Detection of Oral Cancer Using Deep Neural Based Adaptive Fuzzy System in Data Mining Techniques

K. Lalithamani, A. Punitha

***Abstract---** Cancer has the highest growth rate among all diseases globally. Oral cancer is one of the most dangerous cancer which affects and originates from the oral cavity and neck. Overuse of tobacco and smoking cigarettes are the primary risk factor for developing oral cancer. Another habit which has a strong association with oral cancer is the consumption of alcohol. A large number of patient deaths were recorded from oral cancer as a result of lack of its identification and late treatment. Researchers in the medical community are making an effort to provide a system for effective diagnosis and prevention of the serious disease. In the present research, oral cancer patients can be identified through the use of data mining technology which includes detection, classification and clustering. A Deep Neural Based Adaptive Fuzzy System (DNAFS) is proposed in this paper which uses machine learning for the detection and identification of oral cancer. The two techniques which are part of DNAFS are based on fuzzy logic and neural networks. DNAFS and methods for data mining are explored for the identification of suitable techniques which are helpful in classifying data efficiently. The stages included in the proposed mechanism include data collection, pre processing, Fuzzy C-Means for clustering data, feature selection, classification and identification. Meaningful relationships can be extracted effectively from data using data mining techniques. About 96.29 % accurate results are available from experiments. There is less than 5 ms incidence of error in the result. The datasets are required to be investigated further in daily clinical practice.*

***Keywords---** Data mining, feature selection, Machine Learning, Fuzzy C-means, Oral cancer, medical data set.*

I. INTRODUCTION

Cancer leads to the most deaths globally. It is possible to easily reduce deaths caused by cancer, by detecting it early and enforcing preventive measures. Environmental and genetic factors play a significant role in devising new methods for the detection and prevention of cancer. The approach for detecting and preventing cancer which is described here is based on an analysis which uses association rule mapping. The data for analysis is related to history of addiction, clinical symptoms, survivability and co-morbid conditions related to cancer patients. It is possible to discover patterns in large sets of data through data mining which is a computational process. Data mining uses a combination of techniques related to database systems, statistics, Artificial Intelligence and machine learning. Data mining helps to extract and transform data in a data set.

Large and complex data in medical data sets can be analyzed using machine learning techniques and algorithms. In the recent times, a number of researchers have employed machine learning approaches in the Healthcare industry.

One of the data mining technologies in use is clustering which helps to detect and classify potential non-cancer and oral cancer patients. The focus of the present paper is to design and execute an oral cancer identification and detection system using Deep Neural based Adaptive Fuzzy System (DNAFS). DNAFS uses both fuzzy logic and deep neural network based algorithm. The processing is handled by first initiating data collection by gathering online information for the medical data set. This is followed by preprocessing which helps in data cleaning, removal of unwanted data and empty values. Next, data is clustered by Fuzzy C-Means clustering for further processing. Data patterns can be identified by feature selection. This process helps highlight similarities and differences. Next, the methods in DNAFS and data mining are explored for identifying appropriate techniques and methods for classifying data efficiently. Finally, data is classified using DNAFS, a machine learning technique which is a very popular classifier in the field of machine learning related to detection of oral cancer.

A popular classifier used in a number of fields pertaining to Medical Research is DNN which is simple as well as easy to implement. Neural networks are based on the functions and structure of biological neural networks which consists of neurons. Fuzzy logic pertains to the classification of huge amounts of data for obtaining accurate results. In the current research, a rule-based classification technique is proposed for detection of oral cancer. This is one of the fastest fuzzy techniques based on setting rules for clinical processes and attributes pertaining to differentiation between oral cancer and normal cases. In the present study, about 94% accuracy in classification was obtained.

The organization of the paper is done in the following manner: In Section 2, Works by previous authors are explained in brief. Machine learning techniques which are proposed are presented in section 3. This section presents an overview of preprocessing phase of data clustering feature selection and approaches in machine learning in the context of classification processes. Section 4 covers a description of experimental results. In section 5, a conclusion is presented and future research is proposed.

II. RELATED WORK

Clustering pertains to the assignment of objects to its two groups known as clusters.

Manuscript received February 01, 2012. (Fill up the Details)

K. Lalithamani, Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India. (e-mail: klalithamani1948@gmail.com)

Dr.A. Punitha, Research Guide, Bharathiar University, Coimbatore, Tamil Nadu, India. (e-mail: apunithaganesh@yahoo.com)

There are two categories of clustering algorithms. These are hard clustering and soft or fuzzy clustering. In hard clustering, data is divided into clusters such that every data element belongs to only one cluster. These algorithms have achieved good results for a number of real world applications [1]. In this phase, the extracted feature is aligned in the form of input to the proposed Fuzzy C-means based Radial Basis Function Neural Network for the classification of brain tumors. K-Means, KNN and Fast Fuzzy C-Means clustering techniques were used to cluster malignant as well as benign tumors. To better understand the result, a comparison was made between the performances of the Fuzzy C Means, Fast Fuzzy C-Means and KNN algorithms. Improved classification result was obtained by the hybrid model proposed in the study in comparison with the conventional algorithms in use in the past [2]. In the medical science field, the classification and clustering of cancer data has shown good results. Accurate results can be obtained through the use of the Fuzzy C-means and K-means algorithms in the proposed solution. The present paper addresses the problem related to classification of cancer data using the methods and information derived from the phases of testing and training [3]. K-means and DBSCAN (Density Based Spatial Clustering of Applications with Noise) clustering algorithms were compared further for their performances using silhouette scores. The K-means algorithm was initially analyzed with the help of different clusters and distance metrics. Later on, the DBSCAN algorithm was analyzed using different minimum points that are required to form clusters and distance metrics. On the basis of this analysis, the comparison of the performance of the K-means algorithm was better than DB scan. Specifically, better values were achieved in terms of execution time and clustering accuracy [4]. The way in which groundwater affects human health was analyzed by Balasubbramanian and Umarani [5] using the K-means clustering algorithm. In this case, the risk factors for content of fluoride in water were analyzed. Meaningful hidden patterns were revealed with the help of this analysis which can assist in reasonable decisions on a community level. Banu and Jamala [6] devised an algorithm for predicting Heart Attack through fuzzy C-means. This algorithm uses unsupervised clustering where one specific data object can belong to more than a single cluster. In this case, the system can help physicians with the efficient diagnosis of heart attack. The bioprofile concept was used in the work of Escudero et al [7]. K-means algorithm for clustering was used for detecting animal disease early and its classification into non-pathologic and pathologic groups. The clustering algorithm K-means-Mode was proposed by Paul et al. using medical data [8]. Researchers indicated that the medical domain knowledge with respect to clustering tends to improve the algorithmic performance. Towards this end, the clustering algorithm performance is evaluated with either Neural Gas [11], Hard C-Means[9] or Fuzzy C-Means [9] for the detection of tumors. Three different steps are used in this case. In the first step, the three algorithms are evaluated under the conditions of noise. Later, tumor segmentation results are compared with region growing algorithm. In the end, results are manually compared with segmentation results.

Experimental data as well as ANFIS (Adaptive Neuro Fuzzy Inference System) values are compared using this approach with respect to bell shaped and triangular membership functions [13]. The basis of the model developed here is the first order Takagi-Sugeno and Kang. This model showed increased accuracy of prediction when using the bell membership function. A novel problem solving approach is devised for air conditioning systems through Sugeno and Mamdani type inference models which use fuzzy logic. This was discussed by Kaur A et al. [12]. The primary differences between Sugeno type FIS (Fuzzy Inference System) and mamdani type FIS are mentioned using this approach. An enhanced form of membership is chosen using the study with respect to FIS in the context of the air conditioning system. On the basis of the implementation of the inference system, the author made a conclusion stating that Sugeno type and Mamdani type conditioning system performs in a similar manner. However, the use of Sugeno type FIS model enables the air conditioning system to operate at full capacity [12]. Tamer introduced the applications of ANFIS with respect to medical diagnosis. The ANFIS system of Sugeno type assists in the prognostication of mycobacterium tuberculosis existence. About 503 different patient records were collected which were part of a private health clinic. There are 30 different attributes and patient records covering medical test and demographic data [14]. About 250 records were used to generate the ANFIS model. Classification of instances of the proposed model is done with 97% accuracy. The rough set algorithm performs a similar calculation but with only 92% accuracy. This information contributes forecasting of patients before medical tests [15].

Multi Ranked Feature Selection Algorithm (MRFSA) is a new algorithm for effectively selecting features for use in the new system for detecting cancer. A Neuro Fuzzy Temporal Classification Algorithm as used to test how the feature selection algorithm performs with respect to the accuracy of classification. The classification accuracy obtained in this manner has promising results in comparison with existing works. These insights are based on the results pertaining to the same data set [16].

The performance of different feature sets can be compared and evaluated using a novel Intelligent Correlated Fuzzy Neural Network (ICFNN). The combination of GLRLM (Gray Level Run Length Matrix), GLCM (Gray Level Co-Occurrence Matrix) and intensity based first order features helps to improve the accuracy of classification. In this context, 61 important features from 16 patients and their 192 images are taken into consideration. These type of malignancies in the context of classification assist during the prognosis and treatment of oral form of cancer. Results of experimentation indicate the performance of the system of diagnosis [17]. A number of data mining algorithms for oral cancer data have been applied by Jatasuji and Rajagopalan [18]. These algorithms have been used to classify the NMDS (Non-Metric Multi Dimensional Scaling) dataset and

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

researchers have accessed the performance of those algorithms. In this respect, researchers have been able to achieve about 100% accuracy with the C4.5 algorithm, 98.7% with Random Forest Tree algorithm, and 99.5% accuracy with MPNN (Memory Pulsed Neutron-Neutron). These algorithms are also capable of separating oral cancer data set as per non-cancer or cancer patient records.

An approach for detecting and preventing oral cancer was introduced by Neha et al. [19]. The basis of this approach was association rule mapping. The data analyzed was related to comorbid conditions, clinical symptoms, history of addiction and survivability of patients diagnosed with cancer through the use of association rules. Association rules extracted in this way are useful for clinical decision making for diseases. Fuzzy rule-based classification was used by Mohammadpour et al. [20] for the prediction of CAD (Computer Aided Design). The authors are of the opinion that fuzzy classification which is based on the if-then rules can achieve an accuracy of 92.8 %. However, only 91.9% classification accuracy was obtained using equation. The evaluation of Fuzzy rules took place to obtain a reduction in space dimension for classifier training. This also helped improve the overall accuracy. Fuzzy rules which are based on medical experts' approach for the detection of CAD were formulated by Pal et al. [21]. The method proposed here helps doctors during identification and risk prediction with respect to CAD patients. Ten rules were made and every block of rule depicts a module. Every module has a single risk factor. The system developed in this way leads to 83.33 % specificity and 95.85% sensitivity with respect to risk computation for CAD.

III. PROPOSED WORK

3.1 Overview

Discovering patterns computationally in large and complex data sets is referred to as data mining. This method is a combination of machine learning, database systems, artificial intelligence, and statistics. The objective of the process of data mining is the extraction of information from datasets and applying transformations to obtain a structure that is meaningful in future. Large medical data sets are the best candidates for this classification method.

Several data mining techniques are implemented in unison to diagnose and prognose oral cancer for a specific patient. The exploration of DNAFS and data mining methods is done for the identification of appropriate techniques and methods that efficiently classify data. In the end, classification is done using Deep Neural Network Based Adaptive Fuzzy System (DNAFS) which is a machine learning technique. The prognosis of the affected patient determines whether successful treatment for the diseases is possible. In this context, information for prognosis is generated with the help of a statement of prognosis.

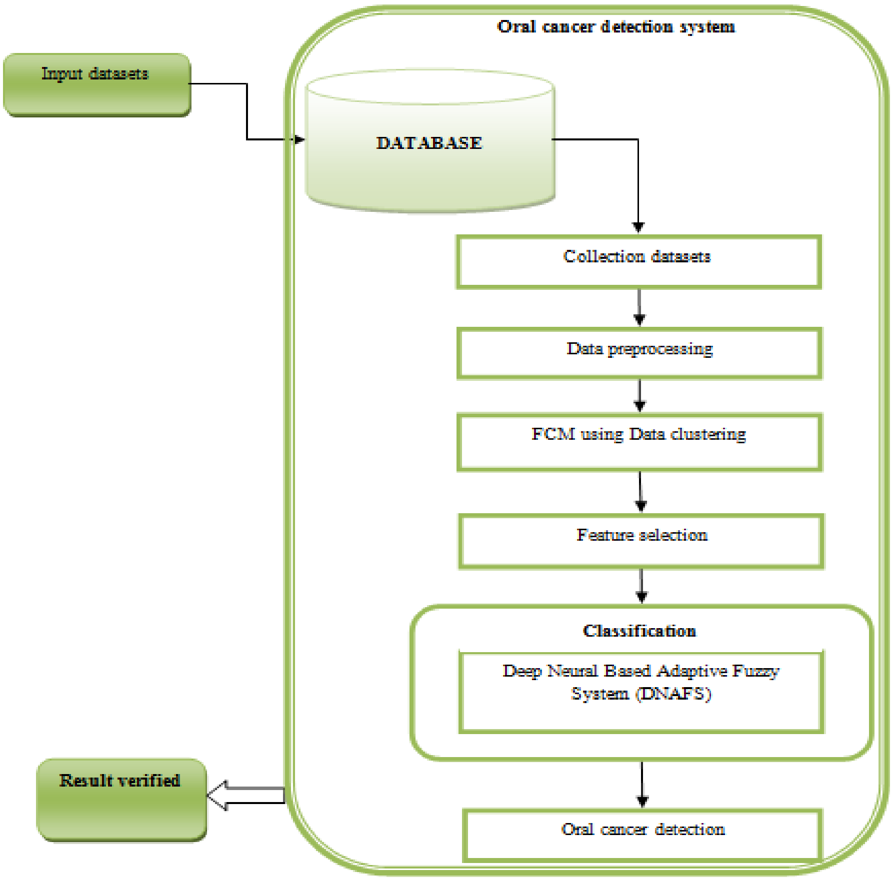

Figure 1: Overall Proposed Architecture

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

3.2 Dataset

There are two sub datasets for this dataset. The first data set has samples from healthy patients and other samples from cancer patients who were diagnosed of cancer of the oral cavity. The second data set has samples of healthy patients and samples of a healthy oral cavity in cancer patients. This division helps obtain an indication.

It contains the standard index of oral information for tolerant points of interest for oral cancer. Standard highlights are incorporated into this information. The oral cancer information is collected from a specific clinic or different disease organizations. A range of oral cancer data sets are involved here such as cerebrum, dental, mouth, and neck.

Table 1: Input of Attributes (Parameters)

Demographic Features	**Clinical Features**
H1: Age	H8: Bleeding
H2: Gender	H9: Mouth Burning Sensation
H3: Ethnicity	H10: Difficulty in Swallowing/Chewing
H4: Quid Chewing	H11: Lose of Appetite
H5: Tobacco Chewing	H12: Lose of Weight
H6: Alcohol Drinking	H13: Loosening of Tooth
H7: Tobacco Smoking	H14: Swelling
Histopathological Features	H15: Site (BM, LA, RMT, LIP, Tongue, UA, Plate)
H20: Squamous Cell Carcinoma	H16: Size in cm (< 2, 2 to 4, ≥4)
H21: Variant of SCC (Verrucous, Adenoca, Basaloid, Plaque Like, Sarcomatoid, Acantholytic , Lympheopithelioma like	H17: Hypertension
H22: Benign	H18: Diabetes
H23: Predisposing factor (Leukoplakia, Erythoplakia, Submucous Fibrosis, Linchen Planus, None	H19: Hoarseness in Voice

Medical data sets pertaining to oral cancers are obtained from UTI medical data sets.

3.3 Preprocessing

Data mining techniques are used in the preprocessing stage to analyze a large data set and identify target data. The preprocessing stage consists of a number of tasks such as cleaning of data, its integration, transformation, reduction and discretization. In the data cleaning stage, noise is eliminated and data is made consistent and coherent.

Absent values are included in this process and outliers are also found.

- ❖ Data Integration: This is performed with the help of data cubes or files, and many different databases.
- ❖ Data Transformation: This refers to normalizing and aggregating data.
- ❖ Data Reduction: This refers to a reduction in volume but the production of similar results of analysis
- ❖ Data Discretization: This refers to reduction of data and replacement of its numerical attributes with nominal attributes.

Data in the real world is incomplete, inconsistent, and noisy. Preprocessing of data is a regular attempt for filling values which are missing and smoothing out noise. In this stage, outliers are identified and inconsistencies in data are corrected. The missing values are filled manually or through the use of a global constant.

3.4 Fuzzy C-Means clustering

Clustering refers to a task when objects are grouped so that similar objects belong to the same group or cluster. This is the primary task when it comes to exploratory data mining. It is a common analysis technique for statistical data and use in many fields such as pattern recognition, bioinformatics, image analysis, information retrieval, machine learning, computer graphics and data compression. One of the common descriptive tasks is clustering, where a finite set of clusters are identified for data description. Grouping elements with similar characteristics is part of the clustering process. The mean of each cluster is used in the technique. Similar data values are grouped in one cluster. The present paper makes use of Fuzzy C- Means clustering technique.

- ❖ FCM or Fuzzy C-Means clustering is also referred to as Fuzzy for the medical data.
- ❖ Fuzzy Partitioning is part of FCM. In this case, specific data can be part of all groups having different grades of membership between 0 and 1.
- ❖ The nature of FCM is iterative. Its objective is to find centroid or cluster centers so as to minimize the function of dissimilarity.

3.5 Algorithm

1. Initialize M=[s_{ij}] matrix, $M^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)} = [c_j]$ with $M^{(k)}$

$$C_j = \frac{\sum_{i=1}^{n} s_{ij}^{m} y_i}{\sum_{i=1}^{n} s_{ij}^{m}}$$

3. Update $M^{(k)}$,$M^{(k+1)}$
4. $d_{ij} = \lVert y_i - c_j \rVert$, $d_{kj} = \lVert y_k - c_j \rVert$,

$$s_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}$$

5. If $\| M^{(k+1)} - M^{(k)} \| < \in$ then Stop; Else

Return to step 2 ;

Here M is any real number greater than 1,
s_{ij} is the degree of membership of y_i in the cluster j,
y_i is the ith of d-dimensional measured data,
c_j is the d-dimension center of the cluster,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The algorithm assigns membership to every data point pertaining to a cluster centre. This assignment is based on the distance between the data find point and the cluster centre. If there is more data in the proximity of the cluster centre, then it means that its membership towards that cluster centre is greater.

3.6 Feature selection

In feature selection, only certain parameters belonging to a set of features are chosen to improve the accuracy of classification. On analyzing, it is found that feature selection is a requirement prior to studies connected to medical data for oral cancer. The technique of feature selection detects attributes which are relevant and result in accuracy of the classifier. Feature selection can be used to identify data patterns and highlight the differences and similarities. This is in contrast with other techniques which use information belonging to different classes.

3.7 Classification

One of the important modules of this project is classification. The data set applied for input is trained and then fed into the process of classification. In the process of classification, the original data is processed from Medical data analysis in Deep Neural Based Adaptive Fuzzy System. In the field of machine learning, one of the simplest classification algorithms is DNAFS.

3.8 Deep Neural based Adaptive Fuzzy System

A feature set which is extract and select from a given data set can use techniques such as data mining and machine learning techniques for the classification of oral cancer. DNAFS has far surpassing classifier capabilities and is in wide use in a number of fields connected with medical research. This is because it is simple and its implementation is easy. The basis of DNAFS is the function and structure of biological neural networks whose building blocks are neurons. A solution is computed using this algorithm similar to the way in which human brain functions. It is observed that the implementation of DNAFS is successful in the context of abnormalities related to oral cancer. The basis of this learning technique is a rule based system devised using fuzzy logic. This system is capable of predicting oral cancer in a non-invasive manner on the basis of risk index of oral cancer and other clinical parameters. The fastest Deep Neural Based Adaptive Fuzzy System was proposed in this context to help predict oral cancer where rules are set through clinical attributes. These attributes are processed for normal and oral cancer differentiation. As much as 96.55% accuracy was obtained with success through the design of the procedure of classification.

Learning techniques that are neuro adaptive provide a method that enables a procedure for fuzzy modeling to help learn the data set information. The parameters of the membership functions are computed so that the inference system associated with fuzzy helps track the data given for input or output. The parameters which are related to the membership functions are modified through the process of learning. A number of approaches are useful in coping with real world problems in an efficient manner. This means that the task related to the learning algorithm pertaining to this architecture assists in tuning with all parameters that can be modified. Other tasks involve formulating output for DNAFS that matches with the training data. This also helps in achieving an improvement in the rate of convergence.

3.9 Design the DNAFS

- Initialization
  - Define number and type of inputs
  - Define number and type of outputs
  - Define number of rules and type of consequents
  - Define objective function and stop conditions
- Collect data
- Normalize inputs
- Determine initial rules
- Initialize network
- Train

3.10 Algorithm Deep Neural based Adaptive Fuzzy System

This is an inference system wherein every rule output is obtained by applying a linear combination to two variables applied for input and adding a constant term to the result. A weighted average of rule output is taken to compute the final output.

The Deep Neural Based Adaptive Fuzzy system works with multiple layers. Every layer is a different function in this classification technique.

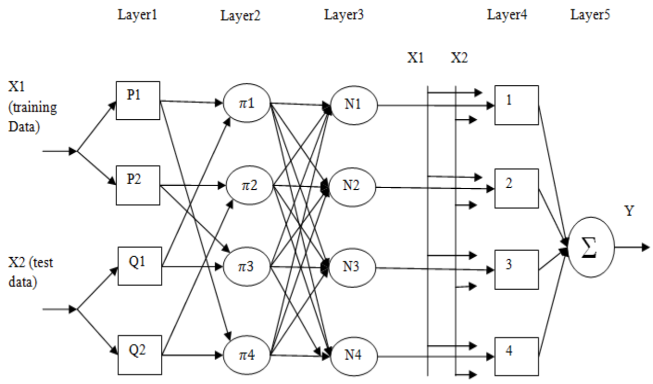

Figure 2: Design of Deep Neural based Adaptive Fuzzy System

Step 1: **Fuzzification (1st Layer):** All nodes in this layer are adapted with respect to unknown function. Every node is depicted by i. Membership values are produced by the nodes.

$$\mu(x) = \frac{1}{1+\left|\frac{x-AM}{SD}\right|^2}$$

Here the input is x which is supplied to the node i. The associated linguistic variable is P_i which is connected to the node function and the membership function is depicted by μ_i which pertains to P_i . μ_{P_i}→Input parameters, like some attributes.

Step 2: we presented simple formulas to compute similarity between two training datasets and testing datasets. For each formula, we apply an appropriate strategy to compute the overall score:

X1=P(A,B,C….). and X2=Q(A,B,C….). Symptoms of attributes values

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\ \ \sqrt{n\sum y^2 - (\sum y)^2}} \quad , x = X1 \;\&\; y = X2$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Perform the Karl Pearson coefficient of correlation between X1 and X2. Determine: P_i to the term occurrence for all the attributes in training datasets X1 and Q_i to the term occurrence for all the attributes in test datasets X2.

Step 3: **Production (2nd Layer):** Every node belonging to this layer is a fixed node. The node helps calculate firing strength indicated by w_i which pertains to a specific rule. The output for every node as a product calculated from the incoming data is given by,

$$Y_i^3 = \prod_{j=1}^{k} X_{ji}^3 \quad i = 1,2,..$$

$$Y_{\prod 1}^3 \ (or\ ) w_1 = \mu_{P_1} * \mu_{Q1}$$

Step 4: **Normalization (3rd Layer):** All nodes belonging to this layer are fixed in nature. The ith node helps in the computation of the ratio of the firing strength of the ith rule versus the total of all firing strengths for all rules. The ith node output pertains to the firing strength which has been normalized and is given by,

$$Y_i^4 = \frac{X_{ii}^4}{\sum_{j=1}^{n} X_{ji}^4} = \frac{\mu_i}{\sum_{j=1}^{n} \mu_j} = \overline{\mu_i} \ , \quad i = 1,2..$$

$$Y_{N1}^4 = \overline{w_i} = \frac{\mu_1}{\mu_1 + \mu_2 + \mu_3 + \mu_4}$$

Step 5: **De-fuzzification (4th Layer):** This layer consists of adaptive nodes which consists of a node function and is given by

$$Y_i^5 = x_i^5[s_{i0} + s_{i1}X1 + s_{i2}X2]$$
$$= \overline{w}_i(s_{i0} + s_{i1}X1 + s_{i2}X2), \quad i = 1,2..$$

Here, the output for layer 5 is given by $\overline{w_i}$ and the consequential parameter set is depicted by $\{s_{i0}, s_{i1}, s_{i2}\}$.

Step 6: Summarization (5th Layer): Only a single fixed node is contained in this layer which helps in the calculation of the overall output. This is the sum of all incoming data.

$$Y = Overall\ Output = \sum_{i=1}^{n} X_i^6$$
$$= \sum_{i=1}^{n} \overline{w_i}\, [s_{i0} + s_{i1}X1 + s_{i2}X2]$$

The forward pass of the learning algorithm identifies resultant parameters using the least squares estimation method. The backward pass consists of error signals in the form of derivatives of the squared error. These are computed for the output of every node and propagate in a reverse manner from the output to the input layer.

3.11 Evaluation metrics

Measurement and performance of stability of the system is evaluated by calculating and analyzing some parameters. A few of these parameters are mentioned in the following paragraph.

The Deep Neural Based Adaptive Fuzzy system proposed here is evaluated for performance using the sensitivity, precision, specificity, probability of misclassification error (PME), and F-score. Further, the parameters of mean square error, accuracy of the overall training set, testing set and overall performance is made using Eqs. (1-7). Here, Y_i is the actual parameter and R_i is the result pertaining to the ith diagnosis for the attribute obtained for oral cancer (num). Non-disease patients who were also found to have no disease are true negatives (TN). However, non-disease patients who are found to have a disease are predicted to be False Negative (FN). Patients with the disease and were also found to have the disease are the True Positive (TP). Finally, the predicted patients with disease that were found to have no disease are the False Positive (FP).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - R_i)^2} \qquad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (2)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (3)$$

$$Precision = \frac{TP}{TP + FP} \qquad (4)$$

$$F - score = 2 * \frac{Recall * Precision}{Recall + Precision} \qquad (5)$$

$$PME = \frac{FP + FN}{TN + FP + TN + FN} \qquad (6)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (7)$$

IV. RESULT AND DISCUSSION

There are a number of benefits of using techniques in data mining in Medical Systems. These techniques have contributed to a large degree in the field of Medical Science. The present review gives an explanation of the oral cancer classification based on artificial intelligence (AI) and data mining methods. Further, a review of literature indicates that several improvements can be made in the classification of oral cancer. A suitable method for extraction of hidden factors must be devised from the data sets of oral cancer. Since medical data sets have irregular attributes, this process is very complex. Data mining technologies such as clustering, prediction, and classification are used in this research to identify potential patients who may have cancer. The Deep Neural Based Adaptive Fuzzy System (DNAFS) is developed which can predict and detect pathological conditions for oral cancer accurately. UCI data sets were used to collect the required data for this experiment.

Table 2: Comparison Of Clustering Techniques

S.No	Clustering Techniques	Accuracy (%)
1	Artificial immune system	84.5
2	Gaussian mixture models	87.12
3	Fuzzy C-Means	91.04

In table 2, the proposed techniques are compared with existing techniques for their performance. Existing techniques include Gaussian mixture models and the artificial immune system.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fuzzy C-Means is a proposed system which gives efficient results based on the overall performance and is better than other techniques.

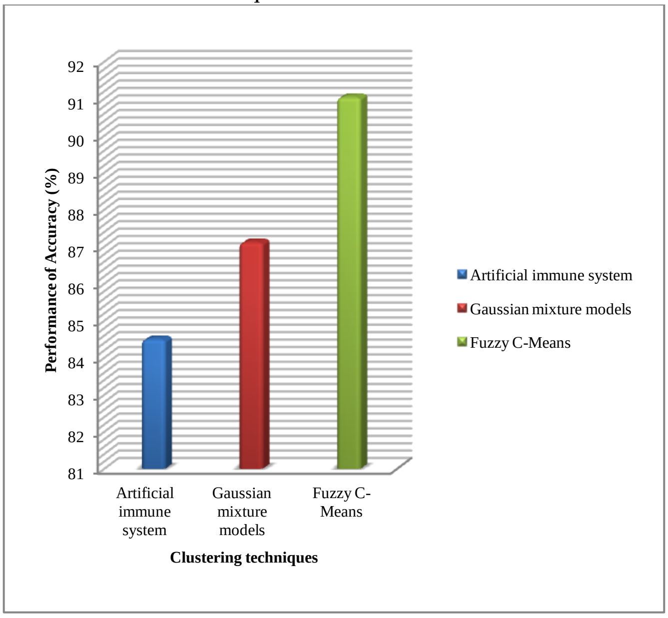

Figure 3: Performance Results of Proposed and Existing Techniques

The figure 3 mentioned above produces evaluation of the clustering techniques which are compared with the Gaussian mixture model and artificial immune system techniques. A high level of clustering accuracy in the results can be achieved by using Fuzzy C-Means clustering.

Table 3: comparison of classification techniques

S. No	Classification Techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)	Time (Ms)
1	SVM	88.30	90.30	93.00	0.43
2	Bayes Net	88.54	91.80	95.90	0.78
3	Decision Stumb	83.76	86.00	91.43	0.67
4	AdaBoostM1	85.34	84.00	89.06	0.945
5	KNN	89.49	85.67	87.00	0.988
6	DNAFS	96.29	94.9	93.4	0.31

In table 3 and figure 4 the performance of SVM, Bayes Net, Decision Stumb, AdaboostM1 and KNN are compared with the performance of Deep Neural Based Adaptive Fuzzy system. The proposed system is clearly better than other techniques existing today. Improved results are obtained with respect to time values, specificity and sensitivity.

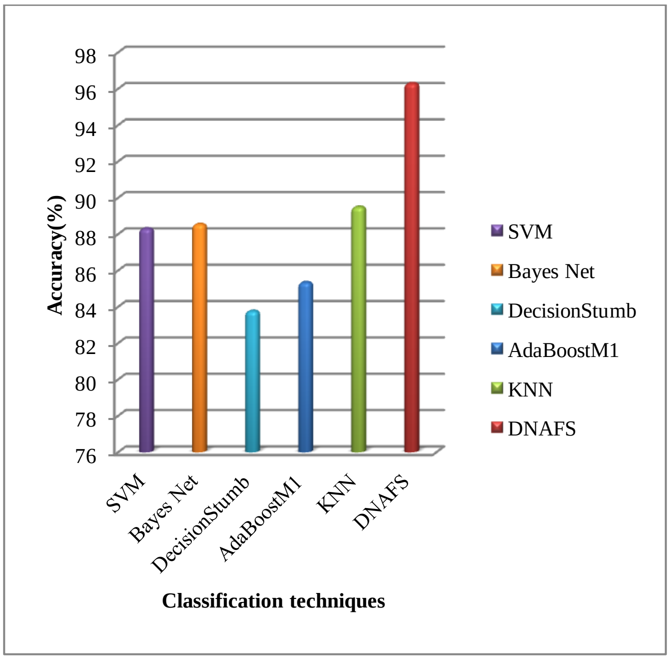

Figure 4: Performance of accuracy result

V. CONCLUSION

An algorithm for the diagnosis of oral cancer is proposed in this paper which is based on medical data sets of oral cancer. The proposed algorithm is trained and tested using the medical database of oral cancer. The obtained features are used to train DNAFS. A high level of accuracy is achieved and the system is capable of handling low energy events, using large data sets, and tackling noise. The classifier has certain limitations and the diagnosis and detection of oral cancer may be affected in terms of its accuracy. The performance of the DNAFS classifier is more efficient than any other classifier which can detect oral cancer automatically. A number of results were achieved from the algorithm on the basis of accuracy, cost, performance, and speed. Further, classifying data effectively assists in finding the right treatment plan for a patient. A superior method can be used in future for the prediction of oral cancer by making improvements in existing methods.

REFERENCES

1. R. Suganya, R. Shanthi "Fuzzy C- Means Algorithm- A Review "International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012 1 ISSN 2250-3153.
2. Gopi Krishna., Sunitha K.V.N and Mishra S," BRAIN TUMOR CLASSIFICATION USING HYBRID Fuzzy C MEANS BASED RADIAL BASIS FUNCTION NEURAL NETWORK", International Journal of Recent Scientific Research Vol. 9, Issue, 3(G), pp. 25119-25125, March, 2018.
3. Mr.S.P. shukla and Mrs. Ritu Dwivedi," Clustering and Classification of Cancer Data Using Soft Computing Technique", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. I (Jan. 2014), PP 32-36.

4. Godwin Ogbuabor, and Ugwoke, F. N," Clustering Algorithm For A Healthcare DATASET USING SILHOUETTE SCORE VALUE", International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018.
5. Balasubramanian, T., & Umarani, R. (2012, March). An analysis on the impact of fluoride in human health (dental) using clustering data mining technique. In Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on (pp. 370-375). IEEE.
6. Banu G. Rasitha & Jamala J.H.Bousal (2015). Perdicting Heart Attack using Fuzzy C Means Clustering Algorithm. International Journal of Latest Trends in Engineering and Technology (IJLTET).
7. Escudero, J., Zajicek, J. P., & Ifeachor, E. (2011). Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 6470-6473). IEEE.
8. Paul, R., & Hoque, A. S. M. L. (2010, July). Clustering medical data to predict the likelihood of diseases. In Digital Information Management (ICDIM), 2010 Fifth International Conference on (pp. 44-49). IEEE.
9. Barrios JA, Villanueva C, Cavazos A, Colas R (2016) "Fuzzy C-means Rule Generation for Fuzzy Entry Temperature Prediction in a Hot Strip Mill". J Iron Steel Res Int 23: 116-123.
10. Krishna PG, Bhaskari DL (2016). "Fuzzy C-Means and Fuzzy TLBO for Fuzzy Clustering." Proc Second Int Conf Comp Comm Technol: 479-486.
11. Ghesmoune M, Lebbah M, Azzag H (2016). "A new growing neural gas for clustering data streams." Neural Networks 78: 36-50.
12. Kaur A., Kaur A.(2012), "Comparison of Mamdani-Type and Sugeno-Type Fuzzy Inference Systems for Air Conditioning System," International Journal of Soft Computing and Engineering, ISSN: 2231-2307, no. 2, pp. 323-325.
13. Efosa C., Akwukwuma N.(2013), "Knowledge based Fuzzy Inference System for Sepsis Diagnosis," International Journal of Computational Science and Information Technology, Vol.1, No. 3, pp. 1-7.
14. Choi H., Yoo H., Jung H., Lim T., Lee K., Ahn K.(2015), "An ANFIS-based Energy Management Inference Algorithm with Scheduling Technique for Legacy Device," International Conference on Artificial Intelligence, Energy and Manufacturing Engineering, pp. 80-82.
15. Uc T., Karahoca A., Karahoca D. (2013), "Tuberculosis disease diagnosis by using adaptive neuro-fuzzy inference system and rough sets," Neural Comput & Applications, Springer, pp. 471-483.
16. R Jaya Suji, SP Rajagopalan," Multi-ranked feature selection algorithm for effective breast cancer detection", Biomed Res- India 2016 Special Issue Special Section: Computational Life Science and Smarter Technological Advancement.
17. R.Jaya Suji , Dr.S.P.Rajagopalan," An Intelligent Oral Cancer Diagnosis System using Texture Analysis based Segmentation and
18. Correlated Fuzzy Neural Classifier Fuzzy Rough Set and LS-SVM", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 6 (2015).
19. Jaya Suji. R, Dr.Rajagopalan S.P, "An automatic Oral Cancer Classification using Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol.2, No.10, 2013.
20. Neha Sharma, Hari Om, "Extracting Significant Patterns for OralCancer Detection Using Apriori Algorithm", Intelligent Information Management, Vol. 6, pp. 30-37, 2014.
21. R. A. Mohammadpour, S. M. Abedi, S. Bagheri, and A. Ghaemian, "Fuzzy rule-based classification system for assessing coronary artery disease," Computational and Mathematical Methods in Medicine, vol. 2015, Article ID 564867, 8 pages, 2015. View at Publisher · View at Google Scholar · View at Scopus
22. D. Pal, K. Mandana, S. Pal, D. Sarkar, and C. Chakraborty, "Fuzzy expert system approach for coronary artery disease screening using clinical parameters," Knowledge-Based Systems, vol. 36, pp. 162–174, 2012. View at Publisher · View at Google Scholar · View at Scopus.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication