A BIG DATA APPROACH FOR ILLNESS PREDICTION TO HEALTHCARE COMMUNITIES

Lohith S Y ,Department Of Studies in Computer Science & Engineering University BDT College Of Engineering (A Constituent College Of VTU,Belagavi),Davanagere,Karnataka
Dr.Mohamed Rafi, Professor , Department Of Computer Science & Engineering, University BDT College Of Engineering (A Constituent College Of VTU,Belagavi),Davanagere,Karnataka

Abstract— with huge information development in biomedical and social insurance groups, precise investigation of restorative information benefits early infection recognition, understanding consideration and group administrations. Be that as it may, the investigation precision is decreased when the nature of restorative information is fragmented. In addition, distinctive areas show one of kind qualities of certain local maladies, which may debilitate the forecast of ailment flare-ups. In this paper, we streamline machine learning calculations for compelling expectation of endless infection episode in malady visit groups. We test the changed forecast models over genuine healing center information gathered from focal China in 2013-2015. To conquer the trouble of inadequate information, we utilize an idle factor model to reproduce the missing information. We probe a provincial constant illness of cerebral dead tissue. We propose another convolutional neural system based multimodal disease risk prediction (CNN-MDRP) calculation utilizing organized and unstructured information from healing center. To the best of our insight, none of the current work concentrated on the two information writes in the zone of restorative huge information examination. Contrasted with a few regular forecast calculations, the expectation exactness of our proposed calculation achieves 94.8% with a joining speed which is speedier than that of the CNN-based unimodal ailment hazard expectation (CNN-UDRP) calculation.

Key words: Data Mining ,Prediction, Risk Factors.

I. INTRODUCTION

According to a report by McKinsey, 50% of Americans have one or more chronic diseases, and 80% of American medical care fee is spent on chronic disease treatment. With the improvement of living standards, the incidence of chronic disease is increasing. The United States has spent an average of 2.7 trillion 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 is also very important in many other countries. In China, chronic diseases are the main cause 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, collecting electronic health records (EHR) is increasingly convenient. Besides, first presented a bio inspired high-performance heterogeneous vehicular telemetric paradigm, such that the collection of mobile Users health related real-time big data can be achieved with the deployment of advanced heterogeneous vehicular networks. proposed a healthcare system using smart clothing for sustainable health monitoring. had thoroughly studied the heterogeneous systems and achieved the best results for cost minimization on tree and simple path cases for heterogeneous systems. Patients' statistical information, test results and disease history are recorded in the EHR, enabling us to identify potential data-centric solutions to reduce the costs of medical case studies. Proposed an efficient flow estimating algorithm for the telehealth cloud system and designed a data coherence protocol for the PHR(Personal Health Record)-based distributed system. Bates et al proposed six applications of big data in the field of healthcare. Qiu et al. proposed an optimal big data sharing algorithm to handle the complicate data set in tele health with cloud techniques. One of the applications is to identify high-risk patients which can be utilized to reduce medical cost since high-risk patients often require expensive healthcare. Moreover, in the first paper proposing healthcare cyber-physical system, it

innovatively brought forward the concept of prediction-based healthcare applications, including health risk assessment. Prediction using traditional disease risk models usually involves a machine learning algorithm (e.g., logistic regression and regression analysis, etc.), and especially a supervised learning algorithm by the use of training data with labels to train the model. In the test set, patients can be classified into groups of either high-risk or low-risk. These models are valuable in clinical situations and are widely studied. However, these schemes have the following characteristics and defects. The data set is typically small, for patients and diseases with specific conditions, the characteristics are selected through experience. However, these pre-selected characteristics maybe not satisfy the changes in the disease and its influencing factors.

II LITERATURE SURVEY:

We outline a dispersed information overseeing structure for telehealth framework, which incorporates BSN, cloud framework, and remote healing center end. By breaking down the highlights of information preparing with medicinal applications, we give a decentralized information lucidness convention to take care of the execution issues by current plan. Our model measures the transfer speed utilization between any hub match in cloud so the transmission capacity can be ascertained in every interim. The exploratory outcomes demonstrate that the data transfer capacity foreseeing mistake is restricted in 10%, which furnishes cloud with an adaptable techniques (4 kinds of anticipating calculations) to evaluate the transmission capacity assets to hubs. Moreover, a contextual analysis demonstrates that our strategy can bolster finding the most proper data transfer capacity evaluating calculation for underlining telehealth applications. In future, we intend to apply our way to deal with some genuine tasks and get the criticism from the coordinated effort with healing facilities. Moreover, we will deal with this subject with further developed approach, for example, concealed Markov model to upgrade the execution transmission capacity assessing.

III.REASEARCH METHDOLOGY

EXISTING SYSTEM:

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 achieved very good results. However, to the best of our knowledge, none of previous work handle Chinese 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. First, we used latent factor model to reconstruct the missing data from the medical records collected from a hospital in central China. 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 to extract useful features. For unstructured text data, we select the features automatically using CNN algorithm. Finally, we propose a novel CNN-based multimodal disease risk prediction (CNN-MDRP) 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.

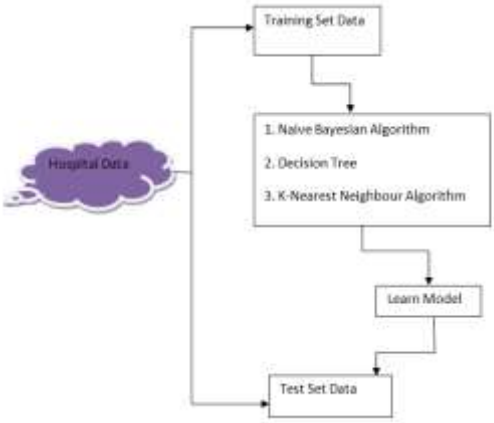
PROPOSED-SYSTEM:

For disease risk modeling, the accuracy of risk prediction depends on the diversity feature of the hospital data, i.e., the better is the feature description of the disease, the higher the accuracy will be. For some simple disease, e.g., hyper lipidemia, only a few features of structured data can get a good description of the disease, resulting in fairly good effect of disease risk prediction. But for a complex disease, such as cerebral infarction mentioned in the paper, only using features of structured data is not a good way to describe the disease. As seen from The corresponding accuracy is low, which is roughly around 50%. Therefore, in this paper, we leverage not only the structured data but also the text data of patients based on the proposed CNN-MDPR algorithm. We find that by combining these two data, the accuracy rate can reach 94.80%, so as to better evaluate the risk of cerebral infarction disease.

We propose a new convolutional neural network based 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. 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.

IV.SYSTEM IMPLEMENTATION

Implementation is the stage of the project where theoretical design turned into a working system. The new system may be a totally new, replacing an existing manual or automated system r it may be a modification to an existing system.



MODULES:

For dataset, according to the different characteristics of the patient and the discussion with doctors, we will focus on the following three datasets to reach a conclusion.

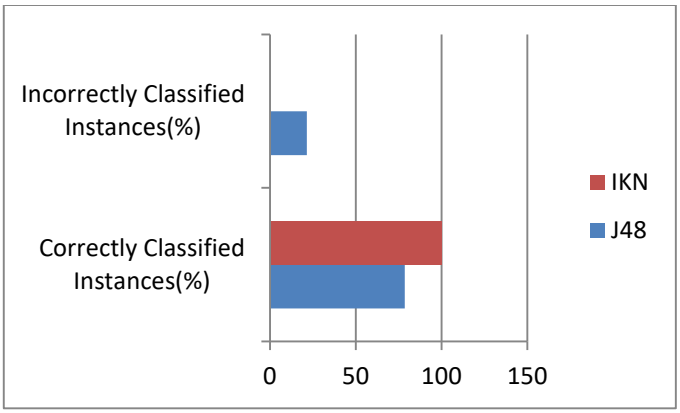
- ✓ Structured data (S-data): use the patient's structured data to predict whether the patient is at high-risk of cerebral infarction.
- ✓ Text data (T-data): use the patient's unstructured text data to predict whether the patient is at high-risk of cerebral infarction.

V.RESULTS

VI.CONCLUSION:

Algorithms	Time Taken To Build a model(sec)	Correctly Classified Instances (%)	F-Measure
J48	0	78.5714	0.842
IKN	0	100	1.00

Table: Comparisons of IKN and J48



I carried out an experiment to find the predictive performance of different classifiers using above training set and class attribute is smoke. I selected six popular classifiers considering their qualitative performance for the experiment. Instance K Nearest classifier is the best in performance. In order to compare the classification performance of six machine learning algorithms, classifiers are applied on same data set and results are compared on the basis of misclassification and correct classification rate and according to experimental results in table.

REFERENCES

[1] P. Groves, B. Kayyali, D. Knott, and S. V. Kuiken, "The'big data'revolution in healthcare: Accelerating value and innovation," 2016.

[2] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," Mobile Networks and Applications, vol. 19, no. 2, pp. 171–209, 2014.

[3] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," Nature Reviews Genetics, 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 Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3033–3049, 2015.

[5] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, "Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System," IEEE Communications, Vol. 55, No. 1, pp. 54–61, Jan. 2017.

[6] 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.

[7] M. Chen, P. Zhou, G. Fortino, "Emotion Communication System," IEEE Access, DOI: 10.1109/ACCESS.2016.2641480, 2016.[8] M. Qiu and E. H.-M. Sha, "Cost minimization while satisfying hard/soft timing constraints for heterogeneous embedded systems," ACM Transactions on Design Automation of Electronic Systems (TODAES), 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," Journal of Systems Architecture, vol. 72, pp. 69–79, 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 tele-health in cloud computing," in Smart Cloud (SmartCloud), IEEE International Conference on. IEEE, 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 Systems Journal, 2015.

[13] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, "Localization based on social big data analysis in the vehicular networks," IEEE Transactions on Industrial Informatics, 2016.

[14] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, "Enhanced fingerprinting and trajectory prediction for iot localization in smart buildings," IEEE Transactions on Automation Science and Engineering,vol. 13, no. 3, pp. 1294–1307, 2016.

[15] 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.