ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse288952020.pdf
https://doi.org/10.30534/ijatcse/2020/288952020

Dual Classification approach for Prediction of Coronary Artery Disease using Gaussian Noise Modeling

Varun Sapra[1]
[1] School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India
varun.sapra@ gmail.com

ABSTRACT

There are many ways both invasive and non-invasive for the identification of coronary artery disease (CAD). Due to the cost, complexity and requirement of highly skilled medical practitioner with sophisticated laboratory, non-invasive methods like echocardiogram, stress-testing electrocardiogram gained popularity among people. Echocardiography is considered as one of the established technique used for diagnosing heart problems like hardening of heart muscles, blood clots and damaged cardiac tissue. It can also reveal some important estimates to physicians that help them to take correct decisions like ejection fraction, cardiac output and diastolic function. This work presents a dual classification approach where the first phase deals with the clinical data of the patients to diagnose the patients with cardiovascular problems and the second phase deals with the echocardiogram data to identify the probability of survival of the patient. In this dual approach we have used deep learning based models and the results reveals that proposed model is more promising.

Key words : About four key words or phrases in alphabetical order, separated by commas.

1. INTRODUCTION

Cardiovascular disease (CVD) can be categorized as the diseases that deal with the conditions that affect the heart or blood vessels. It includes cerebrovascular disease, coronary artery diseases (CAD), and congenital heart disease. One of the palpable reason for CVD is narrowing of the blood vessels due to the accumulation of plaque [1-2]. In one of the reports published by World Health Organization (WHO), around 17.9 million deaths are caused by CVDs in 2016 only and around 85% of all the deaths are due to heart attack and stroke [3]. The severity of the disease can be reduced by identifying the high-risk individuals at early stage and by ensuring that they receive correct and early treatment to avoid premature deaths. There are both invasive and non-invasive methods for the successful identification of CAD, but due to the complexity, cost and requirement of skilled clinical practitioner and laboratory setup, non-invasive methods gained a lot of attention [4-7]. One of the non-invasive method that has been adapted or used by physicians is echocardiography.

Echocardiogram is a non-invasive method that makes use of sound waves to make moving pictures of the heart to monitor how good heart's chamber and valves are working [8]. It also shows the vital statistics about the shape and size of the heart and as well as the heart muscles with poor blood flow. Figure 1 shows an electrographic image of heart.

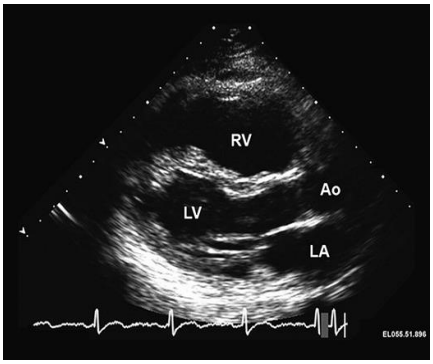

Figure 1: echocardiographic image of the right ventricle (RV) and the left ventricle (LV) [9]

There are different types of echocardiograms:

1. Transthoracic echocardiogram: It is one of the most common type of non-invasive echocardiogram which is just like an X-ray used to visualize a moving image of heart [10].
2. Transesophageal echocardiogram: In this type of echocardiography, the transducer is inserted in the throat into the esophagus as it is nearer to the heart and can give clearer picture [11].

3. Stress echocardiogram: This test is performed while exercising either on a treadmill or while cycling. It is used to observe the heart's motion when it is under stress. Basically, it is done to monitor the blood flow.
4. Dobutamine stress echocardiogram: It is another form of echocardiogram which is used when patient is not able to exercise. In this test, a patient is given a drug called dobutamine to feel like a patient's heart is working hard and doctors can monitor the vital statistics.

Another method that is widely accepted and used by clinical experts is analysis of clinical data. Many researchers have formulated various non-invasive methods/models to analyze the clinical parameters and produce a prediction for the positive CAD cases. Some of the major clinical parameters that are considered for the successful identification of CAD are hypertension, diabetes mellitus, smoking, age, cholesterol and chest pain [12-13]. Most of the models discussed by researchers used single classification technique or proposed their classification technique and compared with the existing ones.

In this paper, dual classification approach for diagnosing coronary artery disease using Deep learning has been proposed. The first classification model deals with the clinical data to diagnose the presence of CAD in the patients. Two separate Deep Neural Network (DNN) architectures are designed, where in the first model normal DNN layers have been used and in the second model Gaussian noise has been injected, as the dataset used is not a huge one, so there are chances of overfitting as well as represent complex mappings to learn. Adding noise to the datasets can produce a regularizing effect and can help in reducing overfitting. The second classification model deals with the echocardiogram data of the patients. This model helps in achieving the survivability of the patients based on echocardiogram data.

2. LITERATURE REVIEW

In the last decade deep learning has been implemented in various domains such as natural language processing, image classification, electronic health records and speech recognition. In healthcare domain artificial natural network based models yield better performance as compared to tradition modeling approaches and automate the feature extraction process.

Electronic Heath records can be organized into structured form into relational database. Various authors explored the power of artificial neural network in structured data [14]. Edward Choi et al. in their research work implemented recurrent neural network for the prediction of heart failure in Electronic heath records (EHRs). They further compared the model performance with tradition approaches of machine learning such as K-nearest neighbor, support vector machine, regularized logistic regression and multilayer perceptron (MLP). The performance is measures using area under the ROC curve. The value of AUC of the proposed model is 0.777 for one year window and for eighteen months window was 0.883. These measures are higher as compared to tradition approaches of Machine learning [15].

Hsiao et al. tried to study the association between the occurrence of cardiovascular disease by analyzing outpatient clinical records of subjects and environment monitoring parameters. softmax and autoencoder models are applied to perform the classification task to analyze the risk of four specific categories of the disease using deep learning approach [16].

Huang et al. proposed a deep learning based prediction model for coronary heart disease from huge volume of electronic health records. They employed stacked denoising auto-encoder deep learning architecture. The data consists of real clinical data of 3463 discharged patients. The deep learning based model outperformed traditional machine-learning models such as random forest, Logistic Regression and MLP [17].

J. Wang et al. proposed an automated detection of Breast arterial calcifications with twelve layer convolution neural network in 840 digital mammograms. The model is validated using three fold cross validation method. [18].

Lih Oh Shu, et al. in their study proposed a model for the successful identification of three categories of cardiovascular diseases. Their model was able to achieve the accuracy of 98.5%. Convolution neural network (CNN) was used to analyze electrocardiographic signal of the patients with long short term memory models. They validated the model using ten-fold cross validation with sixteen layers [19].

Liu et al. in their work proposed a model based on CNN along with bidirectional long short term memory models. The presented model was implemented with 5-fold cross validation method. The proposed system was able to achieve an accuracy of 99.90% with the intra-patient strategy [20].

Kwon, Joon myoung, et al suggested a model for the prediction of mortality of admitted heart patients based on echocardiography data of 30245 subjects. The proposed model consists of 362 nodes with 3 hidden layers. They also

compared the results with the other traditional machine learning approaches. The Area under curve of deep learning model for coronary hear disease and Heart Failure were 0.958 and 0.913 [21].

Gao, Xiaohong et al. in their study proposed a data driven learning framework with convolutional neural network on hand crafted features based on video images classification of echocardiography videos. The propsed framework with improved CNN architecture was able to achieve 92.1% precision [22].

M Kachuee et al. proposed the use of ECG data for the classification of heart beats. Authors have implemented deep CNN for categorization of the heart beat in different arrhythmia classes. For experiment purpose they used MIT-BIH and PTB Diagnostics dataset for training and validation purpose. Their proposed model was able to achieve 93.4% accuracy in detection of arrhythmias classes [23].

K. C. Lin et al. [24] discussed in their paper that how an accurate feature selection can improve the quality of the prediction model and can increase the accuracy. Feature selection is a method of extracting most significant features that contribute maximum to the output. There are many algorithms / techniques, which can be used for optimum, feature selection process. In their work, the authors have presented a hybrid algorithm for feature selection, which is an integration of artificial bee colony, and particle swarm optimization algorithms.

3. DATA AND METHODOLOGY

For experiment purpose, Z-Alizadesh Sani dataset has been used from one of the most popular repository of the University of California at Irvine. The dataset of 303 patients with 54 attributes such as gender, age, height, hypertension, smoking, renal failure, body mass index, dyslipidemia, cerebrovascular accident, blood sugar, hemoglobin, platelet, lymphocyte, triglyceride, low density lipoprotein, sodium, white blood cells, ejection fraction, blood urea nitrogen, potassium, rhythm, Poor R progression, Left ventricular hypertrophy, chest pain, dyspnea, lung rales, pulse rate, diastolic murmur is organized into four categories i.e. Demographic features, laboratory features, ECG features and symptom and examination features.

Preprocessing on the data is carried out to handle missing values and to reduce ambiguity in the data, which may arise due to transmission error, errors while doing manual data entry and faulty data collection tools. The data dimensionality was reduced as it improves the execution speed of a learning scheme; enhance data quality, performance and exploration of results. The feature subset is the collection of least number of attributes that are highly significant and contribute maximum in terms of accuracy and efficiency of algorithm.

Correlation based feature subset (CBFS) method is used to evaluate a subset that is highly correlated with the outcome class label [25-26]. Out of fifty-four features only sixteen features have been selected for the construction of model. The extracted features are given in Table 1.

Table 1: Attribute description of Z-Alizadesh Sani dataset

S.no	Attribute Description	Range Min-max	Mean	Standard Deviation
1	Age (in years)	30-86	58.89	10.392
2	Diabetes Mellitus 0 – No 1 – Yes	0 – 1	0.297	0.458
3	Hyper tension 0 – No 1 – Yes	0 - 1	0.591	0.493
4	Chronic Renal failure 0 – No 1 – Yes	0,1	-	-
5	Blood pressure (in mmHg)	90 - 190	129.554	18.938
6	Chest Pain (Typical) 0 – No 1 – Yes	0,1	-	-
7	Chest pain (Atypical) 0 – No 1 – Yes	0,1	-	-
8	Nonanginal Chest pain 0 – No 1 – Yes	0,1	-	-
9	Q-Wave	0-1	0.053	0.224
10	Tinversion	0-1	0.297	0.458
11	Fasting Blood Sugar (mg/dl)	62 - 400	119.185	52.08
12	Erythocyte Sedimentation Rate	1 - 90	19.462	15.936
13	Potassium	3 - 6.6	4.231	0.458
14	Ejection fraction	15 - 60	47.231	8.927
15	Region with RWMA	0 – 4	0.62	1.133
16	CAD 0 – No 1 – Yes	0, 1	-	-

For Phase II echocardiogram data is obtained from the repository of the University of California at Irvine. The dataset contains data of 132 subjects with 11 attributes. The data contained a lot of missing values, which are handled by replacing the missing values with the average of the class mean. Detailed description of the attributes are given in Table 2.

Table 2: Attribute description of echo datset

S.no	Attribute Description	Range Min-max	Mea	Standard Deviation
1	Survival (in months)	0.03 – 57	22.1	15.74
2	Still Alive (After survival period) (0 – dead 1 – Alive)	0 – 1	-	-
3	Age (in years) at the time of cardiac arrest	35 – 86	62.78	8.15
4	Pericardial-effusion 1 – Fluid 0 – No Fluid	0 – 1	-	-
5	Fractional Shortening	0.01–0.61	0.216	0.104
6	Epss (E-point septal separation,)	0 – 40	12.16	6.95
7	Wall motion score	2 – 39	14.4	4.94
8	Lvdd (left ventricular end-diastolic dimension.)	2.32 – 6.78	4.76	0.77
9	Wall motion index	1 – 3	1.37	0.45
10	Mult	0.14 – 2	0.776	0.19

3.1 Data Insights

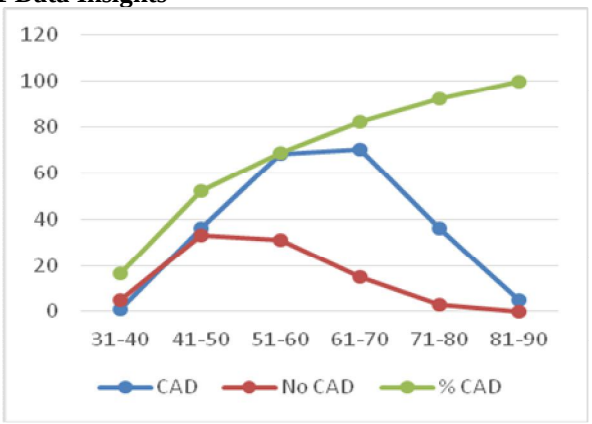

Figure 2: CAD, NO CAD and % of CAD patients in each age group

Figure 2. shows the steep rise in the percentage of CAD patients as the age increases. The figure demonstrates the percentage of sick and healthy patients in each age group.

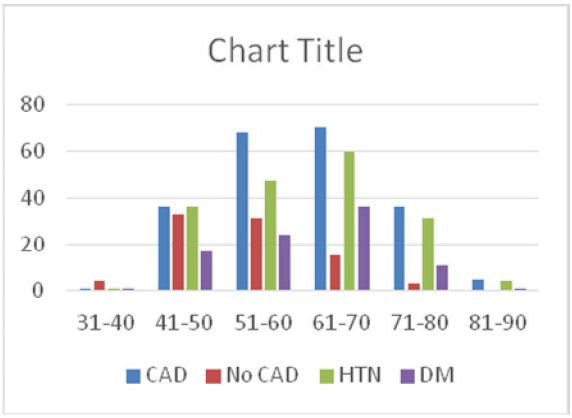

Figure 3: No. of CAD and NO CAD patients with hypertension and Diabetes

Figure 3 depicts the count of CAD and No CAD Patients with hypertension and diabetes in each age group. The figure clearly indicates that in the middle age group 51-60, 61-70 there is a sharp increase in number of CAD patients with the increase of hypertension and diabetes mellitus. So these two factors can be considered as significant contributors to the disease.

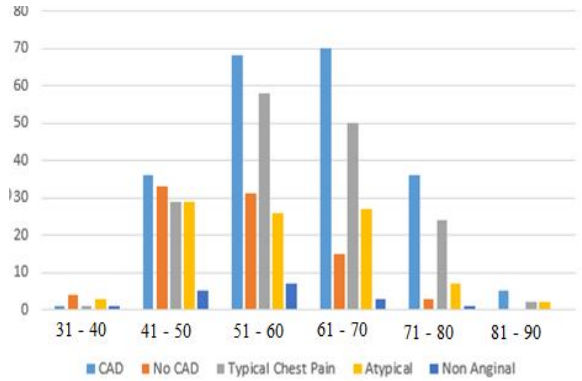

Figure 4: Ratio of CAD and No CAD vs type of chest pain

Type of chest pain is the major risk factor for the identification of disease. In the dataset there are three types of chest pain i.e typical chest pain, atypical chest pain and non-anginal chest pain. Figure 4 shows that chest pain typical and atypical can be symptom of disease.

4. PROPOSED WORK

In this paper a two phase model has been proposed for identification of CAD using deep neural network as shown in Figure 5.

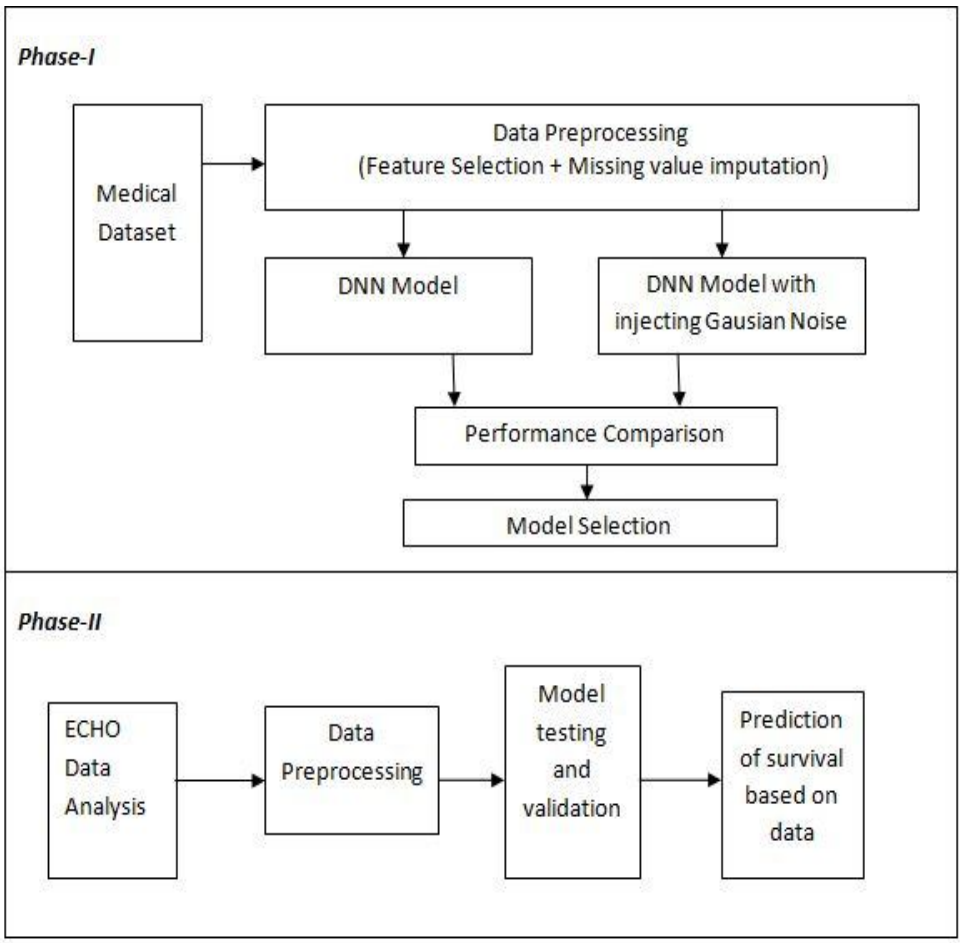

Figure 5: Two phase proposed model

4.1 Phase-I

Model –I: Regular deep neural network model for Identification of CAD**.**

In this phase, a sequential DNN has been implemented with an input layer, output layer and 3 hidden layers. The input layer is responsible to receive the input where the input neurons depends upon the no. of attributes responsible for generating the output. In the proposed model the input layer contains fifteen neurons. Hidden layers performs necessary computations to transform input into output, which is depicted by the output layer [27].

The dataset is divided into 67 :33 ratio for training and testing. Batch size of 14 and 400 epochs were used to develop the model. The sigmoid activation function is used for limiting the output into range between 0 and 1 [28]. The binary cross entropy is used as a loss function as it is the default loss function for binary classification problems. The optimizer used is adam optimizer. Figure 6 shows the architecture of regular DNN model.

Model-II : Deep neural network model with Gaussian noise for identification of CAD

The another model has been explored in order to identify the disease. To improve the performance of regular deep learning model Gaussian noise has been injected, as the dataset used is not a huge one, so there are chances of over fitting as well as represent complex mappings to learn. Adding noise to the datasets can produce a regularizing effect and can help in reducing over fitting. Relu activation function is used with 400 epochs. Neural network has been designed with seven hidden layers with one input and one output layer. Also binary cross entropy is used as a loss function with adam optimizer. The architecture of deep neural network with Gaussian noise is shown in Fig 7.

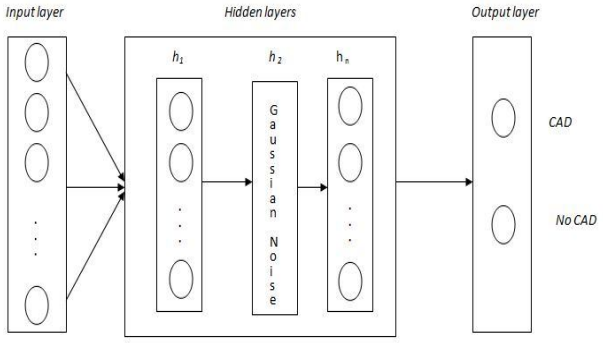

Figure 6: Regular deep neural network model for Identification of CAD.

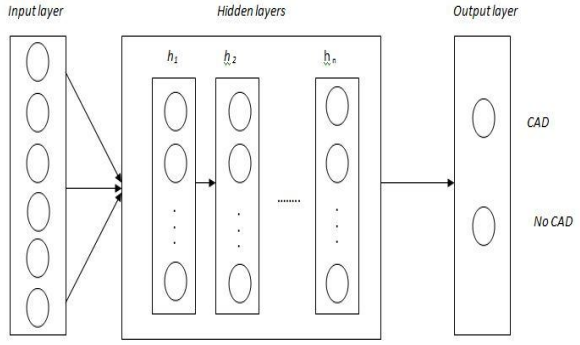

Figure 7: deep neural network model with Gaussian noise injection for Identification of CAD

4.2 Phase-II

In Phase II of the proposed model, the echocardiogram data of the subjects was taken for predicting the survivability of the subjects. The preprocessing was carried out to handle the missing values. The missing values were replaced by the mean of the class. In order to predict the survivability based on the echocardiography data, a model has been constructed using logistic regression and 10 fold cross validation has been implemented.

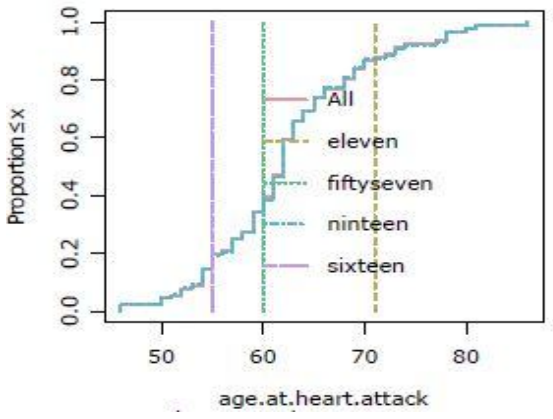

Figure 8: Proportion of survival and age at heart attack.

Figure 8 shows the age at the time of heart attack versus the survivability of the patients while considering the echocardiogram data.

5. RESULTS

The Table 3 shows the performance measures such as mean absolute error (MAE), mean square error (MSE), accuracy, recall, precision and F-score of the regular deep neural network. The network model achieves the accuracy of 92.08% and error rate of 7.92%. The average precision, recall and F-Score measures are 92%. The MSE, MAE are 7.03 and 15.67 respectively.

Table 3: Performance of model using Regular Deep Learning Network Model

Performance Measures	Regular Deep Neural Network				
	Precision	Recall	f – score	MSE	MAE
0	0.84	0.90	0.87	7.03	15.67
1	0.96	0.93	0.94		
Average	0.92	0.92	0.92		
Model accuracy	92.08.%				

Table 4. shows the performance measures of the improved model with Gaussian noise. The improved model achieved the classification accuracy of 94.1%, error rate of 5.9 %, average precision, recall and f-score are 94% each, mean square error of 4.5%. Figure 9 to 11 shows the accuracy, MAE and MSE of regular DNN model.

Table 4: Performance of model using Deep Neural Network with Gaussian Noise

Performance Measures	Deep Neural Network with Gaussian Noise			
	Precision	Recall	f – score	MSE
0	0.89	0.92	0.90	4.5
1	0.96	0.95	0.96	
Average	0.94	0.94	0.94	
Model accuracy	94.1%			

The improvement of prediction accuracy is noted i.e 2.02 % by applying Gaussian noise. Also the 2 % improvement in the value of precision, recall and F-Score. The mean squared error get reduced to 2.8%. Figure 12 shows the Training and testing accuracy of DNN model with Gaussian noise injection.

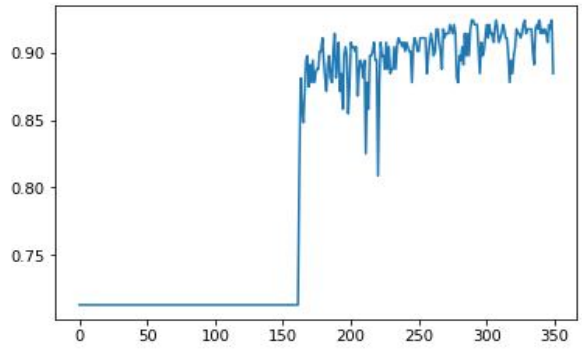

Figure 9: Accuracy with Regular deep neural network

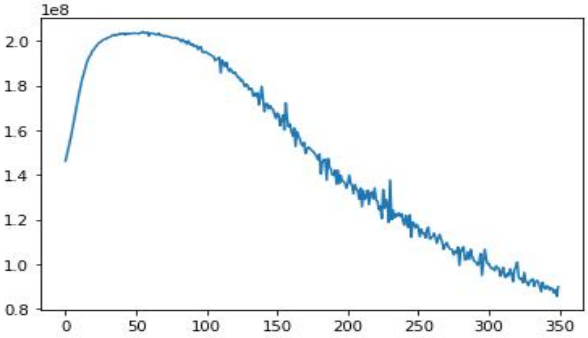

Figure 10: MAE of Regular deep learning model

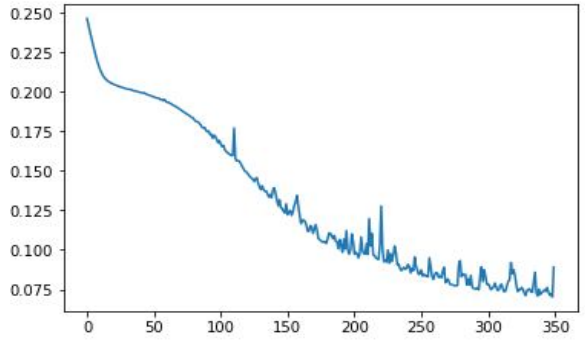

Figure 11: MSE of Regular deep learning model

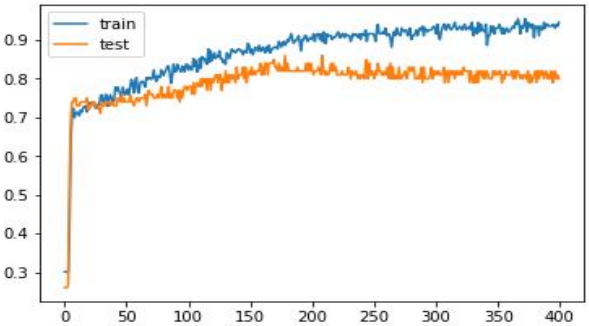

Figure 12: Accuracy (training and testing) of Deep learning model with Gaussian noise

Phase-II

In phase-II we have applied logistic regression in order to predict the survival of the patients based on echocardiogram data. Ten fold cross validation is used to train, test and validate the model. The model is able to achieve the accuracy of 93.1298 %. Table 5 shows the results of phase-II.

Table 5: Performance Measures of model by applying Logistic Regression

Error Rate	Kappa statistic	MAE	Root Relative Squared Error
6.87 %	-0.0208	0.0407	170.9%
RMSE	Relative absolute error	Accuracy	
0.184	118.7%	93.1%	

6. CONCLUSION

Coronary artery disease is the one of the foremost reason of death in all over the world. Invasive and non invasive methods are considered to diagnose the disease. In this work we applied machine learning methods in order to identify the disease and patients survivability using non invasive clinical parameters along with echocardiogram data of the subjects which are routine clinical parameters and can be easily obtained from the hospitals. Results are promising and reproducible. The proposed model can be considered as an adjunct tool in clinical practices.

REFERENCES

1. Nabel, Elizabeth G., and Eugene Braunwald. **A tale of coronary artery disease and myocardial infarction** *New England Journal of Medicine* Vol 366.1 pp. 54-63, Jan 2012
2. Abdar, Moloud, et al. **A new machine learning technique for an accurate diagnosis of coronary artery disease.** *Computer methods and programs in biomedicine* 179, Oct 2019.
3. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on 23-06-2020)
4. Alizadehsani, Roohallah, et al. **Machine learning-based coronary artery disease diagnosis: A comprehensive review.** *Computers in biology and medicine* pp.111 Aug 2019
5. Choi, Harry, et al. **Application of Non-invasive Imaging in Inflammatory Disease Conditions to Evaluate Subclinical Coronary Artery Disease**. *Current Rheumatology Reports* 22.1 Jan 2020.
6. Verma, Luxmi, Sangeet Srivastava, and P. C. Negi. **A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data.** *Journal of medical systems 40.7* July 2016.

7. Verma, Luxmi, Sangeet Srivastava, and P. C. Negi. **An intelligent noninvasive model for coronary artery disease detection** *Complex & Intelligent Systems* 4.1 pp.11-18, March 2018.
8. Rallidis, Loukianos S., Georgios Makavos, and Petros Nihoyannopoulos. **Right ventricular involvement in coronary artery disease: role of echocardiography for diagnosis and prognosis.** *Journal of the American Society of Echocardiography* 27.3 pp 223-229, Mar 2014.
9. Oh, Jae Kuen. **Echocardiography in heart failure: beyond diagnosis.** *European Journal of Echocardiography* 8.1 pp 4-14, Jan 2007.
10. Mitchell, Carol, et al. **Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography.** *Journal of the American Society of Echocardiography* 32.1 pp 1-64, Jan 2019.
11. Maxwell, Cory, Ryan Konoske, and Jonathan Mark. **Emerging Concepts in Transesophageal Echocardiography.** *F1000Research 5* 2016.
12. Nowbar, Alexandra N., et al. **Mortality from ischemic heart disease: Analysis of data from the World Health Organization and coronary artery disease risk factors From NCD Risk Factor Collaboration**. *Circulation: Cardiovascular Quality and Outcomes* 12.6 June 2019 e005375.
13. Malakar, Arup Kr, et al. **A review on coronary artery disease, its risk factors, and therapeutics** *Journal of cellular physiology pp* 16812-16823 234.10 Oct 2019.
14. Bizopoulos, Paschalis, and Dimitrios Koutsouris. **Deep learning in cardiology** *IEEE reviews in biomedical engineering* pp 168-193, Dec 2018.
15. Choi, E., Schuetz, A., Stewart, W. F., & Sun, J **Using recurrent neural network models for early detection of heart failure onset.** *Journal of the American Medical Informatics Association*, 24(2), pp 361-370, Mar 2017.
16. Hsiao, Han CW, Sean HF Chen, and Jeffrey JP Tsai. **Deep learning for risk analysis of specific cardiovascular diseases using environmental data and outpatient records** *IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)* pp. 369-372
17. Huang, Zhengxing, et al. **A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records**. *IEEE Transactions on Biomedical Engineering* 65.5 pp 956-968 July 2017.
18. J. Wang *et al*., **Detecting Cardiovascular Disease from Mammograms With Deep Learning** in *IEEE Transactions on Medical Imaging*, 2017 vol. 36, no. 5, pp. 1172-1181.
19. Lih, Oh Shu, et al. **Comprehensive electrocardiographic diagnosis based on deep learning** Artificial Intelligence in Medicine 103 Mar 2020.
20. W. Liu, F. Wang, Q. Huang, S. Chang, H. Wang, and J. He, **MFB-CBRNN: a hybrid network for MI detection using 12-lead ECGs** *IEEE J. Biomed. Heal. Informatics*, p. 1, April 2019.
21. Kwon, Joonmyoung, et al. **Deep learning for predicting in hospital mortality among heart disease patients based on echocardiography** *Echocardiography*, pp 213-218, Feb 2019.
22. Gao, Xiaohong, et al. **A fused deep learning architecture for viewpoint classification of echocardiography** *Information Fusion* Vol 36: pp 103-113, July 2017.
23. Kachuee, M., Fazeli, S., &Sarrafzadeh, M.. **Ecg heartbeat classification: A deep transferable representation**. EEE International Conference on Healthcare Informatics (ICHI) 2018 (pp. 443-444).
24. Lin, K. C., & Hsieh, Y. H.. **Classification of medical datasets using SVMs with hybridevolutionary algorithms based on endocrine-based particle swarm optimization and artificial bee colony algorithms**. *Journal of medical systems*, Vol 39(10), pp 119, June 2015.
25. Saqlain, Syed Muhammad, et al. **Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines**. *Knowledge and Information Systems* 58.1 pp 139-167, Jan 2019.
26. Singh, Surender, and Ashutosh Kumar Singh. **Correlation-based feature subset selection technique for web spam classification** *International Journal of Web Engineering and Technology* 13.4 pp 363-379, 2018.
27. Feng, Shuo, Huiyu Zhou, and Hongbiao Dong. **Using deep neural network with small dataset to predict material defects** *Materials & Design* 162 pp: 300-310, Jan 2019.
28. Mourgias-Alexandris, George, et al. **An all-optical neuron with sigmoid activation function** *Optics express* 27.7 pp: 9620-9630, Apr 2019.