As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology
Volume 10, Issue 03, pages 307-312.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol10issue03.html
DOI: https://doi.org/10.46501/IJMTST1003053

Machine Learning for Real Time Heart Disease Prediction

Dr. D Suneetha | B Bhavya | D Aashika | G Srinidhi | B Vamsi Krishna

Department of Computer Science and Engineering, Department of CSE, NRI Institute of Technology, AP-521212 India.

To Cite this Article
Dr. D Suneetha, B Bhavya, D Aashika, G Srinidhi and B Vamsi Krishna, Machine Learning for Real Time Heart Disease Prediction, International Journal for Modern Trends in Science and Technology, 2024, 10(03), pages. 307-312.https://doi.org/10.46501/IJMTST1003053

Article Info
Received: 08 February 2024; Accepted: 28 February 2024; Published: 07 March 2024.

ABSTRACT

Heart disease is a leading cause of mortality in the modern world, responsible for around one death per minute. This ratio takes into account persons in the age range of 25 to 69 and accounts for both males and females; however it may vary by area. This in no way rules out the possibility of heart disease affecting persons of various ages. Predicting the etiology and progression of this illness is a big difficulty in the modern world, and it may manifest itself at any age. In this paper, we've looked at a number of different methods for predicting cardiac issues.

Keywords: Decision tree, Data mining, Heart Disease Prediction, Naïve Bayes, K-means, Machine learning

1. INTRODUCTION

This paper primarily focuses on data mining approaches that are helpful in predicting heart disease using a variety of available data mining technologies. A malfunctioning heart may cause problems for the human body as a whole, including the brain, kidneys, and other organs. When the heart is afflicted by illness, it is called heart disease. Heart disease has surpassed all other causes of mortality in modern times. The World Health Organization (WHO) estimates that cardiovascular disease kills 12 million people annually. Diseases of the heart include cardiovascular disease, coronary disease, heart attack, and knock. Knock is a kind of cardiac disease that develops when blood arteries that supply the brain are either expanded, constricted, or weakened; high blood pressure may also play a role in its onset. In today's healthcare system, achieving technological and infrastructure dominance is the primary obstacle. The quality of care delivered to patients relies on their doctors' ability to accurately diagnose their conditions and treat them successfully.

The devastating consequences of a wrong diagnosis are often disregarded. Medical history records and data are extensive, yet they come from a wide variety of sources. Doctors' interpretations are a crucial part of this information. Due to the potential for noise, incompleteness, and inconsistency in real-world data, data preprocessing will be necessary in directive to fill in the missing values. Despite the fact that cardiovascular disease has been identified as a leading cause of mortality worldwide, it has also been declared to be one of the most preventable and treatable conditions. Proper and timely diagnosis of an illness is crucial to its complete and correct care. There seems to be a critical need for a reliable and systematic technique of identifying individuals at high risk and data mining for rapid investigation of heart infection. It's important to remember that the signs and symptoms of heart disease might vary from person to person. Pain in the back, jaw, neck, stomach, shortness of breath, chest, arms, and shoulders are common complaints. Heart disorders come in many forms, with heart failure, stroke, and coronary artery disease being just a few examples [3]. It's true that heart disease is the most common kind of chronic illness worldwide, but it's also the most preventable. Main prevention of heart disease may be achieved via adopting a healthy lifestyle, and secondary prevention can be achieved through prompt diagnosis. The importance of regular checkups (inferior preventive) in diagnosing and preventing heart disease issues early on cannot be overstated. Several diagnostic procedures, including angiograms, chest X-rays, echocardiograms, and an activity tolerance test, corroborate the seriousness of the problem. However, these examinations may be quite costly and need precise medical equipment. Doctors specializing in the heart maintain a large and thorough database of patient information. It also provides a fantastic opportunity for extracting useful information from large databases.

The variables that put people at risk for heart disease are the subject of intensive study, with researchers using a wide range of statistical methods and data mining software. Smoking, advanced age, hypertension, diabetes, high cholesterol levels, high blood pressure, a family history of heart disease, being overweight, and not getting enough exercise are all recognized as risk factors for cardiovascular disease, according to statistical analyses. Awareness of cardiac illnesses is crucial for the prevention and treatment of those at risk of developing the condition. In order to aid experts and doctors in the diagnosis of heart disease, researchers have turned to a variety of data mining approaches. Decision trees, k-nearest neighbors, and Naive Bayes are frequent methods of choice. Other methods for classifying data include the bagging algorithm, kernel density, sequential minimum optimization, neural networks, a straight Kernel self-organizing map, and support vector machines (Support Vector Machine). The methods that were implemented in this research are described in depth in the next section. Coronary heart disease (CHD), cerebrovascular disease (Stroke), congenital heart disease (CHD), provocative heart diseases (PHD), hypertensive heart diseases (HHD), and peripheral artery disease (PAD) are all examples of cardiovascular disease. Tobacco use, poor nutrition, lack of exercise, and excessive alcohol use are the main contributors to cardiovascular disease. It's no secret that the research of cardiac illnesses makes use of several categories of current mathematical data mining technologies.

2. LITERATURE SURVEY

Mohammed Abdul Khaleel has[2] presented a paper in the Survey of Data Mining Techniques on Medical Data for Detecting Common Diseases in the Area. This study focuses on medical information mining techniques necessary for identifying locally visit diseases like heart ailments, lung cancer, breast disease, and so forth. To study and diagnose cardiac illness, Vembandasamy et al. conducted information mining to extract information for discovering inactive instances. This application made use of the Naive Bayes method. In the Naive Bayes algorithm, Bayes's theorem was employed. This means that Naive Bayes has a lot of leeway to form inferences on its own. The dataset utilized comes from a top research center for diabetes in Chennai, Tamil Nadu. The collection includes information from more than 500 individuals. Weka is utilized for categorization, and 70% Percentage Split is employed to carry it out. Naive Bayes guarantees an 86.419% success rate. Predicting Remote Health Monitoring Outcomes with Data from the First Month and Baseline Interventions, a work written by Costas Sideris, Nabil Alshurafa, Haik Kalantarian, and Mohammad Pourhomayoun ([3]).

Saving money and lowering disease rates are two of the main benefits of RHS systems. In this research, the authors provide an improved RHM framework called Wanda- CVD, which is mobile phone based and designed to provide distant teaching and social support to participants. Social security organizations all throughout the globe see CVD preventative intervention measures as a fundamental priority. A presentation titled "Prediction for Similarities of Disease Using ID3 Algorithm in Television and Mobile Phone" was presented by L.Sathish Kumar and A. Padmapriya [4].

This research presents a disguised and automated method of handling cardiac disease cover-up recognize designs. Information extraction techniques, such as the ID3 algorithm, are included into the provided framework. The suggested strategy not only increases public awareness of illnesses but also has the potential to cut down on the number of fatalities and afflicted individuals. [5].

In their study "Disease Predicting System Using Data Mining Techniques," M.A.Nishara Banu and B.Gomathy present their methodology for doing just that. They discuss the MAFIA (Maximum Frequent Item set algorithm) and K-Means clustering methods in this work. Because illness forecasting relies heavily on proper categorization. When used together, MAFIA and K-Means provide an accurate categorization. In their article titled "Intelligence System for Diagnosis Level of Coronary Heart Disease using K-Star Algorithm," [6] Wiharto and Hari Kusnanto present their system. In this study, they provide a prediction framework for heart infection based on a computation made using a Neural Network trained with Learning Vector Quantization. This framework's neural system is able to recognize 13 clinical features as information, making predictions about the presence or absence of coronary sickness in the patient and other performance metrics.

[7]. Using the Learning Vector Quantization Algorithm, D.R. Patil and Jayshril S. Sonawane have presented a paper titled Prediction of Heart Disease. In this study, they provide a prediction framework for heart infections based on the calculation techniques of the Learning vector Quantization neural system. The neural system in this framework is able to recognize 13 clinical features as information and make predictions about the presence or absence of cardiac sickness in the patient, in addition to other performance indicators.

3. METHODOLOGY

3.1 Data Preprocessing:

Data that we wish to analyze is not likely to be clean, in the sense that it may include noise or have values that are missing, therefore in order to achieve decent results, we need to eradicate all of these problems, a process known as data cleaning. Using methods like inserting the most frequent value possible in the blank to eliminate noise, we will complete the data.

Data normalization, smoothing, generalization, and aggregation are all examples of transformations that may be applied to raw data to make it more digestible.

Integration: Data that we don't need to process may not come from a single source; instead, it may come from multiple sources, each of which could cause complications later on in the processing pipeline. As a result, integration is a crucial step in the data pre-processing pipeline, and a variety of factors are taken into account in this section.

As we deal with data, it is possible that it may be complicated and difficult to grasp; in these cases, we will do reduction to simplify the data and make it more amenable to the system, allowing us to get better outcomes.

3.2 Proposed System:

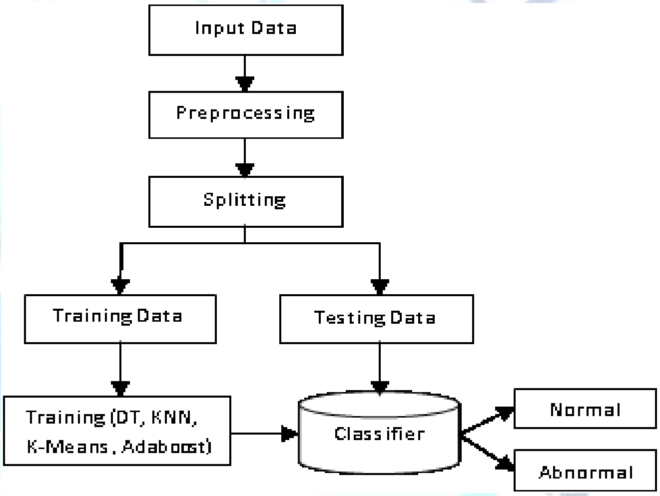
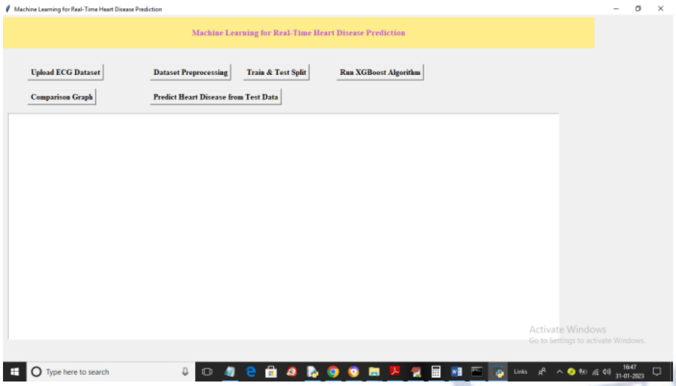


Fig. 1: System we suggest for the problem

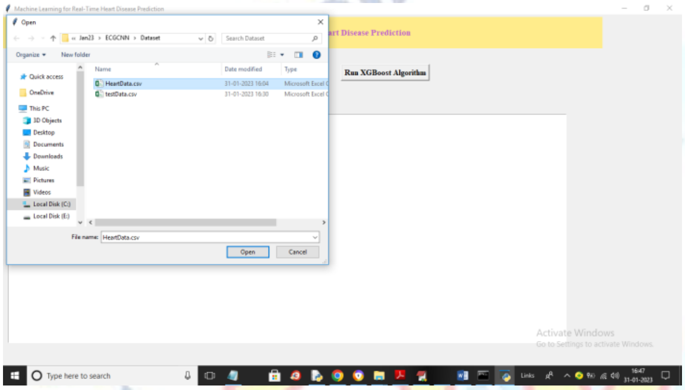
Based on the results of the experiment described above, we can conclude that the accuracy of the predictions made using the Naive Bayes algorithm and the decision tree algorithm may vary. As a result, it is not necessary to make a comparison of the two algorithms before making any predictions in order to obtain accurate results. In the same vein, if we use only a single algorithm that is unable to pre-process data, we will not be able to obtain accurate results.

4. RESULTS:

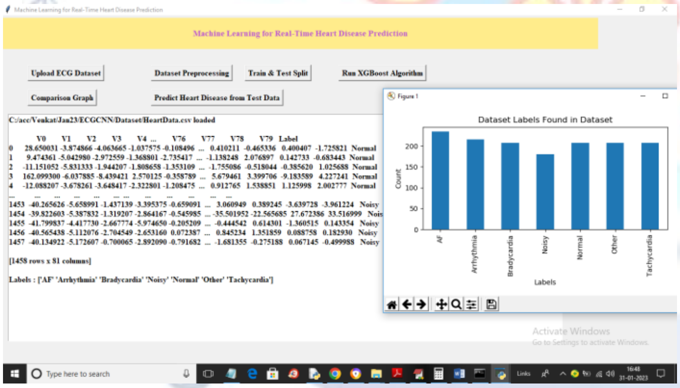
To run project double click on run.bat file to get below screen



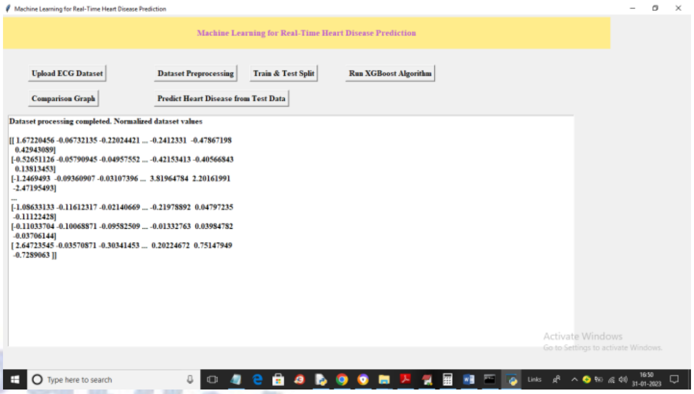
In above screen click on 'Upload ECG Dataset' button to upload dataset and get below output



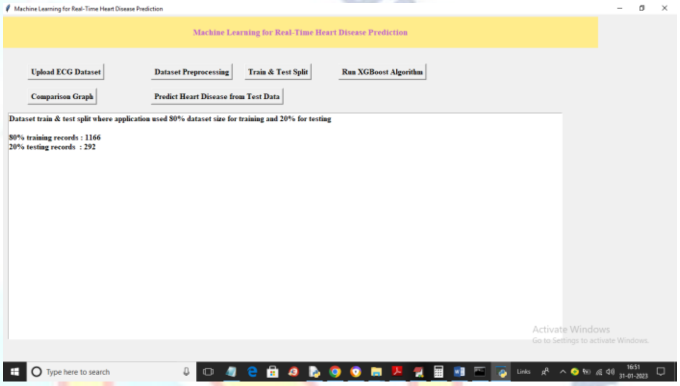
In above screen selecting and uploading 'HeartData.csv' file and then click on 'Open' button to load dataset and get below output



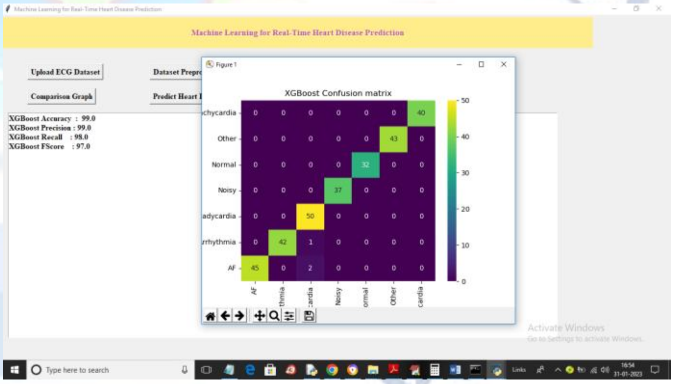
In above screen dataset loaded and we can see dataset contains numeric and non-numeric data but machine learning algorithm will take only numeric data so we need to preprocess data to convert to numeric and in above graph x-axis represents heart disease name and y-axis represents count of that disease records found in dataset. Now close above graph and then click on 'Dataset Preprocessing' button to process dataset and get below output



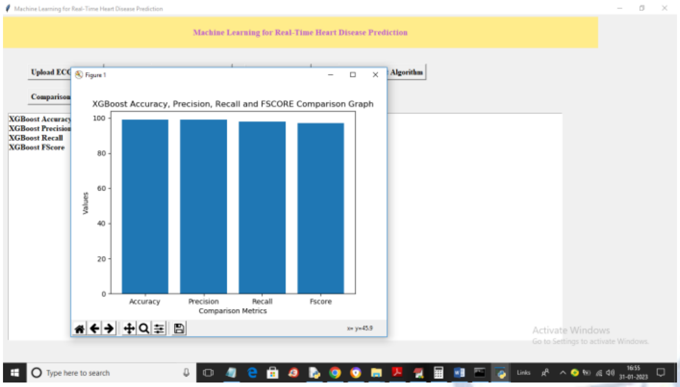
In above screen all values converted to numeric format and now click on 'Train & Test Split' button to split dataset into train and test and get below output



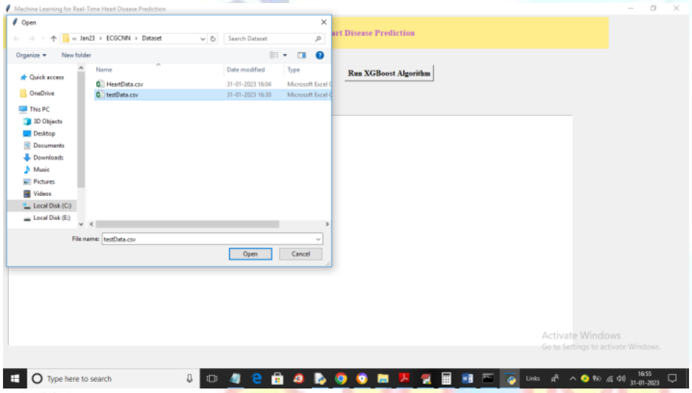
In above screen dataset using 1166 records for training and 292 for testing and now click on 'Run XGBoost Algorithm' button to train XGBOOST and get below output



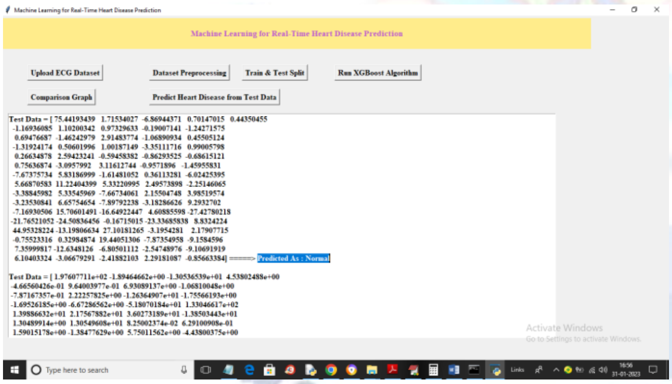
In above screen with XGBOOST we got 99% accuracy and in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and all different colour boxes represents correct prediction count and blue boxes contains incorrect prediction count which are very few. Now close above graph and then click on 'Comparison Graph' button to get below graph



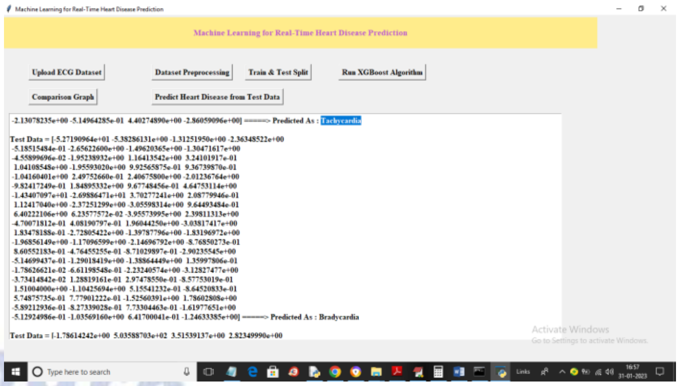
In above graph x-axis represents XGBOOST metrics and y-axis represents values which are closer to 100%. Now click on 'Predict Heart Disease from Test Data' button to upload test data and get below prediction output

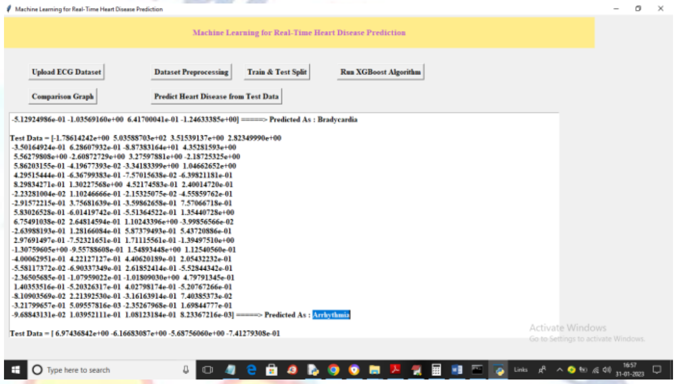


In above screen selecting and uploading 'testData.csv' file and then click on 'Open' button to load dataset and get below output



In above screen in square bracket we can see ECH signal test data and after arrow symbol =➔ we can see predicted output and scroll down above output screen to view other prediction output





5. CONCLUSIONS:

In this study, we present a novel convolutional neural network-based multimodal disease risk prediction (CNN-MDRP) method that uses structured and unstructured data from hospitals. The publication is titled "Using Structured and Unstructured Data from Hospitals." In the field of medical big data analytics, none of the previously published work seems to have focused on both kinds of data to the best of our knowledge. The suggested method has a prediction accuracy of 94.8% and a convergence speed that is quicker than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm. This is in comparison to the accuracy of predictions made by various conventional prediction algorithms.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] Sonam Nikhar, A.M. Karandikar "Prediction of Heart Disease Using Machine Learning Algorithms" in International Journal of Advanced Engineering, Management and Science (IJAEMS) June-2016 vol-2

[2] Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal

[3] Nabil Alshurafa, Costas Sideris, Mohammad Pourhomayoun, Haik Kalantarian, Majid Sarrafzadeh "Remote Health Monitoring Out- come Success Prediction using Baseline and First Month Interven- tion Data" in IEEE Journal of Biomedical and Health Informatics

[4] Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial" in JMIR Cardio 2017, vol. 1, issue 2, pg no:1

[5] DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem "The Utilisation of Machine Learning Approaches for Med- ical Data Classification" in Annual Conference on New Trends in Information & Communications Technology Applications - march- 2017

[6] Applying k-Nearest Neighbour in Diagnosing Heart Disease Pa- tients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3,

[7] June 2012

[8] Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dha- vachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.

[9] Alapatt, B.P., Kavitha, A., Amudhavel, J., "A novel encryption al- gorithm for end to end secured fiber optic communication", (2017) International Journal of Pure and Applied Mathematics, 117 (19 Special Issue), pp. 269-275.

[10] Amudhavel, J., Inbavalli, P., Bhuvaneswari, B., Anandaraj, B., Vengattaraman, T., Premkumar, K., "An effective analysis on har- mony search optimization approaches", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2035-2038.

[11] Amudhavel, J., Kathavate, P., Reddy, L.S.S., Bhuvaneswari Aadharshini, A., "Assessment on authentication mechanisms in dis- tributed system: A case study", (2017) Journal of Advanced Re- search in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1437-1448.

[12] Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattatraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) International Journal of Applied Engineering Re- search, 10 (3), pp. 2058-2061.

[13] Amudhavel, J., Kathavate, P., Reddy, L.S.S., Satyanarayana, K.V.V., "Effects, challenges, opportunities and analysis on security based cloud resource virtualization", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1458-1463.

[14] Amudhavel, J., Ilamathi, R., Moganarangan, N., Ravishankar, V., Baskaran, R., Premkumar, K., "Performance analysis in cloud au- diting: An analysis of the state-of-the-art", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2043-2046.