International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 08, August-2018

ANALYZING BIG DATA TO DETECT THE FUTURE HEALTH CONDITIONS USING CNN-UDRP ALGORITHM

Shrinivas[1], Swaroopa Shastri[2], Mahesh Reddy[3]

[1]*Department of Studies in Computer Applications (MCA), Visvesvaraya Technological University Centre for Post-Graduation Studies, Kalaburagi.*

[2]*Department of Studies in Computer Applications (MCA), Visvesvaraya Technological University Centre for Post-Graduation Studies, Kalaburagi.*

[3]*Department of Studies in Computer Applications (MCA), Visvesvaraya Technological University Centre for Post-Graduation Studies, Kalaburagi.*

Abstract— In therapeutic administrative organization, patient data is collected from the patient's history, authority's notes and wearable body sensors. The valuable stages are the examination social protection arguments and the desire for coming after something in time further prosperity conditions. To separate the composed and unstructured information made through restorative administrations organization structures, a cloud enabled tremendous information explanatory stage is the good way. In the existing system we used the map reduce algorithm which is the creative background for text presentations that manner bulky extents of organized and amorphous records warehoused. The future prediction implemented by using CNN-UDRP algorithm. By using the CNN-UDRP algorithm we can predict the patient unstructured data at last we get the work incorporates of original convolutional neural framework created many modal disorder chance estimate process by the structure and shapeless records after specialist's office. In this work we are focusing on the two data writes in the region restorative tremendous data examination.

Keywords— Big Data, Prediction, CNN-UDRP, Structure, Unstructured.

I. INTRODUCTION

Big data in healthcare is an analytic environment to handle the massive volume of the structured and unstructured patient data. The health data are attributed as big data, which is defined by Vs in terms of Volume, Velocity, Variety, Value, and Veracity. The collected patient data are of peta or zeta bytes, which describe the volume. The velocity is expressed in terms of data arrival rate from the patients. Variety explains the diversified data sets with respect to the structured, semi-structured and unstructured data sets such as clinical reports, EHRs, and radiological images and veracity explains the truthfulness of the data sets with respect to data availability and authenticity. The collected data are transformed into meaningful insights, which explain the value in Vs. It's safe to say there are too many manual processes in medicine. Usually they used to write lab values, diagnoses, and other chart notes on paper. But this was an area in which technology could help improve the workflow and hoped it would also improve patient care. Since then, advancements in electronical medical records have been remarkable, but the information they provide is not much better than the old paper charts they replaced. If technology is to improve care in the future, then the electronic information provided to doctors needs to be enhanced by the power of analytics and machine learning. All hospitals allow Wireless sensor network and mobile networks. Outdoor patients monitoring through Internet of things (IoT) in which patients are equipped with different smart devices such as in-plant pacemaker, ECG (Electrocardiogram), EMG (Electromyography), and motion sensors. These devices collects health related data of patients like body temperature, pulse rate, blood pressure. In advanced healthcare systems, the patient data are collected through wearable devices equipped with different types of sensors. Recently, the advancement in mobile devices such as multi-sensor equipped smart phones are also used as the data collection devices. Hence, colossal amounts of patient data are generated within a hospital network, which needs to be stored and analysed efficiently. Therefore, a cloud computing enabled distributed storage and processing environment is essential to store and process the healthcare data, which can be accessed anywhere and anytime. Now-a-days various data intensive applications are emerged, which need some efficient analytic models. Many stochastic approaches are considered by different authors in the recent past for healthcare parameter analysis. Moreover, the similarity between health parameters of a patient is considered by the physicians for better decisions. Big data analytic is applied in healthcare to identify the clusters of patients, diseases and future predictions with the help of various machine learning tools . In a learning healthcare system, data are analysed and used as insights continuously for patient care. During this process, the patient data are combined with the clinical reports for better suggestions and decisions. 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 in order to successfully store, analyze, and visualize the data. Healthcare is a prime example of how the three Vs of data, velocity (speed of generation of data), variety, and 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 Vs, the veracity of healthcare data is also critical for its meaningful use towards developing translational research.

II. RELATED WORK

In [1] Presently present therapeutic dataset can anticipate crises up to certain level yet can't create better outcome. A superior another for this is huge information. In this paper another novel strategy stands presented which gather results and forecast information from numerous long range interpersonal communication locales like twitter, Google look. Specific classes can be made relying upon maladies. A summed up strategy is examined in this paper which will anticipate the fore-coming strokes of various infections relying upon information assembled by long range interpersonal communication destinations. The Data mining systems can be utilized to extricate the valuable data from enormous information. A Multi rank calculation can be utilized for expectation of asthma. Precision of process relies upon important discovered information sss in enormous information. 80 % precision can be accomplished in expectation of diseases.

In[2] Expanding interest and expenses for social insurance, exacerbated by maturing populaces and an awesome deficiency of specialists, are not kidding concerns around the world. Thusly, this has created an extraordinary measure of inspiration in giving better human services through quicker witted social insurance frameworks. Administration and handling of medicinal services information are trying because of different elements that are innate in the information itself, for example, high-dimensionality, inconsistency and Sparsity. A long stream of research has been proposed to address these issues and give more proficient and adaptable human services frameworks and arrangements. There are primarily two sorts of EHR information, in particular electronic medicinal records (EMR) and sensor information. There are two noteworthy headings of the headway of Big Healthcare Analytics identified with EMR information and sensor information separately.

In [3] In this specific situation, quicker and inconspicuous wellbeing information can be given by methods for inescapable detecting. The utilization of sensors implies the limit of covering expansive times of constant observing without the requirement for performing sporadic screening, which may just speak to a thin photo of the improvement of an infection. In any case, the reality of conveying persistent detecting over a vast populace will bring about a lot of data that requires both on-hub information reflection and appropriated deduction. From a populace level, one's grievous past can give noteworthy understanding into anticipating and keeping a similar episode from happening in others. Last however not the slightest, the legislative approach and direction are required to guarantee security amid information transmission and capacity, and additionally amid resulting information examination assignments.

In [4] A time where we centre around wellbeing and health as opposed to ailment. A time where innovation not just makes mind more secure, more productive, and higher quality, yet in addition where it enhances access to everybody. Today, the innovation exists—and is being used—to break down the quadrillions of bits of information beforehand siloed and disregarded. Torn between the suspicion and the build-up, it is officeholder on every one of us to understand that the capability of advanced wellbeing is as of now being figured it out. We are amidst the carefully empowered wellbeing insurgency, where the cloud associates individuals, information, and machines, and in which investigation has turned into a basic device to conveying higher quality, more productive care. Brains working nearby machines break even with lifesaving care, regardless of whether in the busiest crisis room in northern California or amidst Sub-Saharan Africa. Enormous information. Examination. Manmade brainpower. Machine and profound Learning. This is the eventual fate of pharmaceutical. Also, in the event that we are available to grasping it, what's to come is presently.

In [5] Enormous information investigation is a promising right bearing which is in its early stages for the social insurance area. Social insurance is an information rich area. As an ever increasing number of information is being gathered, there will expand interest for huge information investigation. Disentangling the "Enormous Data" related complexities can give numerous experiences about settling on the correct choices at the opportune time for the patients. Proficiently using the giant medicinal services information vaults can yield some prompt returns as far as patient results and bringing down care costs. Information with more complexities continues developing in human services therefore prompting more open doors for huge information examination.

In [6] The paper has recorded a few information investigation instruments and methods that have been utilized to enhance human services execution in numerous regions, for example, restorative tasks, reports, basic leadership, and expectation and counteractive action framework. Additionally, the precise survey has demonstrated an intriguing statistic of fields of production, inquire about methodologies, and also sketched out a portion of the conceivable reasons and issues related with social insurance information examination, in light of topographical conveyance topic. This work takes care of give a decent begin for additionally contemplates in human services parts as it exhibits the positive effect of rising between Information Technology field and Healthcare divisions.

In [7] With the guarantees of prescient investigation in huge information, and the utilization of machine learning calculations, foreseeing future is never again a troublesome assignment, particularly for medication in light of the fact that anticipating maladies and envisioning the cure wound up conceivable. In this paper we will display a review on the

advancement of huge information in human services framework, and we will apply a learning calculation on an arrangement of medicinal information. The goal is to foresee incessant kidney maladies by utilizing Decision Tree (C4.5) calculation. The classifier utilized demonstrated its execution in anticipating with best outcomes as far as exactness and least execution time.

In [8] There are a few difficulties in preparing tolerant records which manages assortment of organized and unstructured organization. Inciting BDA in to Healthcare (HBDA) will manage touchy patient driven data for the most part in unstructured configuration involving solutions, reports, information from imaging framework, and so forth., the difficulties will be overwhelmed by enormous information with improved proficiency in getting and putting away of information. In this undertaking, dataset alike Electronic Medical Records (EMR) created from various restorative gadgets and versatile applications will be instigated into Mongo DB utilizing Hadoop system with Improvised handling strategy to enhance result of preparing understanding records.

In [9] Utilizing organized and unstructured information from healing facility it utilizes Machine Learning Decision Tree calculation and Map Reduce calculation. To the best of our insight in the territory of restorative enormous information examination none of the current work concentrated on the two information writes. Contrasted with a few run of the mill assess calculations, the estimation precision of our proposed calculation achieves 94.8% with a meeting speed which is speedier than that of the CNN based unimodal sickness chance expectation (CNN-UDRP) algorithm.

In [10] We propose a PGAR (individual quality examination record) and PHR (individual social insurance record)-based individual medicinal services benefit. We actualized this enormous information examination based medicinal services stage for more exact illness forecast and counteractive action. At the point when the individual keen social insurance ICT benefit is utilized, it altogether decreases individual human services costs in healing facilities. The creators set up a social insurance stage giving customized medicinal services administrations in view of individual quality investigation and wellbeing examination record for better personal satisfaction and health.

III. SYSTEM DESIGN

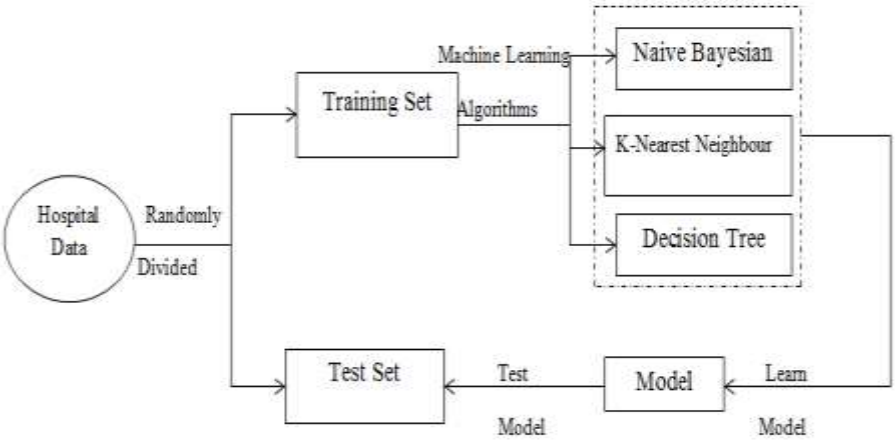

Figure System Architecture Design

In the above system architecture shows the complete architecture of the praposed system. In this paper the hospital data is represented into two sets of records , the test set and training set. The praposed work focuses on the training set of the data. In this training data to predict the future health condition.

IV. METHODOLOGY

A. *Data Imputation:*

For patient's examination data, there is a large number of missing data due to human error. Thus, we need to fill the structured data. Before data imputation, we first identify uncertain or incomplete medical data and then modify or delete them to improve the data quality.

B. *CNN-Based Unimodal Disease Risk Prediction (CNN-UDRP) Algorithm:*

For the processing of medical text data, we utilize CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm which can be divided into the following five steps.

1. *Representation of Text Data*:

As for each word in the medical text, we use the distributed representation of Word Embedding in natural language processing, i.e. the text is represented in the form of vector. In this experiment, each word will be represented as a Rd -dimensional vector, where d=0 . Thus, a text including nwords can be represented as $T=(t,t,\cdots,tn)$, $T\in Rd\times n$.

2. *Convolution Layer of Text CNN:*

Every time we choose s word. In other words, we choose two words from the front and back of each word vector t′i in the text, i.e. use the row vector as the representation, to consist a 0×=0 row vector,

i.e. si=(t′i−,t′i−,t′i,t′i+,t′i+) for s , s , sn− and sn , we adopt an zero vector to fill. The selected weight matrix W∈R00×0 is weight matrix W includes 00 convolution filters and the size of each filter regions is 0. Perform convolution operation on W and si(i=,,⋯,n) . Specific calculation progress is that:hi,j=f(Wi·sj+b)() where i=,,⋯,00 , j=,,⋯,n . Wi is the i-th row of weight matrix. · is the dot product (a sum over element-wise multiplications), b∈R00 is a bias term, and f(·) is an activation function (in this experiment, we use tanh-function as activation function). Thus we can get a 00×n feature graph
h=(hi,j)00×n

3. Pool Layer of Text CNN

Taking the output of convolution layer as the input of pooling layer, we use the max pooling (-max pooling) operation. Select the max value of the n elements of each row in feature graph matrix
h:hj=max≤i≤nhi,j,j=,,⋯,00()

After max pooling, we obtain 00× features h. The reason of choosing max pooling operation is that the role of every word in the text is not completely equal; by maximum pooling we can choose the elements which play key role in the text. In spite of different length of the input training set samples, the text is converted into a fixed length vector after convolution layer and pooling layer, for example, in this experiment, after convolution and pooling, we get 00 features of the text.

4. Full Connection Layer of Text CNN :

Pooling layer is connected with a fully connected neural network the specific calculation process is that:
h=Wh+b() where h is the value of the full connection layer, W and b is the corresponding weights and deviation.

5. *CNN Classifier* **:**

The full connection layer links to a classifier, for the classifier, we choose a softmax classifier.

V. RESULT AND DISCUSSION

In this section we are going to deal with the results of our paper. The main concept of our paper is going to predict the patient diseases risk using the CNN-URDP algorithm. In the existing system we used the MapReduce frame work for the patient predicting the diseases risk but this frame work which only store the patient real time data but we cannot store the patient historical diseases symptoms in the existing system. When we processing, it requires a lots of data to be shuffled over the network MR is not suitable for a large number of short on-line transactions. In the propose work we used a novel CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm for structured and unstructured data. The disease risk model is obtained by the combination of structured and unstructured features the performance of CNN-UDRP is better than other existing methods. This algorithm speed and accuracy is more for the risk classification.

VI. CONCLUSIONS

In this paper the work incorporates of original convolution neural framework created many modal disorder chance estimate process by the structure and shapeless records after specialist's office. In this work we are focusing on the two data writes in the region restorative tremendous data examination. The work can be improved by including heart diseases and cancer severity. In further work we can include SPARK algorithm that consist of built in machine learning procedures and graph examinationprocedures that stay used to complete in matching. This algorithm does not consume more storage space to store the patient details.

References

[1] M Archana Bakare "PREDICTION OF DISEASES USING BIG DATA ANALYSIS".
[2] Chonho Lee "BIG HEALTHCARE DATA ANALYTICS: CHALLENGES AND APPLICATIONS".
[3] Javier Andreu-Perez "BIG DATA FOR HEALTH".
[4] Makary MA "BIG DATA, ANALYTICS & ARTIFICIAL INTELLIGENCE"
[5] Jimeng Sun "BIG DATA ANALYTICS FOR HEALTHCARE",
[6] Mohammad Ahmad Alkhatib "ANALYSIS OF RESEARCH IN HEALTHCARE DATA ANALYTICS",
[7] Basma Boukenze1*, Hajar Mousannif2 and Abdelkrim Haqiq3 "PREDICTIVE ANALYTICS IN HEALTHCARE SYSTEM USING DATA MINING TECHNIQUES",
[8] Antony Basco J and Senthilkumar N C "REAL-TIME ANALYSIS OF HEALTHCARE USING BIG DATA ANALYTICS",
[9] Vinitha S, Sweetlin S, Vinusha H and Sajini S "DISEASE PREDICTION USING MACHINE LEARNING OVER BIG DATA ",
[10] RiWen1, Seung Min Yang1 and Byung Mun Lee 2 "HEALTHCARE PLATFORM AND BIG DATA ANALYSIS BASED PERSONAL FITNESS HEALTHCARE SERVICE MODEL LONG ",