Kafkas Univ Vet Fak Derg
26 (4): 541-549, 2020
DOI: 10.9775/kvfd.2020.23955

Kafkas Universitesi Veteriner Fakultesi Dergisi
ISSN: 1300-6045 e-ISSN: 1309-2251
Journal Home-Page: http://vetdergikafkas.org
Online Submission: http://submit.vetdergikafkas.org

Research Article

# Classification of Raw Milk Composition and Somatic Cell Count in Water Buffaloes with Support Vector Machines

Yalcin TAHTALI [1,a]

[1] Tokat Gaziosmanpasa University, Faculty of Agriculture, Department of Animal Science, TR-60250 Tokat - TURKEY
[a] ORCID: 0000-0003-0012-0611

**Article ID: KVFD-2020-23955 Received: 23.01.2020 Accepted: 21.05.2020 Published Online: 21.05.2020**

**How to Cite This Article**

**Tahtali Y:** Classification of raw milk composition and somatic cell count in water buffaloes with support vector machines. *Kafkas Univ Vet Fak Derg,* 26 (4): 541-549, 2020. DOI: 10.9775/kvfd.2020.23955

## Abstract

The study investigates the classification of milk quality with support vector machines (SVM) using the raw milk composition and somatic cell count (SCC) data on buffalos. For this purpose, 11-variable (dry matter, fat-free dry matter, fat (%), protein, lactose, casein, urea, density, acidity, pH, freezing point) on milk composition and SCC of 288 buffalos were used. SVM is a classifier with a high generalization ability that is based on structural risk minimization with a statistical learning system and can be applied to both linear and non-linear data. The classification successes of some kernel functions used in the SVM (polynomial kernel, normalized polynomial kernel and radial basis kernel) were investigated and their classification performances were compared with a multilayer perceptron algorithm. The results showed that the classification successes of polynomial kernel, normalized polynomial kernel and radial basis kernel were 93.06%, 92.36% and 90.97%, respectively, while the classification success of the multilayer perceptron was 81.60%. The comparison of the results with respect to the root mean square error (RMSE) values revealed that the polynomial kernel had the lowest value (0.263), while the multilayer perceptron had the highest value (0.384). According to this criterion, the best classifier was the polynomial kernel function, while the weakest classifier was the multilayer perceptron (0.384). Considering the receiver operating characteristic (ROC) area values, with respect to the closeness to 1 criterion, normalized polynomial kernel was the best function, while the multilayer perceptron function was the weakest function. The separate evaluation of the precision, sensitivity and F-measure values showed that the polynomial kernel was the most successful function, while the multilayer perceptron was the weakest function.

***Keywords:*** *Support vector machine, Somatic cell count, Kernel model optimization*

# Mandalarda Çiğ Süt Bileşimi ve Somatic Hücre Sayısının Destek Vektör Makinaları İle Sınıflandırılması

## Öz

Bu çalışmada amaç mandalarda çiğ süt bileşimi ve somatik hücre sayısı verilerini kullanarak süt kalitesinin destek vektör makineleri (DVM) ile sınıflandırılmasını araştırmaktır. Bu amaçla, 288 mandaya ait somatik hücre sayısı ve 11 değişkenli (kuru madde, yağsız kuru madde, yağ, protein, laktoz, kazein, üre, yoğunluk, asitlik, pH, donma noktası) süt bileşenleri kullanılmıştır. DVM, istatistiksel öğrenme sistemi ile yapısal risk minimizasyonuna dayanan, hem doğrusal hem de doğrusal olmayan verilere uygulanabilen yüksek genelleme kabiliyetine sahip bir sınıflandırıcıdır. DVM'de kullanılan bazı çekirdek fonksiyonlarının (polinom çekirdeği, normalleştirilmiş polinom çekirdeği ve radyal temel çekirdeği) sınıflandırma başarıları araştırılmış ve sınıflandırma performansları çok katmanlı bir algılayıcı algoritması ile karşılaştırılmıştır. Sonuçlar, polinom çekirdeğinin, normalize polinom çekirdeğinin ve radyal temel çekirdeğin sınıflandırma başarılarının sırasıyla %93.06, %92.36 ve %90.97 olduğunu, çok katmanlı algılayıcı algoritmanın sınıflandırma başarısının %81.60 olduğunu göstermiştir. Çekirdek fonksiyonlarının hata kareleri ortalamasının karekökü (RMSE) değerleri ile karşılaştırılması yapıldığında, polinom çekirdeğinin en düşük değere (0.263) sahip olduğunu, çok katmanlı algılayıcının en yüksek değere (0.384) sahip olduğu tespit edilmiştir. Bu kritere göre, en iyi sınıflandırıcının polinom çekirdek fonksiyonu, en zayıf sınıflandırıcının ise çok katmanlı algılayıcı (0.384) olduğu görülmüştür. ROC eğrisi altında kalan alan değerleri göz önüne alındığında, 1'e yakınlık kriteri açısından, normalleştirilmiş polinom çekirdeği en iyi fonksiyon, çok katmanlı algılayıcının en zayıf fonksiyon olduğu gözlenmiştir. Hassasiyet, duyarlılık ve F-ölçüm değerlerinin ayrı ayrı değerlendirilmesi sonucunda sınıflandırmada en başarılı fonksiyonun polinom çekirdeğini, en başarısız fonksiyonun ise çok katmanlı algılayıcı olduğu belirlenmiştir.

***Anahtar sözcükler:*** *Destek vektör makinesi, Somatik hücre sayısı, Çekirdek model optimizasyonu*

## INTRODUCTION

The studies regarding the solution of classification problems hold an important place in data mining. The generalization performance of an algorithm is an important criterion that should be considered in the selection of the machine

**Correspondence**
+90 356 2521616-2181
yalcin.tahtali@gop.edu.tr

learning algorithm that will be developed for the solution of classification problems. Generalization performance depends on factors such as training data, number/structure of independent qualities, model selection and parameter selection. Considering these factors, retrieving classified and meaningful information from the data and obtaining accurate information are directly proportionate to the generalization success of the algorithm. In other words, the better the generalization performance of the algorithm, the more realistic the retrieved information [1].

In recent years, Support Vector Machines (SVM) have become one of the most successful machine learning algorithms in the solution of the classification problems [1]. The method carries out classification either using a linear or non-linear function. The SVM method is based on estimating the more suitable function for separating the data.

The most prominent advantage of the SVM is they solve the classification problem by turning it into a quadratic optimization problem. Thus, they minimize the number of processes during the learning stage for the solution of a problem and reach solutions faster than other methods/algorithms [1]. They especially provide advantages in large scale datasets thanks to this feature. Moreover, as an optimization-based system, their classification performance, calculation complexity and practicality are better than other methods [2]. During the application of the SVM to the solution of the classification problems for various datasets, the selection of the kernel function and optimization of parameters play important roles. Based on data transformations, the solution of the kernel function finds the most suitable boundary among the possible outputs. This method is used in various fields including the classification of data in animal breeding.

Buffalo milk, the material of the study, contains relatively higher levels of protein, fat and mineral matters (especially calcium and phosphorus) and, thus, is more nutritious than cow milk [3]. Due to its high milk quality and availability for processing into other dairy products add to the demand for buffalo milk. Milk composition and somatic cell count (SCC) are important parameters in the determination of milk quality [4]. The number of somatic cells in normal milk is low and high number of somatic cells indicate that the milk is of low quality. In addition to its function as a quality measure, SCC in milk is an indicator of udder health in herd management and indicator in the diagnosis of mastitis [5,6]. The SCC limit for raw buffalo milk is 400.000 cells/ml according to the European Union (EU) directives (92/46 CEE and 94/71CEE) [7], while it is ≤500.000 cells/mL according to the Turkish Food Codex [8].

The study investigates the data mining applications for classification that is based on the critical SCC limit for animal breeding and especially for buffalo milk content using the SVM method.

# MATERIAL and METHODS

The study material consisted of milk yield recordings on the 288 Anatolian buffalos that gave birth between 2011 and 2013 in Tokat and its counties, Turkey. The milk yields of the Anatolian buffalos were obtained with the help of the National Buffalo Improvement by the Public Project supported by the General Directorate of Agricultural Research and Policies. On the control days, the milk yields of the buffalos were recorded in kilograms in the morning and evening. The buffalo breeding in the research area is carried out under extensive conditions. The breeders usually do not use additional feeding especially during the foraging period, but additional feeding can be carried out during winter depending on available feed types (hay, dry clover, silage, etc.).

Support Vector Machines is controlled classification algorithm that is based on the statistical learning theory. The mathematical algorithms of SVM were first designed for the solution of the classification problem of two-class linear data and, then, generalized for the classification of multiclass and nonlinear data. The working principle of SVM is based on the estimation of the most suitable decision function for the distinction of two classes, in other words, it is based on the identification of the hyperplane in which the distance between two classes is the maximum and most appropriate distance [9,10]. There are two cases in the SVM: data are linearly separable or not linearly separable.

### *Linear Separability*

Let's assume the data that will be used in training and contain N number of elements is $\theta = \{x_i, y_i\}, i = 1, 2\ldots, N$. Here, $y_i \in \{-1, 1\}$ is the label values and $x_i \in R^d$ is the feature vector. In the case of linear separation, these two-value data can directly be separated by a hyperplane. This hyperplane is called the separating hyperplane. The purpose of the SVM is to make sure the hyperplane is at the same distance to the sample groups in both classes.

If the training data comprising k number of samples in a two-class linearly separable classification problem is accepted as $\{x_i, y_i\}$, i = 1, .., k, the equations of the optimum hyperplane will be:

$$w.x_i+b \geq 1, \text{ for each } y= +1 \quad (1)$$

$$w.x_i+b \leq 1, \text{ for each } y= -1 \quad (2)$$

Here, $x \in R^n$ represents an N-dimension space, $y \in \{-1, +1\}$ represents class labels, w represents weight vectors (the normal of the hyperplane) and b represents the bias value [11]. To determine the optimum hyperplane, two hyperplanes that are parallel to this plane and form its boundaries should be created. The points forming these hyperplanes are referred to as support vectors and defined as $w.x_i+b<1$ or $w.xi+b>1$. *Fig. 1* shows the hyperplane for linearly separable datasets.

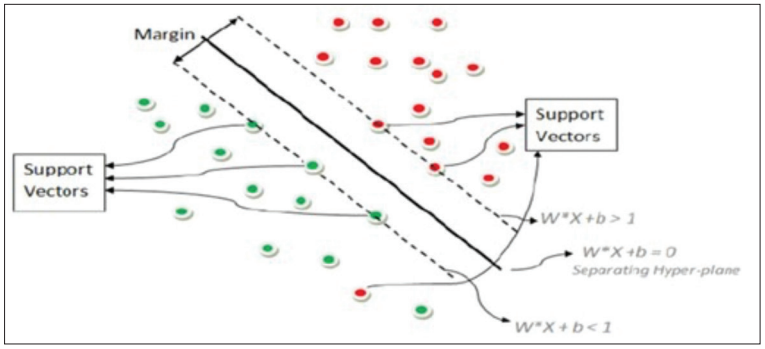
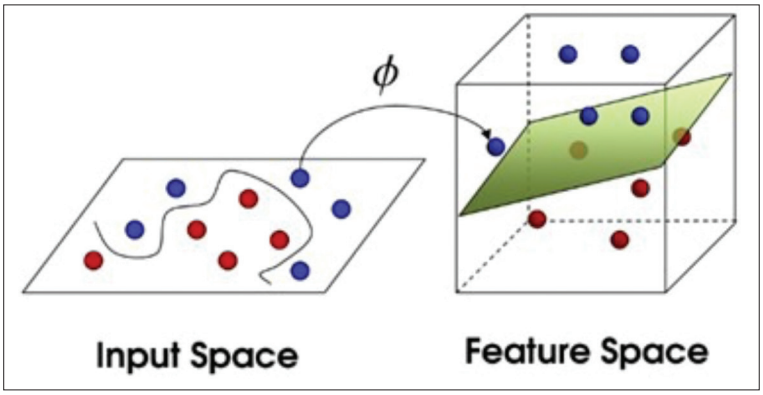



**Fig 1.** Hyperplane for linearly separable datasets

**Fig 2.** Transformation of the data into a higher dimension with the kernel function




To maximize the boundary of the optimal hyperplane, ||w|| should be minimized. In this case, determining the most suitable hyperplane requires solving the following limited optimization problem.

$$min\left[\frac{1}{2}\|w\|^2\right], \quad (3)$$

The corresponding boundaries are defined as [9].

$$y_i(w.x_i + b) - 1 \geq 0 \text{ and } y_i \in \{1, -1\} \quad (4)$$

The optimization problem can be solved using the Euler-Lagrange equations. After this process, equation 5 is obtained.

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{k} \alpha_i y_i (w.x_i + b) + \sum_{i=1}^{k} \alpha_i \quad (5)$$

In conclusion, the decision function for a linearly separable two-class problem can be described as [11].

$$f(x) = sign\left[\sum_{i=1}^{k} \lambda_i y_i (x.x_i) + b\right] \quad (6)$$

### *Non-Linear Separability*

As is the case in the classification of some data, the linear separation of many problems is not possible. In this case, the problem stemming from a portion of the training data that remains on the other side of the optimal hyperplane is solved by defining a positive dummy variable ($\zeta_i$). The balance between maximizing the boundary and minimizing the inaccurate classification errors can be controlled by defining a regularization parameter that has positive values and is shown with C ($0 < C < \infty$) [12]. The optimization problem for the data that cannot be linearly separated using a regularization parameter or a dummy variable is:

$$min\left[\frac{\|w\|^2}{2} + C.\sum_{i=1}^{r} \zeta_i\right] \quad (7)$$

The corresponding boundaries are defined as:

$$y_i(w.\varphi(x_i) + b) - 1 \geq 1 - \zeta_i, \zeta_i \geq 0, \ i = 1, \ldots, N \quad (8)$$

As seen in *Fig. 2*, for the solution of the optimization problem given in Equation 7 and 8, the data that cannot be linearly separated in the input space is displayed in a high-dimension space defined as the feature space. Hence, the data can be separated linearly and the hyperplane between the classes are determined.

The SVM can make nonlinear transformations with the aid of a kernel function that is defined as fallow:

$$K(x_i, x_j) = \varphi(x_i).\varphi(x_j) \quad (9)$$

Thus, it allow the linear separation of the data in a higher dimension.

As a result, the decision rule for the solution of a two-class problem that cannot be linearly separated using the kernel function can be written as [11].

$$f(x) = sign\left[\sum_{i=1}^{k} \alpha_i y_i \varphi(x).\varphi(x_i) + b\right] \quad (10)$$

The most important point of a classification process that

**Table 1.** *Kernel functions and parameters used in support vector machines*

| Kernel Function | Mathematical Statement |
|---|---|
| *Polynomial Kernel (d)* | $K(x,y) = ((x.y)+1)^d$ |
| *Normalized Polynomial Kernel (d)* | $K(x,y) = \frac{((x.y)+1)^d}{\sqrt{((x.x)+1)^d((y.y)+1)^d}}$ |

will be carried out with SVM is the determination of the kernel function and optimum parameters of this function.

*Table 1* shows the most commonly used kernel functions (polynomial, radial basis function and normalized polynomial kernels) in the literature.

A comparison of the kernel functions reveal that the polynomial and radial basis kernels are simpler and more understandable. Although it can appear mathematically simple, the increase in the degree of a polynomial function can complicate the algorithm, which both prolongs the process and reduces the accuracy of classification after a certain point. On the other hand, the effect of the changes in the kernel size parameter (γ) of the radial basis function on classification performance was determined to be relatively lower [13]. The normalized polynomial function was suggested by [14] to normalize the mathematical statement of the polynomial kernel instead of the normalization of the dataset. The normalized polynomial kernel can be viewed as a generalization of the polynomial kernel.

### Multilayer Perceptron

Multilayer perceptron is a feedforward artificial neural network model in which the input data are adjusted on an appropriate output sequence. It is a non-parametric artificial neural network method that carries out various detection and estimation processes. In the multilayer perceptron, each j neuron in the hidden layer takes the sum of the multiplication of the input signals with the connection weight $w_{ji}$ and calculates the $y_j$ output as a function of this sum:

$$y_i = f(\sum w_{ji}x_i) \tag{11}$$

Here, *f* is an activation function that transforms the weighted sum of the signals affecting a neuron into the output value. The activation function can be a simple threshold function, sigmoidal or hyperbolic tangent function.

The sum of the quadratic differences between the calculated and desired values of the output neurons is defined as:

$$e = \frac{1}{2}\sum_j (y_j^* - y_j)^2 \tag{12}$$

Here, $y_j$* and $y_j$ are the calculated and desired values of the j. output neuron, respectively. Each $w_{ji}$ weight is adjusted to reduce the e value as fast as possible. How the $w_{ji}$ value will be adjusted depends on the training algorithm.

### Comparison of the Performances

Comparing the classifiers and determining the best classifier are of great importance in data mining [15]. The classification performances of algorithms are usually compared with respect to accurate classification percentage, accuracy rate, RMSE, ROC area, sensitivity, precision and F-measure.

***Accuracy Rate:*** It gives the accurate classification percentages of observations.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{13}$$

***Root Mean Square Error:*** Also known as the quadratic mean, it is a statistical method used in the measurement of the sizes of changing amounts.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{14}$$

Here, *n*: number of samples, $e_i$: models error.

***ROC Area:*** ROC area curve determines the estimation performances of different classification algorithms. The area below the ROC curve is one of the important measures in selecting the best classification algorithm. When the area below the curve approaches 1, it indicates that the classification is accurate.

***Confusion Matrix:*** The confusion matrix shows the numbers of inaccurate or accurate classifications of data *(Table 2)*. The confusion matrix is used in the calculation of precision, sensitivity and F-measure, which measure the performances of classification algorithms.

The most popular and simple method in measuring the performance of a model is the accuracy and error rates of the model. Accuracy is the ratio of the accurately classified sample number (TP+TN) to the total sample number (TP+TN+FP+FN). This value is completed to 1 by the error rate. In other words, error rate is the ratio of the inaccurately classified sample number (FP+FN) to the total sample number (TP+TN+FP+FN) [16].

**Table 2.** *Classification according to the confusion matrix*

| Accurate Class | Estimated Class | |
|---|---|---|
| | Class 1 | Class 2 |
| Class 1 | a | b |
| Class 2 | c | d |

***a:*** *TP (True-Positive),* ***b:*** *FN (False-Negative),* ***c:*** *FP (False-Positive),* ***d:*** *TN (True-Negative)*

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Error\ rate = \frac{FP + FN}{TP + FP + FN + TN}$$

***Sensitivity:*** Sensitivity is the ratio of the accurately classified positive sample number to the total positive sample number [16].

$$Sensitivity = \frac{TP}{TP + FN}$$

***Precision:*** Precision is the ratio of the number of True Positive samples with a class estimated to be 1 to the total number of samples with a class estimated to be 1 [16].

$$Precision = \frac{TP}{TP + FP}$$

Precision and sensitivity are not separately sufficient to derive a meaningful comparison. Considering both measures together will yield more accurate results. The F-measure is defined for this purpose.

***F-measure:*** F-measure is the harmonic mean of precision and sensitivity.

$$F - measure = \frac{2xSensitivityxPrecision}{Sensitivity + Precision}$$

In the SVM, the WEKA 3.8.3 (Waikato Environment for Knowledge Analysis) software and IBM SPSS 21.0 statistical package program were used for the analysis of the classification results. WEKA is a popular machine learning package written with Java and developed in Waikato University, New Zealand, and contains visualization tools and algorithms for data analysis and estimation modelling and graphical user interfaces for easy access to these functions.

## RESULTS

*Table 3* shows the descriptive statistics for the variables used in the study. Considering the SCC as a criterion of milk quality, the groups with high and low milk quality were formed with respect to these variables. As seen in *Table 3*, in terms of the -DM- and fat variables, there were statistically significant differences between low and high SCC according to the investigated criterion (P<0.05, P<0.01, respectively).

However, in terms of the fat-free dry matter (FDM), protein, lactose, casein, urea, density, acidity, pH and freezing point variables, there were no statistically significant differences between the low and high SCC (P>0.05). Moreover, the significant relationship between SCC and fat and dry matter was supported by the correlation matrix results in *Table 4*.

*Table 4* shows the correlation matrix of the variables used in the study. As seen in *Table 4*, the highest correlation was between fat and -DM- (0.93, P<0.01), while the lowest correlation was between freezing point and pH (0.02, P>0.05). The relationship between the variables and SCC showed that it had the highest correlation with protein (0.69, P<0.01), while the lowest correlation was with urea (-0.02, P>0.05).

For the detailed analysis of the classification results, the performances of the algorithms were statistically compared. The WEKA 3.8.3 (Waikato Environment for Knowledge Analysis) software was used for classification with SVM. The SMO (Sequential Minimal Optimization) algorithm in the software and Multilayer perception algorithm were used for the multilayer perceptron. The 10-fold cross validation was selected and used for the data. *Table 5* shows the statistical results for the classification algorithms.

As seen in *Table 5*, the results for the classification algorithms revealed that, for the polynomial kernel, the accurate classification percentage was 93.06% and ROC area value was 88.7%, which agrees with the accurate classification percentage. Moreover, RMSE, precision, sensitivity and F-measure values were 0.263, 0.913, 0.995 and 0.931, respectively.

The accurate classification percentage of the normalized polynomial kernel was 92.36%. Furthermore, the ROC area value was 89.2%, which agrees with the accurate classification percentage. The RMSE, precision, sensitivity and F-measure values were 0.276, 0.924, 0.970 and 0.922, respectively.

As seen in *Table 5*, the accurate classification percentage of radial basis function kernel was 90.97%. In a similar fashion, the ROC area value was 86.6% and agreed with the accurate classification percentage. In addition, the RMSE, precision, sensitivity and F-measure values were 0.301, 0.903, 0.975 and 0.904, respectively.

The accurate classification percentage for the multilayer perceptron was 81.60%. The ROC area value was 81.2%, which is in keeping with the accurate classification percentage. Moreover, the RMSE, precision, sensitivity and F-measure values were 0.384, 0.843, 0.866 and 0.816, respectively.

**Table 3.** *Descriptive statistics for the milk characteristics*

| Milk Characteristics | Low (n=76) Mean ± Std. Deviation | High (n=212) Mean ± Std. Deviation | P-value |
|---|---|---|---|
| DM | 16.35±2.53 | 17.29±2.78 | **0.011** |
| FDM | 11.07±1.21 | 10.87±0.83 | 0.179 |
| Fat | 5.18±2.43 | 6.25±2.85 | **0.002** |
| Protein | 4.90±1.47 | 4.91±0.95 | 0.981 |
| Lactose | 5.23±0.61 | 5.16±0.53 | 0.366 |
| Casein | 3.54±1.12 | 3.63±0.87 | 0.542 |
| Urea | 0.04±0.02 | 0.04±0.02 | 0.833 |
| Density | 1029.45±8.21 | 1028.97±7.47 | 0.651 |
| Acidity | 7.92±5.40 | 8.21±3.37 | 0.658 |
| pH | 6.56±0.15 | 6.55±0.14 | 0.704 |
| Freezing Point | 0.57±0.16 | 0.59±0.16 | 0.573 |

***DM:** dry matter; **FDM:** fat-free dry matter*

**Table 4.** *Correlation coefficients between the variables and significance test results*

| Milk Characteristics | SCC | DM | FDM | Fat | Protein | Lactose | Casein | Urea | Density | Acidity | pH |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM | 0.68** | 1 | | | | | | | | | |
| FDM | 0.28* | 0.21* | 1 | | | | | | | | |
| Fat | 0.66** | 0.93** | -0.16* | 1 | | | | | | | |
| Protein | 0.69** | 0.58** | 0.83** | 0.29* | 1 | | | | | | |
| Lactose | -0.55** | 0.58** | -0.19* | -0.54** | -0.67** | 1 | | | | | |
| Casein | 0.38** | 0.69** | 0.43** | 0.59** | 0.76** | -0.80** | 1 | | | | |
| Urea | -0.02 | -0.31** | 0.39** | -0.51** | 0.15* | 0.17* | -0.32** | 1 | | | |
| Density | -0.05 | 0.08 | -0.08 | 0.20* | 0.09 | -0.35* | 0.63** | -0.41** | 1 | | |
| Acidity | 0.34** | 0.57** | 0.34** | 0.49** | 0.68** | -0.76** | 0.92** | -0.21* | 0.63** | 1 | |
| pH | -0.56** | -0.23* | -0.16* | -0.19* | -0.37** | 0.44** | -0.32** | 0.08 | -0.06 | -0.34** | 1 |
| Freezing Point | -0.04 | 0.05 | -0.04 | 0.07 | 0.03 | 0.13* | 0.08 | -0.04 | 0.10 | 0.10 | 0.02 |

** significant at the 0.05 level (2-tailed); ** significant at the 0.01 level (2-tailed); **DM:** dry matter; **FDM:** fat-free dry matter*

**Table 5.** *Comparison of the classification algorithms*

| Algorithms | Features | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy | RMSE | ROC Area | Precision | Sensitivity | F-Measure |
| Polynomial kernel | 93.06 | 0.263 | 0.887 | 0.913 | 0.995 | 0.931 |
| Normalized polynomial kernel | 92.36 | 0.276 | 0.892 | 0.924 | 0.970 | 0.922 |
| Radial basis function kernel | 90.97 | 0.301 | 0.866 | 0.903 | 0.975 | 0.904 |
| Multilayer perceptron | 81.60 | 0.384 | 0.812 | 0.843 | 0.866 | 0.816 |

***RMSE:** Root Mean Square Error*

*Table 6* shows the confusion matrix results. As revealed by the table, for the polynomial kernel, the number of accurately classified observations in the high class was 201 and the number of observations that were in the low class while they should have been in the high class was 1; the number of accurately classified observations in the low class was 67 and the number of observations that were in the high class while they should have been in the low class was 19. In a similar manner, for normalized polynomial kernel, the number of accurately classified observations in the high class was 196 and the number of observations that were in the low class while they should have been in the high class was 6; the number of accurately classified observations in the low class was 70 and the number of observations that were in the high class while they should have been in the low class was 16.

**Table 6.** *Confusion matrix of the classifications*

| Accurate Class | | Estimated Class | | Error Rates |
|---|---|---|---|---|
| | | High | Low | |
| Polynomial kernel | High | 201 | 1 | 0.071 |
| | Low | 19 | 67 | |
| Normalized polynomial kernel | High | 196 | 6 | 0.081 |
| | Low | 16 | 70 | |
| Radial basis function kernel | High | 197 | 5 | 0.092 |
| | Low | 21 | 65 | |
| Multilayer perceptron | High | 175 | 27 | 0.181 |
| | Low | 26 | 60 | |

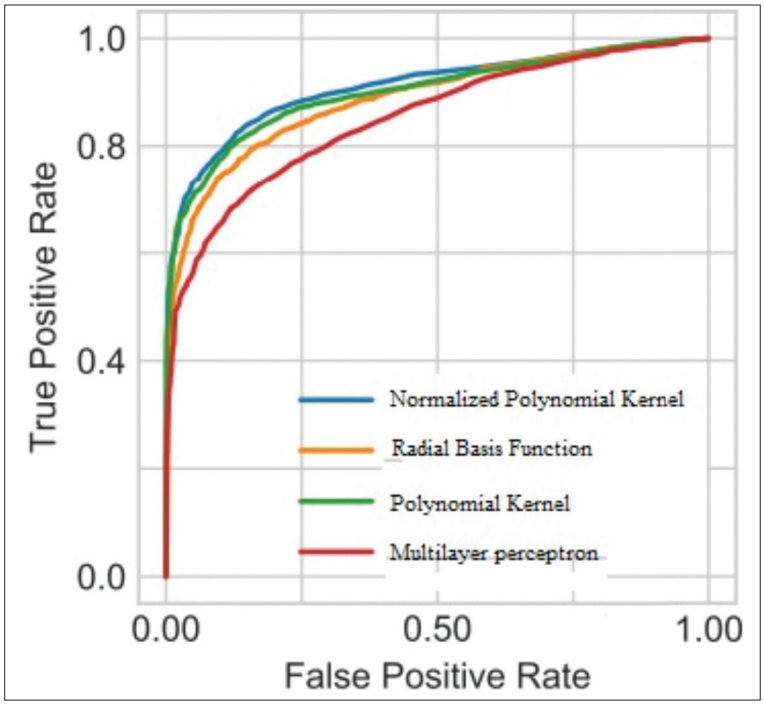



**Fig 3.** ROC curves of of different classification algorithms

For radial basis function kernel, the number of accurately classified observations in the high class was 197 and the number of observations that were in the low class while they should have been in the high class was 5; the number of accurately classified observations in the low class was 21 and the number of observations that were in the high class while they should have been in the low class was 65.

Furthermore, for the multilayer perceptron, the number of accurately classified observations in the high class was 175 and the number of observations that were in the low class while they should have been in the high class was 27; the number of accurately classified observations in the low class was 26 and the number of observations that were in the high class while they should have been in the low class was 60.

The error rates for the polynomial kernel, normalized polynomial kernel, radial basis function kernel and multilayer perceptron were calculated to be 0.071, 0.081, 0.092 and 0.181, respectively. The precision, sensitivity and F-measure values given in *Table 5* were calculated using the accuracy and error rates.

In the analysis using the WEKA data mining software, the classification algorithms were summarized in *Table 5* in light of various statistical criteria. According to the accuracy percentage in *Table 5*, the polynomial kernel function had the highest accuracy rate (93.06%), while the multilayer perceptron had the lowest accuracy rate (81.60%). When the functions were compared considering their RMSE values, the polynomial kernel had the lowest value (0.263), while the multilayer perceptron had the highest value (0.384). According to this criterion, polynomial kernel was the best classifier, while the weakest classifier function was the multilayer perceptron (0.384).

When the ROC area values are evaluated by their closeness to 1, normalized polynomial kernel was determined to be the best function, while the multilayer perceptron was

the weakest function. The separate examination of the precision, sensitivity and F-measure values revealed that the normalized polynomial kernel was the most successful function, while the weakest function was the multilayer perceptron. ROC curve are shown in *Fig. 3*.

## DISCUSSION

Milk composition and the SCC in milk are important parameters in the determination of milk quality [4]. As SCC increases, the shelf-life and quality of milk decrease [17]. In addition to its role as a milk quality parameter, SCC is an indicator of udder health and used in the diagnosis of mastitis [5,6]. The SCC in milk is an important parameter in the determination of milk quality and early diagnosis of subclinical mastitis.

Although the number of studies on the issue is limited, SVM are used to the estimation of clinical and subclinical mastitis in dairy cows [18]. The researchers divided the animals into 2 groups as healthy and infected animals by determining the SCC in milks obtained monthly from Holstein cows for 12 months and reported that SVM achieved classification with an accuracy rate of 91% [18]. In the present study, accuracy rates were in the range of 81.59% and 93.05% for the kernel functions, which are close to the results found in the previous study.

Support vector machines are used to the diagnosis of clinical mastitis [19,20]. SVM are achieved classification with an 83.2% accuracy rate [19], while accuracy rate was 84.6% [20] These results are close to the results obtained with the multilayer perceptron in our study. SVM are used to develop a milk recognition system and classify milk with respect to its content after UHT [1]. In addition SVM are used to the estimation of rumen acidosis in dairy cattle and the classification of SCC [21,22]. Milk fatty acids and rumen pH value were taken as classification variables and linear kernel and radial basis kernel functions were used in classification with SVM [21].

Dry matter, fat, protein and lactose contents of milk obtained from 222 milch Murrah buffaloes were in the range of 16.94%-18.55%, 6.28%-8.38%, 4.05%-4.59% and 4.96%-5.34%, respectively, and the lactose content decreased as the SCC increased [23]. In this study, the dry matter, fat, protein and lactose contents of the milk obtained from 288 Anatolian buffaloes were close to those previously obtained by the researchers. As the SCC in milk increased, lactose content decreased (r=-0.55), while dry matter, fat and protein contents (r=0.68, r=0.66, 0.69) increased. Sekerden and Avsar [24] found that the ash, fat, dry matter, protein, acidity, density, pH and urea content of buffalo milk were 0.47%, 7.67, 17.55, 5.28, 0.17, 1.028, 6.61 and 3.78 mg/100 mL, respectively. Fernandes et al.[25] determined that the dry matter, fat, protein and lactose contents of milk were in the range of 14.5-17.1%, 6.1-6.9%, 3.9-4.2% and 4.5-5.2%, respectively, and stated that SCC did not affect the composition of buffalo milk. Ayasan et al.[26] investigated the effects of SCC on milk urea nitrogen and milk composition. 30 Holstein cows were divided into 2 groups based on their SCC. The researchers found that SCC (Group 1: x<268.000 cells/mL; Group 2: x>268.000 cells/mL) had a significant effect on the milk fat, milk lactose, fat-free dry matter and density (P<0.05) but did not have a significant effect on milk urea nitrogen, milk protein, milk casein, urea, dry matter, acidity, free fatty acid, citric acid and freezing point (P>0.05). In this study, SCC had a statistically significant effect on dry matter and fat (P<0.05), but did not affect other milk components (P>0.05). Furthermore, Ayasan et al.[26] found a significant relationship between SCC and milk fat (r=0.209; P=0.026), fat-free dry matter (r=-0.183; P=0.050), milk lactose (r=-0.196; P=0.037) and density (r=-0.281; P=0.002). In this study, there was a significant relationship between SCC and milk fat (r=0.68, P<0.01), fat-free dry matter (r=0.28, P<0.05), lactose (r=-0.55, P<0.01), but SCC did not have a statistically significant relationship with urea (r=-0.02, P>0.05), density (r=0.05, P>0.05) and freezing point (r= -0.04, P>0.05). Yesilova et al.[27] demonstrated that lactation milk yield in Anatolian buffaloes can be classified using mixture model. In our study, we investigates the classification of milk quality with SVM using the raw milk composition and SCC data on buffalos.

Somatic cell count is a good classifier in the determination of milk quality and mastitis. A review of the scientific literature revealed that although SVM was used in the determination of mastitis, the number of studies on its use in the determination of milk quality was limited. Within this context, in the study, SVM models were developed with polynomial, normalized polynomial and radial basis kernel functions. The classification performances of the kernel functions with SVM were compared with the multilayer perceptron method.

The study on SVMs analyzed the effects of critically important kernel functions on the classification results, i.e. performance, in detail and investigated the classification of milk quality with SVMs with three different kernel functions and multilayer perceptron using raw milk composition and SCC data. Compared with the results obtained with the multilayer perceptron method, the most commonly used kernel functions in SVM were more efficient and successful.

The dataset in the study comprised 12 variables including raw milk composition in buffaloes (DM, FDM, fat, protein, lactose, casein, urea, density, acidity, pH, freezing) and somatic SCC. The SCC in milk were divided into two classes as high and low in accordance with the Turkish Food Codex [28,29].

After the analysis of these variables, the performances of the algorithms were compared with respect to their accuracy, RMSE, ROC area, precision, sensitivity and F-measure and the results showed that the classification performances of SVMs were better than the multilayer perceptron algorithm.

Considering the kernel functions used in the study, the highest accuracy in SVM was obtained with the polynomial kernel function. The values obtained with the normalized polynomial kernel and radial basis function kernel were close to those obtained with the polynomial kernel. The algorithm of the multilayer perceptron had a lower accuracy value and thus, was not useful when compared with the kernel functions.

Considering the precision measure, the best result was obtained with the normalized polynomial kernel function. However, precision should not be interpreted separately; instead, it should be considered together with the sensitivity measure. As seen in *Table 3*, according to the sensitivity measure, the algorithms in descending order were polynomial kernel, radial basis function kernel, normalized polynomial kernel and multilayer perceptron. Evaluation with the F-measure in which the precision and sensitivity measures are considered together will yield better results.

In conclusion, using the performance measures for the models that were developed using different kernel functions for SCC revealed that higher or lower SCC than the specified criteria affected the quality of milk and other products obtained from milk. The systematic analysis and classification of the results are of great importance in the classification of milk quality. Future studies with larger datasets will add to the success of computer-based diagnosis systems.

## REFERENCES

**1. Abukhait J, Mansour AM, Obeidat M:** Classification based on gaussian-kernel support vector machine with adaptive fuzzy inference system. *Prz Elektrotechnıczn*, 94 (5): 14-22, 2018. DOI: 10.15199/48.2018.05.03

**2. Ghafouri-Kesbi F, Rahimi-Mianji G, Honarvar M, Nejati-Javaremi A**: Predictive ability of random forests, boosting, support vector machines and genomic best linear unbiased prediction in different scenarios of genomic evaluation. *Anim Prod Sci,* 57 (2): 229-236, 2017. DOI: 10.1071/AN15538

**3. Damé MCF, Lima CTS, Marcondes CR, Ribeiro MER, Garnero ADV:** Preliminary study on buffalo *(Bubalus bubalis)* milk production in Southern Brazil. *Rev Vet*, 21 (1): 585-587, 2010.

**4. Barth K:** Evaluation of somatic cell count under automatic milking conditions. Physiological and technical aspects of machine milking. *Proceedings of an International Conference*, Nitra, Slovak Republic, 165-169, 26-27 June 2001.

**5. Dhakal IP, Kapur MP, Anshu S:** Significance of differential somatic cell counts in milk for the diagnosis of subclinical mastitis in buffaloes using foremilk and stripping milk. *Indian J Anim Health,* 31, 39-42, 1992.

**6. Singh M, Ludri RS:** Somatic cell count in Murrah buffaloes (Bubalus bubalis) during different stages of lactation, parity and season. *Asian Australas J Anim Sci,* 14, 189-192, 2001. DOI: 10.5713/ajas.2001.189

**7. Sharma N, Singh NK, Bhadwal MS:** Relationship of somatic cell count and mastitis: An overview. *Asian Australas J Anim Sci,* 24 (3): 429-438, 2011. DOI: 10.5713/ajas.2011.10233

**8. Anonymous:** Türkiye İstatistik Kurumu Hayvancılık İstatistikleri. http://www.tuik.gov.tr/hayvancilikapp/hayvancilik.zul"hayvancilikapp/hayvancilik.zul. *Accessed:* 12.05.2019.

**9. Vapnik VN:** The Nature of Statistical Learning Theory. 167-174, Springer-Verlag, New York, 1995. DOI: 10.1007/978-1-4757-2440-0

**10. Vapnik VN:** Statistics for Engineering and Information Science. The Nature of Statistical Learning Theory. 2th ed., 131-137, Springer, New York, 2000. DOI: 10.1007/978-1-4757-3264-1

**11. Radhakrishnan S, Ramanathan R:** A support vector machine with gabor features for animal ıntrusion detection in agriculture fields. *Procedia Comput Sci*, 143, 493-501, 2018. DOI: 10.1016/j.procs.2018.10.422

**12. Udaya Shalika AWD, Seneviratne L:** Animal classification system based on image processing & support vector machine. *J Comput Commun,* 4 (1): 12-21, 2016. DOI: 10.4236/jcc.2016.41002

**13. Wang G:** Machine learning for inferring animal behavior from location and movement data. *Ecol Inform*, 49, 69-76, 2019. DOI: 10.1016/j.ecoinf.2018.12.002

**14. Zhao HT, Feng YZ, Chen W, Jia GF:** Application of invasive weed optimization and least square support vector machine for prediction of beef adulteration with spoiled beef based on visible near-infrared (Vis-NIR) hyperspectral imaging. *Meat Sci*, 151, 75-81, 2019. DOI: 10.1016/j.meatsci.2019.01.010

**15. Amraei S, Mehdizadeh SA, Sallary S:** Application of computer vision and support vector regression for weight prediction of live broiler chicken. *Eng Agric Environ Food*, 10 (4): 266-271, 2017. DOI: 10.1016/j.eaef.2017.04.003

**16. Ahmadi H, Rodehutscord M:** Application of artificial neural network and support vector machines in predicting metabolizable energy in compound feeds for pigs. *Front Nutr*, 4:27, 2017. DOI: 10.3389/fnut.2017.00027

**17. Harmon RJ:** Physiology of mastitis and factors affecting somatic cell counts. *J Dairy Sci*, 77, 2103-2112, 1994. DOI: 10.3168/jds.S0022-0302(94)77153-8

**18. Mammadova N, Keskin İ:** Application of the support vector machine to predict subclinical mastitis in dairy cattle. *Sci World J,* 2013:603897, 2013. DOI: 10.1155/2013/603897

**19. Cavero D, Tölle KH, Buxade C, Krieter J:** Mastitis detection in dairy cows by application of fuzzy logic. *Livest Sci,* 105, 207-213, 2006. DOI: 10.1016/j.livsci.2006.06.006

**20. De Mol RM, Ouweltjes W:** Detection model for mastitis in cows milked in an automatic milking system, *Prev Vet Med,* 49, 71-82. 2001. DOI: 10.1016/s0167-5877(01)00176-3

**21. Colman E, Waegeman W, De Baets B, Fievez V:** Prediction of subacute ruminal acidosis based on milk fatty acids: A comparison of linear discriminant and support vector machine approaches for model development. *Comput Electron Agr*, 111, 179-185, 2015. DOI: 10.1016/j.compag.2015.01.002

**22. Gao X, Xue H, Pan X, Jiang X, Zhou Y, Luo X:** Somatic cells recognition by application of gabor feature-based (2D)$^2$PCA. *Int J Pattern Recogn,* 31 (12):1757009, 2017. DOI: 10.1142/S0218001417570099

**23. Cerón-Muñoz M, Tonhati H, Duarte J, Oliveira J, Muñoz-Berrocal M, Jurado-Gámez H:** Factors affecting somatic cell counts and their relations with milk and milk constituent yield in buffaloes. *J Dairy Sci*, 85 (11): 2885-2889, 2002. DOI: 10.3168/jds.S0022-0302(02)74376-2

**24. Sekerden Ö, Avsar YK:** Milk composition, rennet coagulation time, urea content and environmental factors affecting them in Anatolian Buffaloes. *J Anim Prod,* 49 (2): 7-14, 2008.

**25. Fernandes SA, de Mattos WRS, Matarazzo SM, Gama MAS, Malhado CHM, Ferrão, SPB, Etchegaray MAL, Lima CGD:** Effect of somatic cell count on Murrah buffaloes milk. *Prev Vet*, 21 (1): 552-553, 2010.

**26. Ayasan T, Hızlı H, Yazgan E, Kara U, Gök K:** The effect of somatic cell count on milk urea nitrogen and milk composition. *Kafkas Unıv Vet Fak Derg,* 17 (4): 659-662, 2011. DOI: 10.9775/kvfd.2011.4489

**27. Yesilova A, Yilmaz A, Ser G, Kaki B:** Modeling with Gaussian mixture regression for lactation milk yield in Anatolian buffaloes. *Indian J Anim Res*, 50 (6): 989-994, 2016. DOI: 10.18805/ijar.v0iOF.4545

**28. Anonim:** Türk Gıda Kodeksi Yönetmeliği. T.C. Resmi Gazete, 14 Şubat 2000, Sayı 23964, s. 35. Ankara, 2000

**29. Beykaya M, Özbey A, Yıldırım Z:** Determination of physical, chemical and microbiological properties of milk from some dairy plants in Sivas Province. *TURJAF*, 5 (4): 388-396, 2017. DOI: 10.24925/turjaf.v5i4.388-396.1172