

RESEARCH ARTICLE

# Analysis of The Effect of Region, Province and Breed on Somatic Cell Count in Dairy Cattle in Turkey By Regression Tree Method

Onur Sahin\*

Muş Alparslan University, Faculty of Applied Sciences, Department of Animal Production and Technologies, 49250, Mus, Turkey

## ABSTRACT

This study aimed to determine the effect of various environmental factors on the somatic cell count (SCC) in cow's milk. The material of the study consisted of an SCC analysis result obtained from 207.614 raw milk samples collected from 65.288 head cattle had different breeds, ages, lactation number, and lactation length between May and December in 2019. They were raised on 6,305 dairy farms that are members of the Cattle Breeders Association (CBAT). In order to determine the effect of region, city, and cattle breed on SCC in raw milk of cattle, a regression tree was created using the Chi-square Automatic Interaction Detection (CHAID) algorithm, a Data Mining (DM) method. Accordingly, the mean SCC value of all enterprises was determined as 201.253 cells/mL. In this study, it was found that the region, city and breed groups with the highest average somatic cell counts were CAR, MSR, MR and SAR region group (285.430 cells/mL); C09, C45, C20, C64, C48, C43 city group (266.744 cells/mL); and HF, MC, and BS breed group (295.705 cells/mL). As a result of this study, the variables affecting SCC were determined as region, city, and breed respectively, in order of importance.

**Keywords:** CHAID; Dairy cattle; Somatic cell count

## INTRODUCTION

Somatic cell count (SCC), which is used to measure the biological quality of raw milk, is an important indicator of the health of the mammary gland (Li et al., 2014). Somatic cell count, which is a part of the cow's immune mechanism, increases due to mammary gland infection. SCC is affected by genotype, udder type characteristics, lactation order, lactation stage, season, stress, herd size, and milking management (Skrzypek et al., 2004; Ahn et al., 2005). Especially, during the dry period, the character of the cow, the farm facility's infrastructure in terms of animal welfare, and the herd management strategy affect the somatic cell count in the next lactation. Therefore, it is of great importance to take measures to prevent mastitis in cows during the dry period (Green et al., 2008).

The quality of milk is determined by SCC and bacteria number in raw milk which is in the farm cooling tank and has not yet undergone pasteurization. The most important

factor leading to high levels of SCC in raw milk is the disease of mastitis (Green et al., 2006; Leitner et al., 2000). All over the world, mastitis is one of the most important health problems that the dairy sector face (Atasever and Erdem, 2008).

In milk, SCC increases depending on the high relative humidity and temperature in the environment and the procedures applied during milking. In addition, high SCC values observed in mechanical milking, which is a milking method affect milk quality (Nasr & El-Tarabany, 2017; Reyes et al., 2017; Lima et al., 2019). Although the increase in the number of SCC in the tank milk of dairy farms during the summer is not the only indicator for the incidence of mastitis infection, the incidence of this infection is associated with SCC in cow's milk. Cows having high SCC for a long time increase the incidence of mastitis infection in herds. *Streptococcus uberis* is considered to be the cause of mastitis infection in pastures, while other *streptococci* and *E. coli* are considered to be the cause of mastitis infection in barns (Olde Riekerink et al., 2007).

### \*Corresponding author:

Onur Sahin, Muş Alparslan University, Faculty of Applied Sciences, Department of Animal Production and Technologies, 49250, Mus, Turkey. E-mail: o.sahin@alparslan.edu.tr

Received: 28 November 2022;

Accepted: 28 April 2023

The cow's lactation order (primiparous, multiparous) affects the somatic cell count in milk. Multiparous cows are generally more exposed to factors such as the udder base being closer to or below the knee, unfavorable milking technique, and unfavorable milking hygiene. An increasing trend in the SCC is observed in multiparous cows due to aging. In particular, cows in the 4<sup>th</sup> lactation have higher somatic cell count levels in their milk. (Magalhães et al., 2006; Andrade et al., 2007; Coentrão, 2008).

The determination of SCC is of great importance for struggling with the prevalence of subclinical and clinical mastitis, which causes great economic losses in terms of production and animal health in the dairy industry. Factors such as the number of births (primiparous, multiparous), lactation stage, seasons of the year, and milking management strategies cause differences in the number of somatic cells. (Schepers et al., 1997; Green et al., 2007).

A somatic cell count of <100,000 cells/mL indicates that the mammary gland is healthy. While a count of >200,000 cells/mL suggests a bacterial infection (subclinical mastitis) in the mammary gland. (Dohoo and Leslie, 1991; Sordillo et al., 1997; Schepers et al., 1997; Querengasser et al., 2002; Pyörala, 2003).

The threshold value of 200,000 cells/mL in the number of somatic cells in the milk sample obtained from 4 udder quarters indicates that the cow is susceptible to mastitis infection at a rate of approximately 70% and the presence of mastitis infection of at least 1 udder quarter (Dohoo and Leslie, 1991; Schepers et al., 1997).

In the literature on dairy cow breeding, data mining studies may be an alternative method for determining subclinical mastitis by correlating udder measurements with the SCC value. Moreover, the CHAID is a data mining algorithm that has been also used by many researchers in animal husbandry studies (Koç et al., 2017; Çelik, 2019).

This study aimed to examine the effects of main factors (region, city, season, cattle breed) on somatic cell count (SCC) in cow milk in 38 provinces from seven geographical regions of Turkey. In addition, it was aimed to reveal the performance of cattle breeds in geographical regions and cities in terms of somatic cell count using the regression tree analysis method.

Thanks to the results obtained in this study, it is aimed to provide information to future detailed academic studies and regional or national mastitis infection control programs.

## MATERIAL AND METHODS

### Material of the study

The material of the study consisted of SCC analysis results obtained from 207.614 raw milk samples collected from 65.288 cows between May and December 2019. Those cattle were raised in 6.305 dairy farms registered with the Cattle Breeders' Association of Turkey (CBAT). Raw milk samples were collected and analyzed by the provincial associations of cattle breeders.

### Method of the study

Turkey is divided into 7 geographical regions: Marmara Region (MR), Aegean Region (AR), Mediterranean Sea Region (MSR), Central Anatolia Region (CAR), Black Sea Region (BSR), Eastern Anatolia Region (EAR) and Southeastern Anatolia Region (SAR) (Table 1). There are a total of 81 cities in these regions, and these cities are classified according to their traffic codes. In the classification system, the first city is Adana, while the last city is Düzce. Traffic codes of the cities vary between 01 and 81. In the city classification system in this study, the initial letter of the word city (C) and the traffic code of the city were used together. For example, the codings for Adana and Düzce cities were expressed as C01 and C81, respectively.

In this study, somatic cell count (SCC) performances of Brown Swiss (BS), Holstein Friesian (HF), Montbeliarde (MC), Red Holstein (RH), Simmental (S), and Swedish Red (SR) breeds were evaluated by the seasons (Spring, Summer, Autumn) (Table 5). In addition, to investigate the effects of seasons and regions on young and older animal, cows were divided into two groups as first time calving (primiparous) and multiple calving cows (Table 6 and Table 7).

In this study, by taking into account the SCC value (SCC<200.000 cells/mL) that should be present in healthy cow's milk and the SCC value (SCC<400.000 cells/mL) allowed for processing into products by EU norms and Turkish food Codex, cow's milk was divided into 5 groups in terms of the SCC: Group 1 ( $\leq$ 100.000 cells/mL), Group 2 (from 101.000 to 200.000 cells/mL), Group 3 (from 201.000 to 300.000 cells/mL), Group 4 (from 301.000 to 400.000 cells/mL), and Group 5 (>400.000 cells/mL).

### Collection of milk analysis data

Collection and analysis of raw milk samples had carried out according to the instruction of Application Principles for Collection, Analysis, and Evaluation of Samples for Determination of Cow Milk Components which was prepared according to ICAR standards by the Cattle Breeders' Association of Turkey. Based on the instruction, an average of 3 raw milk samples were taken cows each

**Table 1: Descriptive statistics for SCC (cells/mL) by cities in Turkey**

Region	City	City code	N	X±SE	Min.	Max.
MR	Balıkesir	C10	731	270.482±13.930	2.000	1.972.000
	Bursa	C16	3035	294.035±7.082	2.000	1.995.000
	Çanakkale	C17	6	340.500±200.364	76.000	1.333.000
	Kırklareli	C39	158	350.538±41.604	2.000	1.862.000
	Sakarya	C54	13	190.923±44.654	17.000	485.000
	Tekirdağ	C59	61	396.951±61.784	2.000	1.953.000
AR	Afyon	C03	39	185.564±39.869 <sup>a</sup>	1.000	1.514.000
	Aydın	C09	321	412.707±19.830 <sup>e</sup>	1.000	1.932.000
	Denizli	C20	1159	239.335±11.830 <sup>bc</sup>	1.000	1.953.000
	İzmir	C35	194.076	196.740±364 <sup>ab</sup>	1.000	2.000.000
	Kütahya	C43	152	234.928±27.217 <sup>bc</sup>	8.000	1.765.000
	Manisa	C45	374	251.976±15.753 <sup>c</sup>	1.000	1.941.000
	Muğla	C48	162	221.160±23.233 <sup>abc</sup>	5.000	1.903.000
	Uşak	C64	67	306.403±35.608 <sup>d</sup>	5.000	1.608.000
	MSR	Antalya	C07	109	292.101±27.054	12.000
Burdur		C15	700	271.303±10.487	1.000	1.806.000
Hatay		C31	6	217.833±30.346	152.000	334.000
Maraş		C46	6	57.000±11.668	32.000	93.000
CAR	Ankara	C06	24	456.000±115.618 <sup>b</sup>	16.000	1.859.000
	Eskişehir	C26	74	294.189±56.092 <sup>ab</sup>	7.000	1.983.000
	Kayseri	C38	192	330.500±29.449 <sup>ab</sup>	3.000	1.732.000
	Kırşehir	C40	140	196.671±28.466 <sup>a</sup>	1.000	1.820.000
	Konya	C42	743	268.217±11.310 <sup>ab</sup>	1.000	1.993.000
	Niğde	C51	12	185.917±85.643 <sup>a</sup>	15.000	1.070.000
	Sivas	C58	123	122.504±5.054 <sup>a</sup>	16.000	226.000
	Yozgat	C66	9	191.222±37.707 <sup>a</sup>	26.000	338.000
	Aksaray	C68	3056	290.411±6.832 <sup>ab</sup>	1.000	1.999.000
	Karaman	C70	9	178.778±79.643 <sup>a</sup>	8.000	646.000
BSR	Amasya	C05	8	66.625±12.433 <sup>a</sup>	15.000	108.000
	Bolu	C14	9	284.556±44.595 <sup>b</sup>	33.000	442.000
	Çorum	C19	9	266.000±34.115 <sup>b</sup>	143.000	427.000
	Gümüşhane	C29	51	96.961±9.446 <sup>a</sup>	7.000	218.000
	Samsun	C55	16	118.938±23.125 <sup>a</sup>	31.000	398.000
EAR	Erzincan	C24	1928	179.683±6.365	2.000	1.967.000
	Erzurum	C25	6	246.333±89.683	11.000	589.000
	İğdır	C76	3	383.333±116.985	217.000	609.000
SAR	Diyarbakır	C21	2	408.000±318.000	90.000	726.000
	Gaziantep	C27	25	243.560±58.927	17.000	1.117.000

(N: Sample size; X: Least square mean; SE: Standard error).

<sup>(a,b,c)</sup>Values within a row with different superscripts differ significantly at  $P<0.01$ **Table 2: Descriptive statistics for SCC (cells/mL) by regions of Turkey**

Region	N	X±SE	Min	Max.
Marmara Region	4.004	293.267±6.247 <sup>c</sup>	2.000	1.995.000
Aegean Region	196.350	197.535±371 <sup>b</sup>	1.000	2.000.000
Mediterranean Sea Region	821	272.107±9.658 <sup>c</sup>	1.000	1.975.000
Central Anatolia Region	4.382	280.948±5.525 <sup>c</sup>	1.000	1.999.000
Black Sea Region	93	132.645±11.141 <sup>a</sup>	7.000	442.000
Eastern Anatolia Region	1.937	180.204±6.345 <sup>ab</sup>	2.000	1.967.000
Southeastern Anatolia Region	27	255.740±57.682 <sup>c</sup>	17.000	1.117.000
Total	207614	201.253±397	1.000	2.000.000

(N: Sample size; X: Least square mean; SE: Standard error).

<sup>(a,b,c)</sup>Values within a row with different superscripts differ significantly at  $P<0.01$ 

30 days (A4 method). Raw milk samples were taken equally from the beginning to the end of milking with the help

of a tool manufactured and calibrated in Turkey. The raw milk sample collection tool (raw milk sampler) consists of

two parts: a spiral pipe system (which separate the milk sample from the beginning to the end of milking) and a 500-milliliter container for collecting the milk samples (Fig 1).

**Table 3: Descriptive statistics for SCC (cells/mL) by cattle breeds**

Breed	N	X±SE	Min.	Max.
BS	1339	205.113±5.540 <sup>a</sup>	2.000	1.747.000
HF	195864	202.265±407 <sup>a</sup>	1.000	2.000.000
MC	260	169.076±15.729 <sup>a</sup>	1.000	1.743.000
RH	804	199.016±5.282 <sup>a</sup>	5.000	1.202.000
S	9298	180.011±2.002 <sup>a</sup>	1.000	1.967.000
SR	49	290.040±45.526 <sup>b</sup>	11.000	1.654.000
Total	207614	201.253±397	1.000	2.000.000

(N: Sample size; X: Least square mean; SE: Standard error; BS: Brown Swiss; HF: Holstein Friesian; MC: Montbeliarde; RH: Red Holstein; S: Simmental; SR: Swedish Red).

(<sup>a,b</sup>Values within a row with different superscripts differ significantly at  $P<0.01$ )

**Table 4: Descriptive statistics for SCC (cells/mL) values by categories**

SCC Category	N	X±SE	SD	SCC %
≤100	59.973	53.159±115	28.108	28.90
101 to 200	63.809	150.962±114	28.902	30.70
201 to 300	46.824	243.911±132	28.691	22.60
301 to 400	20.364	342.938±200	28.580	9.80
>400	16.644	634.322±2.368	305.454	8.00
Total	207.614	201.253±397	180.991	100.00

(N: Sample size; X: Least square mean; SE: Standard error; SD: Standard deviation)

**Table 5: Descriptive statistics for SCC (cell/mL) by cattle breed for seasons**

Season	Breed	N	X±SE	Min.	Max.
Spring	BS	204	191.750±8.869 <sup>b</sup>	2.000	507.000
	HF	16033	142.917±1.092 <sup>b</sup>	1.000	1.918.000
	MC	42	62.262±16.279 <sup>a</sup>	2.000	463.000
	RH	146	165.301±9.720 <sup>b</sup>	5.000	702.000
	S	1259	136.954±3.123 <sup>b</sup>	1.000	1.777.000
	SR	6	507.333±237.978 <sup>c</sup>	24.000	1.620.000
	Total	17690	143.173±1.028	1.000	1.918.000
Summer	BS	250	131.240±8.132	6.000	763.000
	HF	53815	145.150±617	1.000	1.981.000
	MC	108	89.074±4.559	1.000	178.000
	RH	167	168.988±13.684	7.000	1.202.000
	S	1955	159.958±4.668	2.000	1.967.000
	SR	6	165.333±40.416	37.000	245.000
	Total	56301	145.567±614	1.000	1.981.000
Autumn	BS	885	229.062±7.644 <sup>a</sup>	3.000	1.747.000
	HF	126016	234.207±537 <sup>ab</sup>	1.000	2.000.000
	MC	110	288.409±33.213 <sup>c</sup>	4.000	1.743.000
	RH	491	219.255±6.543 <sup>a</sup>	14.000	1.157.000
	S	6084	195.365±2.563 <sup>a</sup>	1.000	1.903.000
	SR	37	275.027±45.370 <sup>b</sup>	11.000	1.654.000
	Total	133623	232.405±524	1.000	2.000.000

(N: Sample size; X: Least square mean; SE: Standard error; BS: Brown Swiss; HF: Holstein Friesian; MC: Montbeliarde; RH: Red Holstein; S: Simmental; SR: Swedish Red)

After milk samples were taken homogeneously in a sample container, they were sent under a cold chain to raw milk analysis laboratories located in 9 cities (Ankara, Adana, Aksaray, Bursa, Balikesir, Burdur, Erzincan, Izmir, and Usak) in Turkey. Analyses were carried out with Foss or Bentley brand raw milk analyzers in accordance with the International Committee for Animal Recording standards (ICAR, 2020). The device in each laboratory is linked online to the herdbook database. The results of the analysis were transmitted online to the database (Fig 2).

### Statistical analysis

In this study, Independent-Samples T-Test was used to compare the SCC performances of primiparous cows (PP) and multiparous (MP) cows by geographical regions and seasons. ANOVA was used to examine the effects of geographic region, cities, and cattle breeds on the SCC. Also, Duncan's multiple range test ( $p<0.05$ ) was used to compare the mean values of the groups.

In this study, the Chi-Square Automated Interaction Detection (CHAID) algorithm was also used in the evaluation of the environmental factors affecting SCC. CHAID is one of the Data Mining (DM) algorithms that have begun to be used with the advancement of information technology to reveal important information that remains hidden in large data sets (Küçükönder et al.,



**Fig 1.** Raw milk sampler



**Fig 2.** Milk analyzer



**Table 6: Descriptive statistics for SCC (cell/mL) by cow group for seasons**

Season	Cow group	N	X±SE	Min.	Max.	p-Value
Spring	PP	3.184	145.326±2.492	1.000	1.620.000	0.318
	MP	14.404	142.646±1.135	1.000	1.918.000	
	Total	17.588	143.131±1.033	1.000	1.918.000	
Summer	PP	12.180	149.431±1.365	1.000	1.967.000	0.001
	MP	43.651	144.409±693	1.000	1.981.000	
	Total	55.831	145.505±618	1.000	1.981.000	
Autumn	PP	30.944	232.525±1.137	1.000	2.000.000	0.851
	MP	103.251	232.291±593	1.000	1.999.000	
	Total	134.195	232.346±526	1.000	2.000.000	

(N: Sample size; X: Least square mean; SE: Standard error; PP: Primiparous cow; MP: Multiparous cow)

**Table 7: Descriptive statistics for SCC (cell/mL) by cow group for region of Turkey**

Region	Cow group	N	X±SE	Min.	Max.	p-Value
AR	PP	42.617	200.863±817	1.000	2.000.000	0.000
	MP	153.733	196.499±418	1.000	1.999.000	
	Total	196.350	197.454±372	1.000	2.000.000	
BSR	PP	14	95.143±15.273	11.000	202.000	0.158
	MP	79	139.291±12.716	7.000	442.000	
	Total	93	132.645±11.141	7.000	442.000	
CAR	PP	1.230	257.624±9.833	1.000	1.999.000	0.008
	MP	3.152	290.711±6.767	1.000	1.993.000	
	Total	4.382	281.256±5.595	1.000	1.999.000	
EAR	PP	706	179.810±10.047	3.000	1.967.000	0.970
	MP	1.231	179.321±8.199	2.000	1.888.000	
	Total	1.937	179.502±6.362	2.000	1.967.000	
MR	PP	1.423	277.904±10.288	2.000	1.982.000	0.065
	MP	2.581	302.009±7.868	2.000	1.995.000	
	Total	4.004	293.432±6.254	2.000	1.995.000	
MSR	PP	318	243.003±12.149	1.000	1.527.000	0.016
	MP	503	290.507±13.712	1.000	1.975.000	
	Total	821	272.107±9.658	1.000	1.975.000	
SAR	PP	0	-	-	-	-
	MP	27	255.741±0.00	17.000	1.117.000	
	Total	27	255.741±0.00	17.000	1.117.000	

(N: Sample size; X: Least square mean; SE: Standard error; PP: Primiparous cow; MP: Multiparous cow)

2015), CHAID algorithms are used to be able to include continuous and categorical data at the same time and to be able to show the independent variables that are effective on the dependent variables on a tree diagram which is very easy to understand. In order to calculate relative importance values of factors affecting SCC by regression tree analysis, SPSS statistical package program was used (SPSS, 2019).

Decision trees form classification and regression models similar to a tree structure by asking questions and creating decision rules according to the framework of the datasets that pose a problem. Using an algorithm that minimizes variance, regression tree models repeatedly partition the data in order to determine the subgroups, which are becoming increasingly homogeneous, based on the independent variable splitting criteria (Zheng et al., 2009).

For this process, questions began to ask at the root node, which is the primary element of the tree structure, and the tree grows by branching until it reaches the leaves, which are the last element of the tree structure (Kurt et al, 2008; Kocakoç & Keser, 2019).

In a regression tree created using the CHAID algorithm, when the terminal nodes reach sufficient homogeneity, no other branching occurs in subsequent stages, and are not split depending on any other independent variable (Topal et al., 2010). In this study, in the diagram obtained by the CHAID algorithm regarding the SCC, no splitting occurred again since nodes 3, 5, 6, 8, 9, 10, and 11 were homogeneous in themselves.

Splits are selected based on the principle of minimizing the sum of the squared deviations from the mean in each

node. The related function can be summarized and applied as follows (Topal et al., 2010):

$$D(T) = \sum_{i=1}^n (y_i - \bar{y})^2$$

where,  $\bar{y}$  is the mean of observations in  $T$  node and  $y_i$  is the value of the  $i^{\text{th}}$  observation (Larsen and Speckman, 2004; Questier et al., 2004). The root node contains all of the observations. At each stage, the root node is split into two-child node; (node  $T$ ) left ( $T_{\text{left}}$ ) and right ( $T_{\text{right}}$ ). During the splitting process (Larsen and Speckman, 2004):

1. If the independent variable ( $\chi_i$ ) is continuous, left and right child nodes  
 $T_{\text{left}} = \{i \in T: \chi_i \leq t\}$  and  $T_{\text{right}} = \{i \in T: \chi_i > t\}$ ,  $t$  constant value
2. If the independent variable ( $\chi_i$ ) is ordinal, left and right child nodes  
 $T_{\text{left}} = \{i \in T: \chi_i \leq t\}$  and  $T_{\text{right}} = \{i \in T: \chi_i > t\}$
3. If the independent variable ( $\chi_i$ ) is  $k$  level categorical, there are  $2k-1$  splits.

## RESULTS AND DISCUSSION

In terms of cities, the lowest and highest SCC values were determined in Kahramanmaraş ( $n=6$ ) with  $57.000 \pm 11.668$  cells/mL and in Ankara ( $n=24$ ) with  $456.000 \pm 115.618$  cells/mL, respectively (Table 1).

In terms of SCC values, whereas the differences between cities belonging to AR, CAR, and BSR regions were found statistically significant ( $p < 0.01$ ), they were not significant for MR, MSR, EAR, and SAR regions (Table 1).

It was found that the lowest and highest SCC in terms of regions were in the Black Sea region ( $n=93$ ) with  $132.645 \pm 11.141$  cells/mL and in the Marmara region ( $n=4154$ ) with  $293.267 \pm 6.247$  cells/mL, respectively. The difference between the groups of regions in terms of mean SCC was found to be statistically significant ( $p < 0.01$ ) (Table 2).

The mean SCC values for the Brown Swiss, Holstein Friesian, Montbeliarde, Red Holstein, Simmental, and Swedish Red breeds were  $205.113 \pm 5.540$  cells/mL,  $202.265 \pm 407$  cells/mL,  $169.076 \pm 15.729$  cells/mL,  $199.016 \pm 5.282$  cells/mL,  $180.011 \pm 2.000$  cells/mL, and  $290.040 \pm 45.526$  cells/mL, respectively. The difference among the groups of breeds in terms of mean SCC was significant ( $p < 0.01$ ) (Table 3).

Distributions for Group 1 ( $\leq 100.000$  cells/mL), Group 2 (from 101.000 to 200.000 cells/mL), Group 3 (from 201.000 to 300.000 cells/mL), Group 4 (from 301.000 to 400.000 cells/mL), and Group 5 ( $> 400.000$  cells/mL) were determined as 28.90%, 30.70%, 22.60%, 9.80%, and 8.00%, respectively (Table 4).

The effect of the seasons on the SCC values of cattle breeds was found to be statistically significant ( $p < 0.01$ ). While the lowest SCC value was detected for spring ( $143.173 \pm 1.028$  cell/mL), the highest ( $232.405 \pm 524$  cell/mL) was detected for autumn (Table 5). It was determined that in the spring, the SCC values of HF, MC, and S breeds were lower than the seasonal average. For the summer season, lower values compared to the seasonal average were determined for BS, HF, and MC breeds. On the other hand, SCC values of MC and SR breeds were found above the average of the autumn season values (Table 5).

While the effects of spring and autumn seasons on SCC of PP and MP cows were not significant ( $p > 0.05$ ), the effect of summer season on SCC of both cow groups was found to be statistically significant ( $p < 0.01$ ). For all seasons, the SCC values of MP cows were lower than those of PP cows. SCC values increased in both cow groups in the autumn season due to the udder tissue wear caused by the progression of the lactation period (Table 6).

While the effects of AR ( $p < 0.01$ ), CA, and MSR regions ( $p < 0.05$ ) on SCC values of PP and MP cows were found to be statistically significant, the effects of BSR, EAR, and MR regions were not statistically significant. ( $p > 0.05$ ). While SCC values of PP cows in BSR, CAR, MR, and MSR regions were lower than MP cows as expected, SCC values of PP and MP cows in AR and EAR regions were close to each other (Table 7).

In the context of the study, to create the somatic cell map in Turkey, regions, cities, and cattle breeds were considered as independent variables affecting SCC. The factors in question are shown in a tree diagram by using CHAID analysis (Fig 3).

Moreover, to be able to make a comparison between regions, cities, and breeds, the descriptive statistics of these factors are summarized in Tables 1-4. The descriptive values of the SCC are given in the root node (node 0) of the regression tree. As a result of the risk analysis, it was found that the estimated value and the standard error value were closest for the 2:1 parent-child node ratio. Therefore, the 2:1 parent-child node ratio was used in the CHAID algorithm. The regression tree generated by the CHAID algorithm adjusted to estimate the number of SCC is shown in Fig 3.

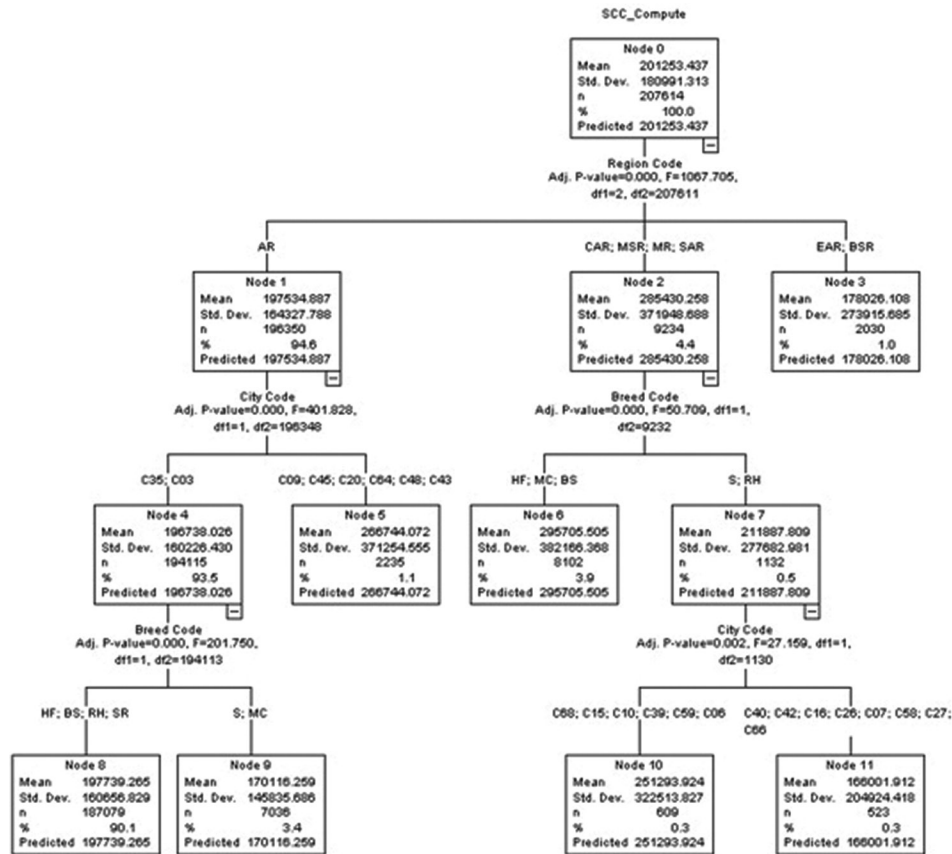


Fig 3. Regression tree analysis of the factors (cities, regions, and cattle breeds) affecting SCC

In this study, using the CHAID algorithm, SCC was determined based on the regions and cities where dairy cattle registered in the herdbook system are raised, as well as breeds. 94.6% of the data were classified in Node 1 (AR region), 93.5% in Node 4 (C35 and C03), and 90.1% in Node 8 (HF, BS, RH, and SH; Fig 3).

As seen in Fig 3 and tables, it was found that the overall mean value calculated for SCC (201.253 cells/mL) was close to the SCC value range (SCC<200.000 cells/mL) defined for healthy cow's milk by EU norms and the Turkish Food Codex.

In this study, SCC values were examined considering cattle breeds as well as regions and cities where they were raised. In terms of cattle breeds, the lowest and highest SCC values were found in the Montbeliarde breed and the Swedish Red breed, respectively (Table 3). Among cattle breeds raised for dairy purposes, there are differences in terms of SCC of milk and resistance to mastitis (Litwińczuk et al., 2015). It has been reported that whereas clinical forms of mastitis are less common and the SCC in milk is low in Montbéliarde, Simmental, and Brown Swiss cattle breeds, the SCC is higher in Holstein Friesian breed (Rupp and Boichard, 2003). It has been reported that the Montbéliarde

breed had 23-38% lower SCC than the Holstein Friesian breed and this breed was more resistant to mastitis (Gołębiewski and Brzozowski, 2007).

The average SCC in Brown Swiss cows as 270.000±39.000 cells/mL (Harmandar and Kaygısiz, 2019). In different studies, the mean SCC values in Holstein Friesian cows have been determined as 419.880±67.500 cells/mL (Yavuz and Kaygısiz, 2015), 169.000±30.000 cells/mL (Harmandar and Kaygısiz, 2019), 199.022 cells/mL (Koç, 2011), and 564.000±54.780 cells/mL (Kuczyńska et al., 2012). In the Montbeliarde breed was found the mean SCC as 138,644 cells/mL (Koç, 2011), it was found as 357.000± 5.760 cells/mL (Kuczyńska et al., 2012). The average SCC in Simmental cows as 140.000±72.000 cells/mL (Harmandar and Kaygısiz, 2019).

In this study, SCC rates in Groups 1,2,3,4, and 5 (≤100.000 cells/mL, 101.000-200.000 cells/mL, 201.000-300.000 cells/mL, 301.000-400.000 cells/mL, and >400.000 cells/mL) were determined as 28.90%, 30.70%, 22.60%, 9.80%, and 8.00%. Considering the EU norms and the Turkish Food Codex, it can be said that 92% of the examined cow's milk samples are appropriate, and 8% need rehabilitation in terms of reducing the SCC of farms.

In a different study, for the same groups, SCC values in 538 dairy cattle herds in the United States were found to be %7.02, % 15.86, % 36.40, % 33.17, and 7.54%, respectively (DeLong et al., 2017). When the results of the study conducted in the United States were assessed in terms of EU norms, it was seen that while 92.46% of milk samples were appropriate, 7.54% needed rehabilitation. Accordingly, the results obtained in this study show similarities to the results obtained in the study conducted in the United States.

Especially in specific studies, it is important to estimate the effects of independent variables or groups of variables on a dependent variable. For this purpose, various statistical analysis methods can be used. In this study, the regression tree method was used as an alternative to general methods in order to determine the effects of various factors on raw milk SCC values in cows. The most important advantage of the regression tree analysis over other methods is that it does not contain assumptions since it is a nonparametric method (Topal et al., 2010; Orhan et al., 2016).

Looking at the regression tree diagram created for the SCC, it was seen that the most important predictor on the SCC was the city factor, while it was followed by the region and the breed factors. In this study, the mean SCC value of the regression tree was determined as 200.826 cells/mL. In a different study, determined of the mean SCC value as 254.479 cells/mL in dairy cattle raised in 538 enterprises in the southern United States. Accordingly, the value obtained in this study was found to be lower than the value determined in the United States (DeLong et al., 2017).

The first important factor affecting the SCC value was the region where cows were raised. In the study, for SCC, 207,614 animals were divided into seven region groups (Marmara (MR), Aegean (AR), Mediterranean Sea (MSR), Central Anatolia (CAR), Black Sea (BSR), Eastern Anatolia, and Southeastern Anatolia (SAR) Regions). It was seen in the decision tree that 94.6% of the animals were classified in the AR group (Node 1), 4.4% in the CAR, MSR, MR, and SAR region groups (Node 2), and 1% in the EAR and BSR groups. It was found that the most important determinant among the regional groups was the AR group and the SCC value in this group was 197,535 cell/mL. While the lowest SCC value was observed in EAR and BSR regions (Node 3) with 178.026 cells/mL, the highest SCC value was detected in MR, CAR, MSR, and SAR regions (Node 2) with 285.430 cells/mL.

In a study conducted in 7 states (regions) where Class A milk production is allowed in the southeastern United States, It has been reported that the highest tank SCC was in the state of Mississippi with 383.314 cells/mL and the lowest SCC was in the state of North Carolina with

212.940 cells/mL (DeLong et al., 2017). In another study, it has been reported that the lowest and highest mean SCC values in different Holstein-Friesian cow herds in the Mediterranean region of Turkey as 296.483 cells/mL and 688.811 cells/mL, respectively (Koç, 2008).

The second important factor affecting SCC was the city factor in which cows were raised. While the lowest SCC was determined in the city groups of C40, C42, C16, C26, C07, C58, C27, and C66 (Node 11) with 166.001 cells/mL, the highest SCC was determined in the city groups of C09, C45, C20, C64, C48, and C43 (Node 5) with 266.744 cells/mL (Fig 3).

In a study conducted in Turkey, it was reported that there were significant differences between cities in terms of the somatic cell count ( $p < 0.05$ ) (Patir et al., 2010). Similarly, it has been reported that there was no statistical difference between the tank SCC values for the cities located in the Thrace region in the northwest of Turkey (Önal and Özder, 2007).

In this study, the third important factor affecting SCC was the breed of cows. Node 9 included only the subgroup of all Simmental and Montbeliarde cows (only those from the İzmir and Afyon cities). Another subgroup of these cows (from the CAR, MSR, MR, and SAR regions) was included in Nodes 6 and 7. While the lowest SCC was determined in S (Simmental) and MC (Montbeliarde) breeds (Node 9) with 170.116 cells/mL, the highest SCC was determined in Holstein Friesian (HF), Brown Swiss (BS), Red Holstein (RH), and Swedish Red (SR) breeds in Node 8 group with 197.739 cells cells/mL (Fig 3).

One of the factors that make cows susceptible to udder infections is breed. In many studies, the relationship between the incidences of catching udder infections in cows of different breeds has been examined and the differences between breeds have been found to be statistically significant (Busato et al., 2000; Welper et al., 1992). In a study conducted in Kahramanmaraş, the effect of the breed on SCC was found to be significant ( $p < 0.05$ ) for Holstein, Brown, and Simmental cattle raised in a private dairy farm (Harmandar and Kaygısız, 2019).

In this study, when the results of the analyses performed using the regression tree method were examined, it was determined that the variables affecting the SCC were the region code, city code, and breed code, respectively, in order of importance. SCC value in simmental and Montbeliarde cow breeds in İzmir (C35) and Afyonkarahisar (C03) cities in the Aegean Region (AR) was found to be lower than cows in other regional groups. On the other hand, SCC values of Holstein Friesian (HF), Brown Swiss (BS), Red Holstein



(RH) and Swedish Red (SR) breeds were determined as high for the same cities.

In terms of SCC, the breed was the secondary factor in the groups of Central Anatolia Region (CAR), Mediterranean Sea Region (MSR), Marmara Region (MR), and Southeastern Anatolia Region (SAR) (Node 2). While the highest SCC value was detected in the Aksaray (C68), Burdur (C15), Balıkesir (C10), Kırklareli (C39), Tekirdağ (C59), and Ankara (C06) cities group (Node 10) cows from the Simmental (S) and Red Holstein (RH) breeds group (Node 7), Kırşehir (C40), Konya (C42), Bursa (C16), Eskişehir (C26), Antalya (C07), and Sivas (C58), The highest SCC value was determined in Gaziantep (C27) and Yozgat (C66) city group (Node 11).

## CONCLUSION

In conclusion, the average SCC was found to vary between 150.000 cells/mL to 635.000 cells/mL in 71.0% of the cow population studied. In 40.4% of this rate, the mean somatic cell count was  $\geq 200,000$  cells/mL. These cows were considered as cows with suspected subclinical mastitis and it was thought that at least one udder quarter had subclinical mastitis infection suspicion (Table 4).

In milk, SCC increases depending on the high relative humidity and temperature in the environment and the procedures (technical and hygienic conditions) applied before, during, and after milking. In this study, it was found that SCC increased in the autumn season, especially in multiparous cows. Carrying out a mastitis eradication program in autumn and before the dry period positively affects the next lactation period. Therefore, it is recommended to screen for mastitis in geographical regions, cities and breeds with high SCC values in the autumn season. Cows with mastitis detected in SCC screening should be treated on time and all cows should start the dry period after intramammary preventive drug application. Also awareness of breeders on milking technical, milking hygiene and mastitis disease should be ensured.

In addition, it is recommended to conduct more detailed studies on the factors affecting the SCC based on the results obtained in this study for the region (CAR, MSR, MR, and SAR), city (C09, C45, C20, C48, and C43), and multiparous cows of all breeds with high SCC values.

## ACKNOWLEDGEMENTS

The use of this study data was authorized by “the Cattle Breeders’ Association of Turkey (CBAT)” on 21.08.2019

with the decision of the Board of Directors No. 2019/10. We would like to thank CBAT.

## CONFLICT OF INTEREST

The author declare that there is no conflict of interest

### Author’s contribution

The design of this study, the literature search, the analysis of the data, the writing of the article, and the reading of the article were all done by the OS

## REFERENCES

- Ahn, B. S., K. S. Baek, S. J. Park and H. J. Lee. 2005. The effects of various factors on milk yield and variation in milk yield between milking, milk components, milking duration and milking flow rate in Holstein dairy cattle. *J. Anim. Sci. Technol.* 47: 919-924.
- Andrade, L. M., L. El Faro, V. L. Cardoso, L. G. Albuquerque, L.D. Cassoli and P.F. Machado. 2007. Genetic and environmental effects on milk production and somatic cell count in Holstein cows. *Rev. Bras. Zoo.* 36: 343-349.
- Atasever, S. and H. Erdem. 2008. Relationships between mastitis and electrical conductivity of raw milk in dairy cows. *J. Fac. Agric.* 23: 131-136.
- Busato, A., P. Trachsel, M. Schallibaum and J. M. Blum. 2000. Udder health and risk factors for subclinical mastitis in organic dairy farms in Switzerland. *Prev. Vet. Med.* 44: 205-220.
- Coentrão, C. M., G. N. Souza, J. R. F. Brito, E. Paiva, M. A. V. Brito and W. Lilenbaum. 2008. Risk factors for subclinical mastitis in dairy cows. *Arq. Bras. Med. Vet. Zoo.* 60: 283-288.
- Çelik, S. 2019. Comparing predictive performances of treebased data mining algorithms and mars algorithm in the prediction of live body weight from body traits in Pakistan goats. *Pak. J. Zool.* 51: 1447-1456.
- DeLong, K. L., M. L. Dayton, S. Schexnayder, P. Krawczel, M. Fly, L. Garkovich and S. Oliver. 2017. Farm business and operator variables associated with bulk tank somatic cell count from dairy herds in the Southeastern United States. *J. Dairy Sci.* 100: 9298-9310.
- Dohoo, I. R. and K. E. Leslie. Evaluation of changes in somatic cell counts as indicators of new intramammary infections. *Prev. Vet. Med.* 1991;10:225-237.
- Golebiewski, M. and P. Brzozowski. 2007. Montbeliarde-economical breed of cattle in Polish. *Prz. Hod.* 2: 7-9.
- Green, M. J., A. J. Bradley, H. Newton and W. J. Browne. 2006. Seasonal variation of bulk milk somatic cell counts in UK dairy herds. Investigations of the summer rise. *Prev. Vet. Med.* 74: 293-308.
- Green, M. J., A. J. Bradley, G. F. Medley and W. J. Browne. 2008. Cow, farm, and herd management factors in the dry period associated with raised somatic cell counts in early lactation. *J. Dairy Sci.* 91: 1403-1415.
- Harmandar, A. and A. Kaygisiz. 2019. Comparison of Holstein, Brown Swiss and Simmental Breed Cattles in Terms of the Somatic Cell Counts. 2nd International Agricultural Congress, Ankara, Turkey. p. 7-16.
- Committee of International Animal Recording (ICAR). 2020. Guidelines for On-line Milk Analysis. Available from: <https://www.>

- icar.org/guidelines/12-milk-analysis.pdf [Last accessed on 2022 Aug 04].
- Koç, A. 2008. A study of somatic cell count in the milk of Holstein Friesian cows managed in Mediterranean climatic conditions. *Turk. J. Vet. Anim. Sci.* 32: 13-18.
- Koç, A. 2011. A study of the reproductive performance, milk yield, milk constituents, and somatic cell count of Holstein-Friesian and Montbeliardee cows. *Turk. J. Vet. Anim. Sci.* 35: 295-302.
- Koç, Y., E. Eyduran and O. Akbulut. 2017. Application of regression tree method for different data from animal science. *Pak. J. Zool.* 49: 599-607.
- Kocakoç, I. D. and I. Keser. 2020. Exploring decision rules for election results by classification trees. *KnE Soc. Sci.* 4: 107115.
- Kuczyńska, B., K. Puppel, M. Gołębiewski, M. Kordyasz, H. Grodzki and P. Brzozowski. 2012. Comparison of fat and protein fractions of milk constituents in Montbeliardee and Polish Holstein-Friesian cows from one farm in Poland. *Acta Vet. Brno.* 81: 139-144.
- Kurt, I., M. Ture and A. T. Kurum. 2008. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Syst. Appl.* 34: 366-374.
- Küçükönder, H., F. Üçkardeş, A. Ceyhan and M. Çınar. 2015. Determination of the effect of somatic cell count on udder measurements and subclinical mastitis with data mining method. *Pak. Vet. J.* 35: 441-445.
- Küçükönder, H., F. Üçkardeş and D. Naric. 2014. A data mining application in animal breeding: Determination of some factors in Japanese quail eggs affecting fertility. *Kafkas Univ. Vet. Fac. J.* 20: 900-908.
- Larsen, D. R. and P. L. Speckman. 2004. Multivariate regression trees for analysis of abundance data. *Biometrics.* 60: 543-549.
- Leitner, G., E. Shoshani, O. Krifucks, M. Chaffer and A. Saran. 2000. Milk leucocyte population patterns in bovine udder infection of different aetiology. *J. Vet. Med. Ser. B.* 47: 581-589.
- Li, N., R. Richoux, M. Boutinaud, P. Martin and V. Gagnaire. (2014). Role of somatic cells on dairy processes and products: A review. *Dairy Sci. Technol.* 94: 517-538.
- Lima, E. R., J. P. F. Oliveira, M. F. Bezerra, L. H. F. Borba, J. S. Bezerra and A. H. N. Rangel. 2019. Influence seasons of the year and somatic cell counts on physical-chemical and sensory parameters of milk and coalho cheese. *Acta Sci. Anim. Sci.* 41: e44666.
- Litwińczuk, Z., J. Król and A. Brodziak. 2015. Factors determining the susceptibility of cows to mastitis and losses incurred by producers due to the disease-a review. *Ann. Anim. Sci.* 15: 819-831.
- Ministry of Agriculture and Forestry, General Directorate of Meteorology (MAAF). 2008. Turkey Climate. MAAF, Ankara, Turkey. Available from: [https://www.mgm.gov.tr/files/genel/makale/13\\_turkiye\\_iklimi.pdf](https://www.mgm.gov.tr/files/genel/makale/13_turkiye_iklimi.pdf) [Last accessed on 2022 Aug 15].
- Magalhães, H. R., L. El Faro, V. L. Cardoso, C. C. P. Paz, L. D. Cassoli and P. F. Machado. 2006. Influence of environmental factors on somatic cell count and its relationship with losses in milk production of Holstein cows. *Rev. Bras. Zoo.* 35: 415-421.
- Nasr, M. A. F. and M. S. El-Tarabany. 2017. Impact of three THI levels on somatic cell count, milk yield and composition of multiparous Holstein cows in a subtropical region. *J. Therm. Biol.* 64: 73-77.
- Olde Riekerink, R. G. M., H. W. Barkema and H. Stryhn. 2007. The effect of season on somatic cell count and the incidence of clinical mastitis. *J. Dairy Sci.* 90: 1704-1715.
- Orhan, H., E. Eyduran, A. Tatlıyer and H. Saygıcı. 2016. Prediction of egg weight from egg quality characteristics via ridge regression and regression tree methods. *Rev. Bras. Zoo.* 45: 380-385.
- Önal, A. R. and M. Özder. 2007. The determination of somatic cell count and some components of raw milk evaluated by a private company in Trakya. *J. Tekirdag Agric. Fac.* 4: 195-199.
- Patir, B., O. P. Can and M. Gurses. 2010. Somatic cell counts in raw cow milk collected from different city. *Firat Univ. Vet. J. Health Sci.* 24: 87-91.
- Pyörälä, S. 2003. Indicators of inflammation in the diagnosis of mastitis. *Vet. Res.* 34: 565-578.
- Questier, F., R. Put, D. Coomans, B. Walczak and Y. Vander Heyden. 2004. The use of Cart and multivariate regression trees for supervised and unsupervised feature selection. *Chemometr. Intell. Lab. Syst.* 76: 45-54.
- Querengasser, J., T. Geishauser, K. Querengasser, K. Fehlings and R. Bruckmaier. 2002. Investigations of milk quality from teats with milk flow disorders. *J. Dairy Sci.* 10: 2582-2588.
- Reyes, J., J. Sanchez, H. Stryhn, T. Ortiz, M. Olivera and G. P. Keefe. 2017. Influence of milking method, disinfection and herd management practices on bulk tank milk somatic cell counts in tropical dairy herds in Colombia. *Vet. J.* 220: 34-39.
- Rupp, R. and D. Boichard. 2003. Genetics of resistance to mastitis in dairy cattle. *Vet. Res.* 34: 671-688.
- Schepers, A. J., T. J. Lam, Y. H. Schukken, J. B. Wilmink and W. J. Hanekamp. 1997. Estimation of variance components for somatic cell counts to determine thresholds for uninfected quarters. *J. Dairy Sci.* 80: 1833-1840.
- Skrzypek, R., J. Wojtowski and R. D. Fahr. 2004. Factors affecting somatic cell count in cow bulk tank milk-a case study from Poland. *J. Vet. Med. Ser. A.* 51: 127-131.
- Sordillo, L. M., K. Shafer-Weaver and D. DeRosa. 1997. Immunobiology of the mammary gland. *J. Dairy Sci.* 80: 1851-1865.
- SPSS. 2019. IBM SPSS Statistics for Windows, Version 25.0. IBM Corp 440. New York, USA.
- Topal, M., V. Aksakal, B. Bayram and A. M. Yaganoglu. 2010. An analysis of the factors affecting birth weight and actual milk yield in Swedish red cattle using regression tree analysis. *J. Anim. Plant Sci.* 20: 63-69.
- Welper, R. D. and A. E. Freeman. 1992. Genetic parameters for yield traits of Holsteins including lactose and somatic cell score. *J. Dairy Sci.* 75: 1342-1348.
- Yavuz, S. and A. Kaygısız. 2015. The relationship between some body and udder measurements with somatic cell count in Holstein cows. *KSU J. Nat. Sci.* 18: 9-18.
- Zheng, H., L. Chen, X. Han, X. Zhao and Y. Ma. 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. *Agric. Ecosyst. Environ.* 132: 98-105.