International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Predictive Healthcare Informatics using Deep Learning - A Big Data Approach

Dr. K. Purna Chand
Associate Professor, Department of CSE, B V Raju Institute of Technology, Narsapur, Telangana, India

ABSTRACT

Big data technologies are increasingly used for biomedical and health-care informatics research. Large amounts of biological and clinical data have been generated and collected at an exceptional speed and scale. The cost of acquiring and analyzing biomedical data is expected to decrease dramatically with the help of technology upgrades, such as the emergence of new deep learning approaches, the development of novel hardware and software for parallel computing, and the extensive expansion of EHRs. Predictive analysis applications in health care can determine the patients who are at the risk of developing certain conditions such as diabetes, asthma and other lifetime illnesses. The clinical decision support systems incorporate predictive analytics to support medical decision making in the domains like health-care. This paper aims to build a predictive system on health care domain using deep learning approaches on Big data.

Keywords : Big Data, Predictive System, Healthcare, Deep Learning, Clinical

I. INTRODUCTION

Big data applications present new opportunities to discover knowledge and create novel methods to improve the quality of health care information. The applications of big data in health care are a fast-growing field, with many new discoveries and methodologies published in the last five years. Big data applications aiming further on four major biomedical sub disciplines: (1) bioinformatics, (2) clinical informatics, (3) imaging informatics, and (4) public health informatics.

Specifically, in bioinformatics, high-throughput experiments facilitate the research of new genome-wide association studies of diseases, and with clinical informatics, the clinical field benefits from the vast amount of collected patient data for making intelligent decisions. Imaging informatics is now more rapidly integrated with cloud platforms to share medical image data and workflows, and public health informatics leverages big data techniques for predicting and monitoring infectious disease outbreaks, such as Ebola.

For example, the new generation of sequencing technologies enables the processing of billions of DNA sequence data per day, and the application of electronic health records (EHRs) is documenting large amounts of patient data.

Predictive analytics is the branch of the advanced analytics which is used to make predictions about unknown future events. Predictive analytics uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future.

Predictive analytics models capture relationships among many factors to assess risk with a particular set of conditions to assign a score, or weightage. By successfully applying predictive analytics can effectively interpret big data and their specific applications like health informatics. Predictive analysis applications in health care can determine the patients who are at the risk of developing certain conditions such as diabetes, asthma and other lifetime illnesses. The clinical decision support systems incorporate predictive analytics to support medical decision making in the domains like health-care.

Big data analytics helps in discovering valuable decisions by understanding the data patterns and the relationship between them with the help of machine learning algorithms. Deep Learning algorithms extract meaningful abstract representations of the raw data through the use of an hierarchical multi-level machine learning approaches, where in a higher-level more abstract and complex representations are identified. Deep Learning algorithms can be applied for extracting meaningful representations and patterns from any specific domain.

II. Related Work

Big Data Analytics will help organizations in providing the applications of advanced analytic techniques to very large data sets. These cannot be achieved by standard data warehousing applications. These technologies are Hadoop, Mapreduce, Massively Parallel Processing Databases, Search Based Applications, Data-Mining Grids, Distributed File Systems, Distributed Databases and Cloud etc.

Qiu et al. [8] proposed an "Optimal Big Data Sharing Algorithm" to handle the complicated data set in telehealth with cloud techniques. One of the applications is to identify high-risk patients those who often require expensive healthcare.

The choice of the best estimator has been typically tackled in Predictive Modeling Tasks (e.g. classification or regression) [8], [4]. Model selection consists in (i) considering a set of alternative families of models (e.g. linear and nonlinear), (ii) assessing the generalization error of these models (e.g. by means of cross validation or resampling techniques) and (iii) proceeding to selection and/or averaging. This procedure is often long and time consuming since it requires to iterate over different families of models and to search over large hyper parameter spaces. At the same time, its outcome is extremely dependent on the accuracy of the estimation of the generalization error.

However, it is well known that the generalization error (e.g. mean squared error in regression) is only an aggregated measure of estimation accuracy since it integrates the bias, the variance of the model and the noise [8]. This is indeed a major problem since the problem of estimation would be solved if we could have accurate measures of the bias and variance.

As described by Shortliffe and Cimin [89] public health has three core functions, (1) Assessment, (2) Policy Development and (3) Assurance. Among these, assessment is the prerequisite and fundamental function. Assessment primarily involves collecting and analyzing data to track and monitor public health status, thereby providing evidence for decision making and policy development. Assurance is used to validate whether the services offered by health institutions have achieved their initial target goals for increasing public health outcomes; as such, many large public health institutions.

Deep Learning algorithms are shown to perform better at extracting non-local and global relationships and patterns in the data, compared to relatively shallow learning architectures [4].

III. Predictive Mechanism using Machine Learning

The healthcare industries have generated large amount of data generated from record keeping, compliance and patient related data. In today's digital world, it is mandatory that these data should be digitized. To improve the quality of healthcare by minimizing the costs, it's necessary that large volume of data generated should be analyzed effectively to answer new challenges. Similarly government also generates petabytes of data every day. It requires a technology that helps to perform a real time analysis on the enormous data set. This will help the government to provide value added services to the citizens.

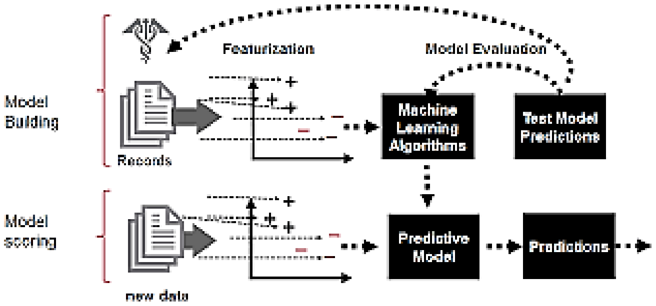

Fig.1. Predictive Mechanism using Machine Learning for Healthcare

This research aims to build a Predictive Mechanism as shown in Fig.1. for tracking Electronic Health Records (EHR) based on any individual's health information.

The following are the main objectives of the proposal.

Health Informatics and Predictions: A Predictive mechanism is needed to build for health information processing and keep track of individual's health information up to date. It involves all the activities like Storing, Accessing, Monitoring and Recommending. It will also provide the needed information to the Patients, Doctors and Hospitals for futuristic examinations.

Solving Emergency Health Issues: This predictive health information is collaborated with insurance agencies or Government schemes to verify the individual bank balances or economical positions in case of emergency situations. It also analyzes history of the patient data in case of unexpected deaths and will help to finds the reasons and can take the precautionary steps further. It also helps to find the patient information immediately in case of accidents and accidental deaths.

Health Check Up and Diagnosis: Based on symptoms present in a person and their matching with Anonymous Patient Database, predicting the person under study has possibility to get an illness, can be used by Ministry of Health to estimate the number of patients and type of illness likely to come in India. It will help the ministry in planning its budget and strategy to control and counter such diseases.

IV. Experimental Analysis

The experiments are made on synthetic data set; it has been generated from the standard PubMed data set.

TABLE-1: Patient Records from Synthetic Dataset

Patient Id	Attributes									
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
3001	1	1	1	115	0	0	1	9	85	0
3002	0	0	0	95	0	0	2	22	85	-9
3003	1	1	1	105	0	0	2	25	85	0
3004	1	1	1	145	0	0	2	26	85	-9
3005	1	1	1	110	0	1	2	26	85	-9
3006	1	1	1	110	0	0	2	28	85	0
3007	1	1	1	110	0	1	2	28	85	0
3008	1	1	1	160	0	0	3	1	85	0
3009	0	0	0	140	0	1	3	1	85	0
3010	1	1	1	125	0	0	3	4	85	0
3011	1	1	1	120	0	1	3	6	85	0
3012	1	1	1	95	0	0	3	6	85	0
3013	1	1	1	120	0	0	3	7	85	0
3014	1	1	1	115	0	0	3	18	85	0
3015	0	0	0	130	0	1	3	19	85	0
3016	1	1	1	115	0	0	3	19	85	0
3017	1	1	1	95	0	0	3	19	85	0
3018	1	1	1	155	0	0	3	20	85	0
3019	1	1	1	125	0	0	3	22	85	0
3020	1	1	1	125	0	0	3	26	85	0

The model is also successful in capturing the correct information for most of the patients and matches with the actual clinical analysis. The results are been captured and compared for each instance at different regions [Table – 2] [16]

TABLE-2: Region Wise Patient Information Analysis

Number of Patients	Region	Patient Information Collected by Theoretical Model (%)	Patient Information described in the Dataset (%)
102	Hungarian	61.9048	47.619
78	Zurich	78.00	71.00
98	Basel	61.9048	47.619
25	Cleveland	41.1765	47.0588

The accumulation results improvement is also been analyzed visually as shown in Fig.2.

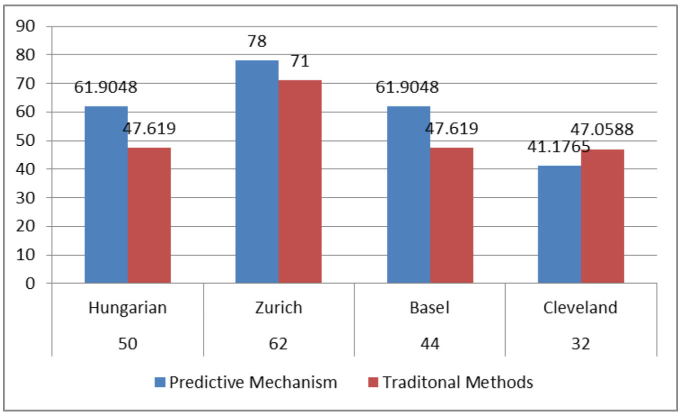

Fig.2.Performance Comparisons between Traditional and Proposed methods

V. CONCLUSION

Motivated by the need for finding the best framework for integrating multiple diagnosis systems along with suitable medical record accumulation framework and storage system, this work analyses the existing methods and frameworks available for the similar purpose. This work compares the benefits and shortcomings of the existing systems. The gained knowledge by this survey helps the researchers to propose the best framework for medical analysis. The outcome of this work is the novel theoretical framework for end-to-end medical information processing. The framework demonstrates the integration with highly popular diagnosis systems such as Coronary Angiography Unit, Stress Testing Unit, Echocardiography Unit, Blood Testing Unit, Cardiac Catheterization Unit, Medical Record Storage with Pre – Processing Unit, Healthcare Insurance System and Healthcare Expert System.

VI. REFERENCES

[1]. Jake Luo, Min Wu et.al. Big Data Application in Biomedical Research and Health Care: A Literature Review Biomedical Informatics Insights 2016:8

[2]. Gianluca Bontempi, Yann-A¨el Le Borgne, Predictive modeling in a big data distributed setting: a scalable bias correction Approach, 978-1-5090-2622-7/16 $31.00 © 2016 IEEE Big DataCongress.2016.17

[3]. Najafabadi et al., Deep learning applications and challenges in big data analytics Journal of Big Data (2015) 2:1, DOI 10.1186/s40537-014-0007-7

[4]. J.Archenaa and E.A.Mary Anita, A Survey Of Big Data Analytics in Healthcare and Government, 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15) doi: 10.1016/j.procs.2015.04.021

[5]. Shortliffe EH, Cimino JJ. Biomedical Informatics. Berlin: Springer; 2014.

[6]. Gerda Claeskens and Nils Lid Hjort. Model selection and model averaging. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, New York, 2008.

[7]. Bengio Y, LeCun Y (2007) Scaling learning algorithms towards, AI. In: Bottou L, Chapelle O, DeCoste D, Weston J (eds). Large Scale Kernel Machines. MIT Press, Cambridge, MA Vol. 34. pp 321–360. http://www.iro.umontreal.ca/~lisa/pointeurs/bengio+lecun_chapter2007.pdf

[8]. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[9]. http://digitalindia.gov.in/content/recent-e-governance-initiatives.

[10]. J. G. Van Manen et al., "The influence of COPD on health-related quality of life independent of the influence of comorbidity," J. Clin. Epidemiol., vol. 56, no. 12, pp. 1177–1184, 2003.

[11]. "Mirror, Mirror on the Wall: An International Update on the Comparative Performance of American Health Care". The Commonwealth Fund. May 15, 2007. Retrieved March 7, 2009.

[12]. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., & Froelicher, V. (1989). International

application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64,304--310.

[13]. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.

[14]. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.

[15]. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.

[16]. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation:Robert Detrano, M.D., Ph.D.