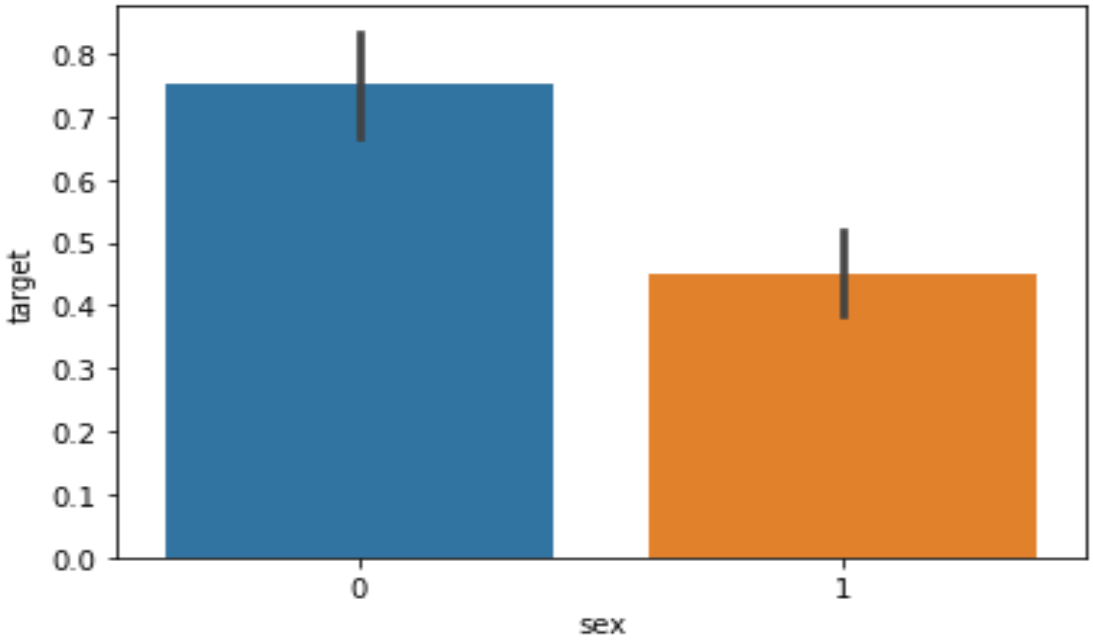

Figure 7: Sex Variable

4.4 Analysing the 'Chest Pain Type' feature

As expected, the CP feature has values from 0 to 3

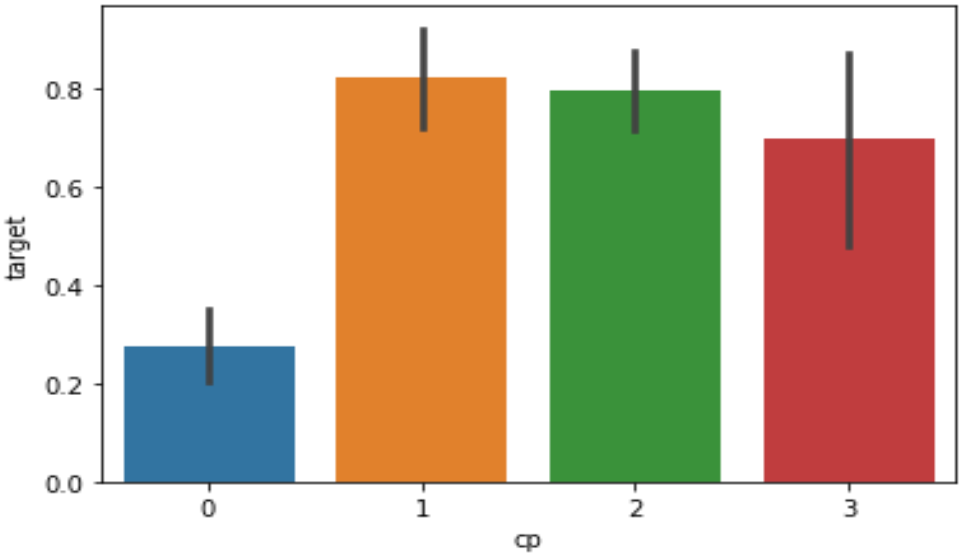

Figure 8: Chest Pain Type Variable

Figure 8 represents the Chest Pain Type Variable of the dataset, chest pain type, 0: typical angina, 1: atypical angina, 2: non-anginal pain, 3: asymptomatic. It was noticed, that chest pain of '0', i.e. the ones with typical angina are much less likely to have heart problems while chest pain of '1' the ones with atypical angina are more likely to have heart problems.

4.5 Analysing the FBS feature

Figure 9 has fasting blood sugar > 120 mg/dl Variable with the value If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false). The value of the True and false are close to each other so Nothing extraordinary here because of the closeness of its value.

Figure 10 has the Exercise induced angina :1 = yes , 0 = no. The value 0 has 0.7 while 1 has 0.25, this has given us a conclusion that people with exang=1 i.e. Exercise induced angina are much less likely to have heart problems while people with exang=0 are more likely to have heart problems

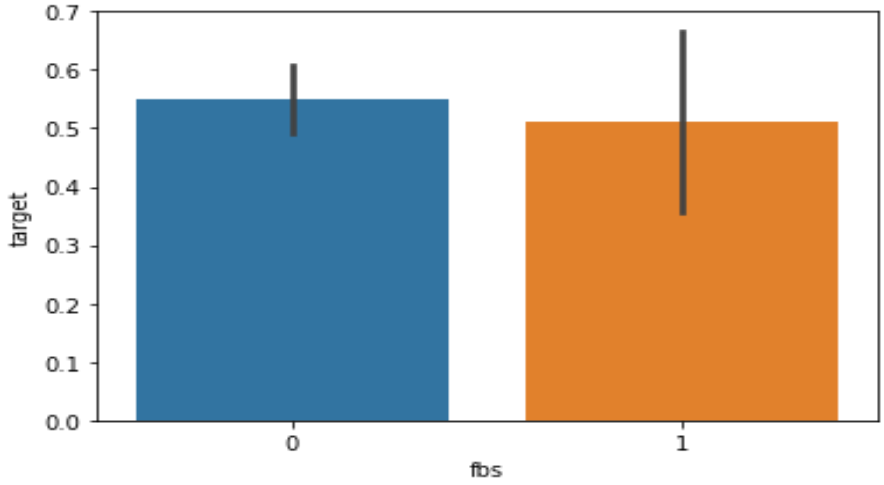

Figure 9: Analysing the fasting blood sugar > 120 mg/dl Variable

4.6 Analysing the Exercise Induced Angina Feature

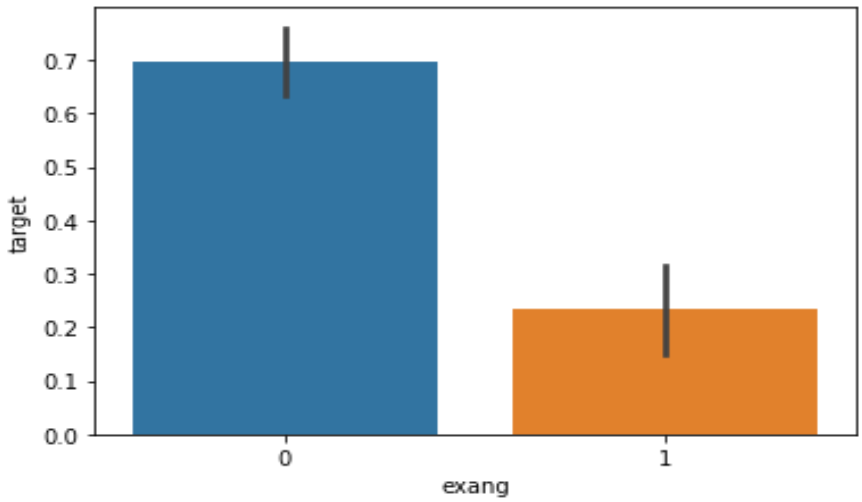

Figure 10: Analysing the Exercise induced angina feature Variable

4.7 Analysing the Peak exercise ST Segment Feature

Figure 11 has the variable Peak exercise ST segment: 0 = up sloping, 1 = flat, 2 = down sloping, with values '0' which has 0.42, '1' which has 0.35 and '2'which has 0.75. We observe, that Slope '2' causes heart pain much more than Slope '0' and '1 because the value 2 = down sloping has the highest value.

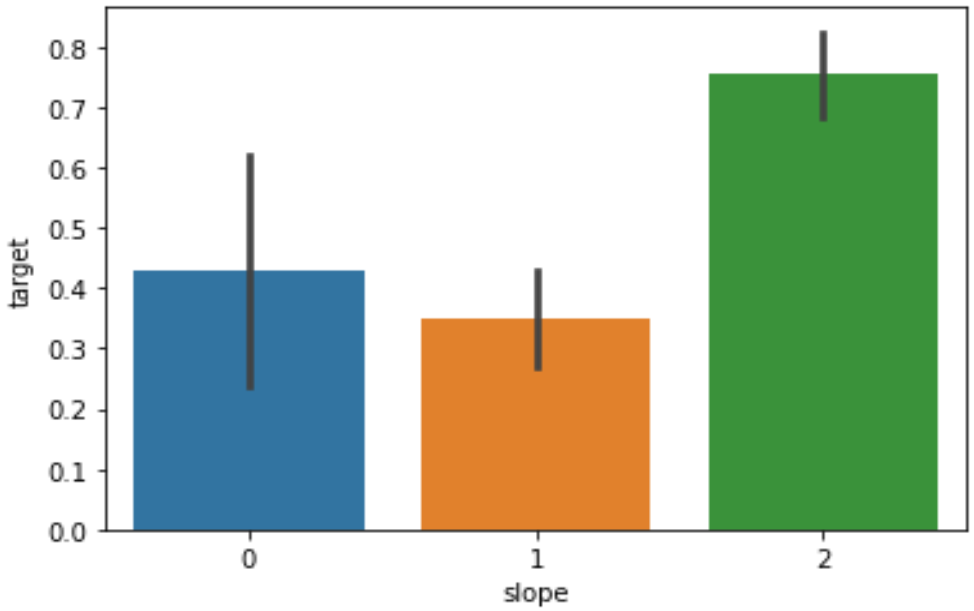

Figure 11: Analysing the Peak exercise ST segment feature Variable

4.8 Analysing the Diagnosis of Heart Disease Feature

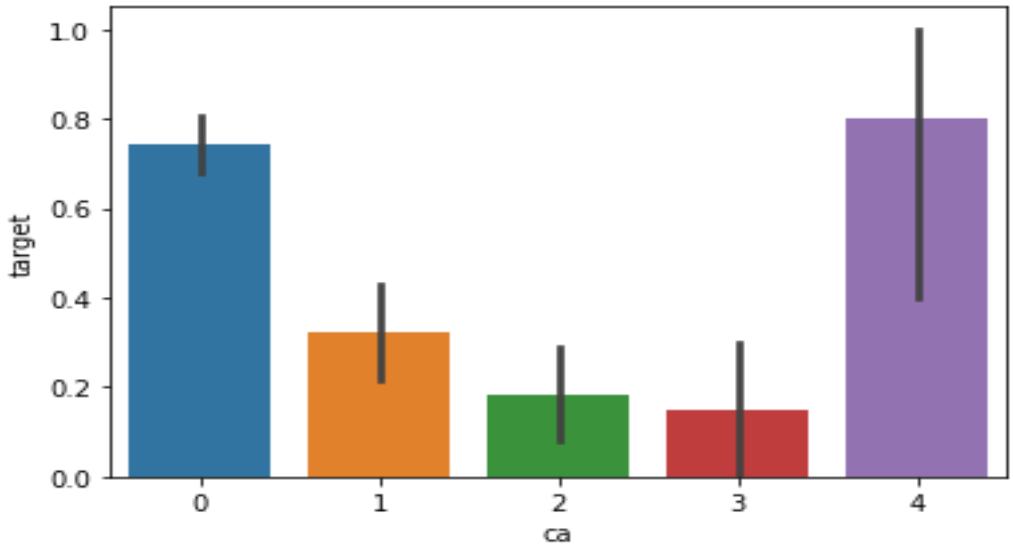

Figure 12: Analysing the Diagnosis of heart disease feature Variable

Figure 12 has the Diagnosis of heart disease: Displays whether the individual is suffering from heart disease or not : 0 = absence, 1, 2, 3, 4 = present. The value 4 has astonishingly large number of heart patients.

4.9 Analysing the thalassemia feature

The figure 13 displays the thalassemia: 0 = normal, 1= fixed defect, 2= reversible defect and 3= defect. The patient with reversible defect '2' has the Highest value

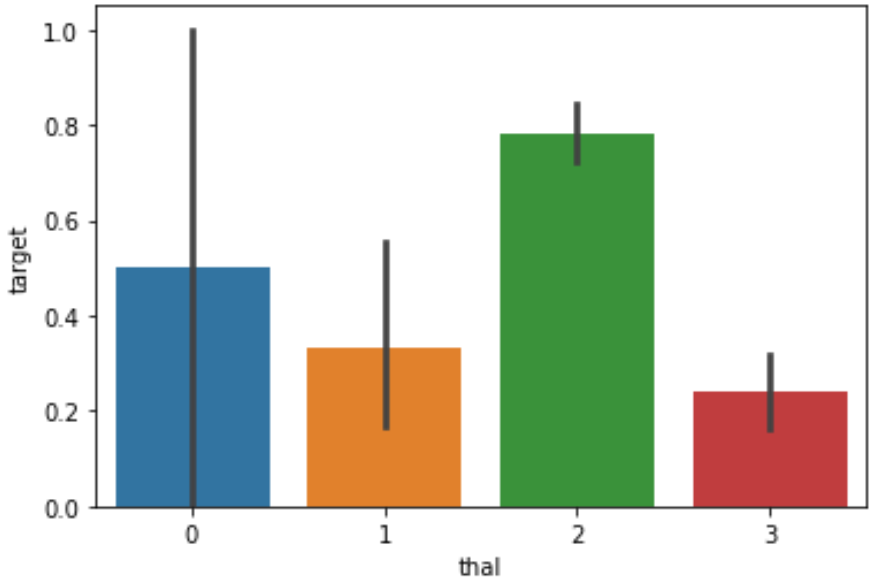

Figure 13: Analysing the thalassemia feature Variable

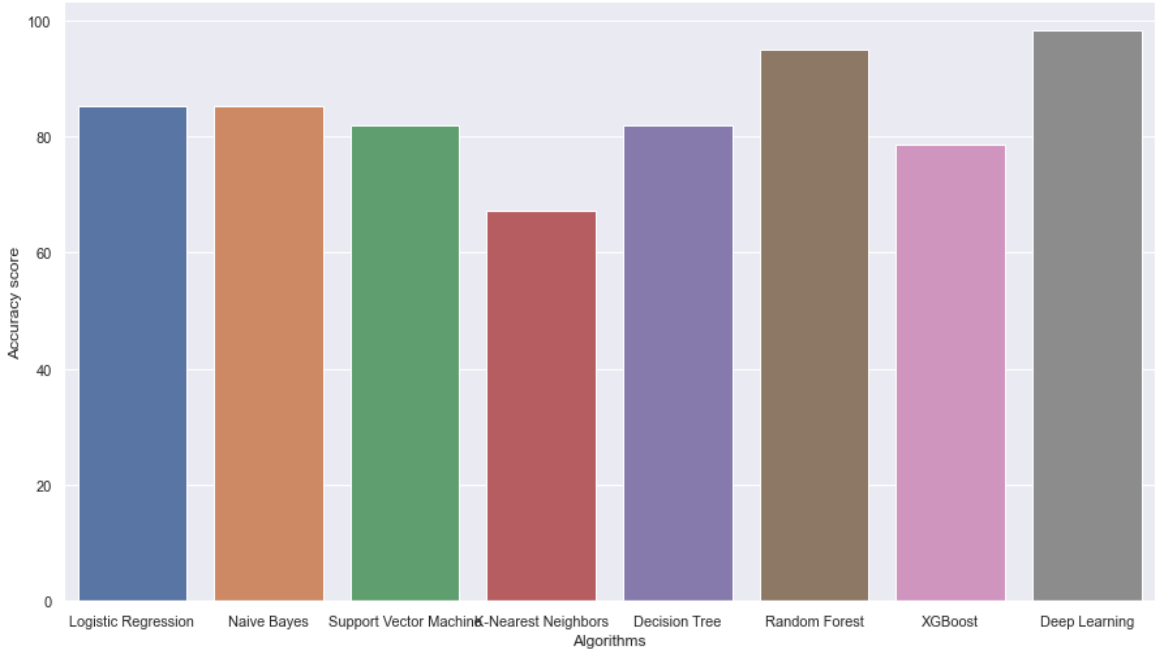

Figure 14: Output Final Score

The accuracy score achieved using Logistic Regression is: 85.25 %
The accuracy score achieved using Naive Bayes is: 85.25 %
The accuracy score achieved using Support Vector Machine is: 81.97 %
The accuracy score achieved using K-Nearest Neighbours is: 67.21 %
The accuracy score achieved using Decision Tree is: 81.97 %
The accuracy score achieved using Random Forest is: 95.08 %
The accuracy score achieved using XGBoost is: 78.69 %
The accuracy score achieved using Deep Learning is: 98.3 %

As it can be seen, deep learning has the best resu lt as compare to other algorithms.

5. Conclusion

This study made use of parameters like cardiac a rrest prediction based on the input attribute like blood pressure, cholesterol, blood sugar, chest pa in, blood sample parameter, ECG results. While in the study that was compared with for this wor k, the prediction is based on the HRV parameter and it can be seen that this study's result is more accurate than the existing method, which is the u niqueness of the study. The accuracy of this rese arch work is rated 98.3%. It could be said that th e use of an Autoencoder to actively learn the dat aset played an important role in the outcome of t he results.

References

[1]. Ahmed (2019). Flavonoids as Prospective Neuroprotectants and Their Therapeutic Propensity in Aging Associated Neurological Disorders. Frontiers in Aging Neuroscience. https://doi.org/10.3389/fnagi.201 9.00155

[2] Anurag, B., Sanjay, K., & Ashutosh, K. (2017). Sudden Cardiac Arrest Prediction Using Predictive Analytics. *International Journal of Intelligent Engineering and Systems*, 8.

[3] Boleslaw (2006). Using Efficient Supanova Kernel for Heart Disease Diagnosis. Proc. ANNIE06, Intelligent Engineering Systems Through Artificial Neural Networks, St. Louis, MO, vol. 16, ASME, New York, NY, 305-310.

[4] Frank Le Duff(2014). The Mini Mental State Examination at the Time of Alzheimer's Disease and Related Disorders Diagnosis, According to Age, Education, Gender and Place of Residence: A Cross-ssectional Study among the French National Alzheimer' Database. https://journals.plos.org/plosone/article/related?id=10.1371/journal.pone.0103630

[5] Kuyong Noh et. al. (2006). Acetoacetate Protects Neuronal Cells from Oxidative Glutamate Toxicity. Journal of Neuroscience Research. Vol. 83, 702-709. https://doi.org/10.1002/jnr.20736

[6] Lars, W., Young, K. W., Maureen, C., Berg.Katherine, Mortensen, S. J., Moskowitz, A., Donnino, M. W. (2015). The Prevalence and Significance of Abnormal Vital Signs Prior to In-Hospital Cardiac Arrest. US National Library of Medicine, 10.

[7] Nachket et. al. (2017). Improving Health Outcomes Through Patient Education and Partnerships with Patients, Baylor University Medical Center Proceedings, vol 30.1, 112-113, DOI: 10.1080/08998280.2017.11929552

[8] Shashikant, R., & Chetankumar, P. (2019). Predictive model of cardiac arrest in smokers using machine learning technique based on Heart Rate Variability parameter. Department of Instrumentation and Control Engineering, 12