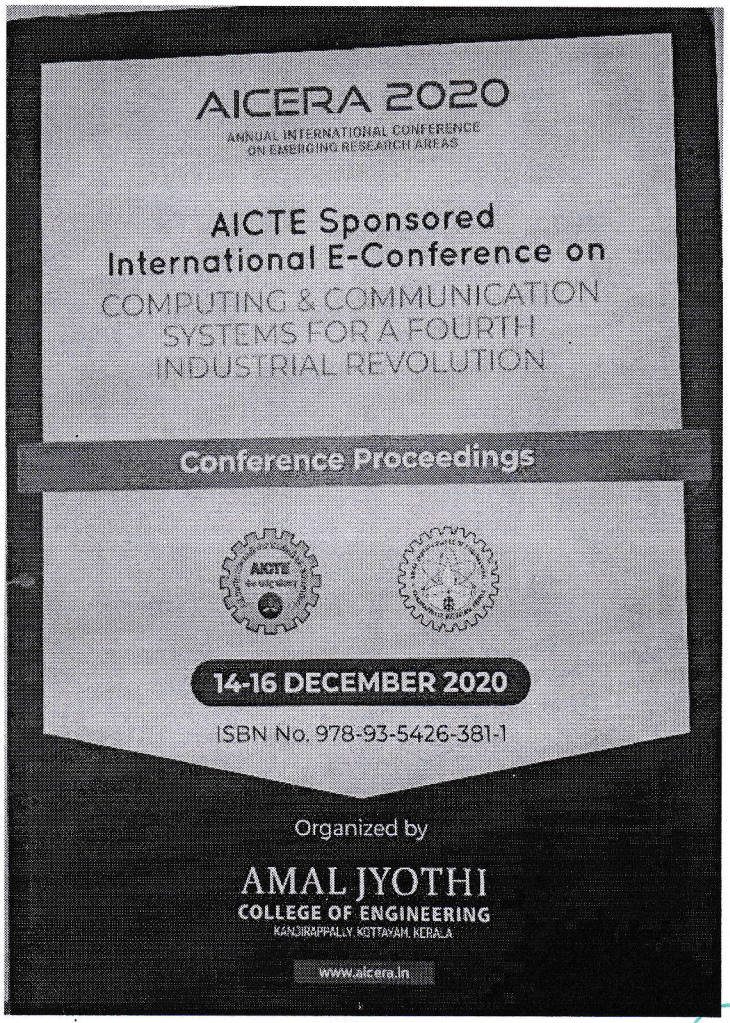
Criterion 3: Research, Innovations and Extension

3.2.1 Number of books and chapters in edited volumes/books published and papers published in national/international conference proceedings per teacher

AMBILY MERLIN KURUVILLA

Certificate of Presentation

Proceedings

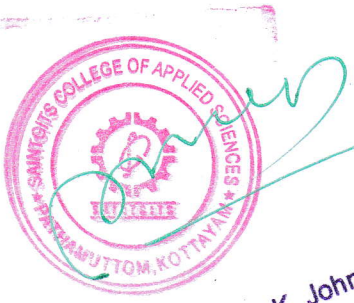


SAINTGITS COLLEGE OF APPLIED SCIENCES
PATHAMUTTOM, Kottayam

Dr. K. K. John
Principal
Saintgits College of Applied Sciences
Kottukulam Hills, Pathamuttom P. O.
Kottayam 686 532, Kerala

Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach.

To cite this article: Ambily Merlin Kuruvilla and N.V Balaji 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1085** 012028



Dr. K. K. John
Principal
Saintgits College of Applied Sciences
Kottukulam Hills, Pathamuttom P. O.
Kottayam 686 532, Kerala

Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach.

Ambily Merlin Kuruvilla[1] and Dr.N.VBalaji[2]

[1] Research Scholar, Department of CS,CA and IT, Karpagam Academy of Higher Education
[2]Dean, Faculty of Arts, Science and Humanitics, Karpagam Academy of Higher Education

E-mail: E-mail:ambilykuruvilla@gmail.com

Abstract. Cardiac disease prediction helps physicians to make accurate recommendations on the treatment of the patients. The use of machine learning (ML) is one of the solution for recognising heart disease-related symptoms. The goal of this study is to suggest a methodology for identifying the most relevant features of cardiac disease characteristics by applying a feature selection technique. The data set used in this study was Framingham heart disease dataset (FHS). It was collected from KAGGLE Machine Learning repository. There are 16 attributes and a mark in the dataset that has been validated by four ML classifiers. There are two feature selection methods, Correlation Based Feature selection (CBFS) and Principle Component Analysis (PCA) was used for the comparison in the study. By using CBFS Method five highly correlated features are selected for the study, and by using PCA thirteen features are selected. The experimental result shows that Correlation Based Feature Selection with Multilayer perceptron (CBFS with MLP) obtained the highest accuracy for this dataset.

1. Introduction

The research concentrates on the two feature selection methods for data reduction before building the predictive models by classification algorithms. These reduced features are then passed into the classification algorithms to design the models for the heart disease prediction. These models are used for the comparison of accuracy of the classifier. Principle Component Analysis and Correlation Based feature selection methods are used for finding out the reduced features. The selected features are inputted to four different classifiers such as Navie Bayes, ADABOOST, MLP and SMO. The accuracy of each model is compared with the other.

2. Background Study

Devansh Shah studied various attributes related to heart disease[1]. The study was conducted with Naïve Bayes, decision tree, K-nearest neighbor, and random forest algorithms[1]. The experimental result proves that K-nearest neighbor algorithm exhibits the highest accuracy.

Hamidreza Ashrafi Esfahani[2] formulated a model to predict cardiovascular disease. The model includes decision trees, Neural Networks, Rough set, Naïve Bayes and SVM for implementation. On comparing the results achieved, it was revealed that the hybrid model of Rough Set, Naïve Bayes and Neural Network obtained the highest accuracy. An ensemble strategy was implemented that allowed for the output to be combined that would result in

Table 3. Comparison of Predictive Accuracy of MLP.

MLP	Precision	Recall	F Measure	ROC
Before FS	0.790	0.842	0.798	0.671
After CBFS	0.805	0.849	0.791	0.649
After PCA	0.784	0.795	0.668	3.62

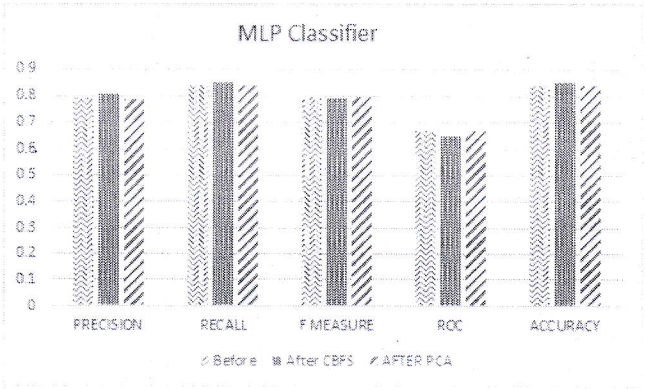


Figure 3. Comparison of Accuracy of MLP Classifier.

14. Conclusion and Findings

During the study the performance of two different feature selection methods CBFS and PCA are evaluated. Eight different classifier models are developed by combining the feature selection and classification algorithms.The performance of each model was evaluated.Performance measures such as Accuracy, Precision, Recall, F Measure and ROC are evaluated for finding out the best classifier. From the result it is proven that the model CBFS with MLP Classifier shows the maximum performance for FHS dataset.

References

[1] Devansh Shah, Samir Patel and Santosh Kumar Bharti 2020 Heart Disease Prediction using Machine Learning Techniques Springer.
[2] Hamidreza and Ashrafi Esfahani 2017 Cardiovascular disease detection using a new ensemble classifier IEEE International Conference on Knowledge-Based Engineering and Innovation (KBEI).
[3] V.V.Ramalingam 2018 Prediction of Heart Diseases Using Machine Learning International Journal of Engineering and Technology.
[4] Amin Ul Haq 2019 A Hybrid Intelligence System Frame Work for Prediction of Heart Disease Using ML, Mobile Information Systems.
[5] Shadman Nashif Heart Disease Detection by Using Machine Learning Algorithm and a Real time Cardiovascular Health Monitoring System World journal of Engineering and Technology,vol.06 no.04,2018,article id 88650.
[6] Poornima Singh 2018 Effective heart disease prediction system using data mining techniques International journal of Nano medicine.
[7] M. Gunay and T. Ensarı 2019 Predictive churn analysis with machine learning methods IEEE 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1- 4.
[8] R. Suguna, M. Shyamala Devi and Rincy Merlin Mathew 2019 Customer Churn Predictive Analysis by Component Minimization using Machine Learning International Journal of Innovative Technology and Exploring Engineering.
[9] Shyamala Devi, Munisamy, Suguna Ramadass and Aparna Joshi 2020 Cultivar Prediction of Target Consumer

Certificate of Presentation

ISBN 978-93-91286-40-8

KRISTU JYOTI COLLEGE OF MANAGEMENT & TECHNOLOGY
IQAC | Department of Computer Applications
RESEARCH HUB

CERTIFICATE
OF PRESENTATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Ambily Merlin Kuruvilla

OF KARPAGAM ACADEMY OF HIGHER EDUCATION
FOR SUCCESSFULLY PRESENTING A PAPER AT THE FIRST INTERNATIONAL CONFERENCE ON ADVANCE MODERN COMPUTING TRENDS AND TECHNOLOGY (ICAMCTT 2021) ON 30TH & 31ST OF JULY 2021

Paper Title : Hyper Parameter Optimization in Stacked Deep Neural Network for Medical Diagnosis

REV. FR. JOSHY CHEERAMKUZHY CMI
Principal

SUSHEEL GEORGE JOSEPH
Conference Secretary

ROJI THOMAS
Conference Director

BINNY S
Conference Convenor

SAINTGITS COLLEGE OF APPLIED SCIENCES
PATHAMUTTOM, KOTTAYAM
SAINTGITS

Dr. K. K. John
Principal
Saintgits College of Applied Sciences
Kottukulam Hills, Pathamuttom P. O.
Kottayam 686 532, Kerala

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

HYPER PARAMETER OPTIMIZATION IN STACKED DEEP NEURAL NETWORK FOR MEDICAL DIAGNOSIS

Ambily Merlin Kuruvilla
Research Scholar
Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India.
ambilykuruvilla@gmail.com

Dr. N V Balaji
Dean, Faculty of Arts, Science and Humanities, Karpagam Academy of Higher Education
Coimbatore, India.
balajinv@karpagam.com

Abstract— Diabetes is a metabolic disease where the blood sugar rate of an individual is consistently above normal. Due to the modern lifestyle and work culture, diabetics is widespread and affects the productivity and quality of life for an individual. A diabetics patient is at a very high risk of various health issues like organ failure and even it can even result in loss of life. An early prediction of this chronic disease can avoid health issues and save many lives. The aim of this article is to develop a better predictive model for diabetics using an automated hyper parameter optimization (HPO) approach in Multilayer Perceptron (MLP). This article provides an efficient way to increase the accuracy of Neural Network to a substantial level through the HPO process using Grid Search Optimization (GSO) through the stacking ensemble model. In order to run the ensemble model at an optimal level and to minimize errors, appropriate hyperparameters must be calculated. Three GSO methods are utilized to tune the hyper parameters. To build the stacking ensemble model, the PIMA data set was used.

Keywords: HPO, MLP, GSO: Hyper Parameter Optimization, Multilayer Perceptron, Grid Search Optimization.

INTRODUCTION

A study carried out by the WHO recently revealed that in 2016, diabetics was one of the leading causes of death worldwide. Diabetes has resulted in 1.6 million fatalities in 2016 and this statistic replaces HIV / AIDS with diabetes as one of the most frequent cause of death [4]. The burden of diabetes disease grew from 108 million in 1980 to 422 million [5] in 2014, and the percentage of diabetic patients amongst adults over 18 years of age rose from 4.7% in 1980 to 8.5% in 2014[5]. 642 million people i.e. (1 in 10 people) are expected to contract diabetes by 2040. 46.5% of people with diabetes have not been diagnosed officially [6]. This makes it necessary to develop techniques and procedures to assist in the early detection of diabetes in order to reduce the number of deaths related to diabetes, as late diagnosis is responsible for a majority of deaths linked to diabetes [7].

There is a need to implement sophisticated information processing to develop cutting-edge strategies for the early detection of diabetes. Data mining tools can also be effectively applied. The ability to remove and uncover previously unseen, secret, yet important patterns from a large database repository is given by data mining [7]. These tools can assist medical evaluation and decision making.

LITERATURE SURVEY

Roshan Birjais [1] conducted a research in many classification algorithms for diabetes prediction and his team found out that the Gradient Boosting algorithm outperform other classifiers and

Regression			
MLP	78.3	75.0	74.0
Stacked Model	90.7	94.0	82.0

Table 1: Performance Comparison of Classifiers

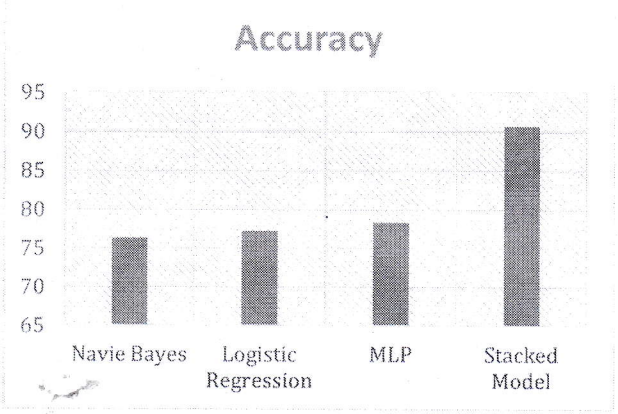


Figure 2: Classifier Performance based on Accuracy

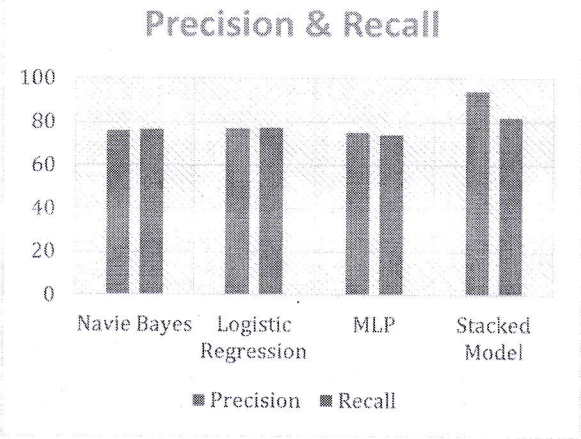


Figure 3: Classifier Performance based on Precision and Recall.

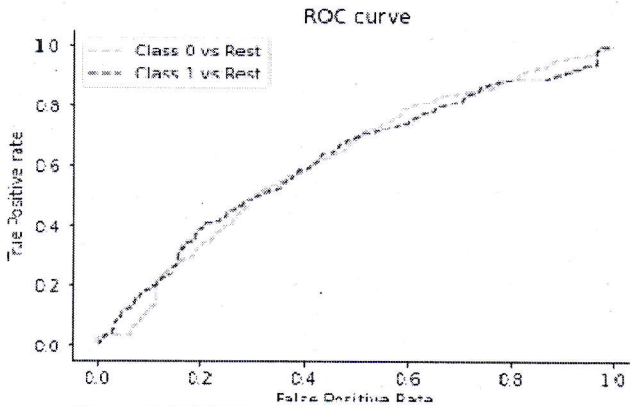


Figure 4: ROC Curve

ROC curve is a graphical plot that shows the system's diagnostic potential as its discriminating threshold is varied. The ROC curve is generated by plotting the true positive rate at different threshold settings against the false positive rate.

CONCLUSION

In this article we are introducing an ensembled hyper parameter tuning mechanism to tackle the deficiencies and improving the accuracy. For this purpose, we have used an automated stacked ensemble method which combines various hyper parameters. Grid Search Optimisation method is used, and three different models were created as base learners using Neural Network by combining various activation functions. The optimum value for each parameter is calculated and stored into an external file by each model. Three output files are created, and these output files are inputted to a logistic regression model which is a learning model. We have used LR Model as the learning model. It is found out that this model improves the accuracy to good extend.

REFERENCES

1.Roshan Birjais, Ashish Kumar Mourya,Ritu Chauhan, Harleen Kaur " Prediction and diagnosis of future diabetes risk : a machine learning approach". Springer.

2.Md.Maniruzzaman,JahanurRahman, BenojirAhammed, Md.Menhazul bedin "Classification and prediction of diabetes disease using machine learning paradigm", Springer

3. Atik Mahabub "A robust voting approach for diabetes prediction using traditional machine learning techniques", Springer.

4.www.geeksforgeeks.org/ml-chi-square-test-for-feature-selection

5.www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes Retrieved.

6.www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, Accessed 27th Jul 2018.

7.www.who.int/news-room/fact-sheets/detail/diabetes retrieved 27/07/2018.

8.www.diabetesdaily.com/learn-about-diabetes/what-is-diabetes/how-many-people-have-diabetes.

Certificate of Presentation

PSN COLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous Institution)

Melathediyoor, Tirunelveli - 627 152.

International Virtual Conference on Computational Engineering (IVCCE- 2022)

Certificate of Appreciation

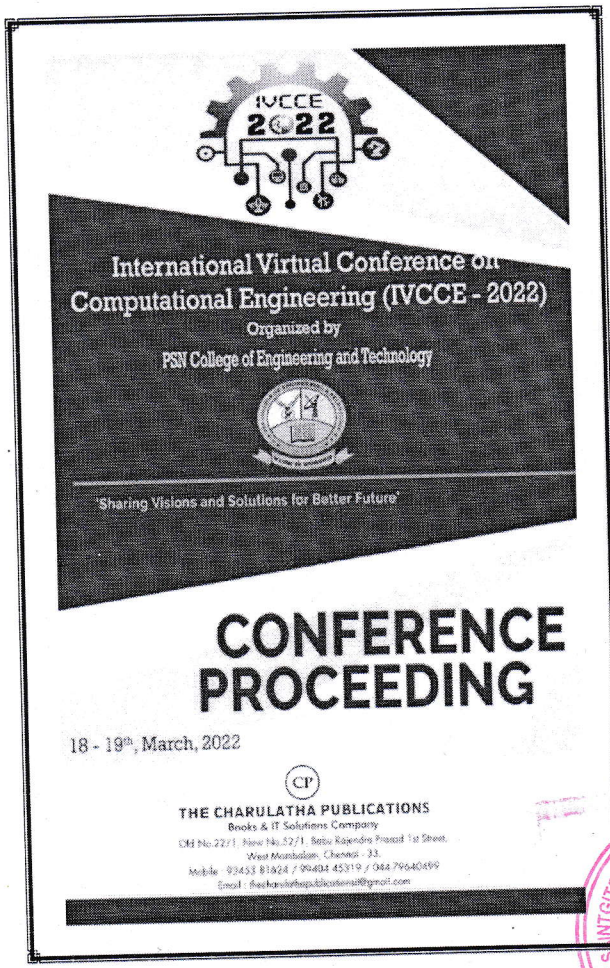
This is to certify that Ambily Merlin Kuruvilla of Saintgits College of Applied Sciences, Kottayam, Kerala,India has presented a paper entitled on Big Data in Cloud Computing Environment in the International Virtual Conference on Computational Engineering (IVCCE - 2022) on 18/03/2022 to 19/03/2022.

Convener
Dr.M.Vignesh kumar
Associate Professor

Executive Director
Dr.P.Selva kumar
Professor

Principal
Dr.V.Manikandan
Professor

Proceedings



IVCCE 2022

International Virtual Conference on Computational Engineering (IVCCE - 2022)

Organized by

PSN College of Engineering and Technology

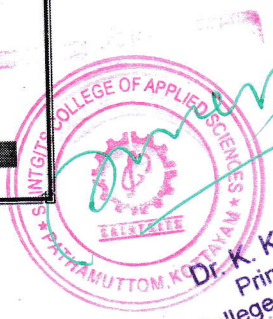
'Sharing Visions and Solutions for Better Future'

CONFERENCE PROCEEDING

18 - 19th, March, 2022

THE CHARULATHA PUBLICATIONS

Books & IT Solutions Company



Dr. K. K. John
Principal
Saintgits College of Applied Sciences
Kottukulam Hills, Pathamuttom P. O.
Kottayam 686 532, Kerala

INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
May- June 2022, Vol 2(03), 01-05.
ISSN No: 2583-0368

Big Data in Cloud Computing Environment

Ansaba R B [1], Ambily Merlin Kuruvilla [2]
[1] *BCA student, Saintgits College of Applied Sciences, Pathamuttom, Kottayam, Kerala, India*
[2] *Assistant Professor & HOD, Department of Computer Applications, Saintgits College of Applied Sciences, Kottayam, Kerala, India*

Article Type: Research

OPEN ACCESS
Article Citation:
Ansaba R B[1],Ambily Merlin Kuruvilla [2], "Big Data in Cloud Computing Environment", International Journal of Recent Trends In Multidisciplinary Research, March-April 2022, Vol 2(03), 01-05.

Accepted date: May 12, 2022

Published date : May 14, 2022

Published by 5th Dimension Research Publication.

Abstract
In today's world, Big Data is an important area that is used in decision making and it processes huge volumes of data to address some query or pattern. Data is analysed through a set of algorithms, which differs depending upon the type of data, business's aim behind the analysis, and also other factors. But bigdata possess many challenges in terms of storing and processing data. Hence cloud computing which is another emerging technology is integrated with big data which provides better infrastructure for processing, storage for enormous data, and networking services.

Key Words: Big Data; Cloud Computing; Hadoop; Hdfs; Map Reduce

1. Introduction

Single Cloud computing is a powerful model and infrastructure that is distributed across the internet which process, manage and store the data. Cloud computing offers services for enterprise applications which centralizes both data storage and perform huge scale complex computing. It can reduce maintenance cost, provide less infrastructure and accelerate automation. [1]
Cloud services enables big data to analyse, manage and process the stored data in a more efficient manner. Through virtualization process integration of big data with cloud is the being achieved. Virtualization denotes the usage and sharing of resources independent of underlying hardware. Microsoft's Cloud Hadoop includes Azure Marketplace which comprise MapR and Azure Data Lake, which comprise Data Lake Store, Azure HDInsight, Data Lake Analytics as Azure cloud services. AWS includes versions of Hadoop, Spark, and Presto which operate on the data stored in Amazon Glacier and S3. Google's managed Hadoop include Cloud Dataproc and Spark cluster which uses GCP cloud services such as Big Query and Bigtable. [2]
Cloud platform provide rich productivity suites for database, data warehouse, collaboration, business intelligence, OLAP, and development tools. Big Data processing has many challenges relating with Data collection, analysis, sharing, research and visualization. Each of these processes need different techniques, infrastructure, and highly skilled professionals. Also, it cannot be done easily with traditional programs because of resource restrictions such as computing power and time, hence we need advanced algorithms and vast databases. And all these difficulties and barriers are much reduced as a result of integrating Big Data within cloud environment. [3]
Big data represents huge amounts of complex data which can be either unstructured or structured generated by multiple sources. The traditional relational databases are not sufficient to process and analyse data from multiple-sources, such as managing data related with record of transactions, customer behaviour, mobile phone and GPS navigation, etc. So, to deal with these kinds of complex data, cloud is employed, which serve as the storehouse where the processed outcome/data will be stored. Cloud computing approach is efficient because of having advanced technologies to handle the vast amount of data. This paper discusses an overall view of cloud computing and big data, their features, Relation and integration of big data & cloud, some **big data management tools in cloud**.

2. Cloud Computing

Cloud computing is a type of service-oriented computing where software and hardware are delivered as a service over the internet. Cloud is a combination of distributed and centralized system which includes virtualized servers, operating systems, applications, etc that are dynamically supplied. It provides services relating to storage, processing and sharing of data through visualized resources over the networks. Cloud platform is completely virtual to its users and require less effort from user to operate and manage its services. Important features associated with cloud includes scalability, on-demand delivery of resources, easy accessibility, cost-effective, flexibility and reliability. [4]

HDFS is a file system used to store or span all the nodes in a Hadoop cluster for data storage. Thereby it improves reliability and support security. HDFS usually splits files into blocks which in turn is stored on the server. Thereby it maintains reliability by duplicating data across multiple hosts combining parallel processing technique. [13]

- **MapReduce**

This is a framework which helps in writing applications that process and generates large datasets on a cluster with parallel or distributed algorithm. At first, breaking Big Data into small subunits takes place which in turn are analysed and processed by Map jobs in parallel. Map () method consists of acquiring, filtering & categorizing datasets. Reduce () method consists of final result generation and locating associated summaries. [14]

- **NoSQL**

NoSQL (Not Only SQL) systems provides systematic way to store and replicate data, giving out retrieval and appending operations from the data. These databases are not bound by the confines of a fixed schema model instead each are deployed as a cluster of nodes. Examples of NoSQL systems include Amazon DynamoDB, Azure Cosmos DB, MongoDB, Cassandra, CouchDB, and HBase.

5. Conclusion

This paper presented how cloud computing helps in analysing, storing and processing big data. Big data and cloud together comprise an integrated model of distributed network technology.Cloud supports big data in terms of security of data, encryption, data integrity, data transformation, data heterogeneity, data quality and others.
Even though there are challenges regarding integration with cloud such as scalability, availability and problems with bandwidth for data transfer, Solutions are constantly being developed by cloud providers for the efficient use of big data on cloud. So, the integration and application of big data in cloud will have a huge impact and continue to grow in the following years.

References

[1] Beri, R. & Behal, V. (2015). Cloud Computing: A Survey on Cloud Computing. Int. J. Comput. Appl., Vol. 111, pp. 19–22.
[2] Sheetal Singh, Vipin Kumar Rathi, Bhawna Chaudhary, "Big Data and Cloud Computing: Challeges and Opportunities",International Journal of Innovations in Engineering and technology, Vol. 5(4), August 2015.
[3] Gupta, H. & Mohania, M. (2012). Cloud computing and big data analytics: What is new from databases perspective? in Big Data Analytics. BDA 2012. Lecture Notes in Computer Science, Vol. 7678, pp. 42–61 (Springer Berlin Heidelberg).
[4] Fonseca, N., & Boutaba, R. (2015). Cloud services, networking, and management. John Wiley & Sons.
[5] K. Kaur, "A Review of Cloud Computing Service Models", International Journal of Computer Applications, Vol.140, No.7, pp.15-18, 2016.
[6] J. Srinivas, K.Venkata Subba Reddy and Dr. A. Moiz Qyser, "Cloud Computing Basics", International Journal of Advanced Research in Computer and Communication Engineering, Vol.1(5), 2012.
[7] Venters, W., Whitley, E.A.: A Critical Review of Cloud Computing: Researching Desires and Realities. J. Inf. Technol. 27, 179–197 (2012).
[8] D.P. Acharjya, Kauser Ahmed P, " A Survey on Big Data Analytics: Challenges, Open Research Issue and Tools", International Journal of Advanced Computer Science and Applications, Vol. 7(2), 2016
[9] N. Elgendy and A. Elragal, "Big Data Analytics: A Literature Review Paper," in Advances in Data Mining.Applications and Theoretical Aspects, 2014, pp. 214–227.
[10] A, Katal, Wazid M, and Goudar R.H. "Big data: Issues, challenges, tools and Good practices.". Noida: 2013, pp. 404 – 409, 8-10 Aug. 2013.
[11] Bautista Villalpando, L. E.; April, A. & Abran, A. (2014). Performance analysis model for big data applications in cloud computing, Vol. 3, pp. 1–20
[12] Agrawal, Divyakant & Das, Sudipto & Abbadi, Amr. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. ACM International Conference Proceeding Series. 530-533. 10.1145/1951365.1951432
[13] K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions.". JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.
[14] L. Zhao, Z. Zhou," Cloud computing model for big data processing and performance optimization of multimedia communication", Computer Communications (2020)