SHIV SHAKTI

International Journal in Multidisciplinary and Academic Research (SSIJMAR)

Vol. 7, No. 2, April 2018 (ISSN 2278 – 5973)

# *Disease Prediction by Machine Learning Over Big Data From Healthcare Communities*

*K.N.Sneha*
Department of CSE
Hitam,Jntuh
Hyderabad,India
kesapragadasneha@gmail.com

*B.Revathi*
Department of CSE
Hitam,Jntuh
Hyderabad,India
revathi.bodagala123@gmail.com

*G.Sumanth*
Department of CSE
Hitam,Jntuh
Hyderabad,India
sumanthreddygopu1234@gmail.com

*A.Harisaikrishna*
Department of CSE
Hitam,Jntuh
Hyderabad,India
harisaikrishna007@gmail.com

*Manikantha Desu M.Tech*
Assistant professor
Department of CSE
Hitam,Jntuh
Hyderabad,India
manikanthadesu@gmail.com

***ABSTRACT*** – With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. In this paper, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data collected from central China in 2013–2015. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a regional chronic disease of cerebral

infarction. We propose a new convolutional neural network based multimodal disease risk prediction algorithm using structured and unstructured data from the 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. Compared with 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 uni-modal disease risk prediction algorithm
.
***INDEX TERMS*** -Big data analytics, machine learning, healthcare.

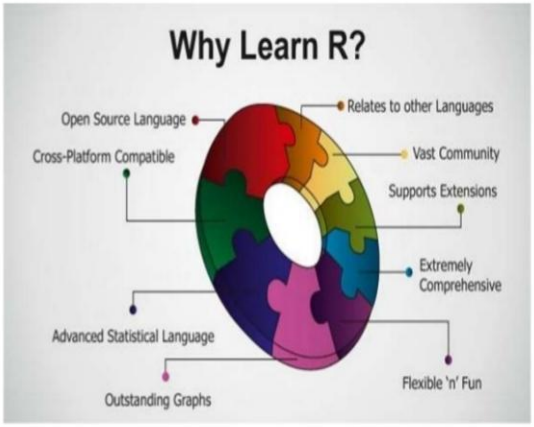
## *INTRODUCTION*

Accordingto a report by McKinsey [1], 50% of Americans have one or more chronic diseases, and 80% of American medical care fee is spent on chronic disease treatment.Withtheimprovement of living standards, the incidence of chronicdisease is increasing.The United States hasspentanaverage of 2.7trillion USD annually on chronic disease treatment. This amount comprises 18% of the entire annual GDP of the United States. The healthcare problem of chronic diseases isalso very important in many other countries. In China, chronic diseases are the maincause of death, according to a Chinese report on nutrition and chronic diseases in 2015, 86.6% of deaths are caused by chronic diseases. Therefore, it is essential to perform risk assessments for chronic diseases. With the growth in medical data [2],collectingelectronic health records (EHR) is increasingly convenient [3 ]With the development of big data analytic technology, moreattention has been paid to disease prediction from the perspective of big data analysis, various researchers have been conducted by selecting the characteristics automatically from a large number of data to improve the accuracy of risk classification [4], [5], rather than the previously selected characteristics. However, those existing work mostly considered structured data. For unstructured data, for example, using the convolutional neural network to extract text characteristics automatically has already attracted wide attention and also achieved very good results. To solve these problems, we combine the structured and unstructured data in the healthcare field to assess the risk of disease. First, we used a latent factor model to reconstruct the missing data from the medical records collected from a hospital. Second, by using statistical knowledge, we could determine the major chronic diseases in the region. Third, to handle structured data, we consult with hospital experts toextract useful features. For unstructured text data, we select the features automatically using CNN algorithm. Finally, we propose a novel CNN-based multimodel disease risk prediction algorithm for structured and unstructured data. The disease risk model is obtained by the combination of structured and unstructured features. Through the experiment, we draw a conclusion that the performance of CNN-MDPR is better than other existing methods.

## *LITERATURE SURVEY*

Programming:R language: R is a protest arranged programming dialect that is a variety of the S dialect and was composed by Ross Ihaka and Robert Gentleman (subsequently the name R), the R Core Development Team, and a multitude of volunteers. R is a programming dialect and

programming condition for measurable processing and designs bolstered by the R Foundation for Statistical Computing. The R dialect is generally utilized among analysts and information mine workers for creating factual programming and information analysis.[6] Polls, reviews of information excavators, and investigations of insightful writing databases demonstrate that R's fame has expanded generously as of late.




## *EXISTING WORK*

The authors used a latent factor model to reconstruct the missing data.The team worked on the regional chronic disease of cerebral infarction.Convolutional Neural Network(CNN)- based multimodel disease riskprediction algorithm using structured and unstructured data from the hospital.

## *PROPOSED WORK*

The development of big data analytictechnology, more attention has been paid to disease prediction.We propose our model in such a way that for the missing data we use linear regression model, which deals with a more accurate data.We use the best prediction approach for analyzing the patient's data in a globalized manner The approach is to gather patients data from hospital apply linear regression and predicting methodologies on the data which in return fetches efficient results.

## *DATASETAND MODELDESCRIPTION*

In this section, we describe the hospital datasets we use in this study. Furthermore, we provide disease risk prediction model and evaluation methods. The hospital data set used in this study contains real-life hospital data, and the data are stored in the data center. To protect the patient's privacy and security, we created a security access mechanism. The data provided by the hospital include EHR, medical image data, and gene data. The inpatient department data is mainly composed of structured and unstructured text data. The structured data includes laboratory data and the patient's basic information such as the patient's age, gender and life habits, etc. While the unstructured text data includes the patient's narration of his/her illness, the doctor's interrogation records, and diagnosis, etc, As shown in Table1

TABLE 1. Item taxonomy in China hospital data.

| Data category | Item | Description |
|---|---|---|
| Structured data | Demographics of the patient | Patient's gender, age, height, weight, etc. |
| | Living habits | Whether the patient smokes, has a genetic history, etc. |
| | Examination items and results | Includes 682 items, such as blood, etc. |
| | Diseases | Patient's disease, such as cerebral infarction, etc. |
| Unstructured text data | Patient's readme illness | Patient's readme illness and medical history |
| | Doctor's records | Doctor's interrogation records |

## *CONCLUSION*

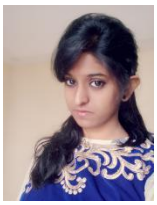
In this paper, we propose an application that helps the users to create a platform for predicting their data.This results in the best growth in terms of all aspects of the organization.As we are implementing with the best prediction methods the accuracy and efficiency of the system will be high and more accurate.

## *REFERENCES*

[1] P. Groves, B. Kayyali, D. Knott, and S. van Kuiken, The'Big Data'Revolution in Healthcare: Accelerating Value and Innovation. USA: Center for US Health System Reform Business Technology Office, 2016.
[2] M. Chen, S. Mao, and Y. Liu, ''Big data: A survey,'' Mobile Netw. Appl., vol. 19, no. 2, pp. 171–209, Apr. 2014.
[3] P. B. Jensen, L. J. Jensen, and S. Brunak, ''Mining electronic health records: Towards better research applications and clinical care,'' Nature Rev. Genet., vol. 13, no. 6, pp. 395–405, 2012.
[4] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, ''A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics,'' IEEE Trans. Intell. Transp. Syst., vol. 16, no. 6, pp. 3033–3049, Dec. 2015.
[5] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, ''Wearable 2.0: Enable human-cloud integration in next generation healthcare system,'' IEEE Commun., vol. 55, no. 1, pp. 54–61, Jan. 2017.
[6] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, ''Smart clothing: Connecting human with clouds and big data for sustainable health monitoring,'' ACM/Springer Mobile Netw. Appl., vol. 21, no. 5, pp. 825–845, 2016.
[7] M. Chen, P. Zhou, and G. Fortino, ''Emotion communication system,'' IEEE Access, vol. 5, pp. 326–337, 2017, doi: 10.1109/ACCESS.2016.2641480.
[8] M. Qiu and E. H.-M. Sha, ''Cost minimization while satisfying hard/soft timing constraints for heterogeneous embedded systems,'' ACM Trans. Design Autom. Electron. Syst., vol. 14, no. 2, p. 25, 2009.
[9] J. Wang, M. Qiu, and B. Guo, ''Enabling real-time information service on telehealth system over cloud-based big data platform,'' J. Syst. Archit., vol. 72, pp. 69–79, Jan. 2017.
[10] 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.
[11] L. Qiu, K. Gai, and M. Qiu, ''Optimal big data sharing approach for telehealth in cloud computing,'' in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2016, pp. 184–189.
[12] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, ''HealthCPS: Healthcare cyber-physical system assisted by cloud and big data,'' IEEE Syst. J., vol. 11, no. 1, pp. 88–95, Mar. 2017.

***AUTHORS***

**Mr.Manikantha Desu**is a M.Tech assistant professor in the Computer Science and Engineering Department at Hyderabad Institute of Technology and Management affiliated by JNTUH.



**K.N.Sneha** is a B.Tech student in the Computer Science and Engineering Department at Hyderabad Institute of Technology and Management affiliated by JNTUH

**B.Revathi** is a B.Tech student in the Computer Science and Engineering Department at Hyderabad Institute of Technology and Management affiliated by JNTUH.

**A.HARI SAI KRISHNA**is a B.Tech student in the Computer Science and Engineering Department at Hyderabad Institute of Technology and Management affiliated by JNTUH.

**G.SUMANTH**is a B.Tech student in the Computer Science and Engineering Department at Hyderabad Institute of Technology and Management affiliated by JNTUH.