International Journal of Engineering Technology and Applied Science

(ISSN: 2395 3853), Vol. 4 Issue 11 November 2018

A Review on Health Care Using Data Mining and Machine Learning Techniques

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

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

ABSTRACT
Health is a potential state in which someone performs well, both mentally and physically living within one inhabits environment. Likewise, the provision of health services by utilizing available health resources to fulfill the needs of a target population is termed a healthcare delivery system. There are various techniques are used to predict and diagnosis of health care data such as data mining, machine learning and optimization. In this paper we present comparative literature study for the health care using various techniques.

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

INTRODUCTION
The mining of healthcare database is very critical issue. The healthcare data stored the information about medical diseases and patient's information. For the estimation of patients and disease used some intelligent software for the predication of disease. The prediction of disease and medicine is fundamental issue in health care environment. For the extraction of better information used data mining technique for the healthcare system.

Heart disease prediction and classification is major issue in current lifestyle and diagnosis system. For the prediction and classification used various classification algorithm are used such as support vector machine, neural network and other technique. Associative classification is other technique for classification based on association rule mining. Association rule mining technique generates the association rule for relation data. These association rules define certain number of class for classification. Some other techniques such as neural network will play the role for classify the data into number of class which is further used with back propagation neural network and feed forward neural network. The best classification results we can also used the super set of classifier i.e. support vector machine which provides the best accuracy and other performance parameters evaluation.

Data mining technique gives well know classifier for medical disease data classification. In cluster oriented ensemble classifier is suffered from a selection of optimal number of cluster for ensemble. For this process we used ant colony optimization technique. Ant is meta-heuristic function inspired by biological ants. The objective of ant colony optimization is multiple. Using ant colony optimization we maintain the selection process of clustering technique and noise removal of boundary base class. Noise reduction and selection of optimal number of cluster in ensemble classifier used features sub set selection process using ant colony optimization technique. The algorithms employed by a CDSS can be characterised by the reasoning paradigm employed by the CDSS which is primarily of two types. The first type is the knowledge-driven paradigm, which employs an inference engine that applies the formalized clinical knowledge representations stored in the knowledge base to an instance of the patient's data.

The knowledge base in this case, can contain rules for treatment or diagnosis, probabilistic associations of symptoms to diseases, drug-drug interactions, or clinical workflows with decision steps. The second type is the data-driven reasoning paradigm which employ methods from machine-learning such as neural networks, support vector machines, statistical methods such as regression models, and pattern recognition methods such as k-nearest neighbours (kNN) to detect

patterns in clinical data. Rather than an inference engine, the data-driven approach uses a classifier, and the data contained in the knowledge-base may be thought of as the trained parameters or weights associated with the classifier. CDSSs come in a variety of forms, with various characteristics. In order to understand why some CDSSs are successful and others are not it is important to examine the critical features that describe a CDSS. One of the most common ways to classify a CDSS is by the type of knowledge and reasoning method it uses for arriving at the advice on the clinical decision.

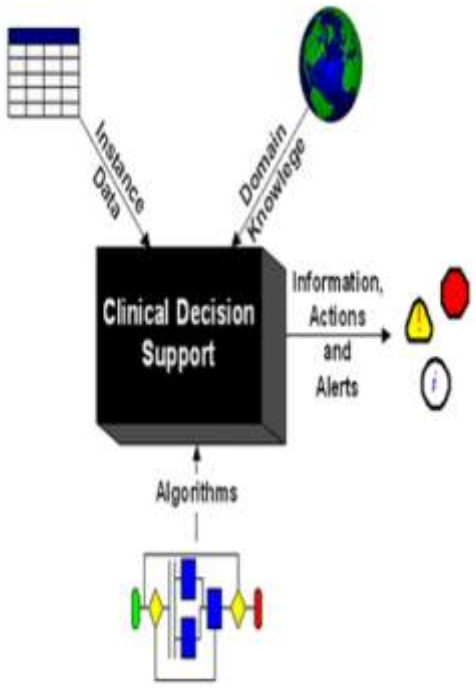

Figure 1: Principle elements of a clinical decision support system (CDSS)

The rest of this paper is organized as follows in the first section we describe an introduction of about Cloud computing and load balancing. In section II we discuss about the rich literature survey for the Health care techniques. In section III we discuss about the machine learning, In section IV discuss about the problem formulation and finally in section V we conclude the about our paper which is based on the literature survey and specify the future scope.

II RELATED WORK

Here we discuss about the health care sector for various diseases over the machine learning and other techniques, in the real world there are various diseases such as heart diseases, liver diseases, cancer, diabetes etc. we can prevent from the all such types of diseases using diagnosis system, here we provide a literature survey for the medical diseases diagnosis using various techniques.

[1] In this paper, author described the health care for the various diseases using machine learning algorithm over big data analytics. They proposed model for the various diseases diagnosis over the big data, big data which is a combination of variety, volume, velocity etc. here the big data analytics provide huge data for the medical science. They provide the model and do their experimental work on the big data of health care sector using machine learning techniques.

[2] They introduce a novel method of using multiple data sources for predicting the number of asthma-related emergency department (ED) visits in a specific area. Twitter data, Google search interests, and environmental sensor data were collected for this purpose. Our preliminary findings show that our model can predict the number of asthma ED visits based on near-real-time environmental and social media data with approximately 70% precision.

[3] They propose in this position paper that big data analytics can be successfully combined with VPH technologies to produce robust and effective in silico medicine solutions. In order to do this, big data technologies must be further developed to cope with some specific requirements that emerge from this application. Such requirements are: working with sensitive data; analytics of complex and heterogeneous data spaces, including non-textual information; distributed data management under security and performance constraints; specialized analytics to integrate bioinformatics and systems biology information with clinical observations at tissue, organ and organisms scales; and specialized analytics to define the “physiological envelope” during the daily life of each patient.

[4] This paper provides an overview of recent developments in big data in the context of biomedical and health informatics. It outlines the key characteristics of big data and how medical and health informatics, translational bioinformatics, sensor

informatics, and imaging informatics will benefit from an integrated approach of piecing together different aspects of personalized information from a diverse range of data sources, both structured and unstructured, covering genomics, proteomics, metabolomics, as well as imaging, clinical diagnosis, and long-term continuous physiological sensing of an individual.

[5] This paper proposes a cyber-physical system for patient-centric healthcare applications and services, called Health-CPS, built on cloud and big data analytics technologies. This system consists of a data collection layer with a unified standard, a data management layer for distributed storage and parallel computing, and a data-oriented service layer. The results of this study show that the technologies of cloud and big data can be used to enhance the performance of the healthcare system so that humans can then enjoy various smart healthcare applications and services.

[6] In this paper, they discuss some of these major challenges with a focus on three upcoming and promising areas of medical research: image, signal, and genomics based analytics. Recent research which targets utilization of large volumes of medical data while combining multimodal data from disparate sources is discussed. Potential areas of research within this field which have the ability to provide meaningful impact on healthcare delivery are also examined.

[7] 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 the future we'll see the rapid, widespread implementation and use of big data analytics across the healthcare organization and the healthcare industry. To that end, the several challenges highlighted above, must be addressed. As big data analytics becomes more mainstream, issues such as guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies will garner attention.

[8] In this paper, they provide an introduction to machine learning tasks that address important problems in genomic medicine. Here, we describe how machine learning can be used to solve key problems in genomic medicine. Genomics is the study of the function and information structure encoded in the DNA sequences of living cells, whereas precision medicine is the practice of tailoring treatment based on all relevant information about the patient, including the patient's genome.

[9] This article presents a comprehensive up-to-date review of research employing deep learning in health informatics, providing a critical analysis of the relative merit, and potential pitfalls of the technique as well as its future outlook. The paper mainly focuses on key applications of deep learning in the fields of translational bioinformatics, medical imaging, pervasive sensing, medical informatics, and public health.

[12] This article presents a comprehensive overview of the challenges, pipeline, techniques, and future directions for computational health, 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. Most popular surveys of big data in health informatics have concentrated on biomedical aspects of big data, while a smaller percentage of papers focus on the computational perspective.

[14] The purpose of this review is to explore what problems in medicine might benefit from such learning approaches and use examples from the literature to introduce basic concepts in machine learning. It is important to note that seemingly large enough medical data sets and adequate learning algorithms have been available for many decades, and yet, although there are thousands of papers applying machine learning algorithms to medical data, very few have contributed meaningfully to clinical care.

III MACHINE LEARNING

Machine Learning is a branch of Artificial Intelligence that has become very popular, and useful, in the last 10 years. One definition of Machine Learning is that it is the semi-automated extraction of knowledge from data. Broadly speaking, machine learning (ML) deals with the question of how to build computer programs that learn from data and, as a result, can generate programs that generalize from that data in the form of a program that reflects concepts implicit in the underlying data. In effect, with machine learning we have programs using data to create new programs. This is in contrast to the traditional way that programs have been generated by human programmers in which they encode the rules that the computer follows in a programming language in order to produce a solution to a specified problem.

Traditional or conventional writing of programs for a computer can be summarized as automating the procedures to be performed on input data in order to create output artifacts. Almost always, they are linear, procedural and logical.

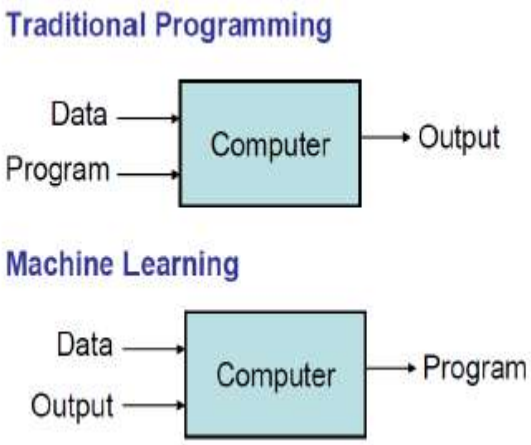

Figure 2: Traditional programming vs. machine learning.

This is not always the case as sometimes, however, there are problems that you can represent in a computer that you cannot write a traditional program to solve. There are two broad classes of machine learning techniques, supervised learning and unsupervised learning.

IV PROBLEM FORMULATION

To predict is to save lives, especially in the domain of healthcare. The more accurately one can predict health related outcomes, the greater are the chances and opportunities to intervene, diagnose and implement pre-emptive measures, thus saving more lives as a result. Similarly, in order to develop a framework for prediction, it is extremely important to identify the factors that directly or indirectly give rise to health-shock. Furthermore, these factors depend greatly on the socio-economic, geographical, and cultural norms of that region so there is a dire need to conduct large-scale health surveys. Health sector mostly prevents possible participation of individuals in the labor force, reducing productivity and hours of work, hence reducing earnings. Unhealthy individuals, especially the elderly, become a liability and burden for the household, and take away large part of household income for their medical expenses. Moreover, it prompts households to engage in low return activities which hinder the ability of the poor to grow their incomes and escape the poverty cycle. The aftermath of health-shock forces the household to deal with financial crises for a long time, possibly for their entire life.

V CONCLUSIONS AND FUTURE WORK

To predict is to save lives, especially in the domain of healthcare. The more accurately one can predict health related outcomes, the greater are the chances and opportunities to intervene, diagnose and implement pre-emptive measures, thus saving more lives as a result. Machine learning play big role in pattern recognition, the recognition of pattern faced the series of training process. In this paper we shows comparative study for the various diseases using various techniques in future we implement this model and improve the prediction and classification rate of diseases.

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.