ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

A Survey on Disease Prediction Using Machine Learning Over Big Data from Healthcare Communities

Shraddha Shirsath, S. R. Patil

M. E Student, Department of CE, SIT Lonavala, Pune, India

Professor, Department of CE, SIT Lonavala, Pune, India

ABSTRACT: Big Data has changed the way we manage, analyze and leverage data in any industry. One of the most promising areas where big data can be applied to make a change is healthcare analytics have the potential to reduce costs of treatment, predict outbreaks of epidemics avoid preventable diseases and improve the quality of life in general. Big Data Analysis has recently been applied to help the process of delivering attention and exploration of diseases. However, the rate of adoption and development of research in this space is still hampered by some fundamental problems inherent in the big data paradigm. In this paper, we have optimized automatic learning algorithms for the effective prediction of chronic disease epidemics in frequent disease communities. I propose a new convolutional neural networks (CNN-MDRP) multimodal disease prediction algorithm using structured and unstructured hospital data.

KEYWORDS: Data analytics; Machine Learning; Healthcare data.

I. INTRODUCTION

The concept of big data is not new; however the way it is defined is constantly changing. Various attempts at defining big data essentially characterize it as a collection of data elements whose size, speed, type, and/or complexity require one to seek, adopt, and invent new hardware and software mechanisms for archiving, analyzing and displaying data successfully. Healthcare is a prime example of how the three V's of data first is velocity, second is variety, and third one is volume are an innate aspect of the data it produces. This data is spread among multiple healthcare systems, health insurers, researchers, government entities, and so forth. Furthermore, each of these data repositories is siloed and inherently incapable of providing a platform for global data transparency. To add to the three V's, the veracity of healthcare data is also critical for its meaningful use towards developing translational research.

With the development of big data technology, more attention has been paid to disease prediction from the perspective of big data analysis; various researches have been conducted by selecting the characteristics automatically from a large number of data to improve the accuracy of risk classification rather than the previously selected characteristics. However, those existing work mostly considered structured data. For unstructured data, for example, using convolutional neural network (CNN) to extract text characteristics automatically has already attracted wide attention and also achieved very good results. However, to the best of our knowledge, none of previous work handles medical text data by CNN. Furthermore, there is a large difference between diseases in different regions, primarily because of the diverse climate and living habits in the region. Thus, risk classification based on big data analysis, the following challenges remain: How should the missing data be addressed? How should the main chronic diseases in a certain region and the main characteristics of the disease in the region be determined? How can big data analysis technology be used to analyze the disease and create a better model? To solve these problems, we combine the structured and unstructured data in healthcare field to assess the risk of disease.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

II. RELATED WORK

1] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," Mobile Networks and Applications, vol. 19, no. 2, pp:

In this paper, we review the background and state-of-the-art of big data. We first introduce the general background of big data and review related technologies, such as could computing, Internet of Things, data centers, and Hadoop. We then focus on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. For each phase, we introduce the general background, discuss the technical challenges, and review the latest advances. We finally examine the several representative applications of big data, including enterprise management, Internet of Things, online social networks, medial applications, collective intelligence, and smart grid. These discussions aim to provide a comprehensive overview and big-picture to readers of this exciting area. This survey is concluded with a discussion of open problems and future directions.

2]**M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, "Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System," IEEECommunications, Vol. 55, No. 1, pp. 54–61, Jan. 2017:** With the rapid development of the Internet of Things, cloud computing, and big data, more comprehensive and powerful applications become available. Meanwhile, people pay more attention to higher QoE and QoS in a źterminal- cloudź integrated system. Specifically, both advanced terminal technologies (e.g., smart clothing) and advanced cloud technologies (e.g., big data analytics and cognitive computing in clouds) are expected to provide people with more reliable and intelligent services. Therefore, in this article we propose a Wearable 2.0 healthcare system to improve QoE and QoS of the next generation healthcare system. In the proposed system, washable smart clothing, which consists of sensors, electrodes, and wires, is the critical component to collect users physiological data and receive the analysis results of users health and emotional status provided by cloud-based machine intelligence.

3] W. Yin and H. Sch¨utze, "Convolutional neural network for paraphrase identification." in HLT-NAACL, 2015, pp. 901–911.: We present a new deep learning architecture Bi-CNN-MI for paraphrase identification (PI). Based on the insight that PI requires comparing two sentences on multiple levels of granularity, we learn multigranular sentence representations using convolutional neural network (CNN) and model interaction features at each level. These features are then the input to a logistic classifier for PI. All parameters of the model (for embeddings, convolution and classification) are directly optimized for PI. To address the lack of training data, we pretrain the network in a novel way using a language modeling task. Results on the MSRP corpus surpass that of previous NN competitors.

4] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," Health Affairs, vol. 33, no. 7, pp. 1123–1131, 2014: The US health care system is rapidly adopting electronic health records, which will dramatically increase the quantity of clinical data that are available electronically. Simultaneously, rapid progress has been made in clinical analytics-techniques for analyzing large quantities of data and gleaning new insights from that analysis-which is part of what is known as big data. As a result, there are unprecedented opportunities to use big data to reduce the costs of health care in the United States. We present six use cases-that is, key examples-where some of the clearest opportunities exist to reduce costs through the use of big data: high-cost patients, readmissions, triage, decompensation (when a patient's condition worsens), adverse events, and treatment optimization for diseases affecting multiple organ systems. We discuss the types of insights that are likely to emerge from clinical analytics, the types of data needed to obtain such insights, and the infrastructure-analytics, algorithms, registries, assessment scores, monitoring devices, and so forth-that organizations will need to perform the necessary analyses and to implement changes that will improve care while reducing costs. Our findings have policy implications for regulatory oversight, ways to address privacy concerns, and the support of research on analytics.

5] M. Chen, Y. Ma, J. Song, C. Lai, B. Hu, "Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring,"ACM/Springer Mobile Networks and Applications' Vol. 21, No. 5, pp.825C845, 2016.: Traditional wearable devices have various shortcomings, such as comfortableness for long-term wearing, and insufficient accuracy, etc. Thus, health monitoring through traditional wearable devices is hard to be sustainable. In order to obtain healthcare big data by sustainable health monitoring, we design "Smart Clothing", facilitating unobtrusive collection of various physiological indicators of human body. To provide pervasive intelligence for smart

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

clothing system, mobile healthcare cloud platform is constructed by the use of mobile internet, cloud computing and big data analytics. This paper introduces design details, key technologies and practical implementation methods of smart clothing system. Typical applications powered by smart clothing and big data clouds are presented, such as medical emergency response, emotion care, disease diagnosis, and real-time tactile interaction. Especially, electrocardiograph signals collected by smart clothing are used for mood monitoring and emotion detection. Finally, we highlight some of the design challenges and open issues that still need to be addressed to make smart clothing ubiquitous for a wide range of applications.

6] **Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Healthcps: Healthcare cyber-physical system assisted by cloud and big data:** The advances in information technology have witnessed great progress on healthcare technologies in various domains nowadays. However, these new technologies have also made healthcare data not only much bigger but also much more difficult to handle and process. Moreover, because the data are created from a variety of devices within a short time span, the characteristics of these data are that they are stored in different formats and created quickly, which can, to a large extent, be regarded as a big data problem. To provide a more convenient service and environment of healthcare, 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.

7] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo, "Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review," Age and ageing, vol. 33, no. 2, pp. 122–130, 2004: A small number of significant falls risk factors emerged consistently, despite the heterogeneity of settings namely gait instability, agitated confusion, urinary incontinence/frequency, falls history and prescription of 'culprit' drugs (especially sedative/hypnotics). Simple risk assessment tools constructed of similar variables have been shown to predict falls with sensitivity and specificity in excess of 70%, although validation in a variety of settings and in routine clinical use is lacking. Effective falls interventions in this population may require the use of better-validated risk assessment tools, or alternatively, attention to common reversible falls risk factors in all patients.

8] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, "Localization based on social big data analysis in the vehicular networks," IEEETransactions on Industrial Informatics, 2016: Location-based services, especially for vehicular localization, are an indispensable component of most technologies and applications related to the vehicular networks. However, because of the randomness of the vehicle movement and the complexity of a driving environment, attempts to develop an effective localization solution face certain difficulties. In this paper, an overlapping and hierarchical social clustering model (OHSC) is first designed to classify the vehicles into different social clusters by exploring the social relationship between them. By using the results of the OHSC model, we propose a social-based localization algorithm (SBL) that use location prediction to assist in global localization in the vehicular networks. The experiment results validate the performance of the OHSC model and show that the presented SBL algorithm demonstrates superior localization performance compared with the existing methods.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

III. PROPOSED SYSTEM APPROACH

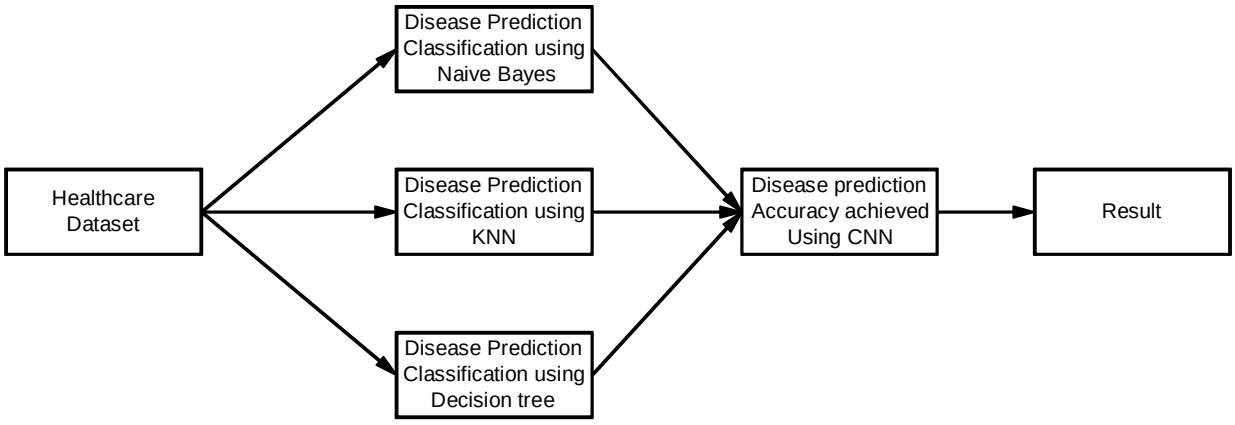

Fig 1. Proposed System Architecture

In Proposed Work, I propose a new convolutional neural network based unimodal disease risk prediction (CNN-UDRP) for structured data and multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data Analytics. With the development of big data analytics technology, more attention has been paid to disease prediction from the perspective of big data analysis, various researches have been conducted by selecting the characteristics automatically from a large number of data to improve the accuracy of risk classification, rather than the previously selected characteristics. However, those existing work mostly considered structured data. For unstructured data, for example, using convolutional neural network (CNN) to extract text characteristics automatically has already attracted wide attention and also achieve.

IV. CONCLUSION

In this paper, propose a machine learning and new convolutional neural network based disease risk prediction algorithm using structured and unstructured data from hospital. In existing work not focused on both data types in the area of healthcare medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNNUDRP) algorithm.

REFERENCES

1] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” Mobile Networks and Applications, 19, no. 2, pp.
2] P. B. Jensen, L. J. Jensen, and S. Brunak, “Mining electronic health records: towards better research applications and clinical care
3] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, “Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System,” IEEECommunications, Vol. 55, No. 1, pp. 54–61, Jan. 2017
4] W. Yin and H. Sch¨utze, “Convolutional neural network for paraphrase identification.” in HLT-NAACL, 2015, pp. 901–911
5] J. Wang, M. Qiu, and B. Guo, “Enabling real-time information service on telehealth system over cloud-based big data platform,” Journal of SystemsArchitecture, vol. 72, pp. 69–79, 2017

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

6] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," Health Affairs, vol. 33, no. 7, pp. 1123–1131, 2014
7] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Healthcps: Healthcare cyber-physical system assisted by cloud and big data
8] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, "Localization based on social big data analysis in the vehicular networks," IEEETransactions on Industrial Informatics, 2016
9] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo, "Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review," Age and ageing, vol. 33, no. 2, pp. 122–130, 2004
10] S. Zhai, K.-h. Chang, R. Zhang, and Z. M. Zhang, "Deepintent: Learning attentions for online advertising with recurrent neural networks