Proceedings of the International Conference on Intelligent Computing and Communication for Smart World (I2C2SW-2018)
14-15 December 2018

Review on Big Data Analytics in Mortality Prediction

A.Vidya
Department of CSE
PSNA college of Engineering and Technology
Dindigul, India.
vidya14051996@gmail.com

D. Shanthi
Department of CSE
PSNA college of Engineering and Technology
Dindigul, India
hodcse@psnacet.edu.in

P. Gokulakrishnan
Department of CSE
PSNA college of Engineering and Technology
Dindigul, India.

K. Manivannan
Department of CSE
PSNA college of Engineering and Technology
Dindigul, India.

***Abstract*— Big data analytics is the process of examining large and varied data sets i.e., big data to uncover hidden patterns, unknown correlations. ICU is for the patients with severe diseases and most of the patients need support from medical equipments for their survival and need to be closely monitored. Recently, big data analysis has been proved to be useful in the ICU mortality prediction to discover patterns hidden in the large clinical data. The significant aim is to predict the fatality rate among the ICU patients of large data by analyzing their medical parameters periodically. Prediction in the large ICU dataset becomes challenging due to various factors such as Time Asynchronization, Imbalanced data and high dimensionality. To overcome these challenges a technique called Principal component analysis is used through Singular value decomposition (SVD-PCA). SVD is a general matrix decomposition method which reduces the high dimensional data sets into fewer dimensions and has greater classification accuracy in both noisy and noiseless environments. In the proposed method the performance of the prediction is improved by the machine learning technique called Extreme learning machine to reach the best performance. Analysis tools are used to give statistical reports which gives easy prediction of mortality. This prediction model can help the healthcare institutions to easily analyse the mortality of patients and proceeds further for the welfare of patients.**

Keywords*— Big data, Mortality Prediction, ELM classifier

I. INTRODUCTION

Big data analytics is the process of exploring varied data sets like to uncover the hidden patterns and unknown relations to improve the welfare of organizations and institutions with enormous amount of data sets. Big data comprises of mainly three types data such as Structured data, Unstructured data and Semi-structured data. The data is of large variety, volume and velocity. As data sets are growing rapidly it is difficult for the traditional techniques to handle this much of data. Therefore, big data analytics have been widely used to overcome the challenges of traditional techniques. With big data, analyzing of insights leads to better business decisions and strategies. The main characteristics of big data are volume i.e. the quantity and the size of data, velocity i.e. processing speed of the data generation, variety i.e. different types and nature of data and veracity i.e. variation of the data quality captured.

An intensive care unit (ICU) is a place where patients with serious diseases will be admitted. Most of the patients needs support from equipment like the medical ventilator to maintain normal body functions and need to be periodically and closely monitored. More recently, the big data techniques have been proved to be useful in the ICU clinical data. It is difficult for human experts to extract such large amounts of information by looking at them manually. The volume of clinical data is increasing every single day and big data techniques can handle huge voluminous amount of data and automatically predicts using certain techniques.

Mortality is the condition of susceptible to death and the mortality data is used for the assessment of public health problems. Mortality Prediction is mainly carried out to analyze which age group has the highest death rate, demographic characteristics of deceased patients, disease prevention and evaluation of medical research with improved medical procedures, early warning of the mortality and to get more knowledge after the preliminary treatment of the patients .Here, large scale ICU database is analyzed for the identification of some critical diseases such as Heart attack, Chronic lung Diseases, Neurological Diseases etc..

ICU mortality prediction plays a significant role for improving the patient survival. Mortality prediction is based on mortality rates which gives a clear outcome of the survival or death of a patient. A death rate measure is a death occurred in a given interval of time. Death prediction is used to derive the risk measure and indicates the severity of the disease. The types of mortality rates in ICU differs such as age-specific mortality rate, infant mortality rate, crude mortality rate. For mortality prediction, Mortality rate is important which is easy to obtain and the obtained information is accurate. It provides a starting point for many epidemiological studies. However mortality prediction suffers from various challenges such as High dimensionality, Time asynchronization and imbalance distribution. High dimensionality refers to a large number of variables which leads to information redundancy, imbalance distribution is more samples from one class and less samples from another class and time asynchronization refers to missing data.

Big data presents better Service delivery and productivity growth of a large number of productive sectors. Although Big data is used in enormous applications it finds difficult to deal with the validation and security of the data as well as dealing with the data growth. It faces difficulty in parsing, interpreting and also faces analytical and technical challenges, Privacy, Security, Data access and in Sharing of

information. In the field of Healthcare it becomes more challenging in cleaning, storage, Querying, Reporting, Updating, Sharing and Visualization and the advantages includes High-Quality care, Early Intervention and Fraud detection. Various number of techniques has been used in big data analytics like machine learning, artificial neural networks etc. In big data analytics, analysis of data sets can find new associations to prevent diseases, to fight against crimes and so on. The architecture of Big data comprises of Map reduce which provides a parallel processing model and reduces the huge data sets into smaller ones. With map reduce data are splitted and processed in a parallel manner. Big data also comprises of Distributed parallel architecture which distributes data across multiple servers to be executed in parallel environments.

The rest of this brief introduces the big data importance in Patient's health in section I. The related works of this brief are considered in Section II. Then, in Section III, the proposed work is presented with the preliminary descriptions. Section IV presents experimental Results and performance analysis to illustrate the productiveness of the hardware security approach. Finally, the conclusions are summarized in Section V.

II. RELATED WORKS

Guang-Bin Huang et al [1] presents as follows:

- ELM offers a combined learning platform which can be applied in regression and classification applications rightly.
- ELM has lighter optimization limitations when compared with Least Square-Support Vector Machine and Parallelizing SVM in the point of optimization;
- ELM when compared with LS-SVM and PSVM in theory has low computational complexity.
- In theory, ELM can estimate any chosen functions which are continuous and classify any number of regions.

Wang, M. D et al [2] is generally based on multi-scaled biomedical data for delivery and development; clinical identification with information based on images for decision support; informations on ICU for better patient health care; offers health care for chronically illed patients. Here it emphasizes the major limitations in health sector which includes feature extraction, decision making, feature selection.

Barajas et al [3] gives a outline for Intensive Care Patients (ICU) with the failure of physiological subsystems. With different types of data. So we develop a latent model to calculate the probability of severity of failure of the physiological subsystems. In order to further model the latent state we use Generalized linear Dynamic model which updates the patient health data periodically. Some health data of patients cannot be obtained numerically, in such cases we use statistical topical models and a noun phrase extraction to get the patient health data.

Sheth et al [4] establishes infected patients with rheumatoid arthritis has a low severity level of lethality when compared to other patients without this disease. It also gives us the information based on the length of stay of patients with rheumatoid arthritis in Intensive Care Units (ICU).

Yun Chen et al [5] points on the prediction of death rates in Intensive Care Units (ICU) using recordings of the patients specifically. It mainly focuses on the risks and diseases that patients faces after each operations along with the prediction of mortality rate of patients who undergoes different critical operations. We use big data analytics to analyze with the various parameters of different patients with different diseases. It also proposes the methods to faces the challenges of variable heterogeneity and high dimensionality of patient's health care data.

Masud et al [6] recognizes and identifies the challenges which arises when using the lab tests results for the prediction of mortality of patients in Intensive Care Units (ICU) and gives us the suitable solutions to overcome these limitations. Therefore, a proposed technique called Feature vector compaction is used. This technique is totally different from feature selection and feature extraction in which features are discarded but in Feature vector compaction we reduce the vacuum.

Pratama et al [7] suggests a meta cognitive Extreme Learning Machine named as Evolving Type-2 ELM(eT2ELM). This ELM type is mainly composed of three basic constituents such as what to learn, when to learn and how to learn. It uses Type-2 Fuzzy Neural Network which uses type-2 multrivariate Gaussian function in the hidden layer.

Ouyang et al [8] presents a solution for the challenges faced in feature extraction and classification. It proposes a numerically image extraction for handwritten based on the techniques such as Histogram of Oriented Gradient (HOG) and Random Forest (RF). With these techniques it improves the accuracy in the recognition and reduces the higher dimensions of the input for classification.

Shi et al [9] proposes a combination of Extreme Learning Machine (ELM) and Map Reduce for further advantages. The proposed method uses the map reduce framework for parallelizing work of large number of data and uses a technique called Distributed Extreme Learning Machine based on Map Reduce Technique (DELMM). It presents a classification of remote sensing images which are hyperspectral which combines the morphological profiles.

Chen et al [10] presents a model to overcome the challenges of scalability and efficiency. A technique of Map Reduce accelerated Extreme Learning Machine called ELM-Map Reduce is used for vast amount of classification of data. ELM-Map Reduce offers van algorithm called ELM ensemble learning algorithm with elevated stability and greater efficiency.

III . PROPOSED SCHEMES

Fig.1 shows the general architecture of the ICU mortality prediction model. Just as mentioned above, the properties of health data like patient's heart rate,

temperature, weight, pressure and also the urine outcome makes the work more complicated. Health data often have characteristics of high dimensionality, imbalance and time asynchronization. In order to overcome these challenges we have to do certain preprocessing methods. The proposed ICU mortality prediction model is made up of four core components, that is data preprocessing, feature extraction, feature selection, and predictive modeling which are performed to improve patient specific prediction of in-hospital mortality [11].

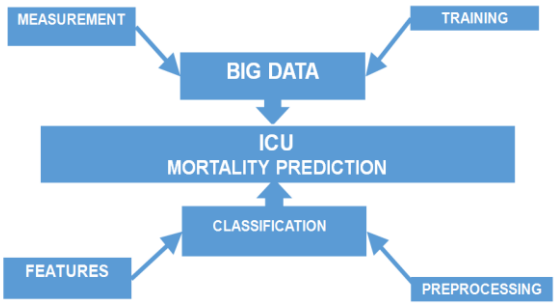

Fig.1. General architecture of ICU morality prediction

A. Big Data preprocessing

The data used in this work comes from biomedical signal research resources and Collected a measured physiological values for mortality prediction. But data in real is incomplete with missing values, redundant attributes, imbalanced and with high dimensions. So, it is important to model the data with certain pre processing steps where all the bugs are removed. The pre processing steps includes

- Data cleaning
- Data integration
- Missing values imputation
- Data Transformation
- Data Reduction

B. Feature extraction

Feature Extraction is carried out to extricate useful information from a wide variety of data. The initial step is to reduce the amount of data by discounting the majority samples. The next step is to extract the features that are useful to carry out the implementation. Principal component Analysis using Singular Value Decomposition which is a general matrix decomposition method. SVD reduces the high dimensional data sets into fewer dimensions which has greater classification accuracy in both noisy and noiseless environments than the modified cost-sensitive principal component analysis (MCSPCA).

C. Classifier

Extreme Learning Machine (ELM) classifier is used for health data and randomly generates the weights between the input and hidden layer. For the estimation of the output layer, ELM introduces a mathematical model to identify the suitable parameters with iteration like the traditional methods. ELM offers a solution for solving the least square optimization problem which is transformed into generalized inverse of matrix.

The training data set is given by

$$\sum_{i=1}^{L} \beta_i g(x_j) = \sum_{i=1}^{L} \beta_i g(w_i x_j + b_i) = t_j, j = 1,2\mathrm{L}\ N \quad \textbf{(1)}$$

D. Extreme Learning Machine (ELM) classifier

1) Preface:

ELM theories Figure 2 exhibits that hidden neurons are significant but need not to be remain as important for many applications such as regression, classification, feature learning and clustering. In ELM theory, these hidden neurons including hundreds of biological neurons can be a piece wise continuous neurons with the correct mathematical modeling not known to the humans. ELM theory was established in biological systems in 2013.

2) Proposed ELM classifier:

The different ELM models are proposed to further improve the quality of the weak classifiers in ensemble learning. Three mechanisms are outlined for enhancing the quality of the weak classifiers. They are as follows:

- First, individually every network of ELM is trained individually which confirms that the hidden nodes which were randomly generated of all ELM networks are not dependent with each other.
- Second, for every ELM network individually ELM model is designed to train it, with different composition of activation functions and hidden node numbers.
- Third, In order to filter out the weak classifiers with low accuracy we use ELM filtering mechanism. This filtering out prevents the roughly fitting weak classifiers from affecting the ensemble accuracy sternly.

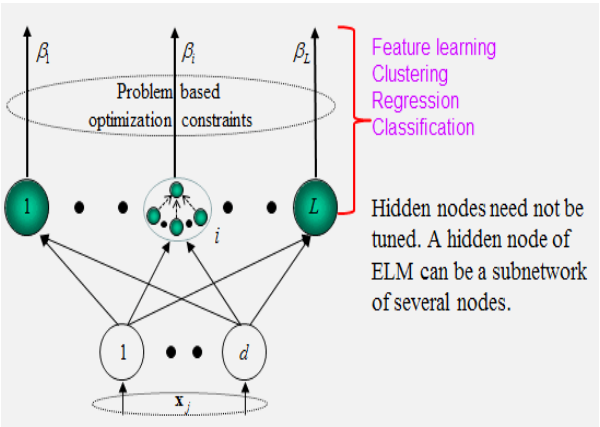

Fig.2. ELM theory*3) Experimental Results:*

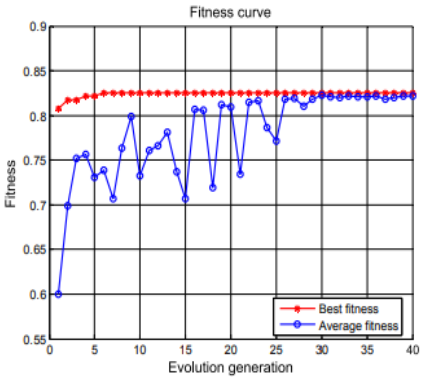

Fig. 3. Simulation result of ELM classifier

3)Experimental Results:

The experimental result of the proposed ELM classifier of various iterations is shown as a graph of fitness. Figure 3 indicates a graph in which the searching speed of the proposed ELM classifier is shown. This represents that the ELM classifier has fair testing speed in minimum number of iterations and the best fitness is 0.82 and this value therefore indicates that our proposed ELM classifier algorithm has good generalization performance which also includes good performance in testing, training and running time.

IV. CONCLUSION

In this paper, with Mortality Prediction, the death rates, survival rates and the severity rates of the patients are analyzed. As the project is done for ICU patients, large amount of data is needed for classification. Therefore, big data is a clear platform for a vast amount of ICU data of patients in a hospital. In this work, we have taken certain analysis parameters of patients like heart rate, Blood pressure, Urine output, Sugar level, creatinine level etc., For a clear classification of these different rates a machine learning technique called Extreme learning machine classifier is used.

The major outcomes of this project are higher accuracy of classification and reduced training time of datasets. This research provides a clear outcome measure whether a patient is died or not and gives the demographic assessment of the diseased patients from which early warning of mortality can be given to the specific location. However, this mortality prediction faces many challenges such as high dimensionality and imbalance distribution which are solved using a series of analytical methods like pre-processing, feature selection, feature extraction and predictive modeling.

REFERENCES

[1] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification,"IEEE, 2012.

[2] Wang, M. D. "Biomedical Big Data Analytics for Patient-Centric and Outcome-Driven Precision Health". IEEE,2015. Barajas, K. C., & Akella, R. (2015).

[3] "Prediction of physiological subsystem failure and its impact in the prediction of patient mortality". IEEE,2015.

[4] Sheth, M. B., Chahin, A., Mark, R., & Markuzon, N. (2014). "Protective effects of rheumatoid arthritis in septic ICU patients".IEEE,2014.

[5] Yun Chen, & Hui Yang. (2014). Heterogeneous postsurgical data analytics for predictive modeling of mortality risks in intensive care units". IEEE,2014.

[6] Masud, M. M., & Harahsheh, A. R. A. (2016). "Mortality prediction of ICU patients using lab test data by feature vector compaction & classification".IEEE, 2016.

[7] Pratama, M., Lu, J., & Zhang, G. (2015)." A Novel Meta-Cognitive Extreme Learning Machine to Learning from Data Streams". IEEE,.2015.

[8] Ouyang, J. (2017). "Combining Extreme Learning Machine, RF and HOG for Feature Extraction". IEEE .2017.

[9] Shi, J., & Ku, J. (2017). "Spectral-spatial classification of hyperspectral image using Distributed Extreme Learning Machine with MapReduce".IEEE,2017.

[10] Chen, J., Zheng, G., & Chen, H. (2013). "ELM-MapReduce: MapReduce accelerated extreme learning machine for big spatial data analysis".IEEE,2013.

[11] Jiankang Liu,"Mortality prediction based on imbalanced high dimensional ICU big data".Elseiver, 2018