Lehality Prediction of Highly Disproportionate Data of ICU Deceased using Extreme Learning Machine

A.Vidya, D.Shanthi, P.Gokulakrishnan, K.Manivannan

Abstract-: ***Big data in mortality prediction is rationed with enormous amount of dataset of patients admitted in ICU for the healthcare providers to clarify and interpret about the status of the patients. However, it is difficult to process these large datasets for which big data is used. Mortality prediction of patients admitted in ICU faces many challenges such as imbalance distribution, high dimensionality etc. This paper focuses on overcoming the challenges that arise during the prediction of mortality of ICU patients through pre-processing, feature selection, feature extraction, and classification have been developed. The performance of classifiers has been affected by the high dimensional and unbalanced data of patients. Therefore, a classifier called Extreme Learning Machine has been used for a generalized performance of the classification.***

In order to predict the rate of mortality for the patients admitted in the ICU by solving the challenges using various methods and tools. For this work, the dataset is collected from a rural hospital that provides medical services in the particular locality. To evaluate the performance of the proposed model, various algorithms have been used and the obtained results are compared. The proposed approach is implemented and experimented in MATLAB software. Various statistical reports are obtained as results and verified. From the results and comparison, it is noticed that the proposed method outperforms than other approaches.

Keywords-Mortality prediction, Extreme Learning Machine, MATLAB.

I. INTRODUCTION

Patients admitted in Intensive Care Units(ICU) suffer from serious illness and injuries and they are at higher risk of dying. Therefore predicting the mortality of ICU is significant in the upcoming years for which we use various techniques. Mortality rates in ICU differs according to the various diseases and their severity of illness. The examination of the severity of diseases can be carried out by analysing their physiological variables such as Blood pressure, cardiac output, Haemoglobin, insulin level etc., clinical data and their locality considerations. In general, the assessment of mortality prediction helps in allocating the resources to various patients and also determines the level of care towards the patients and interaction with their families. Mortality prediction of ICU patients is mainly used in the research and administration of critically ill patients. This research also compares the average severity of illness among the group of patients admitted in the particular hospital. To calculate the classification accuracy the predicted mortality rates can be compared with the observed mortality rates for additional information of the prediction.

Since the medical data generated in ICU section of any hospital is huge in volume. So, that the medical data is considered and measured as bigdata. Thus, prediction in medical bigdata, it requires a bigdata analytics tool. The need for big data is day by day increasing as the data are becoming more complex and high volume.

Thinking about medical bigdata, it is known that the sources of data (IoT medical devices) are different, whereas it makes variations in the data in terms of volume, types, velocity and value it hold. 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 AI, Machine learning and Deep Learning methods. In big data analytics, analysis of data sets can find new associations to prevent diseases, to fight against crimes and so on.

Contribution of the Paper

Mortality in ICU is common but it is important to predict the mortality to the welfare of the patients. Many challenges arises during predicting the mortality rate which are met through certain techniques such as Extreme Learning machines, Singular Value Decomposition principal component analysis. Predicting the mortality is significant as it is useful in analyzing the severity of the illness of the patients and their causes. Mortality prediction gives us a clear view of the death rates and survival rates of the ICU deceased patients. Therefore it is important to predict the mortality by taking the patient attributes such as blood pressure, pulse rate, heartbeat, cholesterol, blood sugar level and urine output.

Revised Manuscript Received on July 18, 2019.

A.Vidya, CSE department, PSNACET, Dindigul, Tamilnadu , India
D.Shanthi, Professor&Head , CSE department, PSNACET, Dindigul, Tamilnadu , India
P.Gokulakrishnan, Associate Professor , CSE department, PSNACET, Dindigul, Tamilnadu , India
K.Manivannan, Associate Professor , CSE department, PSNACET, Dindigul, Tamilnadu , India

II. RELATED WORK

Related works helps to identify the problem statement and it provides more key idea in design and developing proposed approach. For example, in [1], Jianfeng Xu et al. discussed about various methods have been focused on predicting the rate of morality in medical data, which is gathered from ICU patients. Certain features of the patient records like high-dimension, irregular, and inequality characteristics of the ICU data degrades the performance of the prediction accuracy. Improving the performance of ICU mortality prediction is the advantage of this paper. Uncertain sampling and high dimension of data is the challenge to be solved in the future work.

In [2], Silva et al. proposed a novel prediction scheme for obtaining the rate of morality for ICU patients. Various built in software like APACHE, SAPS, and MPM are greatly applied in calculating the variations of population in terms of various medical parameters like medication, caring, surgery and other methods and medicines. The primary objective CinC/PhysioNet (2012) is to design and implement various methods for predicting the rate of morality, which is patient-specific. Early identification of patients with elevated risks is the main advantage of this work. However, it has a disadvantage of Frequent missing and occasionally incorrectly recorded observation.

In [3], Yun Chen et al. proposed a data analytics method to predict the rate of morality over Heterogenous Postsurgical data. It obtained the risk factors of the patient data. It mainly focused on data collection at ICU. It predicts and prepare reports as patient specific method. Surgical data has certain main properties like healthcare recordings, ICU monitoring, and decision support system. Managing and predicting the health condition on ICU data faces lot of challenges, need to be rectified.

In [4], the author Kuang Fangjun et al. described about multiclass SVM method for integrating the PCA kernels and improved PSO algorithm for IDS. The MSVM method employs the actions of the attack and KPCA is used for preprocessing the MSVM where it does dimensionality reduction. Hence the time taken for training process is reduced. Also, MSVM increased the speed of the prediction process and reduces the computational time. The accuracy of the IDS detection rate is also increased by improving the prediction accuracy.

In [5] Yangyang Ding et al. proposed a novel morality prediction model for healthcare industry. The data collected from ICU department in the hospital is used for experiment the proposed method. In the proposed method to increase the accuracy of the prediction by integrating Just in time and Extreme Learning method. The proposed method motivated to increase the prediction accuracy. From the experiment it is obtained that the proposed JITL-ELM method outperforms in prediction process.

In [6] author Diego Gachet et al. proposed a distributed bigdata method for gathering and analyzing medical sensor data. The author used very common methods like sorting, filtering and processing the data comes from medical sensors especially from ICU department. Apart from prediction the proposed method also focused on reducing various complexities regarding time, comparison and computational.

In [7] author Guang-Bin Huan et al. described various stages such as, ELEM incorporated with regression classification, ELM versus SVM, ELM versus LS-SVM, ELM versus PSVM and finally theoretical estimation for choosing and classifying regions.

Limitations

From the above discussion, it is obtained that various earlier research works focused on analysing ICU patient data in medical industry faced various challenges such as less accuracy, massive amount of data, comparison complexity, time complexity and obtaining the morality accurately. Hence this paper mainly motivated to increase the accuracy of achieving rate of morality.

III. DISCUSSION & RESULTS

The objective of this proposed system is to study the mortality rates in Intensive Care Unit (ICU) patients which also includes classifying the death rate and survival rates of patients with different diseases by collecting the original dataset of the ICU patients and to provide a new mortality prediction algorithm for ICU patients using an optimized machine learning approach. Fig.1 shows the general architecture of the ICU mortality prediction model. The prediction of mortality faces many challenges such as dimensionality reduction and imbalanced nature of the datasets which increases the computational complexity which further also reduces the accuracy of the computation of the predicted model. In addition to this, the imbalance nature of the ICU data makes the prediction even more challenging which affects the performance of the classifiers.

The vast ICU data is preprocessed to remove the incomplete data through various steps such as Data cleaning, Data integration, Missing values imputation, Data Transformation and Data Reduction. The preprocessed data is then extracted for suitable features using the algorithm called Principal component Analysis using Singular Value Decomposition which is a general matrix decomposition method which has larger accuracy than the traditional modified cost-sensitive principal component analysis (MCSPCA). After extracting the features are classified using a machine learning algorithm called Extreme Learning Machine (ELM) which arbitrarily develops and assigns weights between the input layer and the hidden layer.

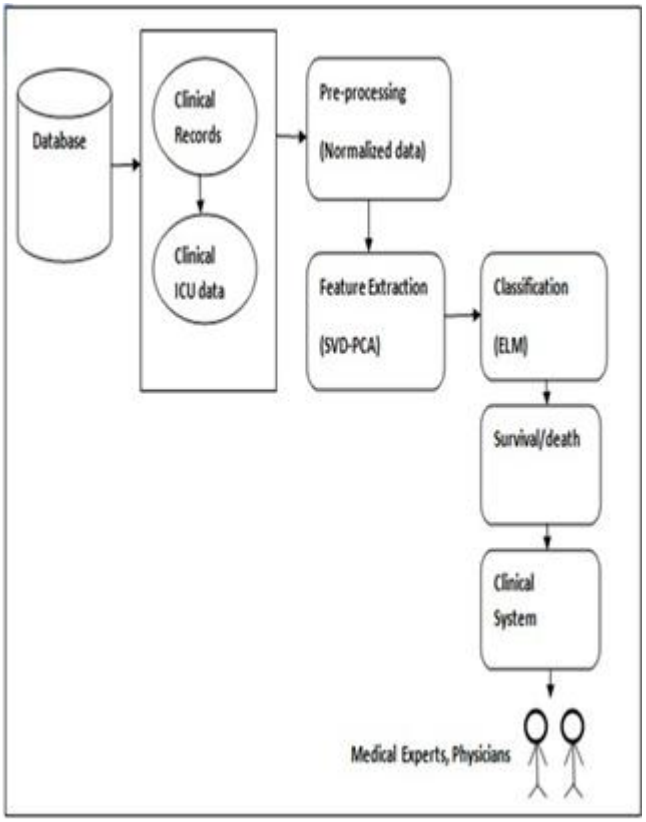

Figure1. General architecture of ICU morality prediction

The training data set is calculated as follows using the Extreme Learning Machine classifier:

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

Then the classified results are taken for the prediction of survival and death rates of the patients which are stored in a database. The results from the classifier gives a statistical report to various physicians, medical officers which they use for the welfare and care of the patients.

3.1. ALGORITHM OF PROPOSED SYSTEM

The algorithm of the proposed system is as follows:

1) The medical dataset is collected with required features.2)The medical dataset is preprocessed using the suitable tool and the redundant data are cleaned. 3) The features are extracted from the preprocessed data using suitable technique and the required features are selected for further classification. 4)The extracted features and the preprocessed data are given as input to the Extreme Learning Machine classifier(ELM) which classifies whether the patient is dead or alive measured by the particular features like BP, Sugar, Cholesterol, etc 5)By parallel processing the Mapreduce framework reduces the data. 6) The data is stored in Excel sheets.7)The Matlab tool generates a statistical report of survival and death rates.8)The predictive model is generated for the lethality rates of Intensive Care Unit (ICU) patients.9)The mortality rates of the admitted patients in ICU are identified by the predictive model.

Figure2. Flow of Proposed system

3.2.EXTREME LEARNING MACHINE (ELM) CLASSIFIER

ELM theory is mainly used for the classification of the ICU patients to classify whether the patient is alive or dead. The theory is 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. Figure 3 represents the structure of ELM classifier.

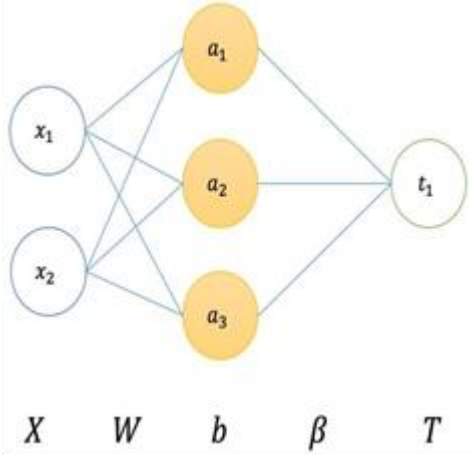

Figure3. Structure of ELM theory

C. ADVANTAGES OF PROPOSED SYSTEM

1. Reduced training time
2. Easy understanding
3. Random assigning of weights
4. More accuracy in reduced time
5. Less complex and increased generalization

IV. CONCLUSION

In this work, the accessibility of dataset of Intensive Care Units (ICU) patients information on predicting the mortality can be useful to governments to improve the health programs of the patients. With this data, the death rates, survival rates and the severity rates of the patients are analyzed. 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 early warning of mortality can be given. 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 modelling. This study is one of the principal activities to examine and support the medical services and the event of well being.

V. ACKNOWLEDGMENT

The authors would like to thank the Management of PSNA College of Engineering and Technology, Dindigul for the support to complete this project inside the campus by providing Lab facility.

VI. REFERENCES

[1] I Silva et al., Predicting In-Hospital Mortality of ICU Patients: The PhysioNet / Computing in Cardiology Challenge 2012 - IEEE, 2012.

[2] Yun Chen et al .,Heterogeneous Postsurgical Data Analytics for Predictive Modelling of Mortality Risks in Intensive Care Units -IEEE , 2014

[3] Kuang Fangjun et al .,A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection- Springer , 2014

[4] Yangyang Ding et al., Mortality prediction for ICU patients using just-in-time learning and extreme learning machine-IEEE, 2016

[5] Diego Gachet et al.,Distributed Big Data Techniques for Health Sensor Information Processing,2016

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

[7] Nikhlesh Lakhmani, Big data and healthcare analytics, (2017).

[8] Fang Ruogu, et al., Computational health informatics in the big data age: a survey, (2016)

[9] Braytee Ali, Wei Liu, Paul Kennedy, A cost- sensitive learning strategy for feature extraction from imbalanced data, International Conference on Neural Information Processing, Springer International Publishing, 2016.

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