International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)
ISSN: 2305-0543 (Online), ISSN: 2411-6173 (Print)
Vol. 12 (43), Jan. 2022, pp. 5106-5111

Performance Evaluation of Regression-Based Machine Learning Algorithms for Myocardial Infarction Prediction

[1]*R.C. Diovu
Department of Biomedical Engineering
Federal College of Dental Technology and Therapy
Trans-Ekulu Enugu, Nigeria
remy.diovu@fedcodtten.edu.ng

[2]B. U. Ugwuanyi
Department of Biomedical Engineering
Federal College of Dental Technology and Therapy
Trans-Ekulu Enugu, Nigeria

***Abstract*—Myocardial Infarction (MI) is a disease condition where the supply of blood to the heart or some parts of the heart is obstructed as a result of an occluded coronary artery. Myocardial Infarction is very common in Europe and the United States and is arguably one of the leading causes of death in those countries. Timely detection of MI can reduce the cost of treatment but early prediction can be very helpful in preventing the development of MI in the first instance. Supervised Machine learning even in the presence of uncertainties can be used to make predictions based on trained models from some known inputs and/or outputs. In this paper, the performance of different regression-based Machine Learning algorithms has been carried out using MATLAB. The performance analysis was done in terms of the ability of those algorithms to predict future response events based on changes from predictors present in a dataset. Root Mean Square Error (RMSE), R-Squared, Mean Squared Error (MSE), and Mean Absolute Error (MAE) are four important metrics utilized in the performance analysis. It was discovered from the analysis that ordinary linear regression trained model outperformed other regression based models with respect to the four metrics mentioned above.**

Keywords—Myocardial Infarction (MI), Supervised Machine Learning, Regression Learning, electrocardiograph signals

I. INTRODUCTION

Cardiac disorders leading to different kinds of cardiovascular diseases account for the majority of deaths all over the world according to World Health Organization (WHO). Myocardial Infarction is a common cardiac disorder caused by partial or complete occlusion which can cause inadequate flow of blood to the coronary arteries. This condition otherwise known as ischemia of the heart may lead to irreversible necrosis of the heart muscles in the case of prolonged ischemia. Sudden deaths can happen within one hour of manifestation of MI [1]. Therefore, early detection of MI can hinder its development to an acute stage Myocardial infarction is normally diagnosed from clinical findings by using laboratory results and/or the use of the electrocardiograph machine (ECG) which is an important diagnostic tool that can be used to record the electrical activity of the heart. Proper analysis of ECG signals can reveal MI. Ideally; deviations from the usual ECG signal shape can suggest the presence of MI as well reveal many other cardiac conditions. Unfortunately, ECG signals are normally captured or acquired as complex signals because of the presence of noise in these signals. The complex nature of such signals poses a lot of challenges to physicians who may not have detailed skills needed for accurate processing and analysis of such complex signals. Therefore, there are good chances that inherent MI signals may not be detected if the appropriate signals processing and analysis tools and methods are not employed. Even when the correct signal processing tools are utilized, unnecessary delays can be experienced and this may have severe consequences in the case of emergencies. Early prediction of MI and/or other related cardiovascular diseases can stop the disease condition from developing in the first place. This recent research trend which leverages on the potentials of data mining brings with it a lot of possibilities. The analysis of healthcare data can be useful in making predictions and other informed decisions which are very useful in healthcare delivery.

Machine learning is becoming a very common and effective approach for solving problems in the field of medicine [2-4] and especially for the detection of MI [5-6]. From a very broad perspective, ML can be seen as the ability of computers to extract information from the input of raw data by learning from experience, thus generating complex inferences based on the relationships within the dataset [7]. Therefore, ML builds predictive models without using predefined coding rules [8], and it is able to deal with large and complex datasets for which statistical analysis would be unfeasible. The contribution of this work is highlighted below-

The authors believe that healthcare data can be exploited for making important predictions of MI which will consequently lead to prevention of MI. This presents a remarkable shift in medical diagnosis which often times are more interested in gathering different symptoms associated with a particular disease condition and then making informed decisions from the symptoms. This research is promoting the idea of watching out for risk factors associated with different disease conditions rather than relying only on the conventional way of gathering symptoms and results from laboratory tests. These risk factors care often dependent on lifestyle, environment, psychological and genetic makeup of individuals. We show that with the help of machine learning algorithms, individuals with increased risk factors for MI for example can be identified and used subsequently for making

Article History:
Received Date: Sep. 14, 2021, Accepted Date: Dec. 03, 2021, Available Online: Jan. 01, 2022

important predictions for a future event happening. A typical example of such future event is a death event. Because most of the risk factors (smoking, cholesterol level, etc.,) present in a typical healthcare data are modifiable, a prediction of death event for an individual can compel the person to adjust his lifestyle with a view to removing entirely or reducing the risk factors that contributed to such prediction. Such lifestyle adjustments can surely lead to the prevention of such disease condition that may have been responsible for such death event happening.

Related works had focused on the detection of MI by using different artificial neural network algorithms to classify time series electrocardiograph signals [9-12]. One striking difference between our research approach and the ones presented in [9-12] is that they utilized one form of ECG data or the other. While this approach may be commendable, there are some drawbacks. Firstly, setting up the laboratory for the recording of ECG signals is cost intensive (as it involves the use of costly machines) and time consuming. Secondly, the recorded raw ECG signals have to be pre-analyzed using computer-aided tools. Finally, the pre-analysis of the ECG signals require appreciable skills and knowledge of signal analysis. The above mentioned drawbacks can easily be avoided by following the research approach adopted in this paper. It is to be noted that preliminary results from this research has be presented in [13]. Feedback gotten from the presented paper had been utilized in improving the overall quality of this paper. The statistical metrics utilized in [13] were only three whereas they have been increased to four in this paper.

The rest of the paper is organized as follows: section two describes the research methodology, section three deals with results and the analysis. Section four describes the conclusion and recommendation of the study.

II. MATERIALS AND METHODS

Study Sample: The study sample comprises a dataset of 299 patients from a publicly available repository (UCI repository).

The data was imported to MATLAB and only 13 clinical features from this dataset were included in this study. Some of the clinical features selected for this study have been summarized in table1.

Method: Different regression based models were trained using the MATLAB app while the trained models were used for making predictions. We applied cross-validation in order to prevent over-fitting in the learner's app. This study then carried out the performance evaluation of different regression based Machine Learning algorithms. After training a model in MATLAB Regression Learner's app, the score values (Root Mean Square Error, Mean Squared Error, R-Squared, and Mean Absolute Error) for of the different algorithms are recorded. The score values were then utilized in estimating the performance of the trained model.

III. RESULTS

In the section, the results generated from this study are presented together with important analysis. The first category of results that are presented in this study is the response plot. There are many ways of presenting the response plot but for this study, the box plot was preferred due to the uniqueness of the clinical features investigated in the study.

Table 1: Important Clinical Features

Clinical Feature	Explanation
Age	Age of the patient (years)
Anaemia	Shows the decrease of red blood cells or hemoglobin (Boolean)
High blood pressure	Shows if the patient has hypertension (Boolean)
Creatinine phosphokinase (CPK)	Shows the level of CPK enzyme in the blood (mcg/L)
Diabetes	Shows whether the patient has diabetes (Boolean)
Ejection fraction	Represents the percentage of blood leaving the heart at each contraction
Serum creatinine	Signifies the level of serum creatinine in the blood (mg/dL)
Smoking	Shows that the patient smokes or not (Boolean)
Sex	Shows if a patient is woman or man (binary)
Serum sodium	Shows the level of serum sodium in the blood (mEq/L)

A box plot as is the case in this situation displays the typical values of the response and any possible outliers. The central mark indicates the median, and the bottom and top edges of the box are the 25th and 75th percentiles, respectively. Vertical lines, called whiskers, extend from the boxes to the most extreme data points that are not considered outliers. The outliers are plotted individually using the '+' symbol. The box plots presented here are for two important clinical features- smoking and ejection fraction. These two features were selected for obvious reasons. The two features are modifiable risk factors for the development of myocardial infarction and can be controlled by adjusting one's lifestyle. Figs. 1 and 2 show the box plot (response plot) of the trained models of the different regression-based algorithms (model 1.1- Linear, 1.2- interaction linear, 1.3- robust linear and 1.4- stepwise linear) when ejection fraction and smoking were the clinical features selected as predictors respectively. The blue and orange colours on the plot represent the true response and the predicted response respectively.

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

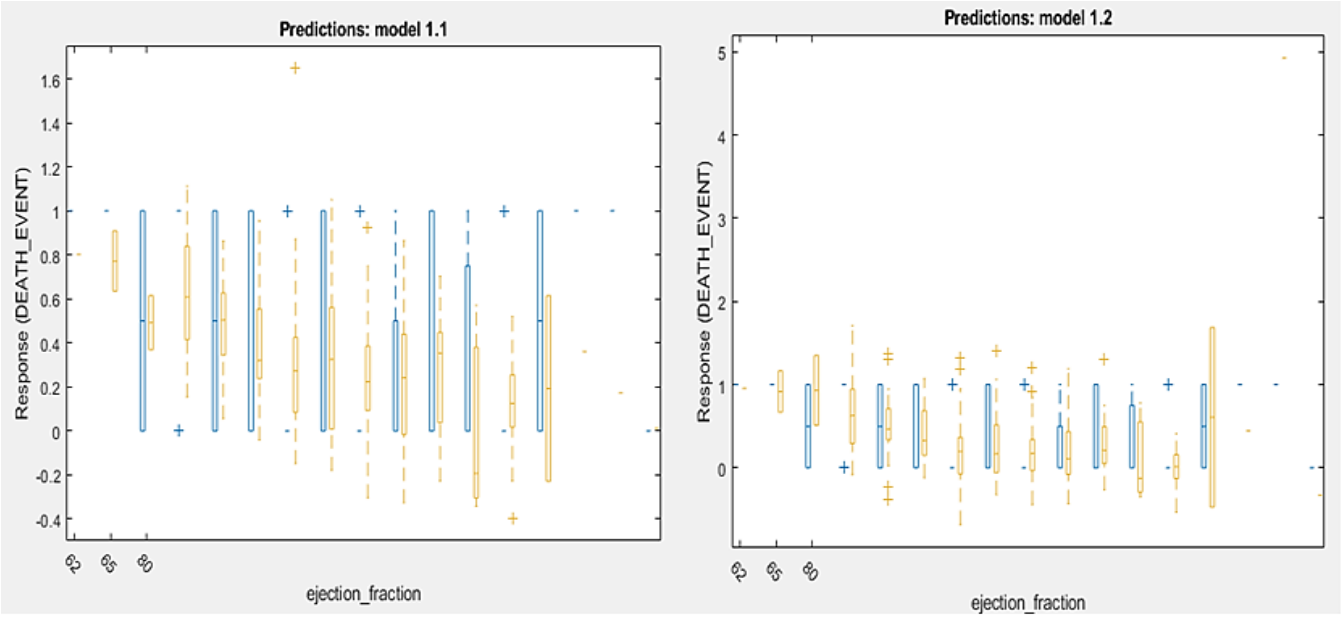

Fig.1: Response plot for ejection fraction as a predictor for Linear and Interaction Linear Algorithms

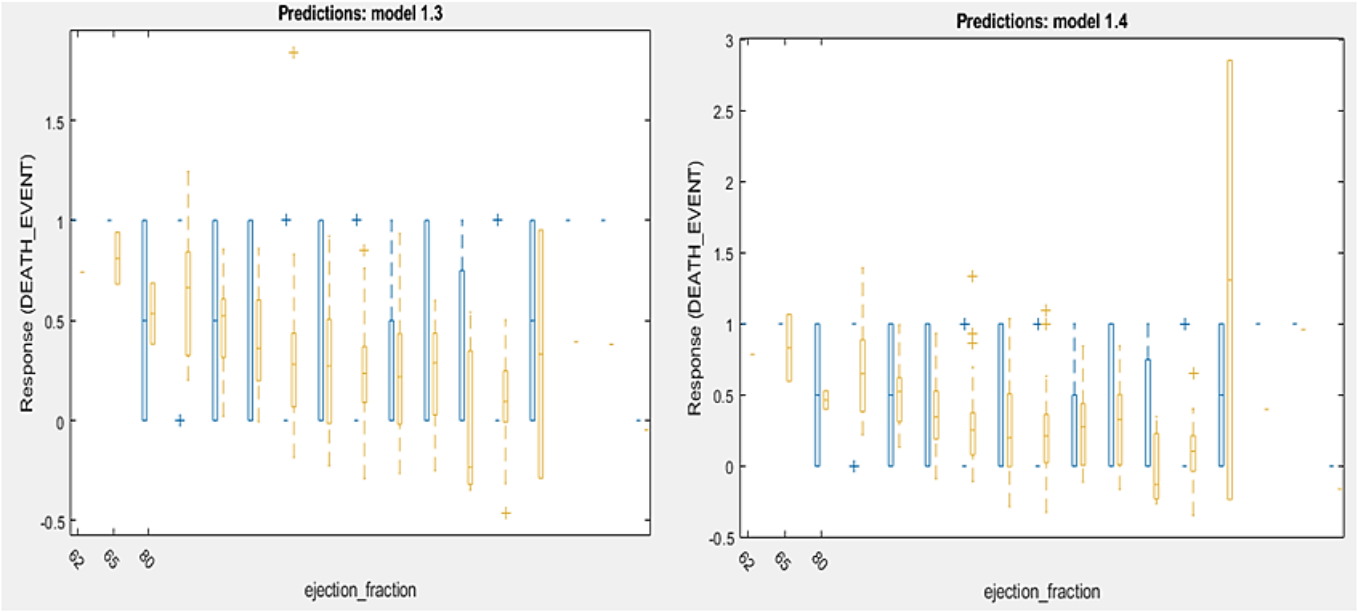

Fig.2: Response plot for ejection fraction as a predictor for Robust Linear and Stepwise Linear Algorithms

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

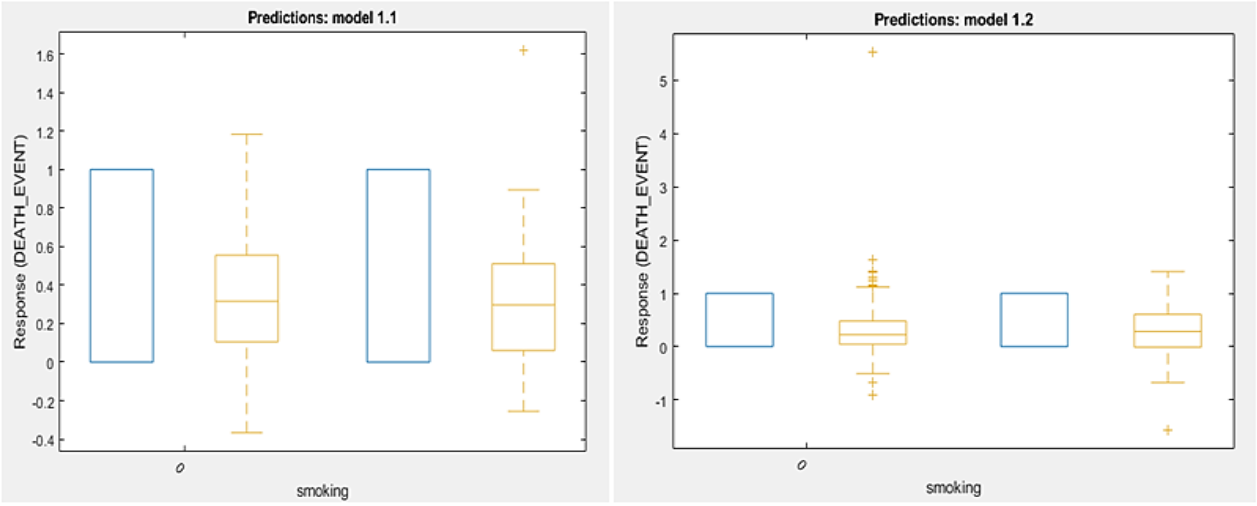

Fig.3: Response plot for smoking as a predictor for Linear and Interaction Linear Algorithms

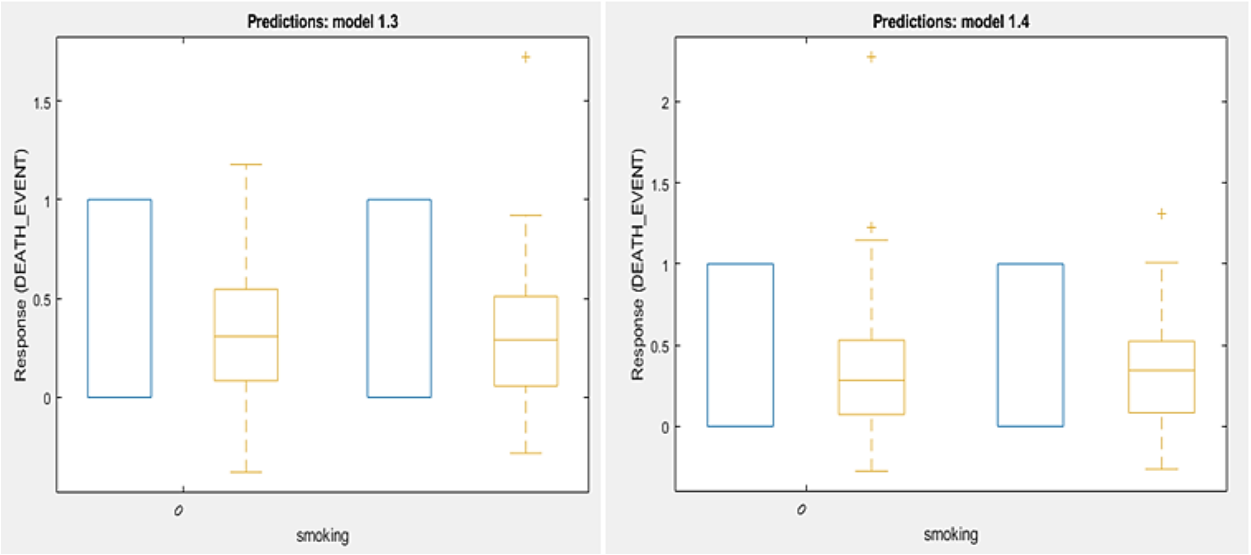

Fig.4: Response plot for smoking as a predictor for Robust Linear and Stepwise Linear Algorithms

In terms of the performance evaluation of the different regression-based machine learning algorithms utilized in Error were utilized. The RMSE score is always positive and its units match the units of the predicted response. On the other hand, R-squared score is always smaller than 1 and usually larger than 0. It compares the trained model with the model where the response is constant and equals the mean of the training response. If the trained model is worse than this constant model, then R-Squared is negative. The MSE is the square of the RMSE. The MAE is always positive and similar to the RMSE, but less sensitive to outliers. The scores for the four statistical metrics have been presented in table 2. From the box plots shown in Figs. 1- 4, and also this study, the statistical score values of Root Mean Square Error, Mean Squared Error, R-Squared, and Mean Absolute from the results presented in table 2, it was discovered that ordinary linear regression trained model outperformed other regression based models with respect to the other three metrics mentioned above. Considering the entire results presented in this paper, the performance of the four studied algorithms in ascending order has been presented in Fig. 5.

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

Table 2: Comparative scores for RMSE, R-SQURED, MSE and MAE for different regression Learner Algorithm

Metric	Linear Regression	Interaction Linear	Robust Linear	Stepwise Linear
RMSE	0.37679	0.51527	0.37854	0.40784
R-SQURED	0.36	-0.2	0.35	0.25
MSE	0.14197	0.2655	0.14329	0.16633
MAE	0.31079	0.38582	0.31093	0.31728

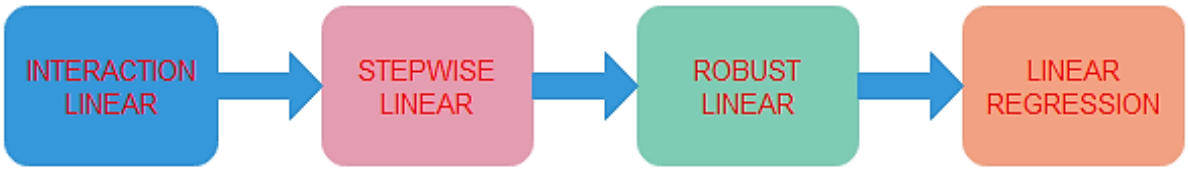

Fig. 3: Performance of the regression-based learner's algorithms in ascending order

IV. CONCLUSION AND RECOMMENDATIONS

This study has investigated a simple but quick approach for the prediction of myocardial infarction. The ability to predict myocardial infarction using machine learning will be a breakthrough for cardiologists. The advantages of this approach cannot be over-emphasized. Among other things, this approach saves a lot of time usually needed for the proper diagnosis of this disease condition which usually involves a lot of laboratory tests that are very expensive and which also involves the gathering of lot of historical data that may be time consuming. The authors believe that the former approach cannot be relied on anymore as the first resort as it cannot produce good results especially with emergency cases. The approach proposed in this study should be encouraged for a number of reasons. Firstly, a lot of data are currently being generated in the healthcare industry and with the help of emerging technologies like data mining; these data can be leveraged on for assisting physicians in making accurate decisions that can affect the quality of care delivered to their patients. This research approach also advocates that people should imbibe a healthy lifestyle since some lifestyles can present themselves as risky factors for the development of certain deadly disease conditions. This study therefore recommends that certain lifestyles like smoking and consumption of saturated fats which can lead to a rise in the cholesterol in the body should be avoided. Since MI results from partial or complete obstruction of blood to the heart or some parts of the heart, regular physical exercises should be encouraged as this to a large extent can improve the flow of blood to the different parts of the body. Future work will explore cloud based computation using FPGA controllers and baseline analytics [14], [15, [16].

REFERENCES

[1] A. Alghamdi, M. Hammad, H. Ugail, A. Abdel-Raheem, K. Muhammad, H.S Khalifa and A.A Abd El-Latif, "Datection of Myocardial Infarction based on Novel Deep Transfer Learning Methods for Urban Healthcare in Smart Cities." Multimedia Tools and Applications, March 2020.

[2] Pławiak P. "Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals." Swarm Evolution Computers vol. 39 pp. 192–208, 2018.

[3] Pławiak P. "Novel methodology of cardiac health recognition based on ECG signals and evolutionary neural System." Expert Syst Appl vol. 92, pp. 334–349, 2018

[4] Tuncer T, Dogan S, Pławiak P, Acharya UR. "Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals." Knowledge.-Based System, 2019

[5] Sadhukhan D, Pal S, Mitra M " Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data" IEEE Transaction Instrumentation Measurement vol. 99, pp. 1–11, 2018

[6] Liu B, Liu J, Wang G, Huang K, Li F, Zheng Y, Luo Y, Zhou F. "A novel electrocardiogra parameterization algorithm and its application in myocardial infarction detection." Comput Biol Med 61: 178–184, 2015

[7] Miotto, R.;Wang, F.;Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. Brief Bioinform. **2018**, 19, 1236–1246.

[8] McBee, M.P.; Awan, O.A.; Colucci, A.T.; Ghobadi, C.W.; Kadom, N.; Kansagra, A.P.; Tridandapani, S.; Auffermann,W.F. Deep Learning in Radiology. Acad. Radiol. **2018**, 25, 1472–1480

[9] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in Neural Networks (IJCNN), 2017 International Joint Conference on. IEEE, 2017, pp. 1578–1585. [Online]. Available: http://arxiv.org/abs/1611.06455

[10] Z. Cui, W. Chen, and Y. Chen, "Multi-scale convolutional neural networks for time series classification," CoRR, vol. abs/1603.06995, 2016. [Online]. Available: http://arxiv.org/abs/1603.06995

[11] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks," JMLR workshop and conference proceedings, vol. 56, pp. 301–318, 2016.

[12] Javad Kojuri**1,** Reza Boostani, Pooyan Dehghani, Farzad Nowroozipour, Nasrin Saki**, "**Prediction of acute myocardial infarction with artificial neural networks in patients with non-diagnostic electrocardiogram," Journal of Cardiovascular Disease Research Vol. 6, Issue 2, Apr-Jun 2015.

[13] Diovu, R.C and Ezema C.N, "Prediction of Myocardial Infarction using Supervised Machine Learning," in 1st International Conference on Health Technology and Biomedical Engineering (ICHTBE), Federal College of Dental Technology and Therapy, Trans-Ekulu Enugu, Nigeria, 18th – 20th August 2021.

[14] K. C. Okafor, M. C. Ndinechi, Sanjay Misra, "Cyber-Physical Network Architecture for Smart City Data Stream Provisioning in Complex Ecosystems", *In Transactions on Emerging Telecommunications Technologies*, United Kingdom. Online ISSN:2161-3915. Vol.32(11), Pp.1-31, 2021

[15] K. C. Okafor, G.C. Ononiwu, Sam G. V.C Chijindu, C. C. Udeze "Towards Complex Dynamic Fog Network Orchestration Using Embedded Neural Switch", *In International Journal of Computers and Applications, (IJCA)*, UK, Print ISSN: 1206-212X Online ISSN: 1925-7074. 2021, Vol.43 (2), Pp.91-108. *Taylor & Francis.*

[16] K. C Okafor, G. N. Ezeh, I.E. Achumba, F.N. Ugwoke, C.C Okezie and U.H.Diala, "Harnessing FPGA Processor Cores In Evolving Cloud Based Datacenter Network Designs", *In Proc.12th International of Conference of Nigeria Computer Society (NCS)-IT for Inclusive Development, Akure, Ondo State, July 22nd-24th, 2015. Pp.1-14.*