International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Efficient Technique for Cancer Prediction Using ANN Classification Based on pH Parameter

Rakesh Godi[1], Dr. Narendra B Mustare[2]

[1]Research Scholar, VTU University, Belgaum, Karnataka

[2]Professor, Department of I.T, PDA Collage of Engineering, Gulbarga, VTU, Karnataka

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. Now 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 ANN 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 ANN Classifier

1. 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 searches 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.

2. Literature 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.

3. 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 behavior 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.

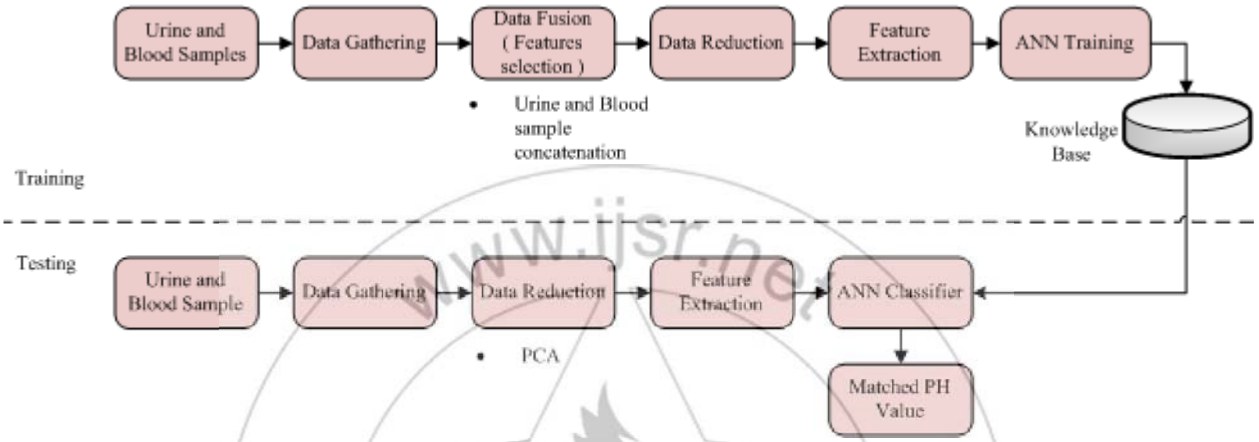

Figure 1: Block Diagram of proposed system

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 a variety of situations, in this case the database collected from the hospital of several kinds of patients like urine and blood samples are both combined together by data fusion technique by concatenating both the data's together to form a single data documents for further process.

3.3 Data Reduction by PCA

In recent medical sector a huge amount data is collected, these high dimensional amount data is need to process for efficient performance of the system. To reduce the high dimensional data into low dimensional level a various data reduction techniques are developed. In this proposed system PCA (Principle Component Analysis) reduction algorithm is used to convert the high dimensional data into lower level.PCA is one of the most famous linear data reduction algorithms. The output of the PCA produced a lower dimensional data from previous and data specify a deviation in the data. There are two principle method are used in PCA they are Matrix method and Data method. In matrix method all the data mentioned in datasets are needed to calculate the variance covariance format and represent in the form of matrix. In data method reduction algorithm directly works on the data.

Considering $\{x_i\}_{i=1}^{N}$ data and $D = [d_{ij}]$ be the pair wise Euclidean matrix whose entries $[d_{ij}]$ denotes the distance between data points x_i *and* x_j i.e high dimensional data points to maximize the cost function multidimensional scaling finds the linear mapping P.

$$\psi(Y) := \sum_{i,j} \left(d_{ij}^2 - |y_i - y_j|^2 \right) \quad (1)$$

Euclidean distance between the low-dimensional data point y_i *and* $y_j = |y_i - y_j|$, y_i is restricted to be x_i A with $|v_j|^2 = 1$ for all column vector v_j of P.

It can be represented that the minimum of the cost function $\psi(Y)$ is given by the eigen-decomposition of the Gram matrix $G = XX^T$ where $= [x_i]$. The Gram matrix by double-centering the pairwise squared Eucliden distance matrix, i.e computing:

$$g_{ij} = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{l} d_{il}^2 - \frac{1}{n}\sum_{l} d_{lj}^2 + \frac{1}{n^2}\sum_{l,m} d_{lm}^2 \right) \quad (2)$$

The multiplication of principal eigenvectors of the double-cantered squared Euclidean distance matrix is considering (*i.e.*, the principal eigenvectors of the Gram matrix) with the square-root of their corresponding Eigen values, this gives us exactly the minimum of the cost function in Equation (1).

The eigenvectors u_i and $v_i$of the matrices X^TX and XX^T are related through $\sqrt{\lambda_i} v_i = X_{ui}$, it turns out that the similarity of classical scaling to PCA. PCA may also be viewed upon as a latent variable model called probabilistic PCA. This model uses a Gaussian prior over the latent space, and a linear-Gaussian noise model.

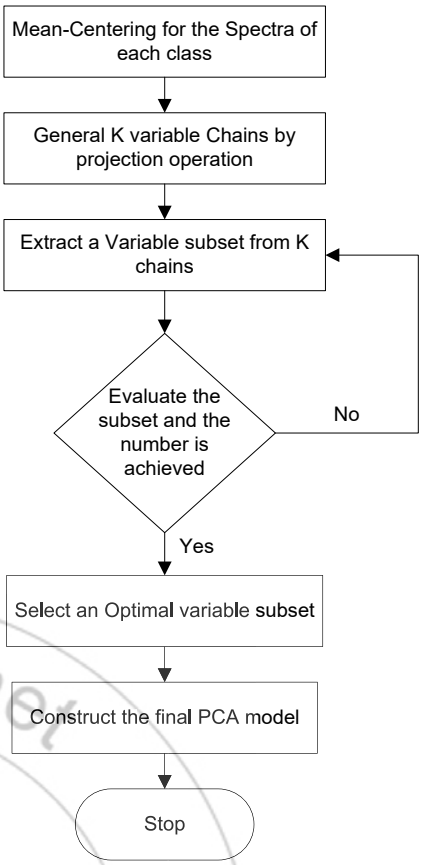

Figure 2: Flow chart of PCA Reduction

The probabilistic formulation of PCA leads to an EM-algorithm that may be computationally more efficient for very high-dimensional data. By using Gaussian processes, probabilistic PCA may also be extended to learn nonlinear mappings between the high-dimensional and the low-dimensional space. Another extension of PCA also includes minor components (*i.e.*, the eigenvectors corresponding to the smallest (eigenvalues) in the linear mapping, as minor components may be of relevance in classification settings. PCA and classical scaling have been successfully applied in a large number of domains such as face recognition, coin classification, and seismic series analysis.

The two main drawbacks are suffered by the PCA and classical scaling. First, covariance matrix size is proportional to the data point's dimension. As a result, the calculation of eigenvectors might be infeasible for very high-dimensional data. In classical scaling computation is done on the number of data points instead of with the number of dimensions in the data. In data set in which $n < D$, this drawback is overcome by performing classical scaling. A simple PCA and probabilistic PCA are employed as a alternative iterative techniques.

Second PCA and classical scaling is mainly focus on retaining large pairwise distances d_{ij}^2 , instead of focusing on retaining the small pairwise distances, which is much more important, it turns out that the similarity of classical scaling to PCA .The cost function is represented in eq. (1). In the probabilistic PCA model, a Gaussian prior over the latent space and a linear-Gaussian noise model is done. Hence by adopting PCA data reduction technique the dimensionality of the data gathered can be reduced successfully. This PCA

technique extended to extract the features of the reduced data so that the accuracy of the classification will be raised.

3.4 Feature Selection and Extraction

Extracting the hidden facts from the data which we are considered as input is referred as data mining; this process must be done to create the knowledge base for the further comparison process and for classification process. The feature selection plays pretty much important role in the data mining; as there will be plenty of huge data present in the data documents, it is very difficult to classify the results by considering all the parameter and attributes present in the data. So by selecting the few features present in the considered data will help in better classification.

In this present work we are considering few attributes from the blood and urine samples considered from the different kind of person they are, pH value, glucose, iron, specific gravity, calcium and etc. The predictive analysis respective to health care analysis is considered for feature extraction process after selecting few features from the data. On the implementation of PCA feature extraction after PCA feature reduction, the accuracy level required in such system is very potential even though they cannot be implemented as an individual systems, all the selected features are extracted and on the basis of these features the data base has been created and training has been dome to ANN classifier to classify the patients according to their blood and urine report.

3.5 ANN Classifier

Data samples are divided and classified into larger target classes. For particular data level, the best classification identifies the target class should be more accurate. In such cases, the best example is, the considered patient can be classified as healthy person or non-healthy person, on the basis of the patient's blood and urine reports using data classification method. The classification process is the standard one with different class categories in it. There are several techniques are present in this classification process. The most important technique is considered as Binary and multilevel classification methods. As followed with the classification method, binary classification having two major levels they are low risk i.e. healthy persons data. Next one is high risk i.e. non-healthy person's data.

In multi-level classification there will be three such states low, medium and high risk status. In our approach the classification which we are going for is binary level along with two output status healthy and non-healthy. Data set is divided as training and testing phase database data. Using training dataset classifier gets trained. Effectiveness of the classifier must be tested by the help of test dataset. Classification is one of the most widely used methods of Data Mining in Healthcare organization. By technical survey, among all types of classification the best rated and used classifier found is ANN classifier. This is the one used in this approach for better and accurate classification results.

Artificial neural networks consist of input layer, hidden layer and output layer. The effectiveness of the methods was illustrated by its capacity to classify the cell line correctly in 100% of classes which the method has declared. The main intention of the training process is to identify the group of weight values that will affect the system output from the neural network to match the actual target values as similar as possible. ANN training has been done with Back propagation algorithm called conjugate gradient algorithm in our proposed model.

Conjugate Gradient Algorithm

Step 1: y_k Pattern from the training set T is to be selected, and pass it to the network.
Step 2: The neurons which are hidden and output layered take them and compute the signals to the output.
Step 3: By doing comparison with the present output neurons to the already generated output neurons the error has to be calculated.
Step 4: To reduce the global error which has been generated the all input neurons are converted into hidden layer neurons and it will be utilized to compute the error produced in hidden layer weights.
Step 5: All the values must be updated from input to hidden layer neurons value which have been
Hidden to output layer weights are calculated as:

$$w_{hj}^{k+1} = w_{hj}^{k} + \Delta w_{hj}^{k} \quad (4)$$

Input to hidden layer weights is calculated as:

$$w_{ih}^{k+1} = w_{ih}^{k} + \Delta w_{ih}^{k} \quad (5)$$

Where $+\Delta w_{hj}^{k}$ and Δw_{ih}^{k} are weight changes computed in Step 4.
Step 6: All the above steps have to be repeated until the results falls to the pre-defined threshold value

4. Experimental Results

Data mining information technology gaining a more imporatance in health care filed. During health analysis, a large number of blood and urine electrolyte sample values are collected from N number patients with different time intervals.

The research represents the link between the pH value of blood and urine and cancer disease. In this research we mainly focused on three different electrolyte i.e Calcium, sodium, and potassium. The Calcium pH value is mainly considered during health analysis in abnormal condition of the patients. The Calcium content exceeds beyond normal range 5.5 mEq/L than it represents the hypercalcemia condition in patient health. This hypercalcaemia condition may lead to a different type of cancer.

The collected medical practical data values are concatenated by using data fusion techniques. The resultant data will be available in large dimension. Hence to process medical data further a PCA data reduction technique is used. The use of PCA reduces the overall dimension of the medical data. For better classification a features selection and extraction techniques are used to selected important data features form huge collected data.

Table 3: Confusion matrix

n=30	Predicted : No	Predicted : Yes	
Actual un healthy person: No	TN=6	FP=4	10
Actual healthy person: Yes	FN=5	TP=15	20
Total	11	19	30

The resultant data is analyzed which is represented in the form of confusion matrix as shown in Table 3. Which represent the respective person is normal or not, if not than clinician may judge the disease, whether it is cancer or not based on electrolyte pH value. Performance evaluation of the entire project is shown in figure 3.The use of ANN classifiers which improve the effectiveness of the project and accuracy up to 70%. The overall use of project makes the medical system easy in health analysis and disease prediction.

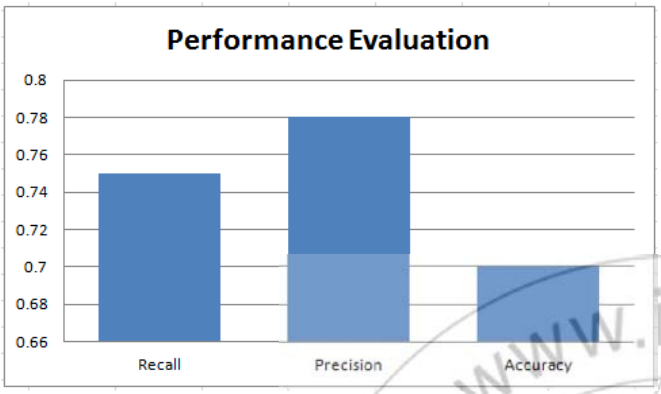

Figure 3: Performance Evaluation Graph

5. Conclusion

Cancer is the major disease in medical field. Analysis or prediction of the cancer in initial stage is very important to save the patient life. Hence the analyzing a pH value of blood and urine using data mining in healthcare sector help the physician predict the disease in early stage itself. This paper aims in providing initial stage of health monitoring, caution to patients with different health condition. This is achieved by gathering data from very large data sets and improvising the outcomes with cost effectiveness. Its capacity is tremendous in regard with health monitoring by analysis the patient's health condition by pH value and other attributes of blood and urine samples. In this paper, data gathering and analysis on the basis of Health Monitoring Systems & the research is based on the clinical reports of blood and urine samples of different kind of patients. The conclusion with the effectiveness of approached soft computing techniques that shows the effective classification based analytical research. To generate reports of patient's, based on health condition. In this work, process of every method used for this particular work is approached and discussed. It may be also analyzed that usability of any methods is decided by the nature of data & application of that particular approach. The experimental result with classification accuracy 70% witnesses the overall effective working of proposed system of health analysis and prediction of the cancer disease.

References

[1] Prof. Dipti Patil, Snehal Andhalkar, Richa Biyani, Mayuri Gund, Bhagyashree Agrawal and Dr. Wadhai, " An Adaptive parameter free data mining approach for healthcare application", International Journal of Advanced Computer Science and Applications, Vol. 3, No. 1, 2012.

[2] Yi Mao, Wenlin chen, Yixin Chen, Chenyang Lu, Marin Kollef and Thomas c Bailey, "An Integrated Data Mining Approach to Real-time Clinical Monitoring and Deterioration Warning", Visiting doctoral candidate from Xidian University, China, 2012.

[3] Ting-Hua Yi, Stathis C. Stiros, Xiao-Wei Ye, and Jun Li, "Structural Health Monitoring-Oriented Data Mining, Feature Extraction, and Condition Assessment", Hindawi Publishing Corporation Mathematical Problems in Engineering, Vol. 2, 2014.

[4] Pooja Mittal and Nasib Singh Gill, "Study and Analysis of Predictive Data Mining Approaches for Clinical Dataset", International Journal of Computer Applications, Vol. 63, pp. 0975, No.3, 2013

[5] Zhongdong Duan, "Technology of data mining for Structural Health Monitoring (SHM)", Vol.3, Issue 8, 2013

[6] Chris Ding and Hanchuan Peng, "Minimum Redundancy Feature Selection from Microarray Gene Expression Data", NERSC Division, Lawrence Berkeley National Laboratory, USA, 2010.

[7] Jiliang Tang, Salem Alelyani and Huan Liu, "Feature Selection for Classification: A Review", Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 67, Issue 2, pp. 301–320, 2005.

[8] Lei Yu Jieping Ye and Huan Liu, "Dimensionality Reduction for Data Mining Techniques, Applications and Trends", 2003.

[9] Dipti Durgesh Patil and Vijay M. Wadhai, "Adaptive Real Time Data Mining Methodology For Wireless BodyArea Network Based Healthcare Applications", Advanced Computing: An International Journal, Vol.3, No.4, 2012.

[10] Ashok N. Srivastava, "Discovering System Health Anomalies Using Data Mining Techniques", Systems Health Technical Area Intelligent Systems Division NASA Ames Research Center, 2012

[11] Divya Tomar and Sonali Agarwal, "A survey on Data Mining approaches for Healthcare", International Journal of Bio-Science and Bio-Technology Vol.5, No.5, pp. 241-266, 2013.

[12] Mohammed Abdul Khaleel, Sateesh Kumar Pradham and G.N. Dash, " A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013

[13] David D. Lewis, "Feature Selection and Feature Extract ion for Text Categorization", Center for Information and Language Studies University of Chicago, 2013

[14] Masoumeh Zareapoor and Seeja K. R, "Classification: A Case Study on Phishing Email Detection", I.J. Information Engineering and Electronic Business, Vol. 2, pp. 60-65 Published Online. 2015

[15] N. Elavarasan and Dr. K.Mani, "A Survey on Feature Extraction Techniques", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 1, 2015

[16] Liang Goh, Qun Song, and Nikola Kasabov, "A Novel Feature Selection Method to Improve Classification of

Gene Expression Data", Discovery Research Institute Auckland University of Technology, 2002

[17] Ashoka H.N and Manjaiah D.H, "Feature Extraction Technique for Neural Network Based Pattern Recognition", International Journal on Computer Science and Engineering, 2012