International Journal of Engineering Technology and Applied Science
(ISSN: 2395 3853), Vol. 4 Issue 12 December 2018

Performance Analysis of Diseases Detection Rate Using Classifier

Bharti Faraiya
M. Tech. Scholar, Department of CSE
ASCT, Bhopal (M.P.), India
E-Mail:-bhartif12@gmail.com

Mr.B.P.S Sengar
Professor, Department of CSE
ASCT, Bhopal (M.P.), India
E-Mail:-bpssengar@gmail.com

ABSTRACT
Spurred by advances in processing power, memory, storage, and an unprecedented wealth of data, computers are being asked to tackle increasingly complex learning tasks, often with astonishing success. By digitizing, combining and effectively using big data, healthcare organizations ranging from single-physician offices and multi-provider groups to large hospital networks and accountable care organizations stand to realize significant benefits. In this paper we presents the comparative experimental study for the various diseases using classification methods and out proposed method show better results than existing methods.

Keywords: Health Care, Data Mining, Supervised Learning, Machine Learning, Medical Science.

INTRODUCTION

The healthcare industry historically has generated large amounts of data, driven by record keeping, compliance & regulatory requirements, and patient care [7]. While most data is stored in hard copy form, the current trend is toward rapid digitization of these large amounts of data. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of data (known as 'big data') hold the promise of supporting a wide range of medical and healthcare functions, including among others clinical decision support, disease surveillance, and population health management.

The key idea behind the probabilistic framework to machine learning is that learning can be thought of as inferring plausible models to explain observed data. A machine can use such models to make predictions about future data, and decisions that are rational given these predictions. Uncertainty plays a fundamental role in all of this. Observed data can be consistent with many models, and therefore which model is appropriate given the data is uncertain [11].

The rapid growth of novel technologies has led to a significant increase of digital health data in recent years. More medical discoveries and new technologies such as mobile apps, capturing devices, novel sensors, and wearable technology have contributed to additional data sources. Therefore, the healthcare industry produces a huge amount of digital data by utilizing information from all sources of healthcare data such as Electronic Health Records (EHRs, including electronic medical records) and Personal Health Records (PHRs, one subset of EHRs including medical history, laboratory results, and medications). Based on reports, the estimation of digital healthcare data from all over the world was almost 500 peta-bytes (10^{15}) in 2012, and it is expected to increase and reach 25 exa-bytes in 2020 [12].

The majority of big data applications deal with data that do not refer to an individual person. This does not exclude the possibility that their aggregated information content might not be socially sensitive, but very rarely is it possible to reconnect such content to the identity of an individual. In the cases where sensitive data are involved, it is usually possible to collect and analyze the data at a single location; so this becomes a problem of computer security; within the secure box, the treatment of the data is identical to that of non-sensitive data. Healthcare poses some peculiar problems in this area. First, all medical data are highly sensitive, and in many developed countries are considered legally owned by the patient, and the healthcare provider is required to respect patient confidentiality. The European parliament is currently involved in a complex debate about data protection legislation, where the need for individual confidentiality can be in conflict with the needs of society [3].

There has been increasing interest in gathering nontraditional, digital information to perform disease surveillance. These include diverse datasets such as those stemming from social media, internet search, and environmental data. Twitter is an online social media platform that enables users to post and read 140- character messages called "tweets." It is a popular data source for disease surveillance using social media since it can provide nearly instant access to real-time social opinions. More importantly, tweets are often tagged by geographic location and time stamps potentially providing information for disease surveillance. Another notable nontraditional disease surveillance system has been a data-aggregating tool called Google Flu Trends, which uses aggregated search data to estimate flu activity [2].

The rest of this paper is organized as follows in the first section we describe an introduction of about Machine learning and Health care. In section II we discuss about the Big data overview in brief for the Health care techniques. In section III we discuss about the experimental comparative study for the various diseases like Cleveland, diabetes etc. , In finally in section V we conclude the about our paper which is based on the health care simulation study and specify the future scope.

II BIG DATA OVERVIEW

The explosive growth and widespread accessibility of digital health data have led to a surge of research activity in the healthcare and data sciences fields. The conventional approaches for health data management have achieved limited success as they are incapable of handling the huge amount of complex data with high volume, high velocity, and high variety. Computational health informatics is an emerging research topic within and beyond the medical industry. It is a multidisciplinary field involving various sciences such as biomedical, medical, nursing, information technology, computer science, and statistics. Using Information and Communication Technologies (ICTs), health informatics collects and analyzes the information from all healthcare domains to predict patients' health status. Big Data also impact more in healthcare. Nowadays, health care systems are rapidly adopting clinical data, which will rapidly enlarge the size of the health records that are accessible, electronically [13].

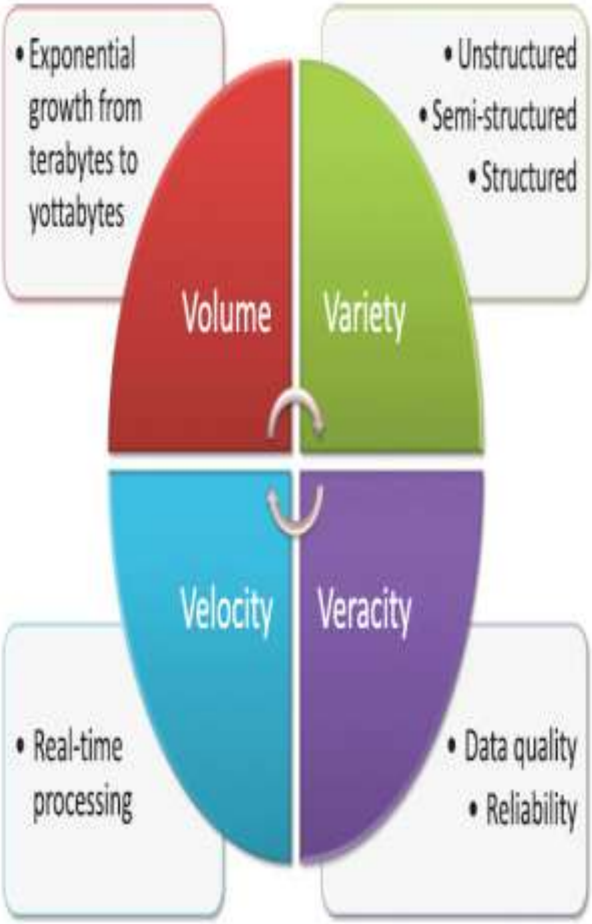

Fig. 1: Four Vs of big data [12].

III SIMULATION RESULTS

In this section, experimental process of we show that the comparative result analysis study for the Health care sector with disease diagnosis of various dataset such as Heart, Liver, Cancer etc.

are performed. This process of disease diagnosis of various dataset is done by using Three methods are such as K-Nearest neighbor classification, Decision tree and Proposed method. For the performance evaluation of ensemble classifier technique and our cascaded model MATLAB software package is used. MATLAB is a software package for high- performance numerical computation and visualization. We discuss about the dataset which we used for the diseases detection in the field of health care. There are all these dataset types will be fetched from the UCI machine learning repository for the research purpose, the datasets are Cleveland, diabetes etc.

Dataset name	Method	Accuracy (%)	Precision (%)
Cleveland	KNN	80.56	85.25
	Decision Tree	89.32	87.17
	PROPOSED	90.72	92.96

Table 1: Show that the comparative result analysis study for the Cleveland dataset with using existing and proposed method.

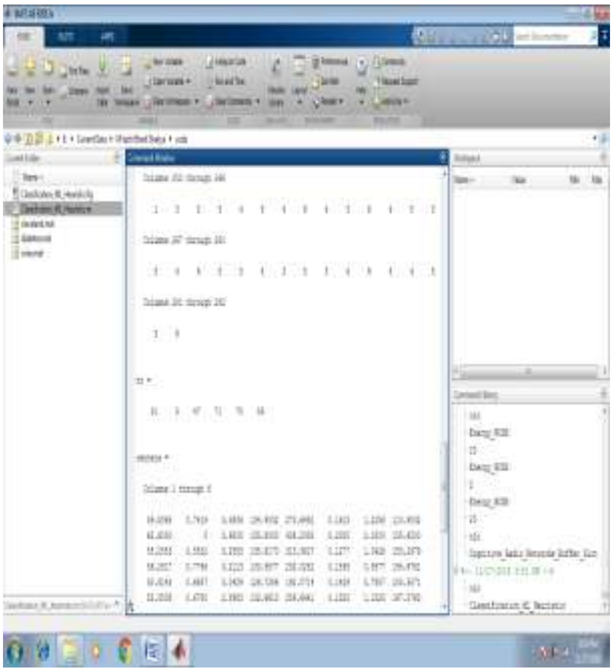

Fig 2: Figure shows that the simulation enviornenr for the KNN methods using Clevelend dataset.

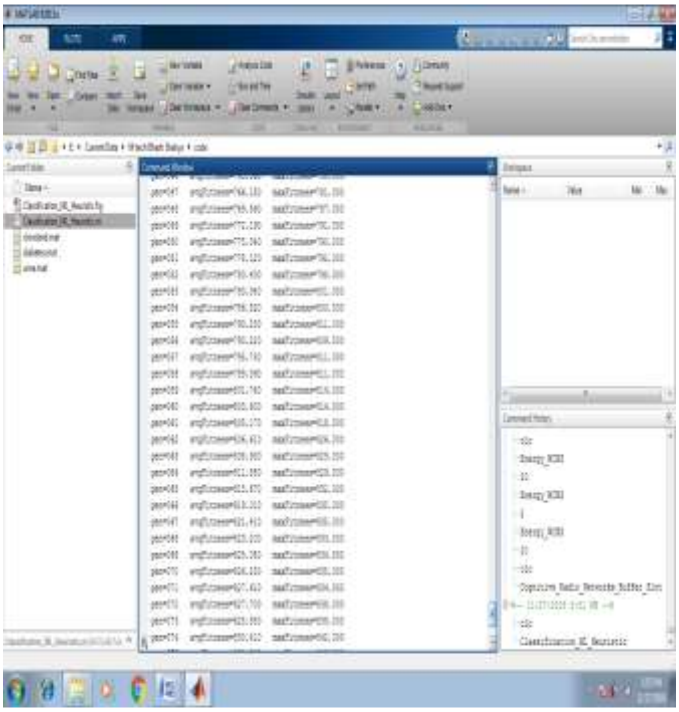

Fig 3: Figure shows that the simulation enviornenr for the proposed methods using Clevelend dataset.

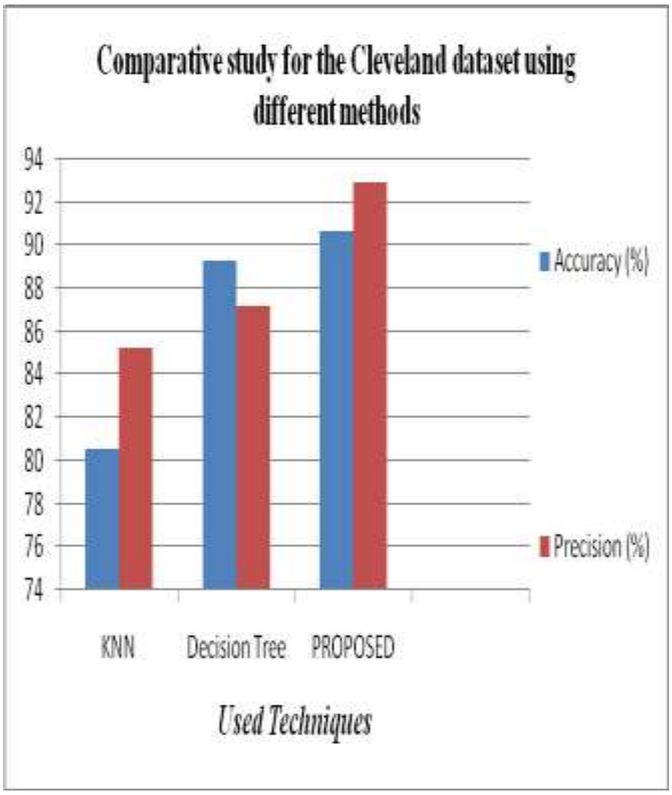

Fig 4: Figure shows comparaive simulation study graph for the existing an proposed methods using cleveland dataset.

IV CONCLUSION AND FUTURE SCOPE

Nowadays, health care systems are rapidly adopting clinical data, which will rapidly enlarge the size of the health records that are accessible, electronically. Big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. In this paper we shows classification and optimization model for the diseases prediction and improve the classification rate. In future we also used some evolutionary and other methods for improving the results.

REFERENCES:-

[1] MIN CHEN, YIXUE HAO, KAI HWANG, LU WANG, AND LIN WANG, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities" SPECIAL SECTION ON HEALTHCARE BIG DATA, Vol-5, IEEE, 2017. Pp 8869-8879.

[2] Sudha Ram, Wenli Zhang, Max Williams, Yolande Pengetnze, "Predicting Asthma-Related Emergency Department Visits Using Big Data", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, 2015. Pp 1216-1218.

[3] Marco Viceconti, Peter Hunter, and Rod Hose, "Big Data, Big Knowledge: Big Data for Personalized Healthcare", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19,, 2015. Pp 1209-1215.

[4] Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang," Big Data for Health", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, 2015. Pp 1193-1206.

[5] Yin Zhang, Meikang Qiu, Chun-Wei Tsai, Mohammad Mehedi Hassan, Atif Alamri, "Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data", IEEE SYSTEMS JOURNAL, 2015. Pp 1-9.

[6] Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, Kayvan Najarian, "Big Data Analytics in Healthcare", Hindawi Publishing Corporation Bio-Medical Research International, 2015. Pp 1-17.

[7] Wullianallur Raghupathi, Viju Raghupathi, "Big data analytics in healthcare: promise and Potential", Raghupathi and Raghupathi Health Information Science and Systems 2014. Pp 1-10.

[8] Michael K. K. Leung, Andrew Delong, Babak Alipanahi, Brendan J. Frey, "Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets", IEEE Vol-104, 2016. Pp 176-197.

[9] Daniele Ravı, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, Guang-Zhong Yang," Deep Learning for Health Informatics", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 21, 2017. Pp 4-21.

[10] MICHAEL J. PAUL, ABEED SARKER, JOHN S. BROWNSTEIN, AZADEH NIKFARJAM, MATTHEW SCOTCH, KAREN L. SMITH,GRACIELA GONZALEZ, "SOCIAL MEDIA

MINING FOR PUBLIC HEALTH MONITORING AND SURVEILLANCE”, Pacific Symposium on Biocomputing 2016. Pp 468-477.

[11] Zoubin Ghahramani, “Probabilistic machine learning and arti_cial intelligence”, 2015. Pp 1-24.

[12] RUOGU FANG∗ , SAMIRA POUYANFAR∗ , YIMIN YANG, SHU-CHING CHEN, S. S. IYENGAR, “Computational Health Informatics in the Big Data Age: A Survey”, ACM Computing Surveys, Vol. 49, 2016. Pp 1-36.

[13] Gunasekaran Manogaran, Chandu Thota, Daphne Lopez, V. Vijayakumar, Kaja M. Abbas and Revathi Sundarsekar, “Big Data Knowledge System in Healthcare”, 2017. Pp 133-158.

[14] Rahul C. Deo, “Machine Learning in Medicine”, 2017. Pp 1920-1931.

[15] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, Joel T. Dudley, “Deep learning for healthcare: review, opportunities and Challenges”, Briefings in Bioinformatics, 2017, Pp 1–11.

[16] Ji-Jiang Yang, Jianqiang Li, Jacob Mulder, Yongcai Wang, Shi Chen, Hong Wu, Qing Wang, Hui Pan, “Emerging information technologies for enhanced healthcare’, Elsevier ltd. 2015. Pp 3-11.

[17] Majid Ghonji Feshki and Omid Sojoodi Shijani “Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network”, IEEE, 2016, Pp 48-53.

[18] Asha Gowda Karegowda and Darshan M “Optimizing Feed Forward Neural Network Connection Weights Using Artificial Bee Colony Algorithm”, International Journal of Advanced Research in Computer Science and Software Engineering, 2013, Pp 452-454.

[19] Sultan Noman Qasem, Siti Mariyam Shamsuddin, Siti Zaiton Mohd Hashim Maslina Darus and Eiman Al-Shammari “Memetic multiobjective particle swarm optimization-basedradial basis function network for classificationproblems”, Elsevier, 2013, Pp 165-190.

[20] Ashraf Osman Ibrahim, Siti Mariyam Shamsuddin, Nor Bahiah Ahmad and Sultan Noman Qasem “Three-Term Backpropagation Network Based On Elitist Multi-objective Genetic Algorithm for Medical Diseases Diagnosis Classification”, Life Science Journal, 2013, 1815-1822.

[21] Mr. Deepak Dhanwani and Prof. Avinash Wadhe “Study Of Hybrid Genetic Algorithm Using Artificial Neural Network In Data Mining For The Diagnosis Of Stroke Disease”, International Journal of Computational Engineering Research, 2013, Pp 95-100.