International Journal of Engineering & Technology, 7 (4.39) (2018) 633-636

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Chronic Conjecture and Uncertainty Detection by Machine Learning

J. Jegan Amarnath[1], Pradeep Kumar Sahoo[2], Suresh Anand.M[3,] A.Mamutha Nisha[4], B. Priyadharshini[5],V.Sumitha[6]

[1,2,3]Associate professor, Department of ComputerScience&Engineering, Sri Sai Ram Engineering College, Chennai
[4,5,6] Department of ComputerScience&Engineering, Sri Sai Ram Engineering College, Chennai
**Corresponding Author Email: jegan.cse@sairam.edu.in*

Abstract

The early disease risk detection and appropriate diagnosis for the chronic disease is possible with the help of analysis of the medical data. The analysis accuracy was reduced due to the incompleteness of medical data. Different areas or regions manifest different chronic outbreak which leads to loss of many lives. In this project, we converge machine-learning algorithms for the effective risk prediction of the chronic heart disease and provide early treatment for the disease in disease-frequent communities. The electronic health record collected from the hospital is categorized based on the clinical attributes of the patient and by using the statistical analysis, the overall risk for the heart disease is found .So for reduce the risk of incomplete datas can be use an observable variable to rebuild the absent data which improves the pre-processing effectively. This project uses a new ConvolutionNeuralNetwork (CNN) supported multi-modal disease-risk prediction procedure for the un-structured data, Decision tree algorithm for structured data to find the risk of the chronic disease. Based on the level of the risk, the treatment plan is provided automatically. In the domain of medical big data-analytics, very few existing works focused on both data types.The accuracy of the proposed prediction algorithm will reach more than 94.8% with convergence speed when compared with many other prediction algorithms. This is faster than the convolution neural network (CNN) based uni-modal and existing multimodal disease-risk-prediction method.

Keywords: *chronic; observable variable; decision; convolution; multimodal; big data*

1. Introduction

Disease prediction was observed as a critical topic. Artificial intelligence and machine learning techniques have been already developed to solve this problem. The most lethal disease is the chronic heart disease. The systems will produce large volume/amounts of data which take the form of numbers, text, charts and images. There raises a significant question: “How we can convert data into useful information that can support healthcare persons to make intelligent clinical decisions?” so we have to find or develop a project to predict the heart disease before it occurs.Machine learning algorithms are generally engaged with the Prediction of traditional disease risk models (e.g. Logistic-regression and regression analysis, etc.), and supervised-learning-algorithm using training data. The data set is too small, for patients and diseases with particular conditions and experiences are needed to select the characteristics . For unstructured data,to extract text characteristics we use convolutional neural network (CNN) . The following challenges will remain based on Risk classification which is depends on big data analytics.How the missing data could be addressed? How the main chronic/long term diseases in a certain area and the important characteristics of the disease in the area could be detected?To investigate the disease and create a better model how the big data analytics technologies could be used? To evaluate the risk of disease and to solve these issues, in healthcare domain we combining structured and un-structured datas. In first step to rebuild the outlier from the medial records which is collaborated from the hospitals we used latent factor model . Second step by using statistical data, we can determine the main chronic/long term diseases in the area. In third step we discuss with the hospital experts to retrieve/extract useful features to handle structured data. For unstructured text data, we select those features by using CNN algorithmic technique. The performance every phase, we saying the general background, discuss the technical challenges & analyze the latest advances. The above said discussions focused to provide a complete overview & big-picture to researchers of this exciting area. This study is concluding as a discussion of open issues and future directions. [1] Clinical data recounting the phenotypes & treatment of patients signifies that less utilized data source which has high research potential than the currently used. EHR(electronic-health-records) Mining has a good potential for creates new patient-stratification principles and for enlightening unknown disease correlations. Fusing genetic data with EHR data will give a better understanding of genotype & phenotype relationships. [2] Modern intelligent-transportation systems area vehicle-to-infrastructure and vehicle-to-vehicle communication needs are increasing and new generation of vehicular telematics surely depends on the cooperative wireless-networks. An novel network selection key for the basic tech requirements multimodal communications in diverse/heterogeneous vehicular of used. In this work selection key for the fundamental technological requirement of multimodal communications in diverse/heterogeneous kind vehicular-telematics. To promise the QoS satisfaction of multiple network users , the competent utilization & fair allocation of heterogeneous/diverse network resources in global view. A dynamic & self-adaptive technique for network selection is proposed. While comparing greedy optimization with utility technique, the results show the efficiency of bio-inspired scheme. And also the proposed network selection tech will give us good performance. [3] With the swift development of the IoT,big data and cloud computing, more powerful and widespread applications become available. After sometime users are getting more focus to high QoS and QoE in terminal cloud computing system. Few advanced cloud systems-smart clothing and advanced terminal system-bigdata analytics cognitive computing were looking to users with reliable and intelligent services. The next important system to improve its QoE and QoS is Wearable 2.0 system [4] Many shortcomings are there in the traditional wearable devices, such as insufficient accuracy and uncomfortableness for long-term wearing, etc. This paper says details about design , key tech and implementation technique of actual working smart clothing application. Generally applications are supported by big-data clouds. Like smart clothing, response for real time tactile interact, disease diagnosis and emotion care ,medical emergency and etc can be done. Particularly, electrocardiograph signals received by smart clothing were used for emotion detection and mood observing . Finally, we highlight open issues design challenges should be solved to make a smart- clothing ubiquitous for a easy usage of applications. [5]

2. Datasets Employed

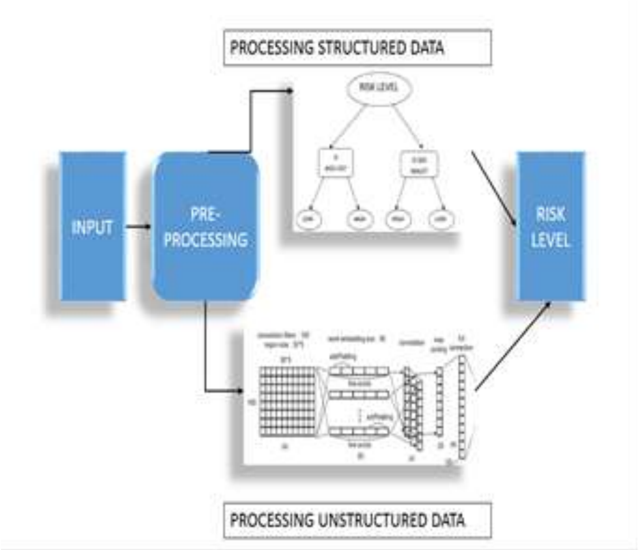
Datasets are the input to this model. More existing works focused on the structured data only because it can be easily retrieved by any search engine and it could be easily arranged in any relational database. Disease risk prediction is made easier in structured data while compared with the unstructured data. In order to provide accuracy in the risk prediction, we must employ unstructured data. The unstructured data is also referred as text data. In this project, we employ the electronic health record collected from Cleveland hospital which mainly focuses on the chronic heart disease. There are nearly 72 attributes. The two types of data are listed below: Structured data: patient’s age, sex, height, weight, etc... Unstructured data: patient’s disease, medical history, health records. Information collected through the electronic health record is highly reliable and provides us prediction accuracy by considering appropriate attributes. It also considers sex ratio because chronic disease affects the men population more while compared with women population. Based on the attributes, the constraints are set in order to evaluate the risk. The model checks whether all the input data collected from the hospital satisfies the constraints or not. Based on the evaluation result, it is found that the patient suffers from the risk of chronic heart disease or not.

3. Model Description

This model mainly comprises of data pre-processing and processing the data through the two algorithms. With the help of this processed result, the risk of the chronic disease is found. The user data collected from the hospital in the form of electronic health record is provided as input to tis model. The model flow is provided in the form of 3 steps.

Step 1: pre-processing the medical data

There are various techniques which are employed in pre-processing. It is very complicated process due to the presence of the incomplete data. The in- complete data prevents us from providing the risk prediction accuracy. The presence of incomplete data may leads to the wrong risk level prediction. The difficulty of incomplete data is overcome by the method of pre-processing. It employs the observable variables directly instead of using the variables compared with other attributes. Applying the observable variables directly leads to more accuracy in prediction.



Step 2: processing the medical data

It involves processing the medical data by two methods based on the type of the data used. The two methods are:

I. Decision tree algorithm for processing the structured data. Decision tree is employed due to its improved accuracy in predicting the result.

II. New CNN multimodal algorithm for processing the unstructured data. This algorithm is used because it could be able to extract the text characteristics and the model can learn by itself by examining the characteristics extracted.

Step 3: Assessing the risk and providing treatment plan

The risk is evaluated by merging the results obtained by processing the medical data by employing two algorithms. Based on the risk level obtained from the previous level, the treatment plan is provided for the patients effectively.

4. Architecture Diagram

The architecture diagram involves three modules. The coding is done by using the java and html programming language. The three basic modules are listed below. They are

I. Data imputation of electronic health record

During examination patient there will a huge number of patient’s data may went missing it is because of human error. And thus ,filling of the incomplete data is needed to be done. Before data assertion,to increase the data quality at first we need to find out uncertain and partial medical data. Then we should modify/delete them .For data pre-processing we make use of data integration.To provide guarantee to data atomicity we can integrate the medical data. Besides employing the latent factor model, we directly apply the observable variables instead of latent variables which increases accuracy and efficiency.

II. Risk prediction for chronic heart disease

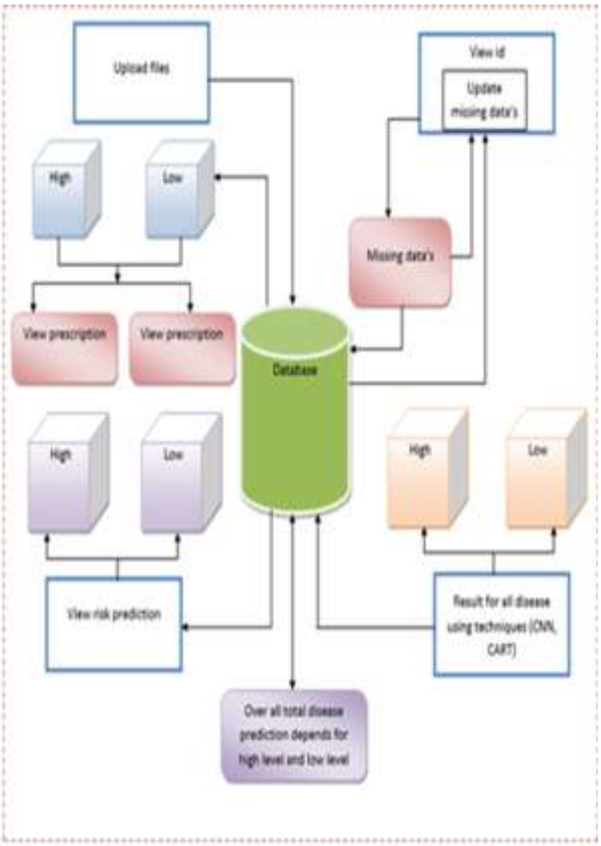
Risk prediction involves processing the structured and unstructured data of the medical data. It involves assessing the risk level for both data types. The algorithms involved are decision tree algorithm for structured data and new multimodal convolution neural network algorithm for unstructured data.

DECISION TREE ALGORITHM:

The decision tree method was the most powerful method for solving classification problems. There are two step available in this technique, one is building a tree and second is applying the tree to the dataset. C4.5, CHAID, CART, ID3 and J48.are well known decision tree algorithms. In this we employ CART (Classification and Regression Technique) algorithm.Object variables is categorical in classification trees . In decision tree generally tree was used to identify a suitable "Class" in the unnamed data , a target object variable would likely come. In regression category tree the target object was continuous. Target object value is identified using a tree. The CART algorithm is formed as a structure of sequence of questions . Answers of structure of sequence will identifies what will be the next question if there might be any questions. The result of above said questions looks like a tree kind of structure where the ends will look like terminal nodes which represent that no more queries.

NEW MULTIMODAL CNN ALGORITHM:

It is a deep feed forward artificial neural network. It involves an input layer, an o/p layer and few hidden layers. These layers comprises of convolution layer, fully connected layer and pooling layer. Each neuron will receive few inputs, performs a dot product and decisively follows it with a non-linearity. The convolution layer involves performing convolution operation. The convolution operation is performed by element wise multiplication of the input and the filter. Finally these multiplications are summed up. It provides an activation map. The output of the convolution-layer is given as the input to the pooling-layer. Then we perform the maximise pooling operation. It involves selecting the maximum value of n elements in a given row. Then the output of max pooling layer is passed as input to the fully connected layer. The fully connected layer involves feature fusion. It fuses the structured data feature and output of the unstructured data max pooling layer.

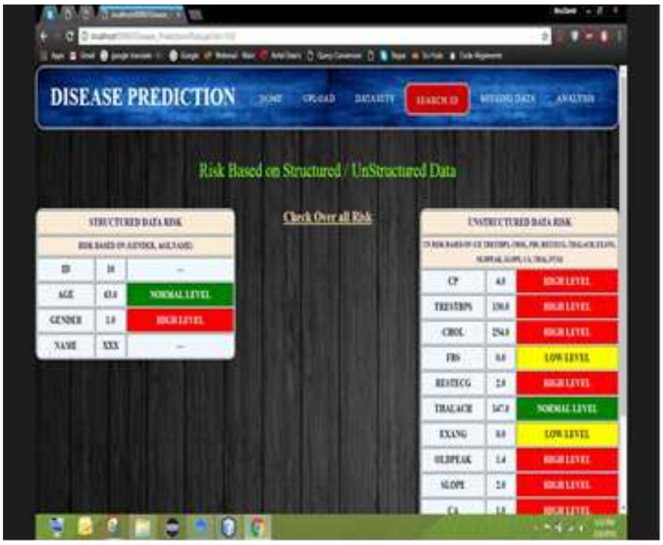


III. Treatment Plan Based on Risk Level and Statistical Analysis

Based on the risk level obtained by processing the unstructured and structured data and with the help of the treatment data provided by the health database, an automatic treatment plan is generated. It helps in providing appropriate treatment for the patients who suffers in high risk and the low risk of the chronic heart disease. Accuracy in providing proper treatment for the patients is achieved by this treatment plan based on the risk level. By the statistical analysis of the electronic health record, chronic risk outbreak of regional disease is found.

5. Working of the system

The below picture describes the risk level of the particular patient based on the structured and unstructured data collected from electronic health record. mise

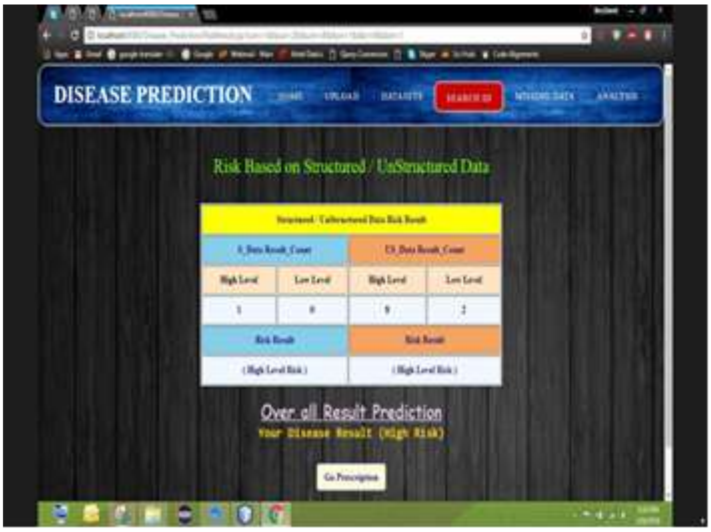


Risk Based on Structured / UnStructured Data

STRUCTURED DATA RISK		
RISK BASED ON GENDER, AGE, NAME		
ID	10	--
AGE	63.0	NORMAL LEVEL
GENDER	1.0	HIGH LEVEL
NAME	XXX	--

UNSTRUCTURED DATA RISK		
[illegible]		
CP	4.0	HIGH LEVEL
TRESTBPS	130.0	HIGH LEVEL
CHOL	254.0	HIGH LEVEL
FBS	0.0	LOW LEVEL
RESTECG	2.0	HIGH LEVEL
THALACH	147.0	NORMAL LEVEL
EXANG	0.0	LOW LEVEL
OLDPEAK	1.4	HIGH LEVEL
SLOPE	2.0	HIGH LEVEL
CA	1.0	HIGH LEVEL

The below picture describes the overall risk level of the particular patient by comparing both the risk level obtained from the unstructured & structured data.



Risk Based on Structured / UnStructured Data

Structured / Unstructured Data Risk Result			
S_Data Result_Count		US_Data Result_Count	
High Level	Low Level	High Level	Low Level
1	0	9	2
Risk Result		Risk Result	
(High Level Risk)		(High Level Risk)	

Over all Result Prediction

Your Disease Result (High Risk)

6. Performance Analysis

The performance could be evaluated in two ways.

I. Run time comparison

The time consumed by the algorithm for the complete execution is less than 1637.2 s. It is quite faster than the existing unimodal and multimodal algorithm.

II. Efficiency calculation

In order to calculate the efficiency, we ought to calculate accuracy, precision, recall, F1 measure. In order to calculate this we need four parameters. They are true +ve, true -ve, false +ve and false -ve. F1 measure provides the entire performance of this model.

- Accuracy = (T P + T N) / (T P + F P + T N + F N)
- Precision = T P / (T P + F P)
- Recall = T P / (T P + F N)
- F1.Measure= (2*precision*recall)/ (precision + recall)

However, the decision tree reaches the efficiency of 63% and the new multimodal algorithm reaches the efficiency more than 95%.

7. Conclusion

The algorithm provides early disease risk prediction, detection of the overall risk outbreak of chronic Regional diseases and proper treatment plan automatically by accurate analysis of both unstructured and structured data. The accuracy prediction of the algorithm will reach more than 94.8% with the convergence speed. It is faster than un-imodal disease risk-prediction algorithms and other machine learning algorithms related to disease prediction.

Future work

The proposed work is quite efficient and consumes less execution time. For future work, efficiency can be further increased by using various machine learning algorithms that employ structured and un-structured data types. The hybrid methodologies can also be implemented in this model. It could be made specialized for many other chronic diseases.

References

[1] M.|Chen, S.Mao, and YLiu,"Big-data:A survey,"Mobile Networks and Applications, 19, no. 2] pp.

[2] PBJensen, LJJensen and SBrunak,"Mining electronic-health records: towards better research applications and clinical care

[3] MChen, YMa, YLi, DWu, YZhang,"Wearable 2.O:Enable Human-Cloud Integration in Next-Generation Heallthcare System," IEEE Communications, Vol. 55, No. 1, pp. 54–61, Jan. 2017

[4] WYin and HSch¨utze, "Convolutional neural-network for paraphrase identification." in HLT-NAACL, 2015, pp. 901–911

[5] JJWang, MQiu, and BGuo,"Enabling real-time information service on tele-health system over cloud-based big-data platform," Journal of Systems Architecture, vol. 72, pp. 69–79, 2017

[6] D. W. Bates, SSaria, LK.Ohno-Machado, A. 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, MQiu, C.-W. Tsai, Hassan, and Alamri, "Healthcps: Healthcare cyber-physical system assisted by cloud and big data

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

[9] D. Oliver, FDaly, FCMartin, and MEMurdo,"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, KChang, Zhang, and Z Zhang, "Deep intent: Learning attentions for online advertising with recurrent neural networks