International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 15(2023) pp. 408-417


Received: 10 February, 2023; Accepted: 6 June, 2023; Publish: 23 June, 2023.

# Prediction of Oral Cancer Treatment Plan using Machine Learning

**Heli Shah[1], Yash Patil[2] and Rutul Patel[3,*]**

[1]Department of Electronics and Communication Engineering, Institute of Technology, Nirma University, Ahmedabad-382481, India.
*18bec098@nirmauni.ac.in*
[2]Department of Mechanical Engineering, Institute of Technology, Nirma University, Ahmedabad-382481, India.
*18bme088@nirmauni.ac.in*
[3]Department of Electronics and Communication Engineering, Institute of Technology, Nirma University, Ahmedabad-382481, India.
[*]Correspondence should be addressed to Rutul Patel; Email:*rutul.patel@nirmauni.ac.in*

*Abstract*: **Machine learning (ML) is a sub-branch of artificial intelligence (AI) that employs statistical, optimization, and fuzzy techniques to learn from past data and detect patterns from large, noisy, and complex datasets. Oral cancer treatment for a patient with mouth or throat cancer is crucial. Oral cancer treatment analyses the development of cancer cells in the tissues of the mouth or throat post-operation. In post-operation care, the mouth opening of the patient depends on the surgery performed. Further, how often should the patient perform the exercise with a jaw stretcher machine and consult the doctor? In this paper, we employed different machine learning algorithms to predict the treatment plan for the patient. We introduced the one-way ANOVA test to identify the optimal feature set. Our proposed approach applies different ML algorithms with ten-fold cross-validations to suggest a post-operation treatment plan for a patient. The experimental results show the highest accuracy with the polynomial regression and the hybrid approach.**

*Keywords*: Oral Cancer, Artificial Neural Networks, Polynomial Regression, Mean Squared Error, K-Nearest Neighbor, Hybrid Model.

## I. Introduction

Cancer is considered a heterogeneous disease consisting of many different sub-types [1]. Oral cancer is one of the most deadly diseases, often diagnosed at an advanced stage and difficult to treat. Oral cancer is the sixth most common cancer in the world, belonging to the larger group of cancers known as head and neck cancer [2]. Oral cancers arise in any part of the oral cavity or oropharynx. Oral cancer ranks in the top three of all the cancers, which accounts for 30% of total cases of cancer in the country, and its control has become a global health priority [3]. The current techniques used for the diagnosis are biopsy, histopathological examination, vital staining, biomarkers, and optical techniques [4]. The treatment of oral cancer includes surgery and radiation therapy. In most cases, due to radiation therapy, the mouth opening of the patients decreases. Tobacco, alcohol, and snuffing are vital factors for the occurrence of oral cancer.

Tobacco plays the predominant role in the occurrence of Oral Cancer. Use of tobacco through cigarettes, cigars, pipes, chewing tobacco, and even snuffing leads to oral cancer [5]. Excessive intake of alcohol causes oral cancer [6]. About 80% of patients who have oral cancer are due to the use of tobacco in different forms. Approximately 10% of patients have oral cancer without consuming tobacco, alcohol, or snuffing. In the UK, the significant risk factor for the cause of oral cancer is cigarette smoking and alcohol misuse. The prevention for oral cancer is reducing exposure to tobacco, alcohol, and betel quid and early detection of oral cancer [7]. Oral cancer occurs in various sites like the tongue, lips, cheeks, palate, pharynx, and mandible bone [6]. The different types of oral cancer considered here are Tongue, Buccal Mucosa, Tonsil, Cheek, Lip, Mandible Bone, and Palate. Most of the cases of oral cancer are of buccal mucosa and tongue. The buccal mucosa is the lining of the cheek and back of the lips inside the mouth. Tongue is the most common intraoral site for oral cancer [8]. Mandible bone cancer is a cancerous tumor. The tumor invasion into the bone changes the microarchitecture of the mandible bone [9]. Patients suffering from mandible bone cancer cannot perform jaw movement. Tonsil is the second most common cancer belonging to the group of head and neck cancers. Tonsil malignancy occurs due to squamous cell carcinoma [10]. Whether the patient will be able to perform the jaw movement or not depends on the surgery performed. Every oral cancer patient is diagnosed at a different stage, so each patient's treatment plan is different. Thus, as a part of the treatment of mouth opening, patients are treated with a jaw stretcher machine.

A jaw stretcher machine is a system that helps to improve jaw mobility and rehabilitate suffering from trismus effectively and efficiently. It is a manually operated device that is entirely under an individual's control. An anatomically designed machine works on the jawline of the patient for anatomically natural mouth opening [11]. Thus, depending on the patient's condition, whether the jaw movement is possible or not, they are given the jaw stretcher machine as part of treatment for mouth opening. The jaw stretcher machine is shown in Fig. 1.

**Figure. 1**: Jaw Stretcher Machine

Tetarbe et al. [6] implemented the REP tree algorithm for the detection of oral cancer and obtained an accuracy of 78.72%. Mohd et al. [12] implemented the support vector machine algorithm with the 14 optimal features and acquired the accuracy of 96.19% for the prediction of oral cancer. Jubair et al. [13] developed a deep convolutional neural network to predict oral cancer through binary classification of oral lesions and achieved an accuracy of 85%. Here one-way ANOVA is used to test the null hypothesis of samples [14] with more than two groups. Shaharum et al. [15] performed a one-way ANOVA test for the feature selection and produced a better classification accuracy of 93.33% for automatic wheeze detection using an artificial neural network. Sheikhan et al. [16] introduced a modular neural-support vector machine algorithm to increase the accuracy of speech emotion recognition by 2.2%, using the ANOVA feature selection method. Elssied et al. [17] implemented ANOVA F-test for the feature selection on the e-mail dataset to improve the limitation of support vector machines in e-mail spam classification. In [18], the authors have examined the oral cancer treatment in depth using principal component analysis. In [19], the authors have discussed the detection of oral cancer at the early stage and followed by appropriate treatment can increase cure rates up to 80%. In [20], the authors developed an artificial neural network for the prediction of oral cancer followed by ten cross-validation folds, obtaining an accuracy of 78.95%.
The key contributions of this paper are summarized as follows:

- We have prepared a dataset for oral cancer treatment plans for patients to train machine learning models to guide patients for future follow-ups.
- We have used a one-way ANOVA test to identify and remove redundant features from the dataset.
- We have proposed a hybrid model based on an ensemble learning approach that improves the classifier's overall performance.
- We have created a web application that guides patients to consult with the doctor for an oral cancer treatment plan.

The rest paper is organized as: Section 2 summarizes the related work for the machine learning method. Section 3 specifies the dataset attributes and corresponding features selection using the one-way ANOVA test. Section 4 discusses the proposed methodology for training the model using machine learning algorithms. Section 5 shows the experimental results of predictions and our design of a web application for oral cancer treatment plans.

## II. Related Work

This section discusses state-of-the-art machine learning algorithms for predicting oral cancer treatment plan for patients.

### A. *Polynomial Regression*

Polynomial regression relates dependent and independent variables of $n^{th}$ degree polynomial. Such a model fits well in cases when the relationship between dependent and independent variables is non-linear. This model minimizes cost function as given in Equation (1) by fitting data points to a polynomial line.

$$Y = \theta_0 + \theta_1 X + \theta_2 X^2 + ... + \theta_m X^m + residualerror \quad (1)$$

Here, $Y$ and $X$ are dependent and independent variables, respectively, and $\theta_0, \theta_1, ..., \theta_m$ are weights to $m^{th}$ degree polynomial. As the degree of polynomial $m$ increases, it will give a better R-square value. However, using polynomial models of degrees higher than 2 or 3 causes overfitting. It uses the same method as linear regression for training the model but supports a higher polynomial degree. After applying the cost function, the output is predicted. Next, find R-square and MSE values for same, analyze the training model, and make necessary changes to get minimum MSE and higher R-square value.

### B. *K-Nearest Neighbor*

A K-nearest neighbor searches for the nearest point to the input point from the given data set [21]. The algorithm classifies the cases based on the majority votes of its K neighbors using test data. In this method, firstly, the data points are transformed into mathematical values. After that, it will find the distance between these data points. The distance between data points and test data is computed to determine the probability of the point being similar to test data. It is classified based on the data points having the highest probabilities. The distance between data points and test data is calculated through euclidean distance. The Euclidean distance $D_{PQ}$ between two points with coordinated $P(x_1, y_1)$ and $Q(x_2, y_2)$ in the same plane is as given in Equation (2).

$$D_{PQ} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

For the given K value, the class with the highest number of data points out of all classes, depending on the nearest K-neighbors, is assigned to the given data point. Based on the distance computed, the input x is assigned to the class that has the highest probability given by Equation (3).

$$P(y = j|X = x) = \frac{1}{k}\sum_{i \epsilon A} I(y^{(i)} = j) \tag{3}$$

### C. Artificial Neural Network

Artificial neural networks (ANN) models are derived from the biological learning process of the human brain and are helpful in classification or pattern recognition tasks. ANN consists of many interconnected neurons called nodes connected through network-making networks [22].

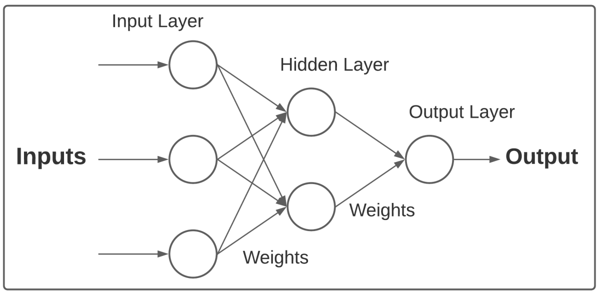



**Figure. 2**: Artificial Neural Network Architecture

As shown in Fig. 2, each node computes inner product of input samples $x_i$ and weights $w_i$ and add bias $b$ into it. Next, a transfer function $f$ normalize this output. Similarly, all other outputs are predicted at each hidden layer node as shown in Equation (4).

$$a = f\left(\sum_{i=1}^{r} w_i \cdot x_i + b\right) \tag{4}$$

The transfer function $f$ is a non-linear function like sigmoid, logistic, soft plus, rectified linear unit (ReLU), 1eaky ReLU, etc. ANN is a supervised learning approach by which the model is trained and compares the predicted output with the desired output. The mean square error (MSE) is calculated for the given ANN model. The backpropagation network minimizes MSE by updating the weights on the neural connection in the multiple layers. The mean square error function $F(x)$ at iteration $k$ is given by Equation (5).

$$F(x) = [t_k - a_k]^2 \tag{5}$$

An $m^{th}$ layer weights and biases at iteration $k$ are updated as shown in Equation (6) and Equation (7).

$$w_{i,j}^{m}(k+1) = w_{i,j}^{m}(k) - \alpha\frac{\partial F}{\partial w_{i,j}^{m}} \tag{6}$$

$$b_{i}^{m}(k+1) = b_{i}^{m}(k) - \alpha\frac{\partial F}{\partial b_{i}^{m}} \tag{7}$$

Here $\alpha$ represents the learning rate and $w_{i,j}$ represents weights of connection between neuron $i$ and neuron $j$.

## III. Attributes and Feature Selection

This section summarizes different attributes of a dataset that includes key features for oral cancer treatment of a patient. Next, a set of useful features are selected based upon their P-values using the ANOVA test.

### A. Attributes

The details mentioned in Table 1 are the case of a particular patient. Attribute name age is the age of the patient in years; Gender is attributed by M (for male) and F (for female); Oral cancer is caused because of consumption of tobacco, alcohol, or snuffing (Yes-1; No-2); Cause of Oral Cancer is because of consuming (Tobacco-1; Alcohol-2; Snuffing-3); Diagnose with which type of cancer (Mandible Bone-1; Palate-2; Cheek-3; Lip-4; Buccal Mucosa-5; Tonsil-6; Tongue-7); On which side the occurrence of cancer is observed (Left-1; Right-2; Upper-3; Lower-4); Patient can do jaw movement or not (Present-1; Absent-2); After operation how much the patient can open mouth in a number of fingers as an average person can open mouth till four fingers; After treatment how much mouth patient can open in a number of fingers; Number of times patient is required to perform the exercise; The doctor should be consulted after the duration of in weeks [23].

*Table 1*: Data Summary

| **Attribute Name** | **Range** |
|---|---|
| Age | 21-83 |
| Gender | M-F |
| Consumption | 1-2 |
| People Consuming | 1-3 |
| Diagnosis | 1-7 |
| Side | 1-4 |
| Jaw Movement | 1-2 |
| Mouth Opening Before Treatment | 0.5-1 |
| Mouth Opening After Treatment | 0.5-3.5 |
| Repetition of Exercise | 10-70 |
| Duration to consult a doctor | 1-10 |

### B. Feature Selection

The feature selection method identifies a redundant set of features and reduces its dimensionality. As a result, feature selection reduces the number of computations for training. In this study, feature selection is done using the analysis of variance (ANOVA) test. ANOVA tests the null hypothesis depending on the statistical differences among the means of two or more groups. The null hypothesis is no difference in mean between two groups and in the alternative hypothesis mean between the groups is not equal. So, the dominance of the feature is decided on the acceptance or rejection of the hypothesis. The variance between the two groups is compared in the ANOVA test, after which P-value and the F-test statistics are obtained.

*Table 2*: P-value of Input Features for Repetition

| Feature | P-value | F-Test |
|---|---|---|
| Gender | 0.76 | 0.094 |
| Age | 0.48 | 1.0 |
| Consumption | 0.41 | 0.65 |
| People Consuming | 0.13 | 2.06 |
| Diagnosis | 4.80e-43 | 51.95 |
| Side | 0.043 | 2.72 |
| Jaw Movement | 2.79e-82 | 530.96 |
| Mouth Opening Before Treatment | 1.65e-24 | 177.50 |
| Mouth Opening After Treatment | 0.0 | 149479.36 |

*Table 3*: P-value of Input Features for Duration

| Feature | P-value | F-Test |
|---|---|---|
| Gender | 0.75 | 0.099 |
| Age | 0.48 | 0.99 |
| Consumption | 0.42 | 0.64 |
| People Consuming | 0.12 | 2.13 |
| Diagnosis | 2.69e-39 | 48.19 |
| Side | 0.045 | 2.69 |
| Jaw Movement | 6.88e-71 | 453.89 |
| Mouth Opening Before Treatment | 3.44e-23 | 170.45 |
| Mouth Opening After Treatment | 0.0 | 47609.58 |

From the P-value statistics, acceptance or rejection of the null hypothesis is done. If the variance obtained between two groups is the same, the corresponding feature has no impact on the response variable.
$H_0$: All means are equal.
$H_1$: All means are not equal.
P-value consists of the probability of random chance, something having equal probability, and probability of something rare. A threshold of 0.05 is decided for P-value for deciding the dominance of the feature. The features having P-value less than 0.05 have the maximum impact on the response variable.
From Table 2 and Table 3, it is observed that diagnosis, side, jaw movement, mouth opening before treatment, and mouth opening after treatment are the most dominating features. Their P-value is less than 0.05, which states that those features are highly dependent on response.

## IV. Proposed Methodology

The proposed method applies state-of-the-art ML algorithms to suggest an oral cancer treatment plan to the patient.

### *A. Methodology*

For experiment purposes, we have divided the dataset for training and testing in the proportion of 80:20. The input data is given, and feature selection is made based on a one-way ANOVA test as discussed in feature selection. The most dominating five features are selected based on the P-value, and different models are trained. Then the testing data is given following the same procedure for the prediction of repetition and duration as shown in Fig. 3. For notation purposes, we will denote model 1 and model 2 for the prediction of repetition and duration in weeks, respectively.

#### *1) Polynomial Regression*

The data is first given to the model in our proposed framework, and the different degree values are applied to train the model. The values of the degree applied are varied from 2 to 10. Then the regression model is fitted, and the R-square value of the model is observed. After seeing the R-Square value for different polynomial degrees, the degree of 3 gives the highest R-square value. As the degree increases, it starts overfitting the data. Therefore, the degree of the polynomial used here is three. The R-square value obtained for prediction of repetition in model 1 is 0.99, and that for prediction of duration in model 2 is 0.98.

#### *2) K-Nearest Neighbor*

In our proposed framework, first, the data is given to the model, and the regression technique is applied. We have two separate models for predicting repetition and duration named model 1 and model 2, respectively. Both models are trained and validated with 10-fold cross-validation. For model 1 the R-square value and MSE are 0.989 and 2.94 and for model 2 are 0.985 and 0.115. For classification using K-nearest neighbor, the choice of K is crucial. Smaller values of K make the decision boundaries unstable that increase the chances of false prediction. As a result, data starts underfitting the model. As the value of K increases, the decision boundary layer becomes smoother, leading to better prediction. We chose K to be 5 for a smooth boundary condition that leads better prediction in our proposed framework. The odd value of K is selected for better prediction between the two groups. The classification performance through 5 nearest neighbors is shown in Fig. 4a. From Fig. 4 of the confusion matrix of repetition and duration, it is observed that from 209 samples of testing data, 198 samples are predicted correctly, and error occurred while predicting for the other 11 samples. These 11 samples actually belong to the other classes; instead, they are predicted as shown in the confusion matrix. The accuracy obtained in model 1 is 95% and model 2 is 96%.

#### *3) Support Vector Machines*

Support vector machines are supervised machine learning algorithms. Support vector machines work based on kernel functions. Support vector machines are used for classification as well as regression analysis. Here the regression technique is applied to the dataset. Different kernels like linear, poly, Radial Bias Function (RBF) and sigmoid are applied, and each case's accuracy is observed. Based on training and testing data, the accuracy of 99% is obtained in the linear kernel, which is the highest among all kernels. Such a linear kernel yields the R-square value of 0.981. Similarly, for the prediction of duration, different kernels are analyzed.

#### *4) Artificial Neural Network*

In our proposed model, the artificial neural network structure consists of an input layer, five hidden layers, and an output

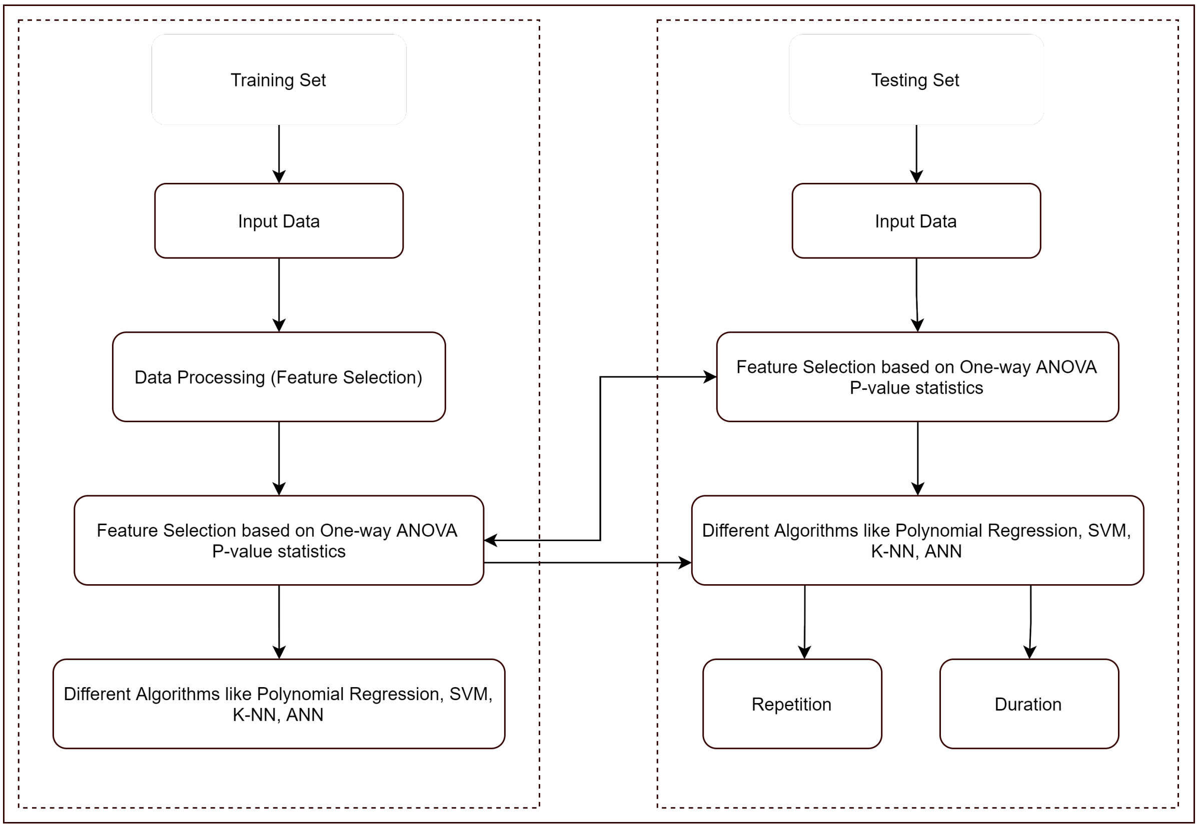



**Figure. 3**: ML approach to suggest oral cancer treatment plan

layer to predict repetition. The number of neurons in the first hidden layer is varied from 8 to 30, in the second hidden layer are varied from 6 to 15, in the third hidden layer are varied from 2 to 15, and similarly for the remaining two hidden layers. After performing a certain number of iterations and observing the MSE for the same, the number of nodes in the hidden layers is obtained. The structure of the ANN giving the best results is n: 13:13:13:13:13:1, where the n represents the numbers of nodes in the input layer, the hidden layers have 13,13,13,13 and 13 nodes, and the output layers had one node. The architecture of the specified artificial neural network is shown in Fig. 5. The model is trained with these specifications and the ReLU activation function, and the R-square value of 0.931 is achieved. The loss obtained during training the model is shown in Fig. 6.

The artificial neural network structure consists of has an input layer, two hidden layers, and an output layer to predict duration. After performing a certain number of iterations and observing the MSE value for the same, the number of nodes in the hidden layers is obtained. The structure of the ANN giving the best results is n:30:30:1, where the n represents the numbers of nodes in the input layer, the hidden layers have 30 and 30 nodes, and an output layer has one node. The architecture of the specified artificial neural network is shown in Fig. 7.

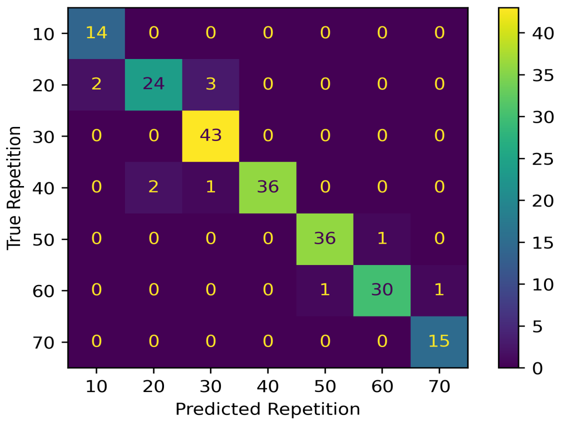
The model is trained with these given specifications, and the sigmoid activation function at the output layer yields an R-square value of 0.865. The loss obtained during training the model is shown in Fig. 8. Also, MSE values for both output repetition and duration are 21.633 and 1.08, respectively. As it is observed from Fig. 8 at each epoch, the loss is decreasing, and the early stopper is used to stop training the model after convergence. Here the numbers of epochs are set to 2000 along with the early stopper. After the convergence is obtained, the model stops learning, and the R-square value for the same is obtained. In this proposed architecture, the converged curve is obtained at the smaller numbers of epochs.

### *B. Principal Component Analysis*

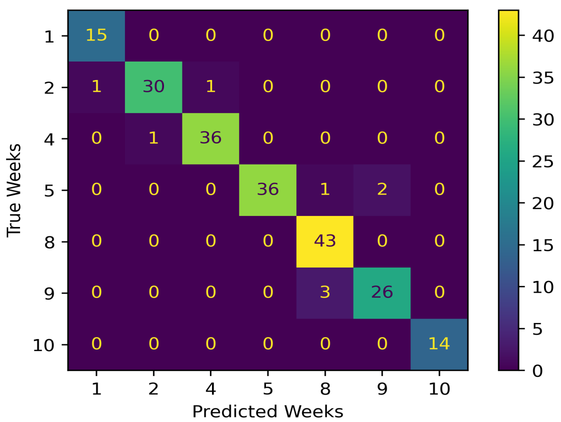
Principal component analysis (PCA) is an unsupervised method that extracts the most relevant information from a large, noisy dataset. It is a linear transformation of the data that minimize the redundancy and maximizes the information. PCA transforms given features into principal components and arranges them into the most significant to the redundant component. PCA eliminates the few last principle components that do not contribute significantly. An 80% of data is given for training the model, and 20% of data is given for testing the model. PCA is applied to both training data and testing data. After applying PCA, polynomial regression is applied to model 1 and model 2 to predict repetition and duration. The R-square value obtained in model 1 and model 2 is 0.99 and 0.98.

### *C. Hybrid Model*

As it is observed that, the ANN algorithm has the lowest accuracy among all the models. Therefore, the hybrid model that combines different models to improve overall accuracy is developed. Here a polling-based hybrid model is developed for oral cancer treatment plan prediction using the




(a) Confusion Matrix for Repetition




(b) Confusion Matrix for Duration

**Figure. 4**: Performance of K-nearest neighbor for to suggest treatment plan

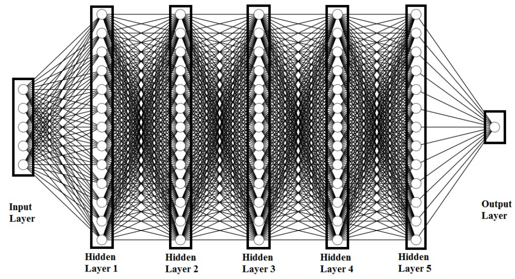



**Figure. 5**: The architecture of ANN for Repetition

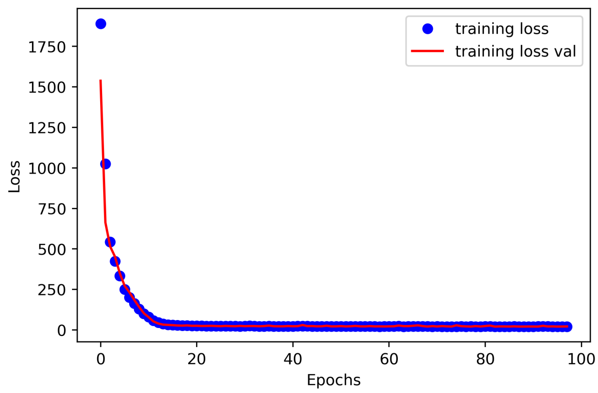



**Figure. 6**: Loss obtained for Repetition

dataset given as discussed above. Polling is the implementation of an ensemble learning algorithm in which the final prediction accuracy can be increased by giving proper importance to the responses obtained and assigning appropriate weight to them. In this method, two outputs are classified as repetition and duration. The implementation uses two different models for the classification ANN and k-nearest neighbor classifier. Fig. 9 shows the methodology proposed and the flow observed. The training dataset must not overfit the model as it leads to high biasing. The trained model is used for the testing process to observe the resultant accuracy. From Fig. 9 it is observed that ANN and k-nearest neighbor classifier models are being trained. For obtaining the final output polling mechanism is used, in which the output is predicted by assigning the appropriate weights to the output obtained from individual models. A one-way ANOVA test is applied for feature selection, and depending on the P-value obtained; the dominating five features are selected.

The architecture consists of two models, ANN and k-nearest neighbor classifier. Initially, 80% of data is given for training the model, and 20% of data is for testing the model. The ANN and k-nearest neighbor classifier models are trained, followed by ten cross-validation folds. After the models are trained, polling is done by providing the appropriate weights to the output obtained from the individual model. The weights for the polling mechanism are decided based on the R-square value obtained. Two-hybrid models are designed for the prediction of repetition and duration. After training the model and assigning the appropriate weights, testing data is given to the hybrid model, and the R-square value for the same is observed. The R-square value obtained in both the hybrid models is 0.98. The R-square value obtained is higher than obtained in the individual trained models of ANN and k-nearest neighbor classifier. Thus, the usage of a hybrid model results in better accuracy for prediction.

## V. Experimental Results

This section illustrates the performance of Machine Learning models to predict oral cancer treatment plans. Further, this section includes a developed web application for predicting treatment plans for an oral cancer patient.

### A. *Performance of Machine Learning models*

The R-square value and mean squared error (MSE) of model 1 and model 2 are shown in Table 4. Model 1 predicts the number of repetitions and model 2 predicts the duration in weeks. From Table 4, it is observed that polynomial regression has the highest value of the R-square of 0.99 for model 1 and 0.983 for model 2, and the lowest value of the mean squared error (MSE) of 0.0123 for model 1 and 0.132 for model 2. The hybrid model also predicts the best with the R-square of 0.987 and 0.986 for model 1 and model 2. Next is the k-nearest neighbor, followed by principal com-

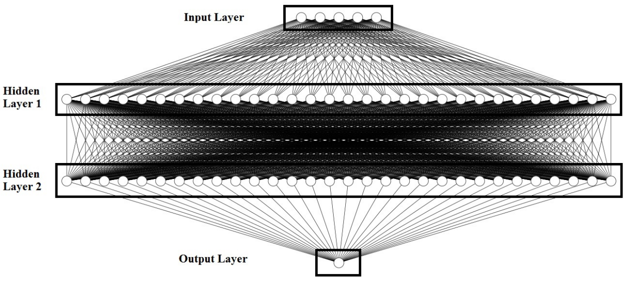



**Figure. 7**: The architecture of ANN for Duration

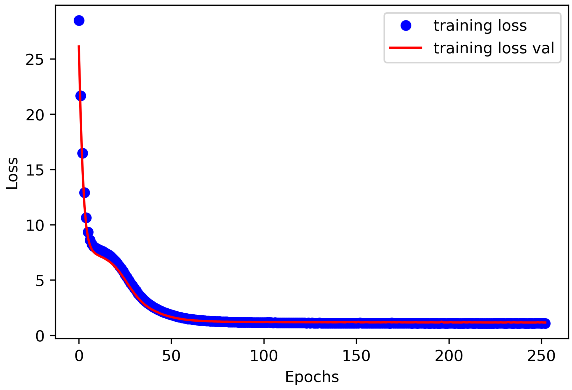



**Figure. 8**: Loss obtained for Duration

ponent analysis with polynomial regression, support vector machines, and artificial neural network. Therefore, polynomial regression is the best algorithm for predicting the oral cancer treatment plan of patients based on the attributes provided in the dataset.

As observed, the R-square value obtained while training the ANN is the lowest of all; the new hybrid approach is proposed by using the polling mechanism. The hybrid model consists of two trained models, ANN and k-nearest neighbor classifier. The k-nearest neighbor classifier model is used along with the ANN model to increase its accuracy, which is the significant advantage obtained. The R-square value obtained in the hybrid model and polynomial regression are almost the same. As a result, any of the algorithms can be used to predict an oral cancer treatment plan.

Table 5 and Table 6 shows that the precision value obtained indicates the accuracy of positive prediction. Recall value indicates the fraction of positive instances that were identified correctly. F1-score is the harmonic mean of precision and recall and is lower than the accuracy obtained. F1-score is maximum when the precision is equal to recall. The best F1-score is 1.00, and the worst is 0.0. In the K-NN classification algorithm, an accuracy of 95% is obtained for repetition and 96% for the duration.

*Table 4*: R-square Value and MSE of algorithms

| Algorithm | Model | R-square value | MSE |
|---|---|---|---|
| Polynomial Regression | Model1 | 0.999 | 0.0123 |
| | Model2 | 0.983 | 0.132 |
| Support Vector Machine | Model1 | 0.981 | 5.312 |
| | Model2 | 0.981 | 0.148 |
| K-Nearest Neighbor Regression | Model1 | 0.989 | 2.94 |
| | Model2 | 0.985 | 0.115 |
| Artificial Neural Network | Model1 | 0.931 | 21.633 |
| | Model2 | 0.865 | 1.08 |
| PCA with Polynomial Regression | Model1 | 0.999 | 0.237 |
| | Model2 | 0.981 | 0.152 |
| Hybrid Model | Model1 | 0.987 | 3.509 |
| | Model2 | 0.986 | 0.107 |

*Table 5*: Classification Report of Repetition for K-NN Algorithm

| Repetitions | Precision | Recall | F1-score |
|---|---|---|---|
| 10 | 0.88 | 1.00 | 0.93 |
| 20 | 0.92 | 0.83 | 0.87 |
| 30 | 0.91 | 1.00 | 0.96 |
| 40 | 1.00 | 0.92 | 0.96 |
| 50 | 0.97 | 0.97 | 0.97 |
| 60 | 0.97 | 0.94 | 0.95 |
| 70 | 0.94 | 1.00 | 0.97 |

*Table 6*: Classification Report of Duration for K-NN Algorithm

| Weeks | Precision | Recall | F1-score |
|---|---|---|---|
| 1 | 0.94 | 1.00 | 0.97 |
| 2 | 0.97 | 0.94 | 0.95 |
| 4 | 0.97 | 0.97 | 0.97 |
| 5 | 1.00 | 0.92 | 0.96 |
| 8 | 0.91 | 1.00 | 0.96 |
| 9 | 0.93 | 0.90 | 0.91 |
| 10 | 1.00 | 1.00 | 1.00 |

### *B. Web Application Development*

A front-end web application was designed using an anvil platform. The user interface is designed with a simple drag, and drop mechanism. Anvil is a platform used for building web apps using python. The end-to-end user application is designed to predict the oral cancer treatment plan, depending on the input entries done per the patient's condition. The treatment plan will predict the number of times the exercise is required to be performed by the patient and consult the doctor for the same after what duration. The link for the same is generated. In the back end, the polynomial regression algorithm of machine learning is linked with the online server connects of the anvil for predicting the output based on the user's entry, as the highest R-square value and minimum MSE are obtained while training the model of polynomial regression. A sample web application page is shown in Fig.

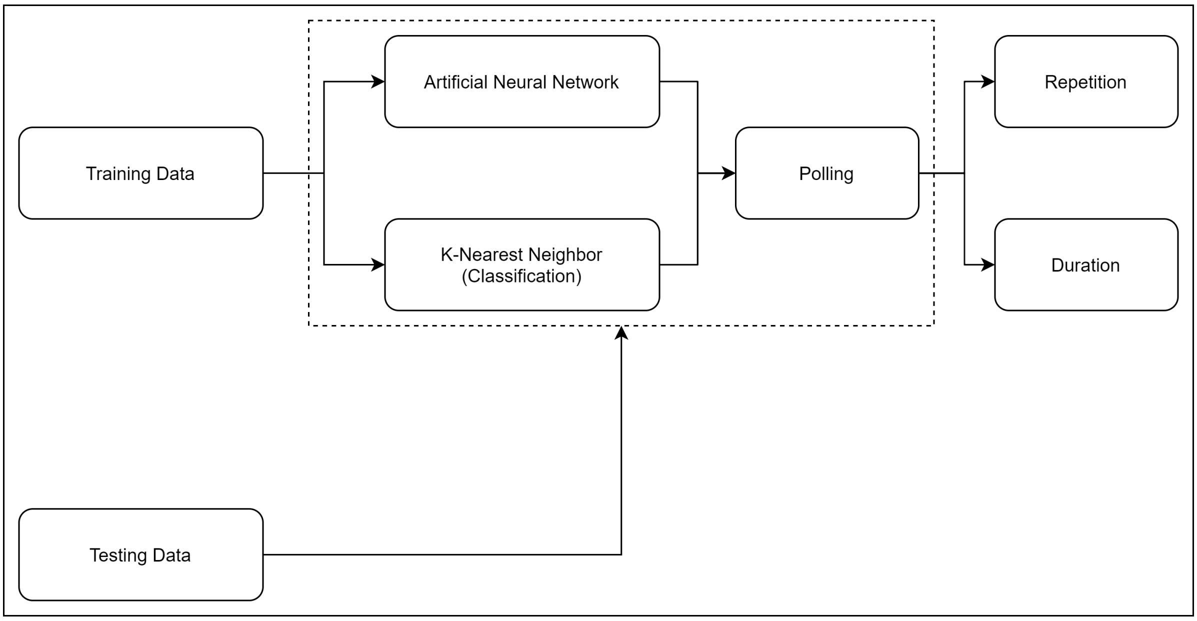



**Figure. 9**: Hybrid Model Architecture

10. The treatment plan predicted for the oral cancer patient by the algorithm is shown in Fig. 11.

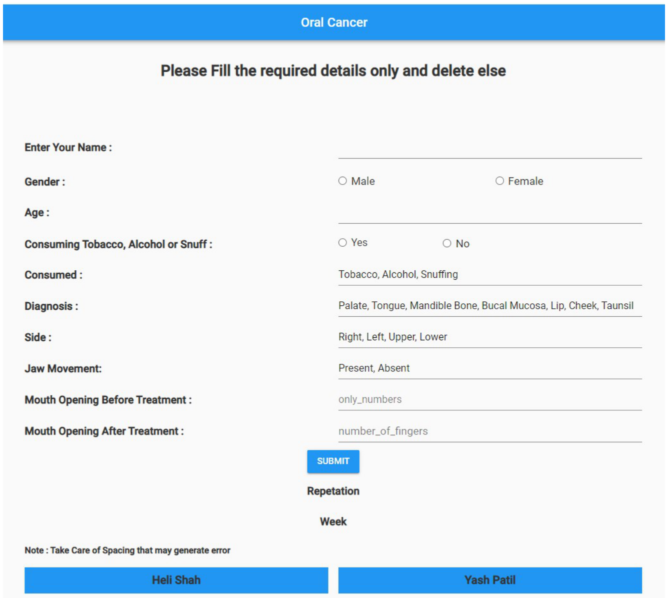



**Figure. 10**: Web Page of Application

The anvil includes complete PythonServer modules, an ordinary CPython interpreter, just like running the own machine. Thus, the web application designed may provide ease to the doctors for consulting the patients. In India, many patients are found clueless regarding their disease; as a result, this application will help and guide them. Thus, the anvil web application will help the patient and make them aware of their disease on a precautionary basis; this will also create ease for the doctors.

## VI. Conclusion

In this paper, we have created a dataset for an oral cancer treatment plan. Different classification and regression tech-

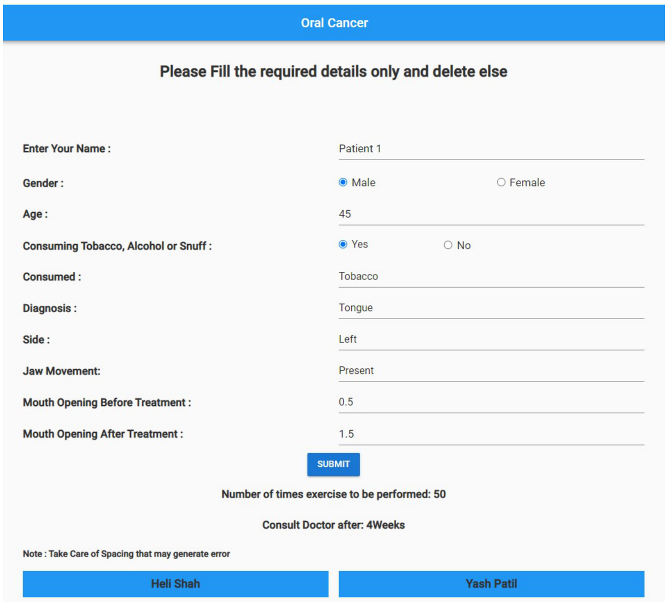



**Figure. 11**: Treatment Plan Predicted

niques like polynomial regression, artificial neural network, k- nearest neighbor, and support vector machines are applied to the generated dataset. We have used a one-way ANOVA test for dimensionality reduction of the feature set using P-value statistics. The most dominating, optimal five features out of nine are selected from the P-value of the features obtained. This optimal feature set is applied to the different models and evaluated using R-square value and MSE. From all the algorithms used, polynomial regression gives the highest R-square value and minimum MSE. Our implemented polynomial regression achieves an MSE of 0.0123 for repetition and 0.132 for the duration of weeks. Similar results are found for principal component analysis applied along with polynomial regression. Our proposed hybrid approach combines an artificial neural network and a k-nearest neighbor classifier to enhance the classifier's performance. Such an ensemble learning approach improves overall R-square value and MSE results. Thus, the proposed hybrid model can also be used as an alternative to polynomial regression.

## Acknowledgment

The authors would like to thank the Department of Physiotherapy, Gujarat Cancer and Research Institute, Civil Hospital, Ahmedabad for providing their data.

## References

[1] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8–17, 2015.

[2] S. Prasanna, K. Govinda, and U. S. Kumaran, "An evaluation study of oral cancer detection using data mining classification techniques." *International Journal of Advanced Research in Computer Science*, vol. 3, no. 1, 2012.

[3] K. R. Coelho, "Challenges of the oral cancer burden in india," *Journal of cancer epidemiology*, vol. 2012, 2012.

[4] C. Scully, J. V. Bagan, C. Hopper, J. B. Epstein *et al.*, "Oral cancer: current and future diagnostic techniques," *Am J Dent*, vol. 21, no. 4, pp. 199–209, 2008.

[5] H.-Y. Sung, Y. Wang, T. Yao, J. Lightwood, and W. Max, "Polytobacco use of cigarettes, cigars, chewing tobacco, and snuff among us adults," *Nicotine & Tobacco Research*, vol. 18, no. 5, pp. 817–826, 2016.

[6] A. Tetarbe, T. Choudhury, T. T. Toe, and S. Rawat, "Oral cancer detection using data mining tool," in *2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. IEEE, 2017, pp. 35–39.

[7] B. K. Joseph, "Oral cancer: prevention and detection," *Medical Principles and Practice*, vol. 11, no. Suppl. 1, pp. 32–35, 2002.

[8] S. Moore, N. Johnson, A. Pierce, and D. Wilson, "The epidemiology of tongue cancer: a review of global incidence," *Oral diseases*, vol. 6, no. 2, pp. 75–84, 2000.

[9] G. D. Rabelo, C. Coutinho-Camillo, L. P. Kowalski, N. Portero-Muzy, J.-P. Roux, P. Chavassieux, and F. A. Alves, "Evaluation of cortical mandibular bone in patients with oral squamous cell carcinoma," *Clinical oral investigations*, vol. 22, no. 2, pp. 783–790, 2018.

[10] M. Guay and P. Lavertu, "Tonsillar carcinoma," *European archives of oto-rhino-laryngology*, vol. 252, no. 5, pp. 259–264, 1995.

[11] L. B. Kaban, L. T. Swanson, J. E. Murray, and W. Sheridan, "Postoperative physiotherapy device for mandibular hypomobility," *Oral Surgery, Oral Medicine, Oral Pathology*, vol. 43, no. 4, pp. 513–516, 1977.

[12] F. Mohd, N. M. M. Noor, Z. A. Bakar, and Z. A. Rajion, "Analysis of oral cancer prediction using features selection with machine learning," in *The 7th International Conference on Information Technology (ICIT)*, 2015.

[13] F. Jubair, O. Al-karadsheh, D. Malamos, S. Al Mahdi, Y. Saad, and Y. Hassona, "A novel lightweight deep convolutional neural network for early detection of oral cancer," *Oral Diseases*, 2021.

[14] R. Palaniappan, K. Sundaraj, and S. Sundaraj, "A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals," *BMC bioinformatics*, vol. 15, no. 1, pp. 1–8, 2014.

[15] S. M. Shaharum, K. Sundaraj, and K. Helmy, "Performance analysis of feature selection method using anova for automatic wheeze detection," *Jurnal Teknologi*, vol. 77, no. 7, 2015.

[16] M. Sheikhan, M. Bejani, and D. Gharavian, "Modular neural-svm scheme for speech emotion recognition using anova feature selection method," *Neural Computing and Applications*, vol. 23, no. 1, pp. 215–227, 2013.

[17] N. O. F. Elssied, O. Ibrahim, and A. H. Osman, "A novel feature selection based on one-way anova f-test for e-mail spam classification," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 7, no. 3, pp. 625–638, 2014.

[18] Y. Matsuda, T. Okui, M. Karino, N. Aoi, S. Okuma, K. Hayashida, T. Sakamoto, and T. Kanno, "Postoperative oral dysfunction following oral cancer resection and reconstruction: A preliminary cross-sectional study," *Oral Oncology*, vol. 121, p. 105468, 2021.

[19] S. Silverman Jr, "Early diagnosis of oral cancer," *Cancer*, vol. 62, no. S1, pp. 1796–1799, 1988.

[20] A. Alhazmi, Y. Alhazmi, A. Makrami, A. Masmali, N. Salawi, K. Masmali, and S. Patil, "Application of artificial intelligence and machine learning for prediction of oral cancer risk," *Journal of Oral Pathology & Medicine*, vol. 50, no. 5, pp. 444–450, 2021.

[21] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[22] C. S. Tyagi, "A comparative study of svm classifiers and artificial neural networks application for rolling element bearing fault diagnosis using wavelet transform preprocessing," *Neuron*, vol. 1, pp. 309–317, 2008.

[23] CivilHospital, "Oral cancer dataset," 2021.

## Author Biographies

**Heli Shah** received her BTech in Electronics and Communication Engineering from Nirma University in 2022. She has been working as an Associate Engineer at Qualcomm since July 2022. Her research area includes machine learning, VLSI Design, Testing and Verification, and Power Optimization.

**Yash Patil** received his BTech in Mechanical Engineering from Nirma University in 2022. He has been working as a Graduate Engineer Trainee at Reliance since August 2022. His research area includes machine learning, deep learning, mechanical design, Propulsion system and maintenance.

**Rutul Patel** received his PhD in Image Processing from Gujarat Technological University in 2022 and a Master's degree in Signal Processing and Communication from A D Patel Institute of Technology, Karamsad, in 2013. He has been working as an Assistant Professor at Nirma University since 2016. His research area includes machine learning, signal processing, and image processing.