

# Knee Osteoarthritis Diagnosis Using Support Vector Machine and Probabilistic Neural Network

Özge Aksehirli<sup>1</sup>, Duygu Aydin<sup>2</sup>, Handan Ankarali<sup>1</sup> and Melek Sezgin<sup>3</sup>

<sup>1</sup> Department of Biostatistics and Medical Informatics, Faculty of Medicine, University of Düzce  
Düzce, 81000, Turkey

<sup>2</sup> Department of Biostatistics, Faculty of Medicine, University of Hacettepe  
Ankara, 06000, Turkey

<sup>3</sup> Department of Physical Therapy and Rehabilitation, Faculty of Medicine, University of Mersin  
Mersin, 33000, Turkey

## Abstract

Support Vector Machines (SVM) and Probabilistic Neural Networks (PNN) that are within the scope of artificial intelligence applications has been widely accepted. Because of strong statistical foundations of these methods, they generate generalizable results. In this study we aimed to compare classification performance of SVM and PNN methods using real and empirical data sets. In the real data set, we investigated the success of diagnosis of some demographic characteristics and some gene polymorphisms on knee osteoarthritis. In addition, success of SVM and PNN methods were compared for classification of samples that have various data structure. We found that the classification performance of PNN method is better than SVM. The most important disadvantage of PNN method is that the analysis time increase with sample size and number of attributes. When the proportion of the observation number to attributes number is greater than 5, SVM and PNN can be used for classification.

**Key Words:** *Optimization, classification, radial basis function, support vector machine, neural networks*

## 1. Introduction

Support Vector Machines (SVM) and Artificial Neural Networks (ANN) that are within the scope of artificial intelligence applications has been widely accepted in many areas. Especially they are used in the areas of learning theory, statistics, biology and engineering. Generally these methods are also called as data mining methods [1-3]. Both of the methods are commonly used for classification, estimation, pattern recognition, handwriting recognition.

A neural network is a powerful data modeling tool that is able to capture and represent complex input/output relationships. The structure of the method can change according to internal and external stimuli and thereby it can learn. Parameters are assigned to the neuron's input. The output of the neuron is combination of nonlinear inputs weighted by parameters [4]. In the literature, there

are many artificial neural network models for various problems. Probabilistic neural network (PNN) is one of the newest artificial neural network models. It is derived from radial basis function and its training phase is very fast than other Networks, and also the classification of a new observation is less time consuming than other models. It is also robust against to outliers [5]. The PNN combines the simplicity, speed and transparency of traditional statistical classification models and the computational power and flexibility of back-propagated neural networks [6]. PNN is attempting to minimize the experimental risk.

SVM is the other method used in this study, and it is aimed to minimize both structural and experimental risk. Thereby its aimed to take similar results from training and test sets. In classification problems, radial basis PNN and SVM with the radial basis kernel function give generally similar results [7-9]. In addition SVM and PNN rely mostly on powerful classification algorithms to deal with redundant and irrelevant attributes [10]. SVM can produce successful estimates for solving nonlinear complex problems [11].

In this study, we aimed to compare performances of SVM and PNN methods by using both empirical data sets obtained from computer simulation study and real data sets with regard osteoarthritis of the knee.

## 2. Methods

### 2.1. Support vector machine (SVM)

The aim of SVM modeling is finding the optimal hyperplane that separates clusters of vector in such a way as to be as far as possible from the closest members of both classes. The original optimal hyperplane algorithm was a linear classifier. Then, nonlinear classifiers was

developed by applying the kernel trick to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. Hyperplanes with larger margin are less likely to overfit the training data [12-14].

Hyperplanes are formulated as follows:

$$\langle w \cdot x \rangle + b = 0 \quad (1)$$

Where  $w$  is the vector of weights for all attributes,  $b$  is a real number representing the y-intercept or bias,  $x$  is the vector of attribute values representing instances in the training data,  $\langle w \cdot x \rangle$  is an inner product and  $\langle w \cdot x \rangle$  equal to  $w'x$ .

Figure 1 below presents an overview of the SVM process.

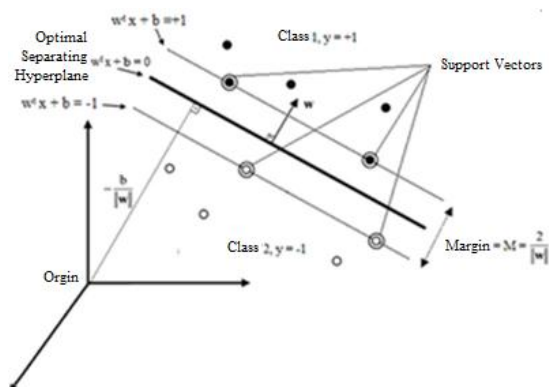


Fig. 1 Overview of SVM classification

$\frac{|b|}{\|w\|}$ : is the perpendicular distance from the hyperplane to the origin,  
 $\|w\|$ : is expressed as the Euclidean norm of  $w$ .

In Figure 1, support vectors are the examples closest to the separating hyperplane and the aim of SVM is to orientate this hyperplane in such a way as to be as far as possible from the closest members of both classes. To determine the optimum hyperplanes to form the boundaries of the two hyperplanes parallel to this plane must be determined [15].

$$\min_{wb} \frac{1}{2} \|w\|^2 \quad (2)$$

$$y_i(w^t x_i + b) \geq 1, \forall_i \quad (3)$$

In order to cater for the constraints in this minimization, we need to allocate them Lagrange multipliers  $\alpha$ , where  $\alpha_i \geq 0, \forall_i$ . As a result linear decision function is obtained as below.

$$f(x) = (\sum \alpha_i y_i x_i x) + b = w^T x_i + b \quad (4)$$

Such that for each training sample, the function yields  $f(x) > 0$  for  $y_i = +1$  and otherwise  $y_i = -1$ . In other words, training examples from the two different classes are separated by the hyperplane  $f(x) = w^T x_i + b = 0$ .

The linear SVM can be readily extended to a nonlinear classifier by first using a nonlinear operator  $\Phi(x_j)$  to map the input pattern  $x$  into a higher dimensional space  $H$ .

To construct a nonlinear support vector decision function, the inner product  $\langle x_i, x_j \rangle$  is replaced by a kernel function  $K(x_i, x_j)$ . The mapping  $\Phi(x_j)$  enters the problem only implicitly through the kernel function  $K(x_i, x_j)$ , thus, it is only necessary to define  $K(x_i, x_j)$ , which implicitly defines  $\Phi(x_j)$ . A kernel function is defined as [13,16]:

$$K(x_i, x_j) = \Phi(x_j)\Phi(x_j) \quad (5)$$

There are number of kernels that can be used in SVM models. Because Radial Basis Function has been found to work well in for a wide variety of applications we used this kernel function in SVM and PNN. SVM models have a cost parameter,  $C$ , that controls the trade off between allowing training errors and forcing rigid margins. The performance of SVM classifier depends on the choice of the cost parameter and the kernel parameter, so grid search was performed first. Once the grid search finished, a pattern search was performed over a narrow search range surrounding the best point found by the grid search.

## 2.2. Probabilistic neural network (PNN)

PNN is conceptually similar to K-Nearest Neighbor (k-NN) models. The basic idea of the method is that an estimated target value of an item is probably to be the same as other items which have close values of the predictor variables. The distance is computed from the point being evaluated to each of the other points, and a radial basis function is applied to the distance to compute the weight (influence) for each point. The further some other point is from the new point, the less influence it has.

The most common RBF is the Gaussian function. The sigma value ( $\sigma$ ) of the function determines the spread of the RBF function; that is, how quickly the function declines as the distance increased from the point. We used the conjugate gradient algorithm to compute the optimal sigma values.

All PNN networks have four layers as illustrated in Figure 2 [17].

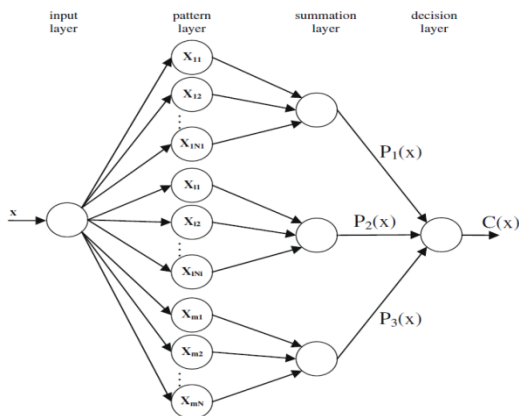


Fig. 2 Probabilistic Neural Network Scheme

The input layer is solely responsible of distribution of neurons to pattern layer. Probability density function is estimated by using multi dimensional kernels in the pattern layer. By summing and averaging, posterior probability density function is computed for each class in the summation layer. Depending on the information taken by all summation layers neurons and using Bayes's decision rule, classification of observation is done in the decision layer [18].

The probability density functions of PNN are evaluated using the Parzen's non-parametric estimator. PNN utilizes one probability density function for each category, as shown by (6).

$$P_i(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d n_i} \sum_{i=1}^{n_i} \exp \left[ -\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2} \right] \quad (6)$$

Where  $d$  denotes the dimension of the  $x$ ,  $\sigma$  is the smoothing parameter and  $x_{ij}$  is the neuron vector,  $n_i$  is the number of observation in  $i^{th}$  category. The process of removing unnecessary neurons is an iterative process. Leave-one-out validation is used to measure the error of the model with each neuron removed. The neuron that causes the least increase in error is then removed from the model. The process is repeated with the remaining neurons until the stopping criterion is reached [18].

### 2.3. Empirical data

Empirical data produced by writing a macro in MINITAB (ver. 15). Six numerical attributes which have low, moderate and high correlation between each other produced from multivariate normal distribution (Table 1).

Table 1: Correlation matrix for six attributes

	x1	x2	x3	x4	x5	x6
x1	1	0.9	0	0	0	0
x2		1	0	0	0	0
x3			1	0.5	0	0
x4				1	0	0.2
x5					1	0
x6						1

Samples were produced from two multivariate normal populations which have different mean vector and same standard deviation (Table 2). Differences between these two populations known as effect size. In this simulation study we took into account two effect size (1 standard deviation and 2 standard deviation, Table 2). The ratio of sample size to number of attribute was determined as 5 and 17. Thereby the size of the selected samples set at 30 and 100 observation. After the combining of effect sizes and sample sizes, we obtained four combinations (each column of Table 2). SVM and PNN were applied 100 times in each combination.

Table 2: Empirical data characteristics in simulation study

p=6 Attribute		$\delta$ (Effect Size)			
		1 SD *		2 SD *	
Sample size		n=30	n=100	n=30	n=100
Mean Vector	1 <sup>th</sup> Sample	$\mu = 0$	$\mu = 0$	$\mu = 0$	$\mu = 0$
	2 <sup>th</sup> Sample	$\mu = 1$	$\mu = 1$	$\mu = 2$	$\mu = 2$
Standard Deviation Vector	1 <sup>th</sup> Sample	$\sigma = 1$	$\sigma = 1$	$\sigma = 1$	$\sigma = 1$
	2 <sup>th</sup> Sample	$\sigma = 1$	$\sigma = 1$	$\sigma = 1$	$\sigma = 1$

\* SD: Standard Deviation

### 2.4 Real data

The real data were obtained from patients who were referred to the Department of Physical Therapy and Rehabilitation of Mersin University School of Medicine. This data set consisted age, sex, weight, height and FAS844, FAS670, FASL377 and FASL124 gene polymorphism of 152 osteoarthritis of the knee and 102 healthy individual.

DTREG (ver. 3.0) packet programme was used for all computation.

### 3. Results

#### 3.1. Simulation results

Two samples which have 1 standart deviation (SD) effect size and contain 30 observations were classified with SVM and PNN methods and classification performances of the methods are given in Table 3.

When the samples which have 1 standart deviation (SD) effect size and contain 30 observations classified separately for train and test sets with SVM, classification performances of the method were found over 80% and error rates were found about 15%. In addition, the results of the training and test samples were compatible with each other. This compatibility shows the generalizability of the model. Analysis time of the method was extremely short with an average 64.42 split-second. When the model parameters were examined, it was found that mean of the sigma, which is RTF parameter, was  $0.923 \pm 1.418$ , minimum value was 0.001 and maximum value was 6.655 for 100 trials. These values can be an estimate that may be used in the analysis of similar data sets. And thus, analysis' run time will be shorter.

When the samples which have 1 standart deviation (SD) effect size and contain 30 observations classified with PNN, classification performances of the method were found over 90% and error rates were found about 5%. In addition, the results of the training and test samples were compatible with each other. This compatibility shows the generalizability of the model. Analysis time of the method was extremely short with an average 62.76 split-second. For 100 trial, mean error, minimum error and maximum error of model obtained with PNN method was calculated as  $0.162 \pm 0.073$ , 0.0007 and 0.280 respectively.

According to results given in the Tablo 3, it can be said that classification performance of PNN method is higher than SVM. However, especially for small sample sizes, significant differences were not observed between the methods in terms of analysis time. These findings show that PNN method should be preferred for defined conditions.

Table 3: The results of the classification performance of SVM and PNN for  $n$  is 30 and ES is 1

<i>SVM results (n=30) ; effect size = 1 SD</i>		
Run Time = $64.42 \pm 11.59$ split-second		
	<b>Training</b>	<b>Validation</b>
Accuracy (%)	$87.3 \pm 4.6$	$82.7 \pm 5.9$
Sensitivity (%)	$87.8 \pm 6$	$83.5 \pm 6.9$
Specificity (%)	$86.9 \pm 6$	$81.9 \pm 6.4$
False Positive (Type 1 error) (%)	$13.1 \pm 5.5$	$18.0 \pm 6.4$
False Negative (Type 2 error) (%)	$12.2 \pm 5.5$	$16.4 \pm 6.9$
Positive Predictive Value (PPV) (%)	$87.2 \pm 4.9$	$82.3 \pm 5.9$
Negative Predictive Value (NPV) (%)	$87.8 \pm 5.6$	$83.4 \pm 6.5$
Area Under ROC Curve	$0.943 \pm 0.033$	$0.895 \pm 0.049$
<i>PNN results (n=30) ; effect size = 1 SD</i>		
Run Time = $62.76 \pm 8.64$ split-second		
	<b>Training</b>	<b>Validation</b>
Accuracy (%)	$95.1 \pm 2.8$	$94.0 \pm 3.3$
Sensitivity (%)	$95.3 \pm 4.7$	$94.1 \pm 5.1$
Specificity (%)	$95.0 \pm 3.7$	$94.0 \pm 4.2$
False Positive (Type 1 error) (%)	$5.0 \pm 3.7$	$6.0 \pm 4.2$
False Negative (Type 2 error) (%)	$4.7 \pm 4.7$	$5.9 \pm 5.1$
Positive Predictive Value (PPV) (%)	$95.1 \pm 3.5$	$94.2 \pm 3.9$
Negative Predictive Value (NPV) (%)	$95.5 \pm 4.4$	$94.3 \pm 4.7$
Area Under ROC Curve	$0.956 \pm 0.028$	$0.939 \pm 0.035$

When the samples which have 1 standart deviation (SD) effect size and contain 100 observations classified with SVM, classification performances of the method were found very similar to the set that contains 30 observations. According to these results, when the proportion of  $n/p$  was greater than 5, SVM gave similar results. Analysis time of the method for 100 observation was found longer than set of 30 observation (Table 4). Also mean, minimum and maximum values of the sigma parameter was calculated as  $0.558 \pm 1.254$ , 0.001 and 6.655 respectively.

When the samples which have 1 standart deviation (SD) effect size and contain 100 observations classified with PNN, classification performances of the method were found over 90% and error rates were found about 7% for both training and test sets similarly. Analysis time of the method was not too long, but according to set of 30 observations and SVM method it was found longer. The differences between analysis times of SVM and PNN methods further increases as the number of observations increases. Mean error, minimum error and maximum error

of PNN model was calculated as  $0.187 \pm 0.041$ ,  $0.115$  and  $0.296$  respectively.

According to these results it can be said that classification performance of PNN method is higher and analysis time is longer than SVM (Table 4). In this case, PNN method can be preferred if the differences of analysis times are not important.

Table 4. The results of the classification performance of SVM and PNN for  $n$  is 100 and ES is 1

<i>SVM results (n=100) ; effect size = 1 SD</i>		
Run Time = $529 \pm 101$ split-second		
	Training	Validation
Accuracy (%)	$85.6 \pm 3.5$	$83.9 \pm 2.8$
Sensitivity (%)	$86.3 \pm 3.3$	$84.5 \pm 2.9$
Specificity (%)	$84.8 \pm 4.1$	$83.3 \pm 3.4$
False Positive (Type 1 error) (%)	$15.2 \pm 4.1$	$16.7 \pm 3.4$
False Negative (Type 2 error) (%)	$13.6 \pm 3.3$	$15.4 \pm 3.3$
Positive Predictive Value (PPV) (%)	$85.2 \pm 3.7$	$83.5 \pm 3.1$
Negative Predictive Value (NPV) (%)	$86.2 \pm 3.4$	$84.3 \pm 2.8$
Area Under ROC Curve	$0.931 \pm 0.021$	$0.914 \pm 0.025$
<i>PNN results (n=100) ; effect size = 1 SD</i>		
Run Time = $3385 \pm 78.7$ split-second		
	Training	Validation
Accuracy (%)	$92.7 \pm 3.1$	$91.8 \pm 2.9$
Sensitivity (%)	$92.2 \pm 3.9$	$91.3 \pm 3.9$
Specificity (%)	$92.8 \pm 3.4$	$92.3 \pm 3.4$
False Positive (Type 1 error) (%)	$7.2 \pm 3.4$	$7.7 \pm 3.4$
False Negative (Type 2 error) (%)	$7.8 \pm 3.9$	$8.7 \pm 3.9$
Positive Predictive Value (PPV) (%)	$93.2 \pm 3.3$	$92.3 \pm 3.3$
Negative Predictive Value (NPV) (%)	$92.3 \pm 3.7$	$91.5 \pm 3.6$
Area Under ROC Curve	$0.956 \pm 0.017$	$0.943 \pm 0.0195$

When the samples which have 2 standard deviation (SD) effect size and contain 30 observations classified separately for train and test sets with SVM, classification performances of the method was found at least 97% and error rates were found about %2. Analysis time of the method was quite short. According to these results it can be said that this classification success was high (Table 5). In addition mean, minimum and maximum values of the sigma parameter for 100 trial was calculated as  $1.11 \pm 1.36$ ,  $0.0$  and  $6.65$  respectively.

When the samples which have 2 standard deviation (SD) effect size and contain 30 observations classified with PNN, classification performances of the method were found over 99% and error rates were found about 1% for both training and test sets similarly. Although analysis time of the method was a bit longer than SVM. Mean error, minimum error and maximum error of PNN method

was calculated as  $0.012 \pm 0.018$ ,  $0.0$  and  $0.075$  respectively. As a result classification performance of PNN method is preferable (Table 5).

Table 5. The results of the classification performance of SVM and PNN for  $n$  is 30 and ES is 2

<i>SVM results (n=30) ; effect size = 2 SD</i>		
Run Time = $43.9 \pm 2.49$ split-second		
	Training	Validation
Accuracy (%)	$98.5 \pm 1.7$	$97.5 \pm 2.1$
Sensitivity (%)	$98.4 \pm 2.3$	$97.7 \pm 2.7$
Specificity (%)	$98.7 \pm 2.1$	$97.3 \pm 2.7$
False Positive (Type 1 error) (%)	$1.3 \pm 2.1$	$2.7 \pm 2.6$
False Negative (Type 2 error) (%)	$1.6 \pm 2.3$	$2.3 \pm 2.7$
Positive Predictive Value (PPV) (%)	$98.7 \pm 2.1$	$97.4 \pm 2.5$
Negative Predictive Value (NPV) (%)	$98.4 \pm 2.2$	$97.7 \pm 2.6$
Area Under ROC Curve	$0.998 \pm 0.003$	$0.996 \pm 0.006$
<i>PNN results (n=30) ; effect size = 2 SD</i>		
Run Time = $71.64 \pm 9.93$ split-second		
	Training	Validation
Accuracy (%)	$99.7 \pm 0.7$	$99.7 \pm 0.7$
Sensitivity (%)	$99.6 \pm 1.3$	$99.6 \pm 1.3$
Specificity (%)	$99.8 \pm 0.6$	$99.9 \pm 0.7$
False Positive (Type 1 error) (%)	$0.1 \pm 0.6$	$0.1 \pm 0.6$
False Negative (Type 2 error) (%)	$0.4 \pm 1.2$	$0.4 \pm 1.3$
Positive Predictive Value (PPV) (%)	$99.9 \pm 0.6$	$99.9 \pm 0.6$
Negative Predictive Value (NPV) (%)	$99.6 \pm 1.2$	$99.6 \pm 1.2$
Area Under ROC Curve	$0.999 \pm 0.004$	$0.998 \pm 0.004$

When Table 6 was examined, it was seen that both of methods had very high classification performance. In addition there was no difference to be concerned according to the analysis time. However it was determined that analysis time of PNN method increases exponentially with increasing sample size.



Table 6. The results of the classification performance of SVM and PNN for n is 100 and ES is 2

<i>SVM results (n=100) ; effect size = 2 SD</i>		
Run Time = 192.6 ± 30.2 split-second		
	Training	Validation
Accuracy (%)	92.5±3.2	97.9±1.0
Sensitivity (%)	92.3±3.8	97.9±1.3
Specificity (%)	92.8±3.5	98.6±1.2
False Positive (Type 1 error) (%)	7.2±3.5	2.1±1.3
False Negative (Type 2 error) (%)	7.6±3.8	2.1±1.3
Positive Predictive Value (PPV) (%)	92.8±3.5	97.9±1.3
Negative Predictive Value (NPV) (%)	92.4±3.6	97.9±1.3
Area Under ROC Curve	0.955±0.019	0.997±0.003
<i>PNN results (n=100) ; effect size = 2 SD</i>		
Run Time = 4975 ± 286.9 split-second		
	Training	Validation
Accuracy (%)	99.6±0.4	99.6±0.5
Sensitivity (%)	99.6±0.7	99.5±0.7
Specificity (%)	99.7±0.5	99.7±0.6
False Positive (Type 1 error) (%)	0.2±0.5	0.2±0.6
False Negative (Type 2 error) (%)	0.3±0.7	0.4±0.7
Positive Predictive Value (PPV) (%)	99.7±0.5	99.7±0.5
Negative Predictive Value (NPV) (%)	99.6±0.7	99.5±0.7
Area Under ROC Curve	0.999±0.002	0.998±0.003

Simulation results showed that when the effect size increased by one standard deviation, the classification accuracy of the two methods has increased at least 10%. PNN method was found better according to classification performance for all results but analysis time of the method increased proportionally with sample size. Classification performance of the methods showed no significant difference with increasing sample size. The results belong to training and test sets were similar. Additionally, the parameter values obtained from this study can be used as a predictive value in the other studies.

### 3.2. Results of real data set

Frequency of categorical attributes and genotypes of gene polymorphisms are given in Table 7. As a result of the univariate analysis there was no statistically significant relationship between fas844, fas670, fas1124, sex and knee osteoarthritis (P values 0.460, 0.723, 0.609, 0.109 respectively). In contrast to there was statistically significant relationship between fas1377 and knee osteoarthritis (P < 0.0001).

Table 7: Frequency of categorical attributes

Category	Patient		Control		P Value	
	Number	%	Number	%		
Fas844	C/C	45	30.4	37	36.3	0.460
	C/T	80	54.1	47	46.1	
	T/T	41	15.5	18	17.6	
Fas670	G/G	27	18.2	21	20.6	0.723
	G/A	63	42.6	46	45.1	
	A/A	58	39.2	35	34.3	
Fas1377	G/G	95	64.2	42	41.2	<0.001
	G/A	51	34.5	60	58.2	
	A/A	2	1.4	0	0	
Fas1124	A/A	107	72.3	68	66.7	0.609
	A/G	37	25	30	29.4	
	G/G	4	2.7	4	3.9	
Sex	Male	116	79.5	72	70.6	0.109
	Female	30	20.5	30	29.4	

There was no statistically significant difference between patients and control groups with regard to height (P = 0.150), however significant difference was found in terms of age and weight (each P < 0.0001).

According to these results, age, weight, and fas1377 are important in the diagnosis of knee osteoarthritis.

In the results obtained from nonlinear SVM, the number of points evaluated during the search for estimation of parameters was 140, the best sensitivity and specificity values optimized by the search was 74.9%, and support vectors used in the model was 215. Stopping criteria ( $\epsilon$ ) was identified as 0.001, and error cost value and sigma parameter was predicted as  $C = 34.109$  and  $\gamma = 0.0816$  respectively.

Classification performances of SVM are given in Table 8. It was seen that sensitivity is higher than the other performance values. In addition, area under ROC curve was 0.936 for training set and 0.852 for test set. When these values were examined together, it was seen that the success of the diagnosis of the model is high.

Table 8: Model performance measures for SVM

Clinical Diagnosis	Training		Validation	
	+	-	+	-
+	136	16	128	24
-	20	82	35	67
Accuracy (%)	85.83		76.77	
Sensitivity (%)	89.47		84.21	
Specificity (%)	80.39		64.69	
Positive Predictive Value (PPV) (%)	87.18		78.53	
Negative Predictive Value (NPV) (%)	83.67		73.63	
Area Under ROC Curve (%)	0.936		0.852	

Classification performances of PNN method are given in Table 9. The results of training and test sets were very similar to each other. In addition, area under ROC curve was 0.951 for training set and 0.914 for test set. The success of the diagnosis of this model was high. It can be said that the small number of individuals in some sub-groups and outliers do not affect adversely success of predict.

Table 9: Model performance measures for PNN

Clinical Diagnosis	Training		Validation	
	+	-	+	-
+	146	6	145	7
-	15	87	17	85
<b>Accuracy (%)</b>	91.73		90.55	
<b>Sensitivity (%)</b>	96.05		95.39	
<b>Specificity (%)</b>	85.29		83.33	
<b>Positive Predictive Value (PPV) (%)</b>	90.68		89.51	
<b>Negative Predictive Value (NPV) (%)</b>	93.55		92.39	
<b>Area Under ROC Curve</b>	0.944530		0.919118	

In addition to these results, analysis times of SVM and PNN were obtained 796 and 15325 split second respectively.

#### 4. Discussion

SVM and PNN methods have become attractive techniques in terms of classification performances especially in recent years. Although both methods achieve a high classification performance, the superiority to each other of the methods with various aspects is tried to prove.

By using real data set, SVM and PNN methods were performed to investigate the success of diagnosis of some demographic characteristics, and FAS and FASL gene polymorphisms on knee osteoarthritis. When the results obtained from the study were examined, it was found that classification performances of PNN method are quite successful for both training and test sets. However classification performances of SVM were obtained lower than PNN.

There are no study comparing classification performances of SVM and PNN methods for the diagnosis of knee osteoarthritis in the literature. However various studies for different diagnosis given below were done.

Muniz et al. (2010) were compared SVM and PNN methods for discriminating between normal and Parkinson disease subjects using a small set of data contains 45 subjects totally and they found that PNN performed better than SVM [19].

In the study done by Kalatris et al. (2003), they purposed to discriminate patients with depression from controls, and as a result of this study, classification performance of SVM was a bit higher than PNN [8].

In addition to these studies, Spyridonos et al. (2006) were performed a comparative evaluation of SVM and PNN exploring their ability to classify superficial bladder carcinomas as low or high-risk and they were found that both classification models resulted in a relatively high overall accuracy [7].

Ubeyli (2007) was compared 4 different classification methods contains SVM and PNN methods for breast cancer classification and explained that SVM has the better classification performance. However these performance levels were very close to each other [9].

In the simulation study, success of SVM and PNN methods were compared using given simulated combinations. As a result of the comparisons, it was found that performance of PNN is higher than SVM method. The most important disadvantage of PNN method is that the analysis time increase proportionally with sample size and number of attributes.

The study done by Iliou and Anagnostopoulos (2010), for PNN classifier with the 10 V-Fold validation method the highest accuracy was found with the Gaussian Kernel function, and Sigma values for each attributes and class. The highest accuracy by SVM using the 10 V-Fold validation method was found with C-SVC SVM Model using RBF Kernel function. In this study also provides the same conditions for the implementation level of PNN and SVM methods [20].

Signh et al. (2010) were compared BDT-SVM and PNN methods for multiple classification and found that the SVM-BDT has superior performance than the PNN. Because of its high generalization performance without the need to add a prior knowledge, even when the dimension of the input space is very high, because of using binary decision tree the less number of SVM are required which improve the classification accuracy [5].

Saiti et al. (2009) were used SVM and PNN methods comperatively in two different data set for diagnosis of thyroid disease and they have seen that both PNN and SVM have given the same efficient results, but according to the second dataset, it is observed that SVM has performed better [10].

In the study done by Lahmiri (2011), it was showed that the superiority of the methods for different combination of predicted variables is changing [6].

Eristi et al. (2007) were done a study about power systems for classification with SVM and PNN methods and they

were found that the best classification was performed by SVM method. Also they were concluded that the best pattern recognition can be done by SVM for similar classification study on power systems [11].

## 5. Conclusion

In previous studies, SVM and PNN methods were compared by using data sets containing different number and type of attributes, and different number of observations. Therefore it can be said that their superiority to each other can be change in various situations, but generally both of the methods have given good results.

In this study, PNN method was found superior for investigated conditions. When the numerical performance values in the other studies were investigated, it was seen that in many of the studies it is not discuss whether there is a statistically significant difference between the performances. In addition, using results based on one or more data sets for selecting the appropriate method may cause erroneous decisions.

Additionally, performances of these two methods can be investigate for data sets which have different distributions and contain several number of predictor variables.

## References

- [1] M. Farrokhrooz, and M. Karimi, "A New Method for Spread Value Estimation in Multi-Spread PNN and its Application in Ship Noise Classification", in 8th International Conference on Signal Processing, 2006, Vol. 4.
- [2] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods", *Artificial Intelligence in Medicine*, Vol. 34, No. 2, 2005, 113-127.
- [3] D. F. Cook, C. T. Ragsdale, and R. L. Major, "Combining a neural network with a genetic algorithm for process parameter optimization", *Engineering Applications of Artificial Intelligence*, Vol. 13, No. 4, 2000, 391-396.
- [4] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*, England: Addison-Wesley, 1991.
- [5] K. Singh, I. Gupta, and S. Gupta, "SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape", *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 3, No. 4, 2010, 67-78.
- [6] S. A. Lahmiri, "Comparison of PNN and SVM for Stock Market Trend Prediction using Economic and Technical Information", *International Journal of Computer Applications*, Vol. 29, No. 3, 2011, 24-30.
- [7] P. Spyridonos, P. Petalas, D. Glotsos, D. Cavouras, V. Zolota, and G. Nikifori, "Comparative evaluation of support vector machines and probabilistic neural networks in superficial bladder cancer classification", *Journal of Computational Methods in Sciences and Engineering*, Vol. 6, No. 5, 2006, 283-292.
- [8] I. Kalatzis, N. Pihouras, E. Ventouras, C. C. Papageorgiou, A. D. Rabavilas, and D. Cavouras, "Comparative evaluation of probabilistic neural network versus support vector machines classifiers in discriminating ERP signals of depressive patients from healthy controls", in *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, ISPA 2003*, Vol. 2, pp. 981-985
- [9] D. E. Ubeyli, "Implementing automated diagnostics systems for breast cancer detection", *Expert Systems with Applications*, Vol. 33, No. 4, 2007, 1054-1062.
- [10] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms using PNN and SVM", in *3rd International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2009*, pp.1-4.
- [11] H. Eristi, A. Ucar, and Y. Demir, "Destek Vektör Makineler ve Dalgacık Dönüşümü Kullanarak Güç Sistemlerinde Geçici Rejim Bozulmalarının Sınıflandırılması", *Elektrik-Elektronik-Bilgisayar-Biyomedikal Mühendisliği 12. Ulusal Kongresi*, 2007.
- [12] W. S. Noble, "What is a Support Vector Machine?", *Nature Biotechnology*, Vol. 24, No. 12, 2009, 21-31.
- [13] C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining And Knowledge Discovery*, Vol. 2, No. 2, 1998, 121-167.
- [14] C. Cortes, and V. Vapnik, "Support-Vektor Networks", *Machine Learning*, Vol. 20, No. 3, 1995, 273-297.
- [15] T. Kavzoğlu, and İ. Çölkesen, "Investigation of the Effects of Kernel Functions in Satellite Image Classification Using Support Vector Machines", *Harita Dergisi*, Vol. 2, No. 144, 2010, 73-82.
- [16] E. E. Osuna, R. Freund, and F. Girosi, "Support Vector Machines: Training and Applications" *Massachusetts Institute of Technology Libraries Artificial Intelligence Laboratory Publications A.I Memo No:1602, Paper No:144*, 1997.
- [17] K. Dutta, N. Prakash, and S. Kaushi, "Probabilistic neural network approach to the classification of demonstrative pronouns for indirect anaphora in Hindi", *Expert Systems with Applications*, Vol. 37, No. 8, 2010, 5607-5613.
- [18] H. Ögüt, R. Aktaş, A. Alp, and M. M. Doğanay, "Prediction of Financial Information Manipulation by Using Support Vector Machine and Probabilistic Neural Network", *Expert Systems with Applications*, Vol. 36, No. 3, 2009, 5419-5423.



- [19] A. M. Muniz, H. Liu, K. E. Lyons, R. Pahwa, W. Liu, F. F. Nobre, and J. Nadal, "Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait", *Journal of Biomechanics*, Vol. 43, No. 4, 2010, 720-726.
- [20] T. Iliou, and C. N. Anagnostopoulos, "SVM - MLP - PNN Classifiers on Speech Emotion Recognition Field - A Comparative Study", *Fifth International Conference on Digital Telecommunications*, 2010, pp.1-6.

**Özge Aksehirli** was born in 1985 at Akçakoca, Turkey. She has acquired her MSc Degree in Department of Biostatistics and Medical Informatics, Faculty of Medicine, University of Düzce and is doing PhD in Department of Biostatistics, Faculty of Medicine, University of Ankara. She has 6 published international/national articles. She has participated to many international/national congress and symposiums have many conference proceedings. Her area of interest is biostatistics and data mining.

**Duygu Aydın** was born in 1985 at Amasya, Turkey. She was graduated MSc Degree in Department of Operation Research from University of Marmara. She is studying PhD in Department of Biostatistics, Faculty of Medicine, University of Hacettepe. She has 5 published international/national articles. She has participated to many international/national congress and symposiums have many conference proceedings. Her area of interest is operation research, biostatistics and data mining.

**Handan Ankaralı** was born in 1971 at Adana, Turkey. She has acquired his MSc and PhD degrees in Department of Biometry and Genetics in Ankara University. She has many published international/national articles. She has participated to many congress and symposiums and have many conference proceedings. Her working areas are data mining, experimental design, survival analysis, ROC analysis, regression analysis and exact tests.

**Melek Sezgin** was born in 1971 at Ankara, Turkey. She is working Department of Physical Therapy and Rehabilitation, Faculty of Medicine, University of Mersin. She has many published international/national articles.