

# ORAL CANCER DIAGNOSIS BASED ON FEATURE SELECTION AND CLASSIFICATION USING E-SVM ALGORITHM

*K. Lalithamani, Research Scholar, Bharathiar University, Coimbatore, India..*

*Dr.A. Punitha, Research Guide, Bharathiar University, Coimbatore, India.*

## ABSTRACT

Oral cancer analysis and its prediction tends to be a crucial area of research when concerning the safety of human life. Various Machine Learning (ML) techniques are adopted for analyzing and predicting the cancer data. The accuracy attained by predicted result is stated within average % by various researches. But by utilizing ML techniques, accuracy of maximum % can be obtained on UCI (Unique Client Identifier) medical oral cancer based dataset. The most significant factor in the diagnosis of any disease is the 'Accuracy'. Therefore, it becomes mandatory to choose best suitable elements for producing optimal accuracy. The main objective of this paper is, to accurately predict the occurrence of oral cancer using minimum number of attributes. Altogether 25 attributes are detected for the prediction of oral cancer staging. The integrated diagnostic model with Principal Component Analysis (PCA) feature selection method is adopted here. With the help PCA feature selection algorithm 25 attributes are being cut down to only 10. Enhanced-Support Vector Machine (E-SVM) algorithm is proposed in this paper for oral cancer detection and identification in medical sector by employing ML approaches. E-SVMs are basically supervised learning models with associated learning algorithms that examines data which is being utilized for the purpose of classification and regression analysis. The components included in the proposed model are Data collection, Pre-processing, Feature selection via PCA, Clustering, Classification and Identification. The techniques of Data Mining (DM) aids in extracting significant patterns and relations from such data. Moreover, the ML and DM are effective techniques to detect and predict oral cancer at an early stage.

**Keywords:** *E-SVM, Oral Cancer, Data Mining, Machine Learning, Feature Selection via PCA, Clustering and Classification.*

## 1. INTRODUCTION

Oral cancer, resembled as a critical cancer of mouth and neck, widely spreading its root across the entire world. Detecting cancer accurately and at an early stage can be extremely valuable for saving patient's lives. Examining the gene expression data aids in identification and classification of cancer that accelerate appropriate selection of treatment and progress. Cancer has become a deadly and rampant disease affecting the entire human population and has heavily increased the death rate. Cancer forms due to ceaseless growth of cells in various body parts or tissues. It can erupt in any body organ and also spread to other organs too [1]. An individual's life can be protected, if cancer is detected at the earliest and is prevented to spread other body parts [2-3]. There exist various parameters by which a person can be susceptible to cancer disease. The most significant indicators being the education along with the involvement of parameters like, occupation and lifestyle [4].

This research incorporates oral cancer with knowledge discovery and DM. Data mining involves implementation of data analysis. Alternatively, Data mining signifies extracting

meaningful patterns or relations from available data, thereby utilizing it for building a predictive model. Knowledge discovery is carried out before the technique of DM and basically performs other processes that being: selection of target dataset, data pre-processing and data reduction (i.e. minimizing number of variables). With the application of ML techniques and algorithm on numerous medical datasets, evaluation of huge and complicated data can be easily automated. Finally, various researchers have adopted the techniques of Machine Learning, helping the health care sector and the medical practitioners in oral cancer diagnosis.

This paper is organized as follows. Section 2 presents work of earlier author. Section 3 illustrates the proposed ML technique for detection of oral cancer and outlook of various levels. Experimental outcome are depicted in Section 4. Finally, Section 5 presents the conclusion and suggests research work for future.

## 2. SYSTEMATIC LITERATURE SURVEY

Tumors developed at any given stage may lead to cancer cells growth. Cancer classification has turn out to be a challenging issue for the researchers till date. Cancer research begins with alterations in the tissues. The malignant growth of cells within the oral cavity leads to oral cancer. The research merges Fuzzy Cognitive Map (FCM) and SVM (Support Vector Machine) to grade oral extracted tumor. The histological attributes are utilizing, thereby the association among the concepts are being determined. For differentiating amidst benign and malignant cases, FCM is employed as a classifier. Thereafter, result from FCM acts as input for the SVM classifier. This leads to improvement in the prediction skills. The classification accuracy achieved for the proposed model is 92.10% in case of malignant and 94.11% in case of benign [5].

Right now, there's no such efficient technique which can diagnose and cure oral cancer at an early stage. A technique is proposed for detecting and preventing cancer depending on analysis using association rule mining. The data analyzed are related to clinical symptoms, history of addiction, co-morbid condition and survival chances of the cancer patients. The extracted rules aids in clinical judgments and taking correct decisions concerning the disease. The results generated are assuring and depicts the possible implementation of the technique in ultimate growth of diagnostic assay and cure along with adequate support and assurance for oral cancer detection at an early stage [6].

When compared, the SVM surpasses various other similar techniques. The research aims to evaluate the prediction-performance of SVM and SVM ensembles concerning datasets of both small and large scale breast cancer. The ROC (Receiver Operator Characteristic) classification accuracy, F-measure and computational times of training-SVM and SVM-ensembles are being compared. The output reveals that SVM ensembles relies upon the bagging method and SVM ensembles (that begin RBF (Radial Basis Function) kernel based) that relies on the boosting

method tends to be ideal options for a small scale dataset, where feature selection is carried out in the data pre-processing level. In case of large scale dataset, SVM ensembles (that being RBF kernel based) that relies on the boosting method yields better performance compared to rest of the classifiers [7].

SapiahSakri proposes that presence of extensive data and advanced methodologies makes the prediction pretty fast and accurate. The present research targets to compare the accuracy of prevailing data mining techniques for the prediction of breast cancer recurrence. The particle swarm optimization is embedded as a feature selection method used in three well-known classifiers, which includes: NB (Naive Bayes), KNN – (K-Nearest Neighbor) and fast Decision Tree learner, aiming to ascend the accuracy factor of the prediction model [8].

In [9] NN (Neural Networks), DT (Decision Tree) and NB (Naïve Bayes), the machine learning technique is adopted for the heart disease diagnosis. For achieving optimized feature selection genetic algorithm is used. The accuracy achieved is 100%, 99.62% and 90.74% correspondingly. In [10] prediction of heart disease detection is carried out by employing neural network using genetic-algorithm based feature extraction. Back-propagation based neural network weight optimization (using Genetic Algorithm) is formulated which yields in accuracy of 89% in the heart disease prediction. In [11] author builds HDPS (Heart Disease Prediction System) based on the mechanism of ANN using LVQ (Learning Vector Quantization). The resultant accuracy is around 80%, sensitivity - 85%, and specificity around 70%.

In [12] proposes diagnosis of heart disease by employing lazy data mining technique with data reduction methods that is PCA being utilized to obtain category association rules. The output exhibits that 14.8 yields 10.26 enhancement and 8.6% enhancement over NB (Naïve Bayes). In [13] put forth a HDPS (Heart Disease Prediction System) by employing DM approach using two additional factors that are (obesity and smoking) for increasing the prediction rate. NN, DT and NB were employed for heart disease prediction yielding in an accuracy of 99.25%, 94.44% and 96.66% respectively.

A web based application is being presented in [14] utilizing NB algorithm, that considers patient's symptoms and provide the diagnosis output to the concerned patient. In [15] the technique of association rule mining is adopted for performing diagnosis of diabetes. It was affirmed by the author that with proper implementation of the data mining approach, both the computation and classification performance can be enhanced and increased. The proposed rules are capable in improvising the expert system and aiding in effective clinical decision making. In order to predict diabetes on WEKA tool, the author in [16] illustrates a comparison among algorithms NB and DT thus obtaining system accuracy of around 79.56% and 76.96% respectively.

In [17], following algorithms of DT, NB, and NB are adopted for detection of liver disease using 10 features. The analysis of result in regard to accuracy of NB Tree algorithm yields in maximum accuracy. On the other hand, NB algorithm yields in better performance in regard to computational time. In [18] a comparative analysis is carried by the author on clustering and classification algorithms. According to the results it's revealed that the classification is superior compared to clustering algorithms projecting around 81% of accuracy.

In [19] the author proposes the techniques of classification with clustering i.e. KNN using Fuzzy c-means clustering and F-KNN using Fuzzy c-means clustering. From the output

generated its revealed that the F- KNN with Fuzzy c-means model yields in enhanced output compared to KNN with Fuzzy c-means model on the datasets of Liver-disorder.

Moreover, the Fuzzy c-means clustering algorithm for datasets pre-processing, improvised the overall output in respect to classification accuracy and speed by minimizing the quantity of tuples from the actual datasets. In [20] it's revealed that KNN along with Fuzzy c-means yields in 97.02% accuracy and Fuzzy KNN with Fuzzy c-means yields in accuracy of 99.25% on PIMA dataset whereas KNN with Fuzzy c-means yields in accuracy of 96.13% and Fuzzy KNN with Fuzzy c-means yields in 98.95% accuracy on liver disorder datasets.

In [21, 22] carried out the chronic disease prediction with the help of data mining approach such as ANN (Artificial Neural Networks), DT and NB and SVM (Support Vector Machine) for performing the diagnosis of heart and diabetes diseases. According to the output SVM yields an increased accuracy of 95.56% pertaining to heart disease and in case of diabetes the NB produces an accuracy of 73.58%.

### 3. PROPOSED METHODOLOGY

#### 3.1 Over View

This research work proposes E-SVM for oral cancer detection and identification in medical sector by employing ML approaches. The integrated diagnostic model with PCA feature selection method is adopted to identify the features helping in oral cancer diagnosis, minimizing the number of attributes gathered from various patient records. Following are the processes involved in the proposed approach: Data Collection, Pre-Processing, Feature Selection via PCA, Clustering, Classification and Identification. Data collection is done from Medical Institute on the basis of doctor's suggestion, data pre-processing involves eliminating unnecessary data like null vales, junk or crap character and empty columns. Then in features selection via PCA method identifies the attributes that hold responsible and influence in the diagnosis of oral cancer. Next in clustering process all the required and essential factors are grouped, thereby reducing the data. The classification process is done by E-SVM technique to detect and predict the oral cancer. E-SVM is adopted for classification since it possess better ability of generalization, performs good with oral cancer dataset and is highly robust. It tends to be an extremely important algorithm, having the ability to carry out classification with highest efficiency and is able to approximate ML in terms of NN (Neural Networks). In additions, the proposed E-SVM can accurately identify the appropriate pathological condition concerning oral cancer. Data mining approaches are surveyed to identity appropriate methods and techniques for effective data classification.

In Fig.1, a common architecture is described with the main components. The techniques of DM aids in extracting significant patterns and relations from huge dataset by utilizing effective data analysis tools. The tools involve: Mathematical Algorithm, Statistical Models and ML approaches for early cancer detection. In classification, the learning scheme involves a group of classified examples for training based on which classification of unseen examples can be performed. A framework is proposed in the research which aids in building of a DM model by employing ML techniques for the early detection and prevention of malignancy of oral cancer. For training and testing of proposed algorithm, oral cancer database were utilized. Making use of retrieved features, E-SVM is trained.

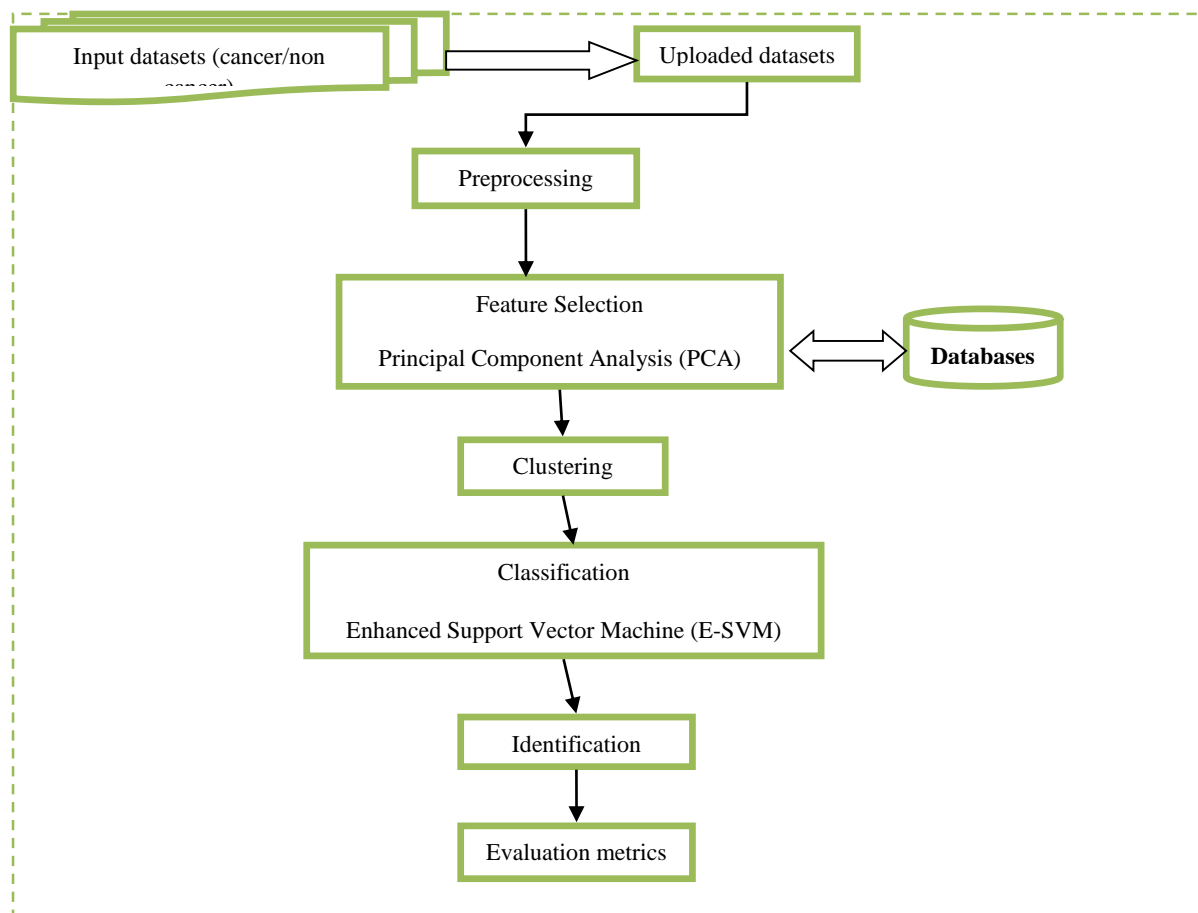


Figure 1: Over all Proposed Architecture

### 3.2 Data collection

It includes generalized oral cancer dataset or can contain oral cancer affected patient information from cancer institute having typical features. Here the dataset having the following attributes are : Gender, Age, Ethnicity, Betel Quid Chewing, Tobacco Chewing, Cigarette Smoking, Alcohol Drinking, Difficulty in Swallowing, Ulcer, Swelling, Painful, Bleeding, Burning Sensation, Loosening of Tooth, Tumor Size, Site, Weight Loss, Lump in Neck, White Patches, Red Patches, Hypertension, Diabetes, Immuno-Compromised, GSTM1, GSTT1. The dataset is split into two segments. The first segment holds healthy patient’s samples and also the samples fetched from cancer affected oral cavity of patients detected with cancer. Second segment holds samples fetched from healthy patients and those fetched from the healthy side of the oral cavity of patients having cancer. The oral cancer medical dataset are assembled online from UCI medical datasets and the entire collection of datasets is managed in databases along with the results generated. Oral cancer information collected from multiple disease organizations or a particular clinic is stored in the database. It also includes variety of oral cancer datasets such as mouth, dental, cerebrum and neck. All the essential data is assembled from the datasets and forwarded to the module of data pre-processing for subsequent processing.

### 3.3 Data preprocessing

The actual data is usually inconsistent, noisy and not complete. With the help of Data pre-processing, the missing values can be filled, noise can be smoothened while detecting outliers and inconsistencies present in the data can be rectified. The values can be filled manually or via global constant. Thereafter, this unnecessary data is eliminated from the original data. Few transactions may contain inconsistencies in the data which can be manually rectified with help of external references. Rest of the inconsistencies may be because of data integration, that is, a given attribute may contain various names in different databases or the same data-value being denoted by various names.

### 3.4 Feature Selection

The process of FS (Feature Selection) involves exhibiting and minimizing weakly relevant, unrelated or redundant dimensions/features present in a dataset. FS aims to identify the optimal subset. Once the dataset gets loaded, FS algorithms are employed to identify the best suitable dataset features. With the

help of FS classification techniques and processes, most appropriate features are generated that aids in building an effective classifier. Moreover, it aids in removing noise and minimize features for obtaining classification with increased accuracy. The process of FS involves picking out selective features from available set of features so as to enhance the classification accuracy. It's being examined that FS being an essential criteria for studies, utilized for oral cancer medical data. It also determines the data patterns, highlighting similarities and variations among them.

**Principal Component Analysis (PCA)**

Using the process of FS, the available feature set can be ranked based on some metric. As a result unimportant features can be eliminated from the problem. PCA represents a method of dimensionality reduction used by F-test, which minimizes a large set of variables to a small form containing most of the information from the large set [Table 1]. PCA tends to be a standard feature selection method in data science.

**Algorithm of PCA:**

*Input: T* → original datasets  
*N* → Total number of attributes  
*M* → dimension feature (valid feature )  
*Output: S* → optimal feature set  
 Initialized each particle in the cancer/ non cancer and whole parameters  
 Apply F-Test between feature in T  
 Calculate the contribution of each feature in T  
 While the termination condition of the iteration is not satisfies do  
     For *i=1* to *N* do  
         Calculate the correct feature in particle datasets(cancer/ non cancer and whole parameters)  
         Update the historical best feature in datasets(cancer/ non cancer and whole parameters)  
     end for  
     Update the optimal features in total datasets(cancer/ non cancer)  
         For *i=1* to *n* do  
         For *j=1* to *M*  
         Update the best feature in datasets in combining the weight(values) of each features  
     end for, end for, end while  
 Output the optimal features → (optimal feature subsets)

Table 1 - Input variables and their descriptions.

Factors	Description
Age	Less than 40 or greater
Gender	Male or Female
Tobacco/Betel Chewing	Quid Chew or not chew
Cigarette Smoking	Smoker or non-smoker
Loose teeth	Yes or No
Bleeding	Yes or No
Difficulty in Swallowing	Yes or No
White/Red Patches	Yes or No

GSTM1	Positive or Negative gene mapped to chromosome 1p13.3
GSTT1	Positive or Negative gene mapped to chromosome 22q11.2

**3.5 Clustering**

Clustering involves grouping a set of object in such a manner that object belonging to similar group (termed as cluster) are in some sense equivalent to each other compared to the one in rest of the groups (clusters). It's a major process of exploratory data mining as well as a popular technique for statistical data analysis adopted in various areas such as pattern recognition, machine learning, information retrieval, image analysis, bio-informatics, computer graphics and data compression. The task of clustering involves determining a fixed set of clusters for elaborating the data. Clustering involves grouping a set of elements possessing equivalent features.

**3.6 Classification**

In ML and Statistics, classification is considered as a supervised learning approach wherein the computer program comprehends from the input data provided to it, thereafter utilizing this learning for classifying new observation. Classification data is split into training and testing. On training, by the input dataset is forwarded to the E-SVM classification process where the original data is classified which is processed from medical data analysis. In testing the new example datasets are classified as benign/ malignant. E-SVM tends to be the simplest classification algorithms for ML techniques.

**Enhanced Support Vector Machine (E-SVM)**

SVM is being utilized for classification as well as regression issues. Mostly it's adopted only for classification problems. Here, each data item is plotted as a point in n-dimensional space (n represents number of available features) where value of each feature represents value of a specific coordinate. Support Vectors are basically the co-ordinates of individual observation. SVM is best utilized for splitting the two classes.

Data classification is a significant section in machine learning. SVM basically generates a hyper plane amidst datasets, so as to denote which class they belong to. The main challenge lies in training the machine to comprehend structure from data and associating it with appropriate class label. To achieve accurate result, the hyper plane takes the largest distance to the closest training data points of any class.

**Algorithm of E-SVM**

*Input: training dataset S, number of base class N; test instance*  
*Training Process:*  
 For *i=1, :N*  
 Use PCA technology on S to generate training subsets *S<sub>i</sub>*;  
 Train the *i*th SVM classifier *SVM<sub>i</sub>*  
     Randomly choose two data from the training data as the initial reduced set  
     Compute the reduces kernel matrix

$$s = \frac{1}{n} \sum_{i=1}^n \varphi(x_i) \varphi(x_j)'$$

For a batch data point not in the reduced sets

Compute their kernel vectors

$$\frac{1}{n} \sum_{i=1}^n \varphi(x_i) \varphi(x_j)' u = \mu u$$

Where  $u \rightarrow$  feature vector sample  $u = \sum_{i=1}^n a_i \varphi(x_i)$  with  $a_i \in R$

Compute the corresponding distance from these kernel vectors to the column space of the current reduced kernel matrix

$$\{ \varphi(x_i), Su \} = \mu \{ \varphi(x_i), u \} \quad i=1,2,\dots,n$$

$$K \in R^{n \times n}, K_{ij} = \{ \varphi(x_i), \varphi(x_j) \} = k(x_i, x_j)$$

For those point distance exceed a certain threshold

Add those point into the reduced set and from the new reduced kernel matrix

Non linear SVM with the obtained reduced kernel

A new data point is classified by the separating surfaces.

Call SVM-PCA technology to adjust the classification hyperplane of SVM<sub>i</sub> to from the SVM-PCA<sub>i</sub>

Testing process:

For  $i=1:N$

Put  $x'$  into the classify SVM-PCA<sub>i</sub> to predict its class label  $y_i'$

End

Use majority voting to get the final class label  $y'$

Output:  $y'$  which is the predicted class label for test instance  $x'$

### 3.7 Identification

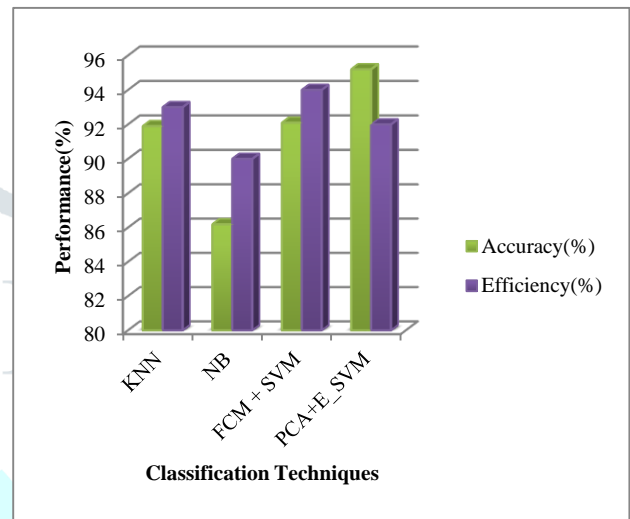
The training set being a subset of a dataset from which the ML algorithm learns relations/links amidst the attributes and the target variable. The validation set being yet other subset of the input data on which the machine learning approach can be imbibed for viewing how precisely the relationships are identified among the known result for the target variable and other features of the dataset. Finally, the available training and testing datasets are determined for classifying the output, if it's oral cancer or non-oral cancer.

## 4. RESULTS AND DISCUSSION

The datasets acquired from various diagnostic centers withholds cancer and non-cancer patient's details. Thereafter the gathered data undergoes pre-processing to find duplicate and missing information. Three proposed classification algorithm are incorporated which being: PCA (Principal Component Analysis) + E-SVM (Enhanced-Support Vector Machine), KNN (K-Nearest Neighbors), NB (Naïve Bayes) and FCM (Fuzzy Cognitive Map) + SVM (Support Vector Machine). In Table 2, the PCA + E-SVM algorithm yields in highest accuracy for available datasets compared with existing techniques for oral cancer detection and prediction. Also the cancer and non-cancer patient's datasets are separately identified. Using data mining algorithms appropriate methods and techniques are being identified for effectively classifying medical datasets. Finally, the proposed approach helps the doctors/practitioners and researchers in making diagnosis decisions and in the process of treatment planning concerning different categories.

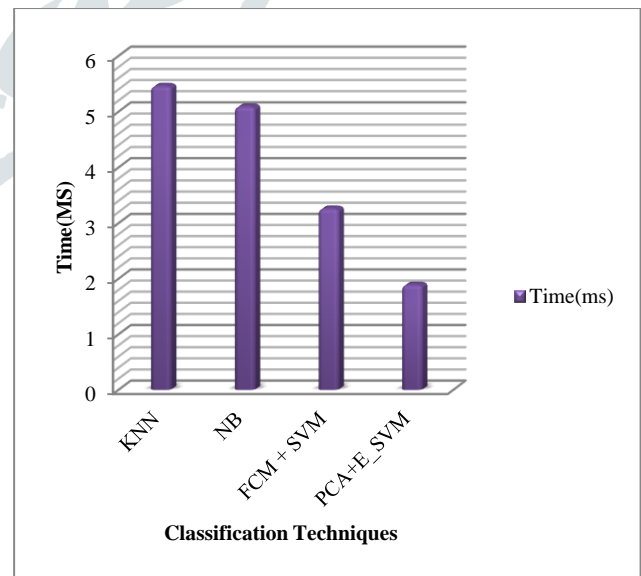
**Table 2:** Comparison of Existing Methods with Proposed Method

S. No	Classification Techniques	Accuracy (%)	Efficiency (%)	Time(MS)
1	KNN	91.9	93	5.46
2	NB	86.19	90	5.09
3	FCM + SVM	92.1	94	3.26
4	PCA+E-SVM	95.2	92	1.89



**Figure 2:** Comparison of Classification Techniques

Figure 2 and 3, illustrates comparative performance of oral cancer detection in healthcare sector that is being adopted in classification methods are PCA+E-SVM, KNN (K-Nearest Neighbors), NB (Naïve Bayes) and FCM (Fuzzy Cognitive Map) + SVM (Support Vector Machine). The proposed PCA + E-SVM algorithm yields the highest accuracy.



**Figure 3:** Comparison of Time

## 5. CONCLUSION

This research proposes E-SVM technique for diagnosis of oral cancer relying upon medical datasets. For the training and testing of the proposed algorithm, oral cancer database is utilized. Making use of retrieved features, E-SVM is trained. To conclude, the proposed PCA + E-SVM algorithm yields to highest accuracy compared with other winning techniques for automated oral cancer detection. There exist multiple approaches for analyzing the oral cancer at an early stage and providing the appropriate treatment for the same. It's revealed by examining that the blend of PCA+E-SVM algorithm yield in better outcome for detecting oral cancer at the earliest. With the output generated its proven that the proposed approach exhibits high capability for diagnosing the clinical data. In future better techniques can be incorporated with the present research work for less complexity and better learning adaptability.

## REFERENCE

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66:7–30.
- [2] Calixto G, Bernegossi J, Fonseca-Santos B, Chorilli M. Nanotechnologybased drug delivery systems for treatment of oral cancer: a review. *Int J Nanomed.* 2014;9:3719–35.
- [3] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65:87–108.
- [4] Tanaka T, Ishigamori R. Understanding carcinogenesis for fighting oral cancer. *J Oncol.* 2011;2011:603740.
- [5] Anuradha.K, Uma K, “Implementation of Fuzzy Cognitive Map and Support Vector Machine for Classification of Oral Cancers “,3rd International Conference on Green, Intelligent Computing and Communication Systems -ICGICCS 2018, 18.5 -19.5.2018, Hindusthan College of Engineering and Technology, India
- [6] Sharma, N. and Om, H. (2014) Extracting Significant Patterns for Oral Cancer Detection Using Apriori Algorithm. *Intelligent Information Management*, 6.
- [7] Huang M-W, Chen C-W, Lin W-C, Ke S-W, Tsai C-F (2017),”SVM and SVM Ensembles in Breast Cancer Prediction”. *PLoS ONE* 12(1): e0161501. doi:10.1371/journal.pone.0161501.
- [8] Sapiyah Sakriet,al,” Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction”, 2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
- [9] Nidhi Bhatla, Kiran Jyoti, “An Analysis of Heart Disease Prediction using Different Data Mining Techniques”,*IJERT*,Vol 1, Issue 8, 2012.
- [10] Syed Umar Amin, Kavita Agarwal, Rizwan Beg, “Genetic Neural Network based Data Mining in Prediction of Heart Disease using Risk Factors”, *IEEE*, 2013.
- [11] A H Chen, S Y Huang, P S Hong, C H Cheng, E J Lin, “HDPS: Heart Disease Prediction System”, *IEEE*, 2011.
- [12] M. Akhil Jabbar, B. L Deekshatulu, Priti Chandra, “Heart Disease Prediction using Lazy Associative Classification”, *IEEE*, 2013.
- [13] Chaitrali S. Dangare, Sulabha S. Apte, “Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”, *IJCA*, Volume 47– No.10, June 2012.
- [14] P. Bhandari, S. Yadav, S. Mote, D.Rankhambe, “Predictive System for Medical Diagnosis with Expertise Analysis”, *IJESC*, Vol. 6, pp. 4652-4656, 2016.
- [15] Nishara Banu, Gomathy, “Disease Forecasting System using Data Mining Methods”, *IEEE Transaction on Intelligent Computing Applications*, 2014.
- [16] A. Iyer, S. Jeyalatha and R. Sumbaly, “Diagnosis of Diabetes using Classification Mining Techniques”, *IJDKP*, Vol. 5, pp. 1-14, 2015.
- [17] Sadiyah Noor Novita Alfisahrin and Teddy Mantoro, “Data Mining Techniques for Optimatization of Liver Disease Classification”, *International Conference on Advanced Computer Science Applications and Technologies*, *IEEE*, pp. 379-384, 2013.
- [18] A. Naik and L. Samant, “Correlation Review of Classification Algorithm using Data Mining Tool: WEKA, Rapidminer , Tanagra ,Orange and Knime”, *ELSEVIER*, Vol. 85, pp. 662-668, 2016.
- [19] Uma Ojha and Savita Goel, “A study on prediction of breast cancer recurrence using data mining techniques”, *International Conference on Cloud Computing, Data Science & Engineering*, *IEEE*, 2017.
- [20] Naganna Chetty, Kunwar Singh Vaisla, Nagamma Patil, “An Improved Method for Disease Prediction using Fuzzy Approach”, *International Conference on Advances in Computing and Communication Engineering*, *IEEE*, pp. 568-572, 2015.
- [21] Kumari Deepika and Dr. S. Seema, “Predictive Analytics to Prevent and Control Chronic Diseases”, *International Conference on Applied and Theoretical Computing and Communication Technology*, *IEEE*, pp. 381-386, 2016.
- [22] Emrana Kabir Hashi, Md. Shahid Uz Zaman and Md. Rokibul Hasan, “An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques”, *IEEE*, 2017.