Cancer Prediction Using ANN Classification Based on pH Parameter and Sodium ions

Dr. Rakesh Kumar Godi

Professor department of Information Science and Engineering PDCEK

*Abstract-*By considering the growth of population in the present situation, there should be a high level and very effective healthcare analysis is the basic criteria. It should be the same situation at both home and hospitals. There are plenty of researches are done on human health level condition, research in data mining for health analysis is most significant in order to provide good services to the patient's health. These researches are becoming a significant opportunity for improving the quality of health care services. In this regard, this paper presents an innovative approach for human health analysis using pH value of blood and urine samples. ow a day it is difficult for the doctor to detect the hypercalcaemia condition which may lead to different cancer disease. So to analysis the patient electrolyte value the doctor come to know that the volume of the electrolytes and made the judgment whether the person is normal or not. To analyze the patient health the blood and urine reports of the patients with different conditions are gathered as input parameters. Both blood and urine values are fused together by concatenation data fusion technique to form as single data. Dimension reduction is also implemented to convert the high dimensional data samples into the low dimensional space, so that the intensive information contained in the data is protected. As the dimensionality of data gets reduced, it encourage improving the robustness of the classifier and decreases computational complexity. In this paper, reduction is carried out by Principal Component Analysis on fused data. Based on feature level and on the basis of PCA technique Feature extraction is applied to extract important features from considered data. For classification A classifier is implemented as it provides best computational speed and accuracy. In accordance with the classification outcome overall health analysis results are estimated.

Keywords: Blood and Urine samples, PCA Reduction, Feature Extraction, Data Fusion and A Classifier

I. INTRODUCTION

Today's most impacting factor which is affecting human life is health imbalance. As there are several health care industries are accomplished for health monitoring. Then also according to survey there are plenty of patients are suffering from severe cardiopulmonary (heat related) and respiratory block and cancer problems. Cancer is one of the major disease increases rapidly in human beings. Even doctor are unable to identify the initial stage of cancer and at last stage it is very difficult to handle the cancer patient. Due this every year a large number of patients die. In final stage, most of the patient's life span can be saved if warning of serious clinical events and health condition could be provided in early stage. In this concern, early prediction based on blood and urine reports

by considering pH value health analysis has become an immense need in many clinical fields.

To increase and maintain the survival rate of patients, there should be a proper clinical data reports must be maintained. This can be achieved by data mining techniques for medical datasets. The data mining is the process where the huge amount of data is mined and gathered from database. As the very important feature of data mining is, it will grab the hidden information from the considered database. This parameter helps in better improvements in information gathering. The results of Data Mining techniques are to give important benefits to healthcare industries, for segregating the patients, suffering from same kind of diseases.

Important factor in present health condition is hygiene and growing population. There should be a very effective healthcare analysis is needed. Analysis is the important parameter in health care industries and self-caring at home. Since there are plenty of researches are done on human health level condition. Research in data mining for health analysis is most significant in order to provide good services to the patient's health and these researches are becoming a significant opportunity for improving the quality of health care services. On the basis of reports gathered from the different kind of people such as alcoholic person, diabetic person and pregnancy women etc has strongly hinges the data driven prediction methods. In many of the cases, every patient ends up with death, only because they are not able to predict their health condition on time. Before the disease gets spread or before the initial level of disease. Thus data mining technique is the best way to extract and predict the health condition of normal and abnormal patients.

In this work, development of combined data mining techniques to find out the health condition of human at early stages and provide them with the early warning of health care to avoid many difficult and complicated disease spread. In particular, this approach develops classification model to monitor real time monitoring of blood and urine values. As according to the survey the parameters in blood and urine which the work considered (Normal range) those are Calcium: 4.5-5.5 mEq/L, Sodium: 133-146 mEq/L, potassium: 3.5-

5.4mEq/L and for urine chloride: 98-106mEq/L, Sodium: 25-100mEq/L and potassium 25-100mEq/L. Similarly a collective data for different kind of people such as pregnancy women, alcoholic person and person suffering from diabetics are collected and the procedure for identifying health condition fallows with the data gathered. This will issue previous alert messages to patients before reaching to the last stage of the

disease which may lead to patient death like cancer. This system enables at-risk patients to be timely checked and consulted by healthcare industrials to avoid unconditional death.

This work comprises of the following method along with steps:

Step 1. Developing a combined data mining technique to give health condition analysis based on pH value of blood and urine. By considering few parameters such as pH value: (Normal range) those are Calcium: 4.5-5.5 mEq/L, Sodium: 133-146 mEq/L, potassium: 3.5-5.4mEq/L and for urine chloride: 98-106mEq/L, Sodium: 25-100mEq/L and potassium 25-100mEq/L. by monitoring the data gathered.

Step 2. Filling the gap among both data mining techniques and biomedical community by implementing booming parameters and methods in both the sectors.

Step 3. This work improves precaution warning by incorporating main data mining techniques such as data fusion, data gathering, feature selection and feature extraction and mainly implementing advanced classifiers to get best computational speed and accuracy.

Step 4. Applying the approached methodology to a huge gathering of real patients and normal people data recorded from blood and urine reports shows the effective improvements upon the existing methods.

Section II surveys the related work in identifying health condition. An overview on our methodology presented here is explained in Section III. Section IV shows the experimental result of our real-time early warning system. Finally, Section V draws the conclusion of the proposed methodology.

II. LITRATURE SURVEY

Prof. Dipti Patil et.al [01] has proposed effective methodology for preventive measures to identifying whether the considered person is fit or unfit on the accordance with respective person's historical and real time reports data, author has achieved above criteria by applying k-means clustering technique and also implemented d-stream technique as clustering methods for the data mining field. Author has applied both the algorithms on every patient's current medical condition and report. As applied both algorithms d-stream algorithm withstands all the drawbacks of k-means and got the best results.

Yi Mao et.al [02] has presented an integrated data mining approach for normal fast deterioration warning, in this author has synthesized a huge feature set that comprises of time-series features for both initial and second order by using Detrended fluctuation analysis (DFA), spectral analysis, cross-signal features, and approximate entropy. Finally they arranged in proper way by apply and evaluate a series of established data mining methods, including forward feature selection, linear and nonlinear

classification algorithms, and exploratory un de-sampling for health imbalance.

Ting-Hua Yi et.al [03] has presented Structural Health Monitoring-Oriented Data Mining, Feature Extraction, and Condition Assessment. In this author has summarized the present methodology incorporation in the area of bridge health analysis with the utilization of the wireless sensor networks. In this they have mainly considered the SHM based data mining for better results.

Pooja Mittal et.al [04] has presented Study and Analysis of Predictive Data Mining Approaches for Clinical Dataset. In this author has mentioned and considered an analytical group on predictive data mining techniques on clinical dataset, where the considered area is very much sensitive and protective under several environments. They implemented KDD with reference to clinical datasets applications and advantages which can be utilized for the predictive data mining in same field. The clinical dataset processing is one of the effective and most sensitive areas which is studied under an expert environment. The effectiveness of this work is proven to the prediction of a person disease, based on early stage behaviors dataset.

Zhongdong Duan et.al [05] has presented Technology of data mining for Structural Health Monitoring (SHM), in this work authors has concentrated on impassive explanation to the background, definition, function, process, methods and advantage of DM technology is carried out. Authors proposed SHM on the basis of DM technology provides the application level review of the technology which the authors has been used throughout the proposed work. Several tasks are included in such work are data monitoring, data gathering, definition of tasks etc. The DM platform and framework both are combined to generate the effective SHM system. The important behavior tasks which need be solved in using DM technology for SHM are analyzed in the proposed work.

III. METHODOLOGY

Figure 1 represents the overall architecture of the proposed methodology. Whole architecture is divided into two main phase they are testing phase and training phase. In training phase the blood and urine reports of different kind of patients like pregnancy women, alcoholic person and diabetic persons are collected. The corresponding data is gathered on the basis of data present in the reports. Data gathering level is very much important for generation of system related data as input to our proposed system. Once data generation is finished, these data of blood and urine reports are fused together. This is done by applying data fusion technique called as concatenation. This technique helps in integrating both blood and urine data reports. The resultant data will be very impressive with data indicating the complete similar real-world object into a consistent, accurate, and usefulness of presenting the data.

The main aim of data fusion is to integrate known data from two or more data sources into a single one that generates an extra accurate explanation than any of the separate data values. Next,

these integrated data are applied with PCA (Linear Dimensional Reduction) method to reduce the dimension of the fused data; this technique will convert the high dimensional data samples into the low dimensional space so that the intensive information contained in the data is protected. As the dimensionality of data gets reduced, it encourage improving the robustness of the classifier and decreases computational complexity. Next, by considering PH value and other different

chemicals present in blood and urine. Feature extraction technique is applied and extracted the considered features from the blood and urine reports. Which are considered as input parameter and these features are given for ANN training to create knowledge base for the better database for comparison.

Similarly in testing phase, input might be one of the blood or urine reports of different kinds of patients. Similarly as considered in training phase. By considering reports of blood or urine data is gathered on the basis of behaviour and parameters present in the report. Thus, after generation of data, this must be dimensionally reduced by applying PCA technique. This technique is for better performance and better computational speed. Next, by considering pH value and other different chemicals present in blood and urine, feature extraction technique is applied. This extracts the considered features from the blood and urine reports. Extracted features are given for ANN classifier to classify the reports. In accordance with different person's reports and by comparing results with knowledge base, ANN classifier will classify whether the person is healthy or not

3.1 DATA GATHERING

The data set contains 7 attributes, which are present in the urine and blood samples. They are pH value: (Normal range) those are Calcium: 4.5-5.5 mEq/L, Sodium: 133-146 mEq/L, potassium: 3.5-5.4mEq/L and for urine chloride: 98-106mEq/L, Sodium: 25-100mEq/L and potassium 25-100mEq/L. These attributes are the main chemicals present in urine and blood report that can help in deciding the patient's health condition. These attributes are gathered from the reports of urine and blood samples taken from the different kind of patient, the above mentioned values are only for the normal health conditioned person's report parameters. These parameters values are different for different type of persons.

Since our database system is still in a lesser scale clinical trial and does not provide useful data at one shot. In this paper, system is worked on the basis of database collected from the hospitals. The database gathering purpose is to collect several kinds of person's reports for health level monitoring. This work provides caution for health condition of particular person's on the basis of the report collected. Since our work concentrating on early warning system, which is based on pH rate, Glucose, calcium and iron for blood report samples. Similarly for urine considered values are pH value, protein, specific gravity and color. These attributes are considered from all type of patients and making them grouping and creating a

class for them. This method also provides that the gathered dataset of most patients in the dataset are from different classes.

Table 1: Normal and Abnormal electrolyte composition in blood sample of a person

pH value of Blood			
	Calcium (mEq/l)	Sodium (mEq/l)	Potassium (mEq/l)
Normal Condition	4.5-5.5	133-146	3.5-4.5
Abnormal Condition	<4.5or > 5.5	<133 or >146	<3.5 or >4.5

Table 2: Normal and Abnormal electrolyte composition in urine sample of a person

pH value of Urine			
	Chloride (mEq/l)	Sodium (mEq/l)	Potassium (mEq/l)
Normal Condition	98-106	25-100	25-100
Abnormal Condition	<98 or > 107	<25 or >100	<25 or >100

3.2 DATA FUSION

The gathered data of both urine and blood has to be combined to form single data for better performance and less memory and space usage. In order to achieve this, the most importantly both data of urine and blood report has to be merged together. To achieve this, the data concatenation fusion method is adopted. This method is one of the basic functionality of the data fusion process. It consisting of combined data residing in different sources and giving users with different kind of data. This process becomes significant in.