ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Kidney Failures Prediction Model for eHealth Platform Integration Using RBFNN.

Vepuri Baby Sneha, Sadum Mounika, and S Gowri*.

Department of Information Technology, Faculty of Computing, Sathyabama University, Chennai, Tamil Nadu, India.

ABSTRACT

Big data is an encompassing difficult or complex large datasets to process on traditional large scale data processing. The main confront of big data processing incorporates the extraction of significant data, from a high dimensionality of a wide assortment of medicinal information by empowering examination, disclosure and elucidation. This data is a useful tool for better understanding of disease and to formulate predictive models in various fields and support numerous tasks, like treatment evaluation, triage, and monitoring. In this work, based on a predictive model using the Distributed radial basis function neural network (DRBFNN) to aiming the estimation of kidney failures is presented. The proposed method exposed it appropriateness to sustain patient & health care professionals (HCP) on clinical decisions and practices.

Keywords: Prediction model, eHealth, Kidney Failure, Platform integration.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

The branch of computer science, which is more aggressively and proficiently involved in medical sciences is Artificial Intelligence. With the assistance of Artificial intelligence numerous Clinical Decision Support Systems (DSS) were built. These system which are in favor are now extensively used in hospitals and clinics. These systems have evidenced to be actual valuable for the medical specialists and even the patients for making the accurate decisions. For the development of these system diverse methods are employed. The technique of accumulating the input data and to extract output information's is not alike in all methods approaches. Any computer program that benefits specialists in making a clinical decision originates under the domain of clinical DSS. A significant characteristic of the Artificial Intelligence is that it can afford the establishment as well as employment of the clinical information.

Patients and health care professionals (HCP) ought to be asked to occasionally associate with the framework so as either to acquire human services data, for example, drug and clinical direction, or to keep up their patients' restorative information up and coming. These clinical data sets have a large or even endless data volume, which makes its computation and management exhaust significant resources. In addition to the large volume, the sources of big data sets can be very diverse and originated on different devices and platforms, which means that these data represents unstructured information and is not typically easy for traditional databases to analyze it.

The rest of the paper is organized as follows. Section 2 gives a brief description of works done which are related to this field, Section 3 shows the proposed system overview, Section 4 described the proposed methodology in detail and Section 5 concludes the work analysis.

RELATED WORKS

Being a vast topic of the present scenario, it is known that Bigdata has a great influence on data, taking this into consideration many works were proposed on prediction and analysis models and data management models impacting the field of medical stream in various scenarios, a few of which are described in this section.

A study big data driven solutions to predict the 30-day risk of readmission for congestive heart failure (CHF) events is done by K. Zolfaghar et al. [3]. They first extracted the factors which are useful from The National Inpatient Dataset (NIS) after which they amplified it with their patient dataset from Multicare Health System (MHS). Using this integrated dataset which they are extracted they have developed scalable data mining models to predict risk of readmission. They demonstrated the effectiveness and efficiency of the open-source predictive modeling framework which they have used by describing the results from various modeling algorithms that they have tested and compared its performance against a baseline non-distributed, non-parallel, non-integrated lesser data outcomes earlier published to demonstrate comparable accurateness over massive number of records.

Mokhairi Makhtar et al., [4] had proposed an original structure The Predictive Toxicology Markup Language (PTML) which offers a representation scheme for predictive toxicology data and models produced by data mining tools, they even showed how this representation offers possibilities to equate models by likeness with their Distance Models Comparison technique. In their presentation they have just calculated PTML distance and reported as of now, as it is an ongoing work. For all these representation they have used Predictive Toxicology as an application field to demonstrate their approach to represent predictive models linked to data for DMG.

Sudha Ram et al., [6] introduced a novel method, which utilizes multiple data sources for predicting the number of asthma-related emergency department (ED) appointments in a specific area. For evaluation of their proposed methodology they have collected data from Twitter, Google search interests and environmental sensor data. Their evaluations helped in predicting the number of asthma ED visits based on near-real-time environmental and social media information with around 70% precision, which in turn would be helpful for the public health surveillance, ED readiness and targeted patient interventions.

A technique was established by Yang Xie et al., [7] which uses a large-scale health insurance claims data, to forecast the number of hospitalization days in a populace. They have used a regression decision tree

algorithm, along with insurance claim data from 242075 people over three years, to deliver forecasts of number of days in hospital in the third year, according to the hospital admissions and procedure claims data. Their proposed methodology had performed well in the general population as well as in subpopulations. According to their evaluation their results had improved estimates over two conventional standard approaches (forecasting a constant number of days for a customer individually and by means of the number of days in the hospital of the preceding year as the prediction for the subsequent year). Their analysis of two subpopulations-namely an old people aged 63 years or older in 2011 and patients hospitalized for at least one day in the preceding year-revealed that the medical evidence (e.g., diagnosis codes) contributed more to predictions for these two subpopulations, in comparison to the population as a whole.

The study made by Yun Chen and Hui Yang [8] focuses on the prediction of mortality rates in Intensive Care Units (ICU), which uses patient-specific healthcare recordings. They stated that postsurgical monitoring in ICU leads to enormous datasets with exceptional properties, e.g., variable heterogeneity, patient heterogeneity and time asyncronization. To cope with the difficulties in ICU datasets, they have established the postsurgical DSS with a series of analytical tools, taking in even data categorization, data pre-processing, feature extraction, feature selection, and predictive modeling. Experimental results using real-world ICU data from 4000 subjects in the database showed that their proposed data-driven methodology performed better than the traditional approaches.

A novel clustering model for the classification of the breast cancer cell have been proposed by S. Gowri et al., [1], in which the clustering was derived from the Crawler's mechanism, this clustering process had shown a better performance than the existing algorithms, resulting in better cluster formation, which in turn helps in better retrieval results due to better organization of data. They have even integrated their clustering algorithm with many other functionalities like preprocessing and segregation of metadata in order to help with the fastness and choosing of relevance in higher level due to this integration [2].

Going through many proposal a case study which is based on a predictive model which utilizes the radial basis function neural network (RBFNN) collaborating with a filtering method targeting the estimation of electrocardiogram (ECG) waveform presented by N. Pombo et al., [5] has trigger an ideology for the proposed methodology stated in our paper. From their study, they have designed their framework to be suitable for supporting health care professionals on clinical decisions and practices. There proposed work had helped in the initiation of this proposal for Kidney Failures Prediction.

PROPOSED SYSTEM OVERVIEW

The model introduced in this work utilizes a prescient model taking into account the Distributed Radial Basis Function Neural Network (DRBFNN) to anticipate the kidney disappointments. The proposed algorithm (Linear classifiers, Feature vectors) DRBFNN (Distributed RBFNN) calculations fundamentally in two classes (1) elucidating or unsupervised learning (i.e., grouping, affiliation, rundown) and (2) prescient or regulated learning (i.e., order, relapse). Notwithstanding, they are missing more profound knowledge into the suitability of the calculations for taking care of the extraordinary attributes of the sensor information in wellbeing observing frameworks. This model uncovered to be precise and suitable when connected on medicinal services and wellbeing setting. The Figure 1 below depicts the Overall system architecture.

PROPOSED METHODOLOGY

Modules:

- Anomaly Detection
- Prediction
- Diagnosis/Decision Making
- Other Big Data Mining Tasks
    a. Preprocessing
    b. Feature Extraction/Selection

ISSN: 0975-8585

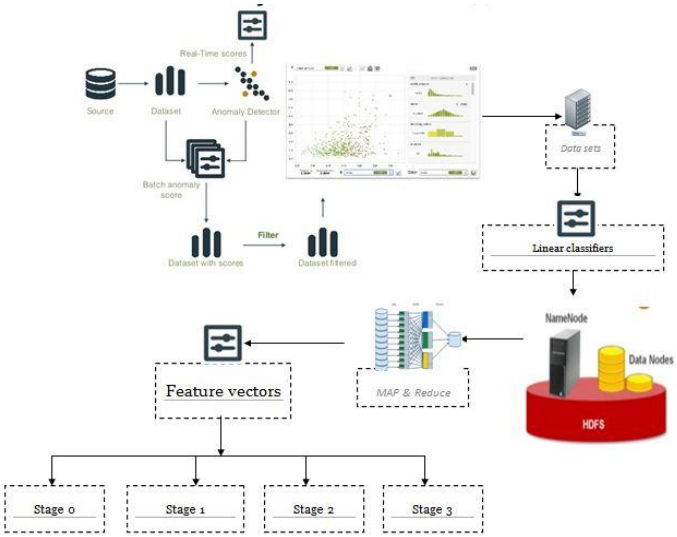

Figure 1: Overall Architecture Diagram

Anomaly Detection

Anomalies are examples in information that don't fit in with a very much characterized idea of typical conduct. Anomaly detection refers to the method of discovering outlines in data that do not follow to predicted performance. These non-conforming patterns are referred to as anomalies, outliers, discordant observations, exceptions, aberrations, surprises, peculiarities or contaminants according to the application domains. Of these, the two terms which are utilized very frequently in the context of anomaly detection are anomalies and outliers; sometimes interchangeable. Due to the fact that anomalies in data translate to significant actionable data in a wide variety of application domains the anomaly detection technique is most significant.

Anomaly detection discovers broad utilization in a wide mixed bag of uses, for example, misrepresentation location for charge cards, protection or human services, interruption discovery for digital security, deficiency identification in wellbeing discriminating frameworks, and military reconnaissance for foe exercises.

Prediction

Anomaly happen because of various reasons, such as i) a processing component is given insufficient resources, ii) the processing capacity of the component is reduced than the input data rate, or iii) the component comprises software bugs (e.g., memory leak, buffer management error). Foresee the most widely recognized irregularity in information handling groups. An information stream handling application commonly comprises of an arrangement of preparing segments. Every part acknowledges info information from its upstream component(s) and produces yield information for its downstream component(s). A bottleneck shows up in the circulated application when the information line of a segment achieves its maximum cutoff.

Diagnosis/decision making

The decision-making / Diagnosis is supported by which inquire different things from the result, and then make various decisions. The decision must be robust and accurate enough that they efficiently produce results to determine concealed things and make decisions. The decision part is significant since any small error in decision-making can degrade the efficiency of the whole analysis. Finally, that any application can utilizes those decisions at real time to do their development. The applications can be any business software, general purpose community software, or other social networks, which require the findings (i.e., decision-making / Diagnosis).

Other Big Data Mining Tasks

Preprocessing

Preprocessing is a procedure to clean and transform the data before it is passed to other modeling procedure. Data cleaning involves removing the noise and outliers in the data set, while data transforming tries to reduce the irrelevant number of inputs, i.e., reducing dimensionality of the input space. As data cleaning is very straightforward of applying standard process of zero mean and unit variance, the concentration is put on data transforming. The following subsections introduce the common data transformation methods.

Feature Extraction/Selection

Feature extraction includes diminishing the measure of assets needed to depict a substantial arrangement of information. At the point when performing examination of complex information one of the significant issues originates from the quantity of variables included. Examination with a substantial number of variables by and large obliges a lot of memory and calculation force or an arrangement calculation which over fits the preparation test and sums up inadequately to new examples. Feature extraction is a common term for approaches of constructing mixtures of the variables to get about these difficulties while still describing the data with adequate correctness.

Feature selection method is that the information contains numerous elements that are either repetitive or immaterial, and can in this way be uprooted without causing much loss of data. Redundant or irrelevant features are two particular ideas, since one significant component may be excess in the vicinity of another pertinent element with which it is unequivocally associated.

CONCLUSION

The proposed system when analysis, has been observed that it is useful for handling large volumes of clinical dataset for storing and retrieve it. The comparison of patient dataset results may promise to yield knowledge about patient records and to find the root cause of the kidney failures. In future it will be used to predict blood clots in veins, brain tumour, weather forecasting, heart attack, Diabetics patient dataset. In Phase II, the complexity evolves due to data reduction can be overcome with the use of Map Reduce. So work of phase two is to implement DRBNN algorithm for big data reduction. Accurate results are produced by predictive modelling and it increases efficient handling of large volumes of patient dataset.

ACKNOWLEDGEMENT

I thank Sathyabama University for providing me with various resources and unconditional support for carrying out this work.

REFERENCES

[1] Gowri S., Anandha Mala Dr. G. S., Mathivanan Dr. G. Classification of Breast Cancer Cells using Novel DPSC Algorithm. Journal of Pure and Applied Microbiology, Vol. 9, No. 2, 2015, 1395-1400.

[2] Gowri S., Anandha Mala Dr. G. S. Strategic enhancement of the collaborative framework for novelty in retrieval from digital textual data corpus by deploying DPSC and RBWM algorithms for forensic analysis. Journal of engineering and research, Vol. 3, No. 4, 2015, 43-58.

[3] Kiyana Zolfaghar, Naren Meadem, Ankur Teredesai, Senjuti Basu Roy Brian Muckian. Big Data Solutions for Predicting Risk-of-Readmission for Congestive Heart Failure Patients. IEEE International Conference on Big Data. 2013, 64-71.

[4] Mokhairi Makhtar, Daniel C. Neagu, Mick Ridley. Predicting Predictive Model Representation and Comparison: Towards Data and Predictive Models Governance. Proceedings of the 10th UK Workshop on Computational Intelligence UKCI. 2010, 1-6.

[5] Nuno Pombo, Nuno Garcia, Virginie Felizardo, Kouamana Bousson. Big Data Reduction Using RBFNN: A Predictive Model for ECG Waveform for eHealth platform integration. IEEE HEALTHCOM 2014- The 2nd International Workshop on Service Science for e- Health SSH. 2014, 66-70.

[6] Sudha Ram, Wenli Zhang, Max Williams, Yolande Pengetnze. Predicting Asthma-Related Emergency Department Visits Using Big Data. IEEE Journal of Biomedical and Health Informatics. Vol. 19, No. 4, 2015, 1216-1223.

[7] Yang Xie, Günter Schreier, David C.W. Chang, Sandra Neubauer, Ying Liu, Stephen J. Redmond, Nigel H. Lovell. Predicting Days in Hospital Using Health Insurance Claims. IEEE Journal of Biomedical and Health Informatics. Vol. 19, No. 14, 2015, 2168-2194.

[8] Yun Chen, Hui Yang. Heterogeneous Postsurgical Data Analytics for Predictive Modeling of Mortality Risks in Intensive Care Unit. IEEE journal of Biomedical and Health Informatics, 2014, 4910-4313.