PREDICTION OF POST-TREATMENT SURVIVAL EXPECTANCY IN HEAD & NECK CANCERS BY MACHINE LEARNING METHODS

H.S. Nogay

Abstract— In this study, survival for head and neck cancer disease was estimated using machine learning methods. Starting from the date on which the head and neck cancer disease was diagnosed, without a maximum time limit, at the end of the minimum 8-month period, it is estimated that the patient will be alive or not. Seven classifying machine-learning predictive methods were used in the study. The main goal of this study is to estimate the survivability of head and neck cancer patients and to provide a decision aid for cancer management with applied estimation methods and results. The results obtained by the application of the designed methods are examined and results with extremely high accuracy rates are obtained.

Keywords— *Machine learning, classification, artificial neural network, support vector machines, decision tree, logistic regression, linear disriminant, nearest neighbor.*

1. Introduction

HEAD and neck squamous cell carcinoma (HNSCC), including upper aerodigestive tract and anatomic regions, is considered the third leading cause of death worldwide. Progression of the HNSCC is a consequence of both the interaction of environmental factors and the genetic inheritance, and is therefore multi-factorial. Smoking and alcohol dependence are the main risk factors for the development of this disease. Human papillomavirus (HPV) is also thought to be a risk factor for the disease at approximately 25%. The annual incidence of head and neck cancers worldwide, the annual incidence of head and neck cancers worldwide; about 300,000 deaths result in about 550,000 cases per year. The male to female ratio ranges from 2: 1 to 4: 1. HNSCC is the sixth most common cancer worldwide incidence. The overall five-year survival rate of HNSCC patients is approximately 40-50%. Approximately one third of patients are suffering from early stage disease (T1-2, N0). Early HNSCC therapy usually involves single modality therapy with surgery or radiation [1-6]. Treatment for head and neck cancer may include surgery, radiotherapy, chemotherapy, targeted therapy, or a combination of these treatments. The treatment plan depends on various factors such as the precise location of the tumor, cancer stage, age of the patient and general health status [7].

Generally, head and neck cancers at the advanced stage result in the death of the patient despite all kinds of treatment. For this reason, the integration of chemotherapy and radiotherapy is crucial to prolong survival and improve the quality of life of patients. It is necessary to know and follow the general conditions of the patients before starting any treatment [8]. Estimation of survival rate and decision-making of treatment are of great importance both for cancer patients and for physicians. The World Health Organization has stated that cancer is the second leading cause of death in the world. With early treatment, early detection of cancer will increase prognosis for cancer. At the same time, the prognosis depends on cancer spreading to lymph node drainage sites and metastasizing in different regions. A cancer staging system called TNM (Tumor, Node, Metastasis) is commonly used to determine the cancer status. Spreading to regional nodes or other nodes and distant metastasis reduce survival. Data-driven predictive models for cancer survival can help in prognosis and cancer management [9].

When deciding on the resection surgeon, a number of factors are considered that will affect the quality of life of the patient, including high rates of morbidity and the likelihood that death will occur rapidly. The morbidity rate is reported to be at least 50% and at least 15% mortality during the operation. The effective use of clinical data through the use of machine learning techniques such as artificial neural networks (ANN) can lead to more accurate diagnosis and prediction results, enabling better understanding of complex procedures and improving patient outcomes. Treatment decisions on the oncology not only directly affect survival, but also affect the quality of life of patients [10]. Survival analysis with machine learning methods provides greater convenience in logic implementation than statistical methods [11]. In machine learning, mathematical algorithms used as computer programs are used to recognize patterns in large data sets and to iteratively refine this recognition with additional data. When a specific medical diagnosis is made, predicting survival is crucial in improving patient care and providing information to patients and clinicians. In a data set of specific demographics (eg, age), diagnostic (eg, tumor size), and procedural (eg, radiation and/or surgery) information, it is very important to know that any of this information is sufficient to predict survival for head and neck cancer. Survival analysis is considered clinically important to evaluate the prognosis of the patient. More accurate results can be obtained by applying a correlational approach through machine learning to predict survival [12].

In recent years, significant progress has been made in the development of machine learning. The machine learning method implements a variety of techniques and approaches to analyze and summarize the data obtained from the databases, thus producing relevant information. Artificial neural networks (ANN) have proven to be very effective in disease prediction and survival analysis. Moreover, the unknown relationship between ANN input and output variables can be effectively predicted by repeating the learning and verification process of an ANN in a computer environment until the desired approach is provided. The quality of life of the patient before the treatment and the possible effect of the treatment on the survival of the patient and the associated quality of life affect the perception of treatment value of doctors and patient. It is

Hidir Selcuk Nogay is with Electrical and Energy Department, Kayseri University, Kayseri, Turkey, (e-mail: nogay@kayseri.edu.tr).

Manuscript received May 15, 2020; accepted May 31, 2020.
Digital Object Identifier:

important that doctors try to protect or improve the quality of life of their patients. Patients generally believe that there is no option other than surgery. Another factor to consider is the potential regret of the doctor or the patient giving a wrong treatment decision [13,14].

To test for survival prediction studies, head and neck cancer disease is the most appropriate disease with some characteristics. Head and neck cancers can be correctly staged using clinical and radiological techniques, distant metastases appear later, and have a short 4 year period of hazard, which facilitates reliable monitoring. In addition, local-regional recurrence in head and neck cancer is easier to detect than other types of cancer and it typically occurs within two years. In addition to predicting survival with artificial neural networks, a hypothesis that ANN produces more successful estimates than other machine learning and classification approaches is tested [15].

In this study, survival for head and neck cancer disease was estimated using machine learning methods. Starting from the date the head and neck cancer disease was diagnosed, the survival of the patient was estimated at the end of a minimum of 8 months of sleep, without a maximum time limit. The data used to perform the study were obtained from the Cancer Imaging Achieve (TCIA) website. The Cancer Imaging Archive (TCIA) is an archive of medical images and clinical data based on the work of Martin Valliere's at the Department of Medical Physics at McGill University. Machine learning methods used in working; Artificial Neural Network, Decision Tree, Linear Discriminant, Logistic Regression, Nearest Neighbor Classifier, Linear Support Vector Machine, and Quadratic Support Vector Machine. The main goal of this study is to be able to provide a decision aid to understand the survivability of head and neck cancer patients and evidence-based cancer management with the applied prediction methods and results. Evidence-based medicine and evidence-based health care are the focus of modern clinical medicine. This study may also contribute to cancer management. While the scope of this article is limited to cases of head and neck cancers, the machine learning algorithms and methodologies used are also suitable for other cancer management practices [16,17].

2. MATERIALS AND METHODS

2.1 Data set

The data were obtained from the Cancer Imaging Achieve (TCIA) website. The dataset consists of FDG-PET / CT and radiotherapy planning CT imaging and clinical data of 300 head and neck cancer (H & N) patients from four different hospitals in Québec province of Canada. Head and neck cancers of 300 patients in the data set are histologically proven. FDG-PET / BT scans were performed between April 2006 and November 2014 on average 18 days before the start of treatment for all patients. In 93 (31%) of 300 patients, radiotherapy treatment was performed by direct radiation oncologists and FDG-PET / BT imaging was performed. These image data were then used for treatment planning. Radiotherapy (16%) was administered alone to 48 of 300 patients. 252 of 300 patients were treated with chemo + radiation (84%) as part of treatment management for remediation. The median follow-up of all patients is 43 months. During the follow-up period, patients with no local or recurrent metastases and less than 24 months of follow-up were removed from the study.

TABLE IA
ENUMERATION OF THE DATASET

Label		Number
Sex	Male	**1**
	Famele	**2**
TNM group stage	Stage I	**20**
	Stage II	**21**
	Stage IIB	**22**
	Stage III	**23**
	Stage IV	**24**
	Stage IVA	**25**
	Stage IVB	**26**
Primary Site	Larynx	**3**
	Nasopharynx	**4**
	Oropharynx	**5**
	Hypopharynx	**6**
	Unknown	**7**
HPV Status	-	**27**
	+	**28**
	N/A	**29**
Therapy	chemo + radiation	**37**
	radiation	**38**
	TBD	**500**

TABLE IB
ENUMERATION OF THE DATASET

Label		Number
T - Stage	Tx	**8**
	T1	**9**
	T2	**10**
	T3	**11**
	T4	**12**
	T4a	**30**
	T4b	**31**
N – Stage	N0	**13**
	N1	**14**
	N2	**33**
	N2a	**15**
	N2b	**16**
	N2c	**17**
	N3	**32**
	N3a	**34**
	N3b	**35**
M – Stage	M0	**18**
	M1	**19**
	Mx	**36**
Survival	Dead	**1**
	Alive	**0**

TABLE II
SUMMARY OF DATA SET

	Data Name	Range
Inputs	Sex	**1…2**
	Age	**18…90**
	Primary Site	**3…7**
	Tstage	**8…31**
	Nstage	**13…35**
	Mstage	**18…36**
	TNM group stage	**20…26**
	HPV status	**27…29**
	Time diagnosis to PET (days) (TDP)	**-203…108**
	Time diagnosis to CT sim (days) (TDCT)	**-210…500**
	Time diagnosis to start treatment (days) (TDS)	**-195…128**
	Time diagnosis to end treatment (days) (TDE)	**-265…458**
	Therapy	**37…38**
	Locoregional	**0…1**
	Distant	**0…1**
Output	**Death**	**0…1**

During the follow-up period, 45 patients (15%) developed locoregional recurrence. Forty patients developed distant metastases. Fifty-six patients (19%) lost their lives [16,17]. 298 of the dataset were used for training and testing in the estimation models. 15% of the data set was used for testing purposes, 15% for verification purposes and the remaining 70% was used to train the models. To use the data set in the models, numbering is done as in Table 1a and Table 1b. The same data set was used in all models in the study. The summary of the data set used in the study is shown in Table 2 together with the input and output data.

2.2 Artificial Neural Networks (ANN)

Artificial neural networks are a mathematical machine learning method that simulates the human brain. ANN is a form of artificial intelligence used for estimation purposes in many application fields. To analyze complex systems, artificial neural networks are generally used. Neural networks are widely used for classification and survival predictions in medical research in the last 20 years. Artificial neural networks are thought to be more influential than statistical methods in that they facilitate not only classification but also decision making [10,15,18].

There are many ANN studies on survival analysis, survival prediction, classification and other diagnostic and therapeutic approaches to diseases. Artificial neural networks provide a more flexible survival time estimate than conventional methods since they can easily account for variable interactions and form a nonlinear prediction model [19-29].

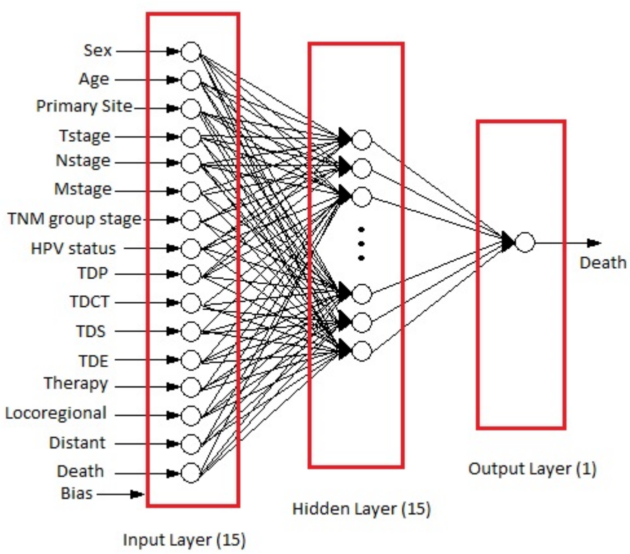


Fig.1. Architecture of the proposed ANN model

The backpropagation learning algorithm is used to train the artificial neural network model used in this study. In the proposed ANN model, the number of neurons in the hidden layer is 15. There are 15 inputs at the input layer and one output at the output layer. The architecture of the proposed ANN model is shown in Figure 1. The transfer function used in the ANN model is the hyperbolic tangent sigmoid transfer function shown in equation (1). 70% of the data set used in the study was used to train the model, 15% of the remaining data was used for model testing and 15% was used for validation purposes. The test and validation data were randomly selected.

$$\text{tansig}(n) = \frac{2}{1+exp(-2n)} - 1 \tag{1}$$

2.3 Decision Tree (DT)

The decision tree is a commonly used data mining approach to classification and estimation. Although other methods such as neural networks can be used for classification, the decision tree provides an advantage for decision-makers in terms of ease of interpretation and intelligibility. Classification of the data using the DT technique is a two-step process, learning and classification. In the learning step, a previously known training data is analyzed by the classification algorithm to form a model. The learned model is shown as a classification rule or decision tree. In the classification step, the test data is used to determine the correctness of the classification rules or decision tree. If accuracy is acceptable, rules are used to classify new data [30-32].

In this study, CART decision tree algorithm is used in the MATLAB © environment. The CART algorithm can be used as a solution to classification and regression problems since it can accept both numerical and nominal data types as input and estimation variables. The CART decision tree has a structure that is divided into two recursively. The CART tree benefiting from the Gini index as a branching criterion grows continuously by dividing, without any stopping in the establishment phase. In the stage where a new division is not going to take place, pruning starts from the tip to the root. The most successful decision tree possible is tried to be determined by evaluating with a test data independently selected after each pruning operation [30-34]. In the study, 15% of the data set was used for the test.

2.4 Linear Discriminant Analysis (LDA)

Discriminant analysis is a classification method. It assumes that different classes produce data based on different Gaussian distributions. To train a classifier, the fit function estimates a Gaussian distribution parameter for each class. To estimate the classes of new data, the trained classifier identifies the class with the lowest false classification cost. Discriminant analysis is a statistical technique that performs the assignment of a unit that is measured over a certain number of known masses. When this assignment is made, an error is made according to the observation value it receives when a unit is assigned to a different mass. In the discriminant analysis, this error is called the error rate or the probability of incorrect classification. The purpose of discriminant analysis is to make the assignment process with a minimum of errors. Linear discriminant analysis is also known as the Fisher separator termed by the inventor Sir R. A. Fisher. Linear Discriminant Analysis (LDA) is a classification method used in statistic, pattern recognition and machine learning to find linear combinations of properties. Although LDA is simple, it is a model that produces good results in complex problems [31,35-39].

LDA is also an important statistical tool for feature extraction and size reduction. The basic tenet of LDA is to reflect the high-dimensional data in a low-dimensional space, to minimize the

 ISSN: 2548-0650

distance within the classroom, to maximize the distance between classes, and then to maximize class separation [39, 40]. A number of discrimination vectors are obtained in the LDA method. These discrimination vectors maximize the 'between classes distribution matrix' (Sb) while minimizing the in-class distribution matrix (Sw) [41].

Suppose that an A data matrix is given as follows;

$$A = [a_1, a_2, a_3, \dots a_n] \in R^{m \times n}$$

, and;

$a_i \in R^m$; $i = 1, \dots, n$; a_i is the ith data sample.

Considering the binary classification example;

Let n_0 be the number of samples with zero class, let n_1 be the number of samples in class 1, and we can express the sum of both classes as; $\sum_{i=0}^{1} n_i = n$

A data matrix; $A = [A_0, A_1]$ and suppose that $A_i \in R^{m \times n_i}$ covers ni data samples of the ith class. In Linear Discriminant Analysis, two matrices called Sb and Sw can be defined as:

$$S_w = \sum_{i=0}^{1} \sum_{a_j \in A_i} (a_j - m_i)(a_j - m_i)^T \tag{2}$$

$$S_b = \sum_{i=0}^{1} n_i (m_i - m_T)(m_i - m_T)^T \tag{3}$$

Where m_i is the mean of the ith class. m_T is the total average of all data samples. LDA uses these two matrices to find the optimal sequence of discriminant vectors that maximize Fisher's criterion [40-42].

In this study, Fisher 's Linear Discriminant Analysis is used in MATLAB Classifier environment. For the LDA prediction model, 15 inputs are assigned as predictors and one data is assigned as outputs or response. 15% of the data set is reserved for testing as it is in other models.

2.5 Logistic Regression Classifier (LRC)

Regression analysis in statistics is a method used to determine the causal relationship between a variable and other variables. The variable is divided into X and Y variables. Variable X (x-axis) is named with various terms such as descriptive variables and independent variables. The Y variable is known as the affected variable and the dependent variable. Both of these variables can be random variables, but the affected variables must always be random variables [40-42].

Regression analysis is one of the statistical methods that have proven to be extremely reliable. Logistic regression is a popular, nonlinear, statistical model in which a flexible logistic function is introduced to form the basic mathematical form of the logistic model. Logistic regression analysis is a regression analysis as well as a differential analysis technique. In the logistic regression model, the dependent (binary) variable is a discrete variable such as 0, 1; risk-indicating case 1, other cases 0. In regression problems, the key value is the mean value of the dependent (result) variable, depending on the value of a given independent variable. This value is called the conditional average and is denoted by E (Y \ x). Here Y is the dependent variable and x is the independent variable. In linear regression analysis, it is assumed that the conditional mean is a linear equation of x. The logistic regression model is very efficient if the outputs are binary. In logistic regression analysis, the corresponding conditional mean function is as follows when the output Y is binary and the variables X are real numbers [42-44].

$$E(Y \setminus X) = x = \frac{\exp(\alpha^* + \beta x)}{1 + \exp(\alpha^* + \beta x)} \tag{4}$$

Here, α^* and β are scalar parameters. A multivariate logistic regression model was used in this study. Logistic models with multiple independent variables are called multivariate logistic regressions. Structurally, this model is not different from many other variable regression models, and interpretation of the regression coefficients is different. Interpretation depends on the type of independent variable. Non-continuous variables in a multivariate logistic regression may be nominal (classifiable) and ordinal (sortable) variables. Design variables can be used in order to put intermittent and nominal scale-independent variables into the equation. The model used in the study was obtained in the MATLAB classifier environment and 15 independent variables were used as predictors [43-45].

2.6 K-Nearest Neighbor Classification (KNN)

The KNN classification method is one of the classical and popular classification approaches. The KNN classification method is used in different areas due to its simplicity and effectiveness. The *K*-Nearest Neighbor classification algorithm, which is briefly referred to as KNN, is based on the principle that "objects close to each other probably belong to the same category". An object that is unknown to which class belongs is called a test example. Pre-classified objects are called learning examples. In the KNN algorithm, the distances from the test sample to the learning samples are calculated and if the nearest k samples belong to which class, the test instance is considered to belong to that class. In the KNN algorithm, the samples in the training set are specified by *n*-dimensional numerical properties. All training samples are held in a *n*-dimensional sample space so that each sample represents a point in n-dimensional space. When an unknown instance is encountered, the class tag of the new instance is assigned by determining the *k* 'instances closest to the relevant instance from the training set, according to the majority vote of the class labels of the nearest neighbour '*k*' [46-49]. In this study, KNN classification method was performed in MATLAB environment. In KNN model, *k* coefficient 1 is taken and Euclidean distance criterion is used for distance. The processing steps of the KNN classification algorithm can be summarized as follows:

Step 1: The distances of the test sample to the learning samples are calculated.

Step 2: Select the closest k samples.

Step 3: If the number of samples belonging to which class is the greatest, the test sample is also assigned to this class.

2.7 Support Vector Machines (SVM)

The SVM has the ability to separate with linear separators in two-dimensional space and planar separators in three-dimensional space two or more classes. The working principle of SVM is to estimate the most appropriate decision function that can distinguish between two classes.

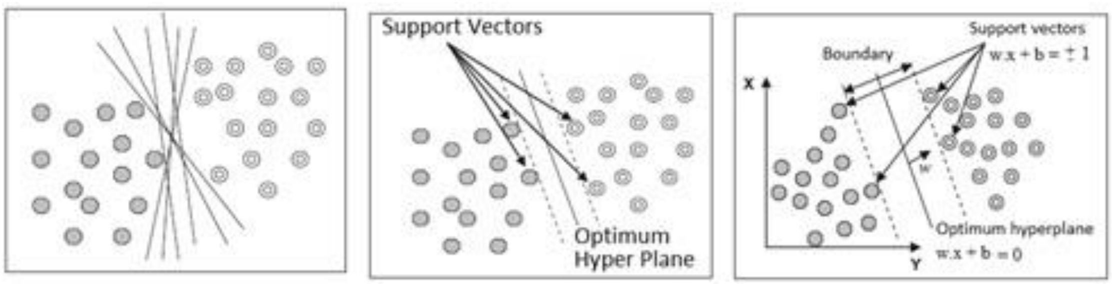
In other words, the basic principle is to define the hyperplane, which can distinguish between the two classes in the most appropriate way [50-54]. SVM is used for classification and estimation purposes. Especially in the field of medicine, many articles have been published for purposes such as diagnosis and classification of diseases [58,59].

2.7.1 Linear Support Vector Machines (LSVM)

In SVM classification, it is aimed to separate samples of two classes, which are usually shown as {-1, + 1}, with the help of a decision function. By using the decision function, it is necessary to find a hyperplane which can best distinguish the training data. As shown in Fig. 2a, many hyperplanes can be plotted which can distinguish two-class data.

However, the SVM's goal is to find the hyperplane that maximizes the distance between its nearest points. The support vectors and the optimal hyperplane are shown in Figure 2b. The 'optimum hyperplane', which makes the most appropriate difference by raising the limit to maximum, is shown in Figure 2c. In Figure 2c, the points that limit the border width are called support vectors. The support vectors are expressed in the form of $\mathrm{w.x_i + b = \pm 1}$. The limits of the optimal hyperplane must be maximized. Lagrange equations are used for this. The decision function for LSVM can be written as in equation (5) [58,59].

$$f(x) = sign\left(\sum_{i=1}^{k} \lambda_i\, y_i (x.x_i) + b\right) \quad (5)$$

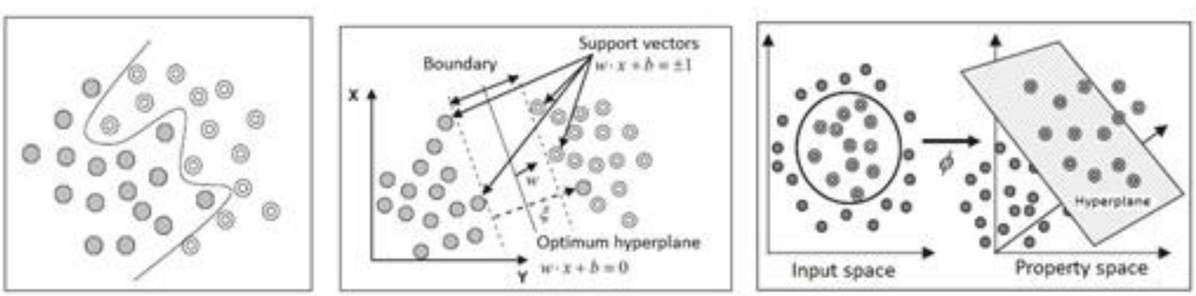


a) b) c)

Fig.2. Support vectors and hyperplanes

2.7.2 Non-linear SVM

For many data sets, the data in the two-dimensional state cannot be separated by the help of linear delimiters. This is seen in Figure 3a. In this case, the problem arising from the fact that some of the training data remain on the other side of the optimum hyperplane is solved by defining a positive artificial variable (ξi) [Fig. 3b]. The balance between maximizing the boundary and minimizing the classification errors is controlled by an adjustment parameter indicated by C. The C adjustment parameter is always positive (0 <C <∞) [60-65]. As can be seen in Figure 3c, data that cannot be linearly separated in the input space is displayed in a high-dimensional space defined as the property space. Thus, the data can be linearly discriminated and the hyperplane between classes can be determined.



a) b) c)

Fig.3. Hyperplanes for non-linear SVM

Equation (6) shows the Kernel function. With kernel function SVM can perform non-linear transformations [58-64].

$$K(x_i, x_j) = \varphi(x).\varphi(x_j) \quad (6)$$

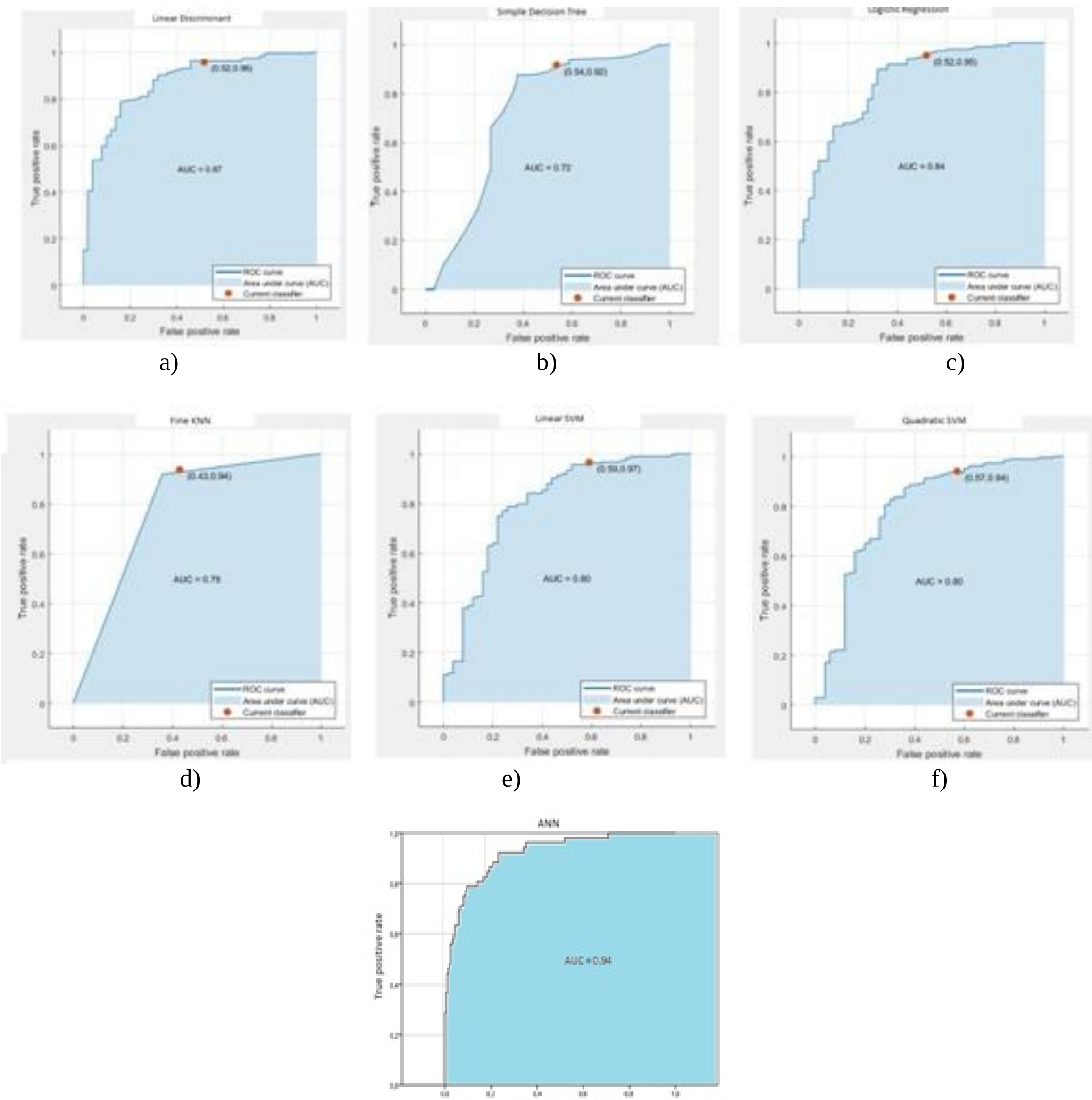
The decision to solve a two-class problem that cannot be linearly separated using the kernel function can be written as in equation (7) [58-64].

$$f(x) = sign\left(\sum_i \alpha_i\, y_i \varphi(x).\varphi(x_i) + b\right) \quad (7)$$

The kernel function to be used for a classification operation with SVM, and the determination of optimum parameters for this function are essential. Besides the parameters specific to the kernel function, the configuration parameter '*C*' for all support vector machines must be specified by the user. For this parameter, if too small or too large values are selected, a serious reduction in the classification accuracy is expected, since the optimal hyperplane cannot be determined correctly. On the other hand, if C = ∞, the SVM model is only suitable for data sets that can be linearly separated. As can be seen here, the selection of appropriate values for the parameters is a factor that directly affects the performance of the SVM classifier. Despite the use of trial and error strategies, the cross-validation approach allows successful results to be achieved. The purpose of the cross-validation approach is to determine the performance of the generated classification model. For this purpose, the dataset is divided into two parts. The first part is used as training data in the modelling which is the basis of classification, while the second part is processed as test data to determine the performance of the model. As a result of applying the model created by the training set to the test data set, the number of correctly classified samples shows the performance of the classifier. Therefore, by using the cross-validation method, the best classification performance is obtained and the model to be the basis of classification is determined by determining the kernel parameters [58,59]. An important issue to be considered in SVMs is the fact that large data groups have more than one cluster, depending on their particular characteristics. In order to be able to use SVM in multiple class situations, the problem must be transformed into a large number of binary class problems. The most commonly used approaches are the “One vs All” approach and the “One vs. One” approach [60-65]. This study was carried out in MATLAB environment and both "Quadratic SVM" and "Linear SVM" classification were realized. The classification approach used is the "One vs One" approach.

3. Results

Seven machine learning models were designed to predict the survival time for head and neck cancer in the study. The ROC (Receiver Operating Characteristic) curves obtained for these seven models are shown in Figures 4a, b, c, d, e, f, and g.



a) b) c)

d) e) f)

g)

Fig.4. The ROC curves; a) LDA, b) DT, c) LRC, d)KNN e) LSVM, f) QSVM, g) ANN.

In Figure 5, ROC curves of all models are shown on the same axis. In addition, the confusion matrices obtained from the classification models designed in Figure 6 are shown. The classification results obtained are shown in Table 3.

TABLE III
CLASSIFICATION RESULTS OF THE MODELS

Model Parameters	Machine learning methods						
	LDA	DT	LR	KNN	LSVM	QSVM	ANN
AUC	0.87	0.72	0.84	0.78	0.80	0.80	0.94
Accuracy (%)	86.9	83.2	86.2	86.9	86.2	84.6	90
Training Time (s)	3.1346	0.8772	11.667	1.9018	4.6418	1.3179	0.4

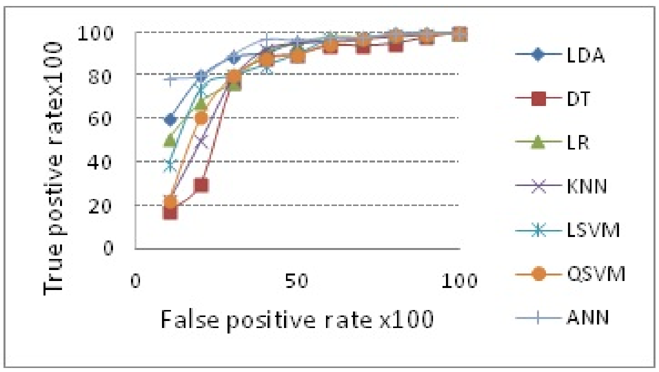


Fig.5. ROC curves of all models

Fig.6. Confusion matrices for all classification models

A performance curve was also drawn to evaluate the success of the designed artificial neural network model. The performance curve drawn is shown in Figure 7b. From this curve, the mean square error can also be observed.

TABLE IV
RESULTS OF ANN MODEL

	Samples	MSE	R
Training	208	7.67845e-2	7.10192e-1
Validation	45	1.58386e-1	6.20647e-1
Testing	45	9.50916e-2	7.04439e-1

In addition, mean square error and predicted values are shown in Table 4. The regression curves of the test and validation data obtained from the artificial neural network model are given in figure 7a. From Figure 8, the success status of the ANN model in the testing process can be interpreted.

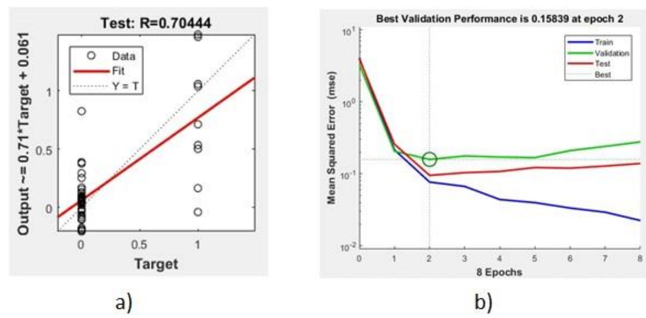


Fig.7. a) Regression and b) performance, curves of proposed ANN model

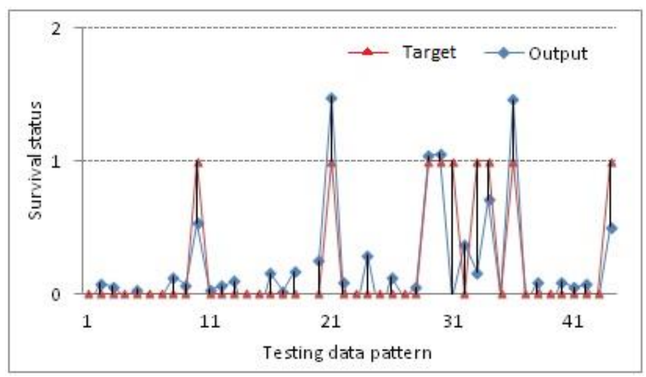


Fig.8. Comparison of the target and output of ANN for testing

4. CONCLUSIONS

When the obtained graphs and confusion matrices were examined, survival estimates were made for 298 cases in total. Separate results can be obtained from the application of each method.

First, from the linear discriminant analysis, it is clear that the prediction that 232 people will survive is correct. It was also correctly predicted that 27 people would die. Despite these correct estimates, the claim that it will survive for 29 people has also come out as a false prediction. In the same way, the prediction that 10 people will die is not true. In this case, for the head and neck cancer patients, survival was estimated with an accuracy of 86.9% by the "linear discriminant classification" method. This rate is extremely satisfactory.

When estimating and classifying by decision tree method was examined, it was estimated that 222 people would survive correctly. It was also estimated correctly that 26 people will die.

Though 20 people were estimated to die, they lived and were estimated incorrectly. Though 30 people were thought to live, they died. In this case, the ratio of correct estimates is 83.2% as can be understood from Table 3. Accuracy and other results may be considered to be generally successful, albeit worse than LDA. It is also very important how many seconds these results are reached and how long these estimates are valid for. As Table 3 reveals, these results were achieved at an extremely short time of 0.8 seconds. These estimates were made for the time until the first control of all patients, i.e. for a minimum of 8 months. In this case, it is easy to say that the decision tree method is also successful.

For another classification method, logistic regression, there are 230 cases that are thought to be alive and predicted correctly. The number of patients correctly estimated to die is 27. In this method, it can be seen that there are mispredictions, as shown in figure 6c. In this case, accuracy is close to LDA, can be said to be extremely successful with 86.2%

The number of surviving patients correctly estimated by the KNN method is 227, and the number of correctly estimated and lost patience is 32. The survival of 39 patients was estimated incorrectly. In this case, it can be seen in Table 3 that the accuracy rate is the same as the linear discriminant with 86.9%. However, when the ROC curve is examined, it can be observed that AUC is 87% in linear discriminant and 78% in KNN. In this case, it is possible to say from the ROC analysis that LDA is more preferable than KNN if it is considered that the KNN method is successful.

Considering the linear SVM, 234 patients in the confusion matrix are correctly predicted to survive. This is the largest number in all other methods. The number of patients who are correctly predicted to lose their lives is 23. This is the smallest number among other methods. The number of patients who are living despite being estimated incorrectly, that is, estimated to die, is only 8. This figure is the lowest and successful figure among other methods. Despite this successful outcome, however, one negative outcome is 33 patients who died despite the fact that they were expected to survive. In this case, the rate of accurate estimates is very successful with 86.2%.

For the Quadratic SVM model, the number of surviving patients correctly estimated is 228. Figure 6f shows that other estimation results are similar to those of other methods. It can be said that the accuracy rate is very good with 84.6%.

When the results obtained from the ANN approach are considered; the number of patients correctly estimated to live is 225, the number of patients who are correctly predicted to lose their lives is 43. Despite being predicted to live, the number of patients who actually lost their lives is 16, which is the most successful value. Despite being estimated to die, the number of patients living is 12. A correctly estimated survival rate of 90% can be said to be the most successful. However, even though the estimates appear to be close to real values, there can be observed differences in the ANN's figure 9 comparative curve. In order to perform ROC analysis, it is necessary to round the estimated values according to the output. The Confusion matrices are the result of these rounds. The accuracy rate is only 70% for the test data, as can be seen from the regression curve in figure 7a and the performance curve in figure 7b without rounding. However, when the outputs are considered to be 1 and 0, the accuracy rate is 90% since values close to these output values are rounded to these values.

These seven classification and prediction models used in the study can be compared with each other. In this case, it can be concluded that the most successful model is ANN and the other methods can be considered successful with at least 80% accuracy. In conclusion, using clinical data of head and neck cancers, survival estimates with machine learning approaches were obtained with at least 83% accuracy and at most 90% accuracy.

ACKNOWLEDGMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
The author would like to thank Dr. Martin Vallières from the Medical Physics Unit of McGill University, and TCIA

REFERENCES

[1] Allison, D. B., Maleki, Z. (2016). HPV-related head and neck squamous cell carcinoma: Anupdate and review, Journal of the American Society of Cytopathology, 5, pp.203-215.

[2] Maund, I., Jefferies, S. (2015). Squamous cell carcinoma of the oral cavity, oropharynx and upper oesophagus, Medicine, 43, 197-201.

[3] McGurk, M., Goodger, N. M. (2000). Head and neck cancer and its treatment: historical review, British Journal of Oral and Maxillofacial Surgery, 38, pp.209–220.

[4] Galbiatti, A. L. S., Junior, J. A. P., Maníglia, J. V., Rodrigues, C. D. S., Pavarino, É. C., Bertollo, E. M. G. (2013). Head and neck cancer: causes, prevention and treatment, Braz J Otorhinolaryngol, 79, pp.239-47.

[5] Young, D., Xiao, C. C., Murphy, B., Moore, M., Fakhry, C., Day, T. A. (2015). Increase in head and neck cancer in younger patients due to human papillomavirus (HPV), Oral Oncology, 51, pp.727–730.

[6] Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., Forman, D. (2011). Global cancer statistics, CA Cancer J Clin, 61, 69-90.

[7] Zini, E. M., Lanzola G., and Quaglini, S. (2017). Detection and Management of Side Effects in Patients with Head and Neck Cancer, IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), pp.1-6.

[8] Drago, G. P., Setti, E., Licitra, L., and Liberati, D. (2002). Forecasting the Performance Status of Head and Neck Cancer Patient Treatment by an Interval Arithmetic Pruned -Perceptron, IEEE Transactions on Biomedical Engineering, 49, 782-787.

[9] Shukla, N., Hagenbuchner, M., Wi, K. T., Yang, J. (2018). Breast cancer data analysis for survivability studies and prediction, Computer Methods and Programs in Biomedicine, 155, 199–208.

[10] Walczak, S., Velanovich, V. (2017). An Evaluation of Artificial Neural Networks in Predicting Pancreatic Cancer Survival, J Gastrointest Surg., 21, pp.1606–1612.

[11] Wróbel, L., Gudy´s A., and Sikora, M. (2017). Learning rule sets from survival data, BMC, Bioinformatics 285 DOI 10.1186/s12859-017-1693-.

[12] Lyncha, C. M., Abdollahib, B., Fuquac, J. D., de Carloc, A. R., Bartholomaic, J. A., Balgemannc, R. N., van Berkeld, V. H., Frieboesc, H. B. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques, International Journal of Medical Informatics, 108, pp.1–8.

[13] Wu, C., Wu, Y., Liang, P., Wu, C., Peng, S. F., Chiu, H. W. (2017). Disease-free survival assessment by artificial neural networks for hepatocellular carcinoma patients after radiofrequency ablation, Journal of the Formosan Medical Association 116, pp.765-773.

[14] Walczak, S., Velanovich, V. (2017). Improving prognosis and reducing decision regret for pancreatic cancer treatment using artificial neural networks, Decision Support Systems, doi: 10.1016/j.dss.2017.12.007.

[15] Walczak, A. S., Taktak, A. G. F., Helliwell, T. R., Fenton, J. E., Birchall, M. A., Husband, D. J., Fisher A. C. (2006). An artificial neural network improves prediction of observed survival in patients with laryngeal squamous carcinoma, Eur Arch Otorhinolaryngology, 263, pp.541–547.

[16] Vallières, M. et al. (2017). Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer, Sci Rep 10117 doi: 10.1038/s41598-017-10371-5.

[17] Cancer Imaging Archieve (2018). http://www.cancerimagingarchive.net/ , last accessed date: February 10, 2018.

[18] Ravdin, P, M., and Clark, G. M. (1992). A practical application of neural network analysis for predicting outcome of individual breast cancer patients, Breast Cancer Research and Treatment, 22, 285-293.

[19] Chi, C. L., Street, W. N., Wolberg, W. H. (2007). Application of Artificial Neural Network-Based Survival Analysis on Two Breast Cancer Datasets, AMIA Annu Symp Proc., pp.130–134.

[20] Dolgobrodov, S. G., Moore, P., Marshall, R., Bittern, R., Steele, R. J. C., Cuschieri, A. (2007). Artificial Neural Network: Predicted vs. Observed Survival in Patients with Colonic Cancer, Diseases of the Colon & Rectum, 50, pp.184–191.

[21] Ahmed, F. E. (2005). Artificial neural networks for diagnosis and survival prediction in colon cancer, Molecular Cancer, 29, doi:10.1186/1476-4598-4-29.

[22] Devi, M. A., Ravi, S., Vaishnavi, J., and Punitha, S. (2016). Classification of Cervical Cancer using Artificial Neural Networks, Procedia Computer Science, 89, 465-472.

[23] Ripley, R. M., Harris A. L., and Tarassenko, L. (1998). Neural Network Models for Breast Cancer Prognosis, Neural Comput & Applic, 7, pp.367 375.

[24] Shukla, R. S., Aggarwal, Y. (2017). Nonlinear Heart Rate Variability based artificial intelligence in lung cancer prediction, Journal of Applied Biomedicine, Vol.16, No.2, pp.145–155. doi:10.1016/j.jab.2017.

[25] De Laurentiis, M., and Ravdin, P. M. (1994). Survival analysis of censored data: Neural network analysis, detection of complex interactions between variables, Breast Cancer Research and Treatment, 32, pp.113-118.

[26] Ochi, T., Murase, K., Fujii, T., Kawamura, M., Ikezoe, J. (2002). Survival prediction using artificial neural networks in patients with uterine, cervical cancer treated by radiation therapy alone, Int J Clin Oncol 7, pp.294–300.

[27] Asria, H., Mousannif, H., Al Moatassime, H., Noel, T. (2016). Using Machine Learning Algorithms for Breast Cancer Risk, Prediction and Diagnosis, Procedia Computer Science, 83, pp.1064 – 1069.

[28] Francis, N. K., Luther, A., Salib, E., Allanby, L., Messenger, D., Allison, A. S., Smart, N. J., Ockrim, J. B. (2015). The use of artificial neural networks to predict delayed discharge and readmission in enhanced recovery following laparoscopic, colorectal cancer surgery, Tech Coloproctol, 19, pp.419 – 428.

[29] Iraji, M. S. (2017). Prediction of post-operative survival expectancy in thoracic lung cancer surgery with soft computing, Journal of Applied Biomedicine, 15, pp.151–159.

[30] Chien, C. F., Chen, L. F. (2008). Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry, Expert Systems with Applications, 34, pp.280-290.

[31] Discriminant analysis (2018). https://www.mathworks.com/help/stats/discriminantanalysis.html, last accessed date: February, 14, 2018.

[32] Zheng, H., Chen, L., Han, X., Zhao, X., Ma, Y. (2009). Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions, Agriculture, Ecosystems & Environment, 132, pp.98-105.

[33] Breiman, L., Freidman, J. H., Olshen R. A., and Stone, C. J. (1984). Classification and Regression Trees, Chapman and Hall, New York, USA.

[34] Stephen, E. F., Hsieh, Y., Rivadinera, A., Beer, T. M., Mori, M., Garzotto, M. (2006). Classification and Regression Tree Analysis for the Prediction of Aggressive Prostate Cancer on Biopsy, The Journal of Urology, 175, pp.918-922.

[35] Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems, Annals of Eugenics, 7, pp.179–188.

[36] Lachenbruch, P. A. (1975). Discriminant analysis, Hafner Press, New York, USA.

[37] Lachenbruch, P. A., and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis, Technometrics 10, pp.1-11.

[38] Elkhalil, K., Kammoun, A., Couillet, R., Al-Naffouri, T. Y., and Alouini, M. S. (2017). Asyptotic Performance of Regularized Quadratic Discriminant Analysis Based Classifiers, 2017 IEEE International Workshop on Machine Learning for Signal Processing Tokyo, pp.25–28.

[39] Cai, J., and Huang, X. (2018). Modified Sparse Linear-Discriminant Analysis via, Nonconvex Penalties, IEEE Transactions on Neural Networks and Learning Systems, Early Acces, pp.1-10.

[40] Lee, Y., Madayambath, S. C., Liu, Y., Da-Ting, L., Chen, R. and Bhattacharyya, S., S. (2017). Online Learning in Neural Decoding Using Incremental Linear Discriminant Analysis, IEEE2017 IEEE International Conference on Cyborg and Bionic Systems Beijing China, pp.173-177.

[41] Lawi, A., La Wungo, S., Manjang, S. (2017). Identifying Irregularity Electricity Usage of ustomer Behaviors using Logistic Regression and Linear Discriminant Analysis, IEEE 3rd International Conference on Science in Information Technology (ICSITech), pp.552-557.

[42] Tsangaratos, P., Ilia, I. (2016). Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size, Catena, 145, pp.164–179.

[43] Geng, P., Sakhanenko, L. (2016). Parameter estimation for the logistic regression model under case-control study, Statistics and Probability Letters, 109, 168–177.

[44] Razanamahandry, L. C., Andrianisa, H. A., Karoui, H., Podgorski, J., Yacouba, H. (2018). Prediction model for cyanide soil pollution in artisanal gold mining area by using logistic regression, Catena, 162, pp.40–50.

[45] Zhou, C., Wang, L., Zhang, Q., Wei, X. (2014). Face recognition based on PCA and logistic regression analysis, Optik, 125, pp.5916–5919.

[46] Duca, A., Bacciu, C., Marchetti, A. (2017). A K-Nearest Neighbor Classifier for Ship Route Prediction, IEEE OCEANS - Aberdeen, pp.1 – 6.

[47] Yu, Z., Chen, H., Liu, J., You, J., Leung, H., and Han, G. (2016). Hybrid, k-Nearest Neighbor Classifier, IEEE Transactions Cybernetics, 46, pp.1263-1275

[48] Li, W., Du, Q., Zhang, F., Hu, W. (2014). Collaborative Representation Based K-Nearest Neighbor Classifier for Hyperspectral Imagery , Hyperspectral Image and Signal Processing: Evolution in Remote Sensing , WHISPERS DOI: 10.1109/WHISPERS.2014.8077601.

[49] Cover, T. M., and Hart, P. E. (1967). Nearest neighbor pattern classification, IEEE Transactions on Information Theory, 13, pp.21–27.

[50] Support Vector Machines (2018). https://www.mathworks.com/help/stats/support-vector-machines-for-binary classification. html, last accessed date February 14, 2018.

[51] Cortes, C., Vapnik, V. (1995). Support-Vector Network, Machine Learning, 20, pp.273–297.

[52] Vapnik, V. N. (2000) The Nature of Statistical Learning Theory, 2nd Edition, Springer-Verlag, New York.

[53] Kavzaoglu, T., Colkesen, I. (2010). Investigation of the Effects of Kernel Functions in Satellite Image Classification Using Support Vector Machines, Map Journal July, 144, 73-82.

[54] Ilias, S., Tahir, N. M., Jailani, R. (2016). Feature extraction of autism gait data using principal component analysis and linear discriminant analysis, IEEE Industrial Electronics and Applications Conference IEACon., pp.275 – 279.

[55] Gao, L., Ye, M., Lu, X., Huang, D. (2017). Hybrid Method Based on Information Gain and Support Vector Machine for Gene Selection in Cancer Classification, Genomics, Proteomics & Bioinformatics, 15, pp.389-395.

[56] Wang, H., Zheng, B., Yoon, W., Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis, European Journal of Operational Research, 267, pp.687-699.

[57] Ghaddar, B., Sawaya, J. N. (2018). High dimensional data classification and feature selection using support vector machines, European Journal of Operational Research, 265, pp.993-1004.

[58] Madadum, H., Becerikli, Y. (2017). The implementation of Support Vector Machine (SVM) using FPGA for human detection, 10th International Conference on Electrical and Electronics Engineering ELECO, pp.1286 – 1290.

[59] Nefedow, A., Ye, J. Ye, Kulikowski, C., Muchnik, I., Morgan, K. (2009). Comparative Analysis of Support Vector Machines Based on Linear and Quadratic Optimization Criteria, IEEE International Conference on Machine Learning and Applications, pp.288 – 293.

[60] Vapnik, V. N. (1995). The Nature of Statistical Learning Theory, 1st Edition, Springer-Verlag, New York USA.

[61] Machine learning methods, (2018). https://www.mathworks.com , last accessed date: February 12, 2018.

[62] Osuna, E. E., Freund, R., Girosi, F. (1997). Support Vector Machines: Training and Applications, Massachusetts Institute of Technology and Artificial Intelligence Laboratory 144, Massachusetts.

[63] Whsu, C., Lin, C. J. (2002). A Comparison of Methods for Multiclass Support Vector Machines, IEEE Transactions On Neural Networks, 13, pp.415-425.

[64] Nogay, H.S. (2018). Classification of Different Cancer Types by Deep Convolutional Neural Networks, Balkan Journal of Electrical and Computer Engineering, Vol.6, pp.56-59.

[65] Zini, E. M., Lanzola G., and Quaglini, S. (2017). Detection and Management of Side Effects in Patients with Head and Neck Cancer, IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), pp.1-6.

BIOGRAPHIES

H. Selçuk Noğay received B.S degrees in Electrical Education from Kocaeli University, and M.S. and Ph.D degrees in Electrical Education from Marmara University respectially 2002, 2003 and 2008. His research interests include Artificial Neural Network, Deep Learning and signal processing technique He has been working as a Professor in Vocational Scholl of Kayseri University in Kayseri, Turkey