



Intelligent Classification & Clustering Of Lung & Oral Cancer through Decision Tree & Genetic Algorithm

Tanupriya Choudhury*

Research Scholar CSE Dept,
Jagannath University, Jaipur, India

Prof. (Dr.) Vivek Kumar

Principal DCTM Palwal
Haryana, India

Dr. Darshika Nigam

Department of Bio Chemistry, SLS
Dr. B.R.Ambedker University, Agra, India

Abstract— *Cancer is a significant health problem all over the world. Most of the oral and lung cancers are detected at a later stage where, treatment is of no use or less significant. It is compulsory to detect such types of cancer at an earlier stage. Due to early detection surgeons gets the aid to identify the disease and to provide necessary therapeutic measures which also benefit the patients. This paper presents a real time experimental methodology to predict and detect Lung and Oral cancer in earlier stage by using Genetic Algorithm and Decision Tree Approach.*

Keywords — *Genetic Algorithm, Data Mining, Lung Cancer, Oral Cancer, Decision tree.*

I. INTRODUCTION

Lung cancer is an uncontrolled cell growth of tissues in the lung. It may lead to metastasis, which is the invasion of adjacent tissue and infiltration beyond the lungs. The treatment depends on the cancer type, its stage, and the performance of the people and their will power. Though there are various possible treatments which include surgery, chemotherapy, and radiotherapy but survival totally depends upon the stage, all over health, and various other factors.

Oral cancer (OC) is the sixth most common cancer and a key health problem in the world. It has been recognized as a huge threat to public health because of its high sickness and mortality. Oral cancer is a heterogeneous assembly of tumors rolling out from various parts of the oral cavity, with typical predisposing issues, prevalence, and outcomes of the treatment. Oral cancer is one of the ten most persistent diseases in worldwide with a yearly incidence of over 400,000 cases, of which 64% rise in developing nations.

Aim of this study is to automate the process of classification for early detection of Lung Cancer and oral cancer in a intelligent way with the help of Genetic algorithm and Decision tree approach.

II. PROBLEM DEFINITION

Cancer is a significant health problem all over the world. Most of the oral cancers are detected at a later stage where, treatment is of no use or less significant. It is compulsory to detect such types of cancer at an earlier stage. Due to early detection surgeons gets the aid to identify the disease and to provide necessary therapeutic measures which also benefit the patients.

Thus there is a need for implementation and development of suitable primordial, primary, and secondary prevention approaches to control this epidemic. To initiate steps in this direction, an urgent and sincere bureaucratic, political, and social will is required. In this section we try to understand the deadly disease, their causes, and how they affect the mankind.

Here the research proposed a pioneering feature selection algorithm constructed on the genetic algorithm exhausting a precisely devised trace-based separability standard. According to the scores of class separability and variable separability, this criterion measures the significance of feature subset, independent of any specific classification. In addition, a mutual information matrix between variables is used as features for classification, and no prior knowledge about the cardinality of feature subset is required. Experiments are performed by using a standard lung cancer, oral cancer dataset. The achieved solutions are verified with two diverse classifiers, including the K-Means, and decision tree and compared with those obtained by the whole feature set, the F-score and the correlation-based feature selection methods. The comparison results show that the proposed intelligent system has a good diagnosis performance and can be used as a promising tool for lung cancer, oral cancer diagnosis. Currently cancer has grown into huge intimidation in human life.

Identification of genetic and environmental factors is very important in developing novel methods to detect and prevent cancer. Therefore a novel multi layered method combining Genetic algorithm, clustering and decision tree techniques to build a cancer risk prediction system is proposed here which predicts lung & oral cancers and is also time saving, cost saving and user friendly. The research customs data mining technology like clustering, classification and prediction. The gathered data is pre-processed, fed into the database and divided to get major patterns using genetic and decision tree algorithm. Then the data is clustered using K-means clustering algorithm to separate cancer and non-cancer patient data. With the help of the classifier it will check the condition of a patient in its early stage whether it is normal or abnormal. The performance is based on the correct and incorrect classification of the classifier. Finally a prediction system is developed to analyze risk levels which help in prediction. The research helps in discovery of a person's predisposition for cancer before going for clinical and lab tests which is cost and time consuming. It is totally a computational procedure, as

before some researchers trying to check and predict cancer using medical image analysis but here only computational methods have been used for best prediction. Evaluation would be done on the basis of correctly classified sample data. For testing and training different datasets has been used.

Aim of this research is to mechanize the classification process for the early detection of lung & oral Cancer. To justify this research, it includes classification algorithm and for optimization GA (Genetic Algorithm) is used.

III. METHODOLOGY

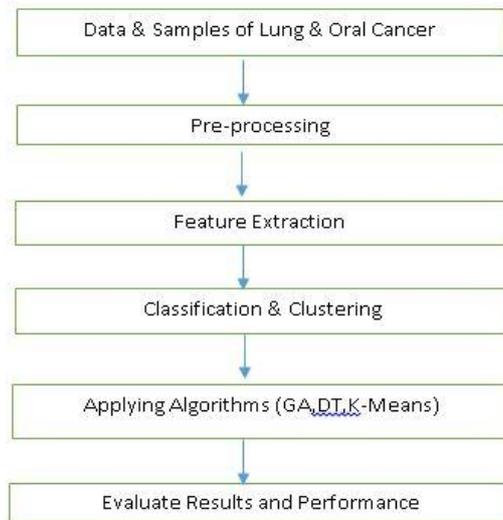


Fig 1: Methodology To Detect Lung & Oral Cancer

In numeric prediction, the outcome to be predicted is not a discrete class but a numeric quantity. In this study, to classify the data and to mine frequent patterns in data set Decision Tree algorithm is used.

A decision tree is a flow chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node holds a class label. The top most node is the root node. The data's attribute value is tested against a decision tree. A path is traced from root to leaf node, which holds the class prediction for that data. Decision trees can be easily changed into classification rules. This decision tree is used to generate frequent patterns in the dataset. The data and item sets that occur frequently in the data base are known as frequent patterns. The frequent patterns that is most importantly related to specific cancer types and are helpful in predicting the cancer and its type is known as Significant frequent pattern. Using these generated significant patterns of decision tree the data set is clustered accordingly and risk scores are given. Clustering is a process of separating dataset into subgroups according to their unique features. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. In K-means clustering, the number of clusters needed is found out and then an algorithm is used to successively associate or dissociate instances with clusters until associations stabilize around k clusters. In this research all the above mentioned Data Mining techniques are implemented together to create a novel method to diagnose the existence of cancer for a particular patient. When beginning to work on a data mining problem, it is first necessary to bring all the data together into a set of instances. Integrating data from different sources usually presents many challenges. The data must be assembled, integrated, and cleaned up. Only then it can be used for processing through machine learning techniques. This developed system can be used by physicians and patients alike to easily know a person's cancer status and severity without screening them for testing cancer. Also it is useful to record and save huge volumes of sensitive information which can be used to gain knowledge about the disease and its treatment.

A. Proposed Model to detect Lung & Oral Cancer in Earlier Stage:

The proposed model is given in the following diagram. The collected data is pre-processed and is stored in the knowledge base to build the model. Seventy five percent of the entire data is taken as training set in order to build the classification and clustering model. The remaining of which is taken for testing purpose. The decision tree model is build using the classification rules, the significant frequent pattern and its corresponding weightage. The clustering model is build using the k-means clustering algorithm. The model is then tested for accuracy, sensitivity and specificity using test data along with merging it to the knowledge base. Finally the model is evaluated and optimized for best prediction result by using Genetic Algorithmic methods. The modified genetic algorithm has been implemented to classify the oral and lung cancer data intelligently. Firstly for oral and lung cancer best attributes possibility has been find it out and then after mining the attributes, find the best reference value for those, which means the attributes must have a maximum value as well as minimum value. The proposed implemented algorithm associate the attributes with it and incorporate the values to perform better intelligent classification. And the output which produce it contains the genetically values in binary format. Here the format has been transferred to decimal value and a threshold value has been calculated. .If the output cross the threshold value that means the person should have been affected by the respective cancer. Same methods has been used to find or detect oral cancer, only the attributes and the respective values for the same has been changed according to the doctor's prescribed value.

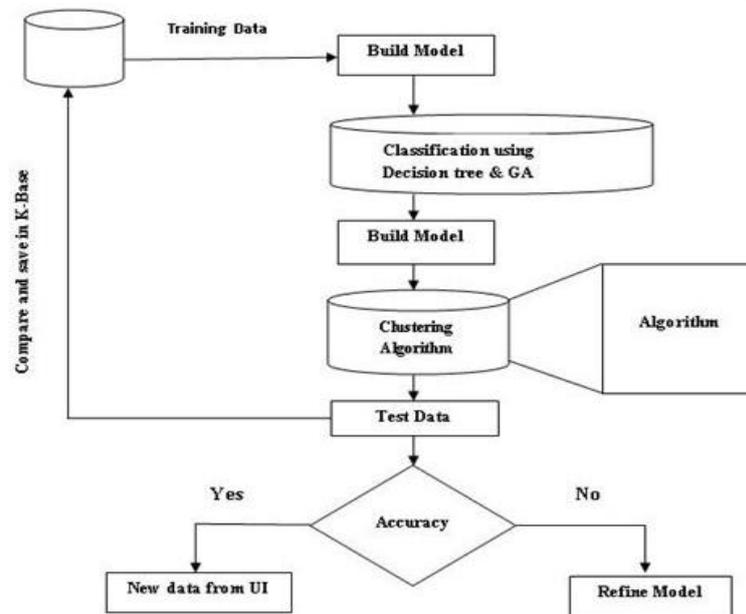


Fig 2 : Proposed Model To Detect Lung & Oral Cancer

B. Data Source:

The data for this study was collected from online repositories and different hospitals surrounding Delhi NCR and Kolkata, consisting of cancer and non-cancer patient’s data and they are pre-processed to suit this research. This data consists of more than 20 attributes such as Age, Marital status, Symptoms relating to cancer, occupational hazards, family history of cancer etc. and the other attributes has been incorporated after screening and reviewing of doctors. These attributes are used to train and develop the system and a part is used to test the significance of the system. These attributes play an important role in diagnosing cancer in all the cases. This data is stored in a knowledge base which has the ability to expand itself as new data enters the system through front end from which new knowledge is gained and thus the system becomes intelligent.

1. Classification and Significant Pattern Generation

Decision tree algorithm is used to mine frequent patterns from the data set. The frequent item sets that occur throughout the data base and have a significant link to cancer status are mined as significant patterns. The data is fed into the decision tree algorithm to obtain the significant patterns related to cancer and non-cancer data sets. In other words the patterns that are mined by the decision tree are well defined and distinguished to be separated as cancer and non-cancer datasets. The following pseudo code is used to generate frequent pattern using decision tree.

A set of candidate attributes III, and R, a set of labelled instances is given as input. The algorithm to generate a decision Tree Q is as follows

Begin 1) If (R is pure or empty) or (III is empty) Return Q. 2) Compute Ps (Ci) on R for each class Ci. 3) For each attribute X in III, compute IIIIG(R, X) based on equation 1 and 5. 4) Use the attribute Xmax with the highest IIIIG for the root. 5) Partition R into disjoint subsets Rx using Xmax. 6) For all values x of Xmax •Qx=NQ (III-Xmax, Rx), •Add Qx as a child of Xmax. 7) Return Q End..

2. Significant Pattern mined using Decision tree algorithm:

Decision Trees is the most popular data mining technique for classification problems. The principal idea of decision tree is to split your data recursively into sub-sets so that each subset contains more or less homogenous states of your target variable. At each split in the tree, all input attributes are evaluated for their impact on the predictable attribute. When the recursive process is completed, a Decision Tree is formed.

Given a database $D=\{t_1, \dots, t_n\}$ where $t_i=\{t_{i1}, \dots, t_{ih}\}$ and the database schema contains the following attributes $\{A_1, A_2, \dots, A_h\}$. Also given is a set of classes $C = \{C_1, \dots, C_2\}$. A Decision Tree (DT) or classification tree is a tree associated with D that has the following properties.

- Each internal node is labelled with an attribute A
- Each arc is labelled with a predicate that can be applied to the attribute associated with the parent.
- Each leaf is labelled with class C_j .

1. Age - gender - living area - family history- anemia symptoms -> none- Cancer Type -> None. Weightage = 99.55
2. Age - gender- marital status-education-smoking-diet symptoms-> Pain in chest, back, shoulder or arm->Shortness of breath and hoarseness-Cancer Type->Lung Weightage = 200.50
3. Gender-Education-Occupational hazards- Alcohol-Family history- Weight loss- symptoms-> severe abdominal pain or bloating-> abdominal pain with blood in stool- Cancer Type ->Stomach Weightage = 190.05
4. Age- gender- no of children- occupational hazards- Family history- relationship with cancer patient- symptoms-> swelling or mass in armpit -> discharge or pain in nipple -> Cancer Type -> Breast. Weightage = 180.55

5. Gender- education- living area- Smoking- Hot beverage Diet- fast food addiction- Earlier cancer diagnosis- symptoms- > Ulcers in mouth or pain of teeth and jaw-> White or red patches in tongue, gums- Cancer Type -> Oral. Weightage = 190.50

Numerical values are given as risk scores to the attributes that have a direct link to the significant patterns mined.

IV. WEIGHTAGE FOR SIGNIFICANT PATTERN & ATTRIBUTES MEASUREMENT

- A. The weightage is designed for every numerous pattern based on the attributes to examine its influence on the output. The numerous patterns mined which satisfies the below situation are taken as significant Numerous Pattern.
- B. $Sw(i) = \sum(Wi * Fi) \longrightarrow$ Equation 1
- C. (1). Where Wi is the weightage of each attribute and Fi represents number of frequency for each rule. And significant Frequent Pattern is selected by using the following Equation (2) $SFP = Sw(n) \geq \phi$ for all values of n
- (2). Where SFP denotes significant frequent pattern and ϕ denotes significant weightage.

Table 1: Lung Cancer Attributes

#no	Attribute	Values
1	Record Number	Continuous
2	Survival Time	Continuous
3	Censored	Binary
4	Medical Condition	Continuous
5	Age	Continuous
6	Time_Diagram Study	Continuous
7	Tumour type	Binary
8	Treatment	Binary
9	Type of Surgery	Binary
10	Cancer Grade	Continuous
11	Radiation Therapy	Binary
12	Cancer Stage	Continuous
13	Diagnostic Confirmation	Binary
14	EGFR	Continuous
15	Bone Age	Continuous
16	Average BP	Continuous
17	Distance to Margin	Continuous
18	Heart Rate	Continuous
19	Platelet Count	Continuous
20	Weight	Continuous
21	Lv-X3	Continuous

Table 2: Oral Cancer Attributes

#no	Attribute	Values
1	Patient_id	Continuous
2	Trt	Continuous
3	Age	Continuous
4	Weights	Continuous
5	Bone_invasion	Binary
6	Totalcin	Continuous
7	Totalcw	Continuous
8	Totalcw4	Continuous
9	Totalcw6	Continuous
10	Stage	Continuous
11	Ethnic_origin	Binary
12	Primary_tumor	Binary
13	Site_radiotherapytreat	Binary
14	Oncologist	Continuous
15	Diagnosis_status	Binary
16	Treatment_offered	Binary
17	Treatment_modality	Continuous
18	Tumor_previoustreatment	Binary
19	Performance_statuspresent	Binary
20	Laterality	Continuous
21	Survival Time	Continuous

Table 3 : Risk Scores For The Attributes That Represent The Significant Patterns.

Serial Number	Attributes	Values	Risk score
1.	Anaemia	Yes	3
		No	1
2.	Weight Loss	Yes	2
		No	1
3.	Family history of cancer	Yes	5
		No	1
4.	Habits	Alcohol	5
		Drinking Hot Beverage	2
		Smoking	3
		Chewing Gutkha	4
5.	Living Area	Urban	5
		Rural	3
6.	Education	Uneducated	5
		School	3
		College	2
7.	Age	$x < 30$	3
		$30 < x < 40$	4
		$40 < x < 60$	5
8.	Occupational Hazards	Radiation Exposure	3
		Chemical Exposure	3
		Thermal Exposure	2

V. RULES FOR DECISION TREE

“If symptoms = none and risk score = $x < x < 60$ then result = you may have cancer, tests = do blood test and x ray to confirm”

“Else if symptom= related to stomach and risk score = $x > 45$ then result = you have cancer, cancer type = stomach, tests = endoscopy of stomach”

“If symptom= related to pelvis and lower hip and risk score= $x > 55$ then result = you have cancer, cancer type = cervix, tests = do Pap smear test If symptom= related to head and throat and risk score = $x > 40$ then result = you have cancer, cancer type = oral, tests = biopsy of tongue and inner mouth”

“Else symptom= other symptoms and risk score = $x > 40$ then result = you have cancer, cancer type = leukemia, tests = biopsy of bone marrow”

“If symptom= related to breast and shoulder and risk score = $x > 45$ then result = you have cancer, cancer type = breast, tests= mammogram and PET scan of breast If symptom= related to chest and shoulder and risk score = $x > 40$ then result = you have cancer, cancer type = lung, tests = take CT scan of chest”.

Grounded on the above cited rules and the designed risk scores the rigorousness of cancer is recognized as well as some tests were prescribed to confirm the presence of cancer..

VI. CLUSTERING USING K-MEANS

The occurrences are now grouped into a number of classes which are called cluster where each class is acknowledged by a sole feature based on the substantial patterns mined by the decision tree algorithm. The main aim of clustering is that the data object is assigned to unknown classes that has a sole feature and hence maximize the intra-class similarity and minimize the interclass similarity. That means through the help of clustering the cancer patient’s information and non-cancer patients information will be separated for better understanding. The weightage scores of the significant patterns mined are fed into K- means clustering algorithm to cluster and divide it into cancer and non-cancer groups. The cancer group is further partitioned into six groups with each cluster representing a type of cancer. At the commencement the data is assigned to a non-cancer cluster and then based on the intensity of the cancer given by its weightage it is either moved to the cancer cluster or gets retained in the non-cancer cluster, further the data object is moved between the subgroups of the hierarchical cancer cluster based on the symptoms the data object contains. To calculate the mean of the cluster centre the symptoms are given certain values the average of which represents each distinguished cluster. The data objects are distributed to the cluster based on the cluster centre to which it is nearest.

It also searches the entire database to find a match to a single input data. The data is moved to that particular cluster if and only if an exact match is found. This technique minimizes the error rate of clustering. The data in the first cluster are all similar with little or no symptoms; no risk factors associated with cancer and low risk scores. Hence the cluster is labelled as Non cancer cluster. The top cluster of the second hierarchical cluster contains all the data that has high risk factors associated with cancer along with distinguished symptoms and high risk scores. The data in the cluster is again fed into k – means clustering algorithm to further subdivide it. The resulting six clusters are separated based on particular symptoms associated with any one type of cancer i.e. lung, cervix, breast, stomach, oral and leukaemia. Finally all the data is partitioned into two types of clusters and six sub clusters of the cancer cluster.

Clustering Algorithm:

The k-means clustering algorithm is used for dividing the data into cancer and non-cancer clusters, where the initial cluster centres are represented by the mean value of the weightage of major patterns.

Input: k: the number of clusters. D: data set containing n objects.

Output: A set of hierarchical clusters

Begin 1) choose two mean values from weightage of significant patterns as the initial cluster centres; 2) assign each object to the cluster to which it is most similar based on the mean value of the weightage. 3) Update the cluster means by calculating mean value of all the objects in the cluster. 4) End

Now two clusters have been generated based on the weightage scores of the significant pattern. The two clusters are named as Non cancer and Cancer clusters. The mean weightage of the non-cancer cluster is significantly lower than the cancer cluster. Again partition the cancer cluster to generate six sub clusters each representing a type of cancer.

Begin 1) arbitrarily chooses k objects from cancer cluster S with distinguished values for its symptoms. 2) Assign each object in S to the cluster whose mean value is closer to its symptom. 3) Update the cluster means and 4) Repeat step 2 and 3 until no change 5) End

VII. GENETIC ALGORITHM DESCRIPTION

Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. As such they represent an intelligent exploitation of a random search used to solve optimization problems.

The general scheme of a GA can be given as follows:

begin

INITIALIZE population with random candidate solutions;

EVALUATE each candidate;

repeat

SELECT parents;

RECOMBINE pairs of parents;

MUTATE the resulting children;

EVALUATE children;

SELECT individuals for the next generation

until TERMINATION-CONDITION is satisfied

end

How the Genetic Modeling & Clustering Finds the Output

It is the modified genetic algorithm that has been implemented in order to classify the oral and lung cancer data intelligently.

Firstly for oral and lung cancer best attributes possibility has been found and then after mining the attributes, find the best reference value for those, which means the attributes must have a maximum value as well as minimum value.

The proposed implemented algorithm associates the attributes with it and incorporates the values to do well intelligent classification.

And the result which produce it contains the genetically values in binary format.

Here the format has been changed to decimal value and a threshold value has been calculated. If the result cross the value that means the person should have been affected by the respective cancer.

After genetically classification through the values getting from oral and lung cancer the process of clustering has been done, to precise the information more logically.

Here as the k-means algorithm gives the best result so K-means has been used for better clustering between these two oral & lung values.

Genetic algorithm used into two datasets such as lung cancer and oral cancer. For this algorithm input will be taken it from dataset in the form of binary digit. The genetic algorithm finally produce or generating result in 10 digit binary format. Here 10 is a constant value as it is constant value for better prediction. Then this 10 digit binary value that is 1111111111 is converted into decimal value, middle range will be taken by dividing the decimal value by 2. From that if input is below the middle range that means it is normal , and if it is not then it is abnormal.

The algorithm is the newest one and it has been produced by random generation. Before less attributes was taken into use, and genetic algorithm is not used for big dataset as well as large number of attributes. Also it is time saving to compile the result. And the execution time has also been measured where it has been proved by comparative study that genetic algorithm produces the best outcome to predict these cancers in earlier stage.

VIII. EXPERIMENTAL RESULTS

The experimental results are segregated into three parts:-

The first is the frequent and significant pattern discovery.

The second is mapping the cancer to its cluster and

The third is prediction by giving risk score as output.

At the beginning all the input data is stored in the non-cancer cluster further it gets classified and clustered by the model. A single user input data is fed into the system and gets classified according to the significant pattern to which it matches through decision tree, gets analyzed for its risk score merged with either one of the non cancer and cancer clusters. This gives the outcome whether the patient has cancer or not. Again the data is merged with any one of the subsequent cancer clusters to which its symptoms belong. The type of cancer the patient has is found out from this step. It is also compared

with the entire database to find its exact or relevant match so that a data with severe cancer related symptoms gets a pair only in the cancer cluster and it cannot get merged with non-cancer cluster even by mistake. This step ensures the accuracy of the model. With each new entry getting appended to the model the process becomes intelligent and ensures accurate results. The front end user interface is designed in a user friendly manner to help people use the system without any hassles.

Fig 3: User Input Screen for Patient value to detect Lung Cancer

Fig 4: Prediction Screen with Genetic Predicted Results for lung cancer

Fig 5: Report Screen with Results for lung cancer

The report shows the cancer status of a patient whether he has cancer or not by matching his data with the entire database. His risk score generated by the significant pattern mined by decision tree, the type of cancer he has which is given as a cluster output, whether his risk status is medium or severe and finally some recommended tests by medical experts to confirm the presence of cancer. This application is directly linked with the knowledge base and the back end model so that it could send the new raw data to the storage unit as well as the model to process it through analyzing the risk scores and also compares the data with existing cases in the knowledge base.

Fig 6: User Input Screen for Patient value to detect Oral Cancer

Fig 7: Prediction Screen with Genetic Predicted Results for Oral cancer

Fig 8: Report Screen with Results for ORAL cancer

IX. SUMMARY & CONCLUSION

The evaluation of the model is done using genetic algorithm to analyze its accuracy rate. The performance of model is evaluated using three statistical measures Accuracy, sensitivity and Specificity.

The correctly and incorrectly classified instances show the percentage of test instances. The percentage of correctly classified instances is called accuracy. Kappa is a chance-corrected measure of agreement between the classifications and the true classes. For a good model the Kappa statistics should be maximum, which shows the model exactly classifies the test instances as true classes. The mean absolute error gives the error percentage of the model. Root mean squared error, Relative absolute error, Root relative squared error is used to assess performance. Root relative squared error is calculated by dividing the Root mean squared error by predicting the mean of the target values.

For this module Genetic module and Decision tree produces the early and best accurate result and also it has found from literature that if p-53 gene loses then the person might be affected by cancer. Earlier GA is not used for big dataset as well as large number of attributes ,here genetic algorithm is used for big dataset and the results are also very much satisfactory than other classifiers and predictors.

In future the disease attributes should be more refined and list of added bio-markers may be incorporated more for more accurate result for prediction for the diseases. And not only cancer, heart disease, arthritis, other chronic disease should also be taken care of for earlier detection.

In future immunotherapy is the big expectation to detect against cancer, heart disease and arthritis. From gut bacteria to nano-drugs, from immune cells to plant derived drugs-Cancer research is throwing up a host of possibilities that are likely to get firmed up in 2016. The most promising is using the body's own immune system to fight against cancer cells by genetic tweaking. This allows body's Special Forces-T-Cells- to start recognizing cancer cells as enemies and targeting them. This has worked for blood cancers and is poised for breakthrough in solid cancers. Meanwhile gut bacteria have been shown to help fight cancer while conventional chemotherapy may be replaced by less toxic targeted nano-particles of drugs. In the war against cancer, 2016 may see significant ground gained.

REFERENCES

- [1] <http://www.cancer8.com/>
- [2] http://www.wcrf.org/cancer_facts/5-most-common-cancers.php
- [3] <http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer>
- [4] <http://www.medicalnewstoday.com/info/cancer-oncology/>
- [5] http://www.wcrf.org/cancer_facts/5-most-common-cancers.php
- [6] <http://www.scribd.com/doc/9522057/an-overview-of-soft-computing-with-its-techniques-and-application>
- [7] <http://dl.acm.org/citation.cfm?id=1539227>
- [8] <http://www.iconip2012.org/s02.pdf>
- [9] <http://www.biomedcentral.com/1471-2105/6/148>
- [10] <http://news.harvard.edu/gazette/tag/cluster-of-cancer-cells/>
- [11] <http://www.healthcommunities.com/cancer-treatment-and-care/cancer-staging.shtml>
- [12] http://en.wikipedia.org/wiki/Oncology#Progress_and_research
- [13] <http://www.cancer.net/patient/All+About+Cancer/Cancer.Net+Feature+Articles/Treatments%2C+Tests%2C+and+Procedures/Understanding+Cancer+Research+Studies%2C+Part+II>
- [14] K. Anuradha and K.Sankaranarayanan "International Journal of Advances in Engineering & Technology, March 2012" Vol. 3, Issue 1, pp. 84-91
- [15] M. E. Futschik, A. Reeve, and N. Kasabov, "Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue," *Artif. Intell. Med.*, vol. 28, pp. 165–189, 2003
- [16] Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm by Latha Parthiban and R.Subramanian in *International Journal of Biological and Life Sciences* 3:3 2007
- [17] Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers by Jyoti Soni, Uzma Ansari, Dipesh Sharma in *International Journal on Computer Science and Engineering (IJCSSE)*
- [18] Decision support system for heart disease diagnosis using neural network *Delhi Business Review* Vol. 8, No. 1 (January - June 2007)
- [19] Using neural networks to predict cardiac arrhythmias by Roland Adams, & Anthony Choi
- [20] Clinical Decision Support System : Risk level prediction of Heart disease using weighted fuzzy rules by P.K .Anuj in *Journal of King Saud University – Computer and Information Sciences* (2012) 24, 27–40
- [21] Modeling & Design of evolutionary network for heart disease detection by KS Kavitha and KV Ramakrishnan in *IJCSI*
- [22] A Data Mining Approach for Prediction of Heart Disease Using Neural Networks by Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte published in *IJCET* in 2012
- [23] Microarray gene expression profile data mining model for clinical cancer research by Rui Xue ; Dept. of Inf. & Comput. Sci., Univ. of Hawaii, Honolulu, HI, USA ; Jianying Li ; Streveler, D.J. published in *IEEE explore* in 2004
- [24] Decision Support in Heart Disease Prediction System using Naive Bayes by Mrs.G.Subbalakshmi, Mr. K. Ramesh, Mr. M. Chinna Rao
- [25] Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction by Jyoti Soni, Ujma Ansari

- [26] Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network by Shantakumar B.Patil, Y.S.Kumaraswamy
- [27] Prediction of ECG Signals Missing Parts Using Artificial Neural Network by Najmeh Mohsenifar, Ali Sadr in Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine Vol. 2 No. 8, November 2011
- [28] “Identification of suspicious regions to detect oral cancers at an earlier stage– a literature survey” by K.Anuradha.
- [29] “Unification of heterogeneous data towards the prediction of oral cancer reoccurrence” by Konstantinos.
- [30] “Survey of Human Cancer Classification using Micro Array Data” by G. Sophia Reena
- [31] “Artificial Neural Networks Applied to Survival Prediction in Breast Cancer” by M. Lundina J. Lundina H.B. Burked S. Toikkanenb L
- [32] “An Extensive Survey on Artificial Neural Network Based Cancer Prediction Using Soft Computing Approach” written by Manaswini Pradhan, Dr. Ranjit Kumar Sahu.
- [33] Khosla, R. and Dillon, T. (1997) Knowledge Discovery, Data Mining and Hybrid Systems. Engineering, IntelligentHybrid MultiAgent Systems, Kluwer Academic Publishers, Norwell, 143-177.
- [34] Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996) Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT Press, 1- 36.
- [35] Cios, K.J. (2001) Medical Data Mining and Knowledge Discovery. Studies in Fuzziness and Soft Computing, 60, 502.
- [36] K.R.Coelho, “Challenges in Oral Cancer Burden in India,” Journal of Cancer Epidemiology, vol. 2012, Article ID 701932, 17 pages.
- [37] Elango, J.K., Gangadharan, P., Sumithra, S. and Kuriakose, M.A. (2006) Trends of Head and Neck Cancers in Urban And Rural India. Asian Pacific Journal of Cancer Prevention, 7, 108-112.
- [38] American Cancer Society (2012) Cancer Facts and figures, Atlanta (GA), The Society