

# Phenotypic Characterization of Hair and Honamli Goats by Using Classification Trees Algorithms and Multivariate Adaptive Regression Splines (Mars)

Yasin ALTAY (✉ [yaltay@ogu.edu.tr](mailto:yaltay@ogu.edu.tr))

Eskişehir Osmangazi Üniversitesi <https://orcid.org/0000-0003-4049-8301>

---

## Research Article

**Keywords:** CART, CHAID, Exhaustive CHAID, Hair, Honamli, MARS, Phenotypic characterization, QUEST

**Posted Date:** May 25th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-522947/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

In order to meet the food demand of the increasing world population, it is very important to define the animal breeds and species raised in tropical and subtropical regions and to organize breeding programs for this. Discrimination animal breeds by morphological classification are a widely used method for a century. Although Honamli and Hair goats are very similar to each other morphologically, they can be subjectively distinguished by experienced breeders with some distinctive morphological markers. In the current study, certain body characteristics of Hair goats, which have a large portion of the population in Turkey, and Honamli goat, which has recently been registered as a new breed were used. Phenotypic characterization of these breeds has been made using data mining methods such as Classification and regression tree (CART), chi-square automatic interaction detector (CHAID), Exhaustive CHAID, Quick Unbiased, Efficient Statistical Tree, (QUEST), and multivariate adaptive regression splines (MARS) algorithms. In other words, the current study is the first data mining algorithms used for phenotypic characterization in Hair and Honamli goat breeds. Goats' morphological characteristics such as live weight (LW), withers height (WH), back height (BH), rump height (RH), chest Depth (CD), body length (BL), chest girth (CG), leg girth (LG), head length (HL), fore head (FH), ear length (EL), and tail length (TL), used in diagnosis of discrimination on breeds, were used as a binary response variable in Honamli and Hair breeds. Here, the independent variables used in data mining algorithms are the morphological characteristics of goats. CHAID, Exhaustive CHAID, CART, QUEST, and MARS were used as data mining algorithms to make an accurate decision in detecting effective morphological traits in breed discrimination. The success of the CHAID, Exhaustive CHAID, CART, QUEST and MARS algorithms in breed discrimination is 87.80%, 85.80%, 87.80%, 77.00%, and 88.51%, respectively, while the area under the ROC curve is 0.880, 0.853, 0.868, 0.784 and 0.942, respectively. As a result, using data mining methods for some body measurements of Honamli and Hair goats, whose morphological distinction is not exactly accurate, phenotype characterization separation was performed with high success in MARS and CHAID algorithms compared with the other methods. The outputs of this study can be used for breeding material by enabling pure Honamli goat breeding. Also, data mining algorithms can be included in gene resource conservation programs.

## Introduction

Identification and classification of breeds within a species according to phenotypic characteristics plays a key role in the basis of breeding and conservation program strategies. In this respect, researches on these strategies have been constantly carried out in tropical and subtropical regions of the world (Nsoso et al. 2004; Yakubu et al. 2011; Hassen et al, 2012). Goat is one of these species and goat breeding is an important form of production especially as livelihood resource for low-income families and barren areas which are not suitable for vegetative production (stony, slope, forest edge, weak pastures, fallow, stubble, and unproductive lands), and animal products such as meat, milk, fleece, hair, and leather are obtained through goat breeding (Kaymakçı et al. 2000). Approximately 97% of the existing goats in Turkey consist of Hair goat. Despite the high adaptability of the Hair goats, they are not preferred by small ruminant breeders (i.e., sheep and goats) due to low profitability. (Taskaya and Kale, 2020). Scientific data are important to solve the problems of the hair goat, which is widely grown. Hair and Honamli goats are very similar to each other morphologically and therefore, they were not separated until the 2000s, and their numerical existence

was evaluated within the Hair goat existence in the statistics. There are no differences between Honamli and Hair goats in terms of care and nutritional conditions (Gok, et al., 2011; Daskiran et al. 2018). However, Honamli breed was stated as superior in terms of birth weight, live weight, lactation milk yield, and breeding criteria (Aktas, et al., 2013; Varol, 2014; Akbas and Saatci, 2016; Tekin and Arli, 2019). Honamli goats have been purely bred for centuries in the Central and Western Taurus Mountains by the Honamli Yoruk tribes (Karadag and Soysal, 2018). The lack of scientific research on the gene resources of the Honamli race until this time is possibly due to continuous migration of the Turkish Yoruks (nomads) (Taskaya and Kale, 2020). Honamli goat, which is defined as a new breed by animal genetic resources, was taken under protection by the General Directorate of Agricultural Research and Policies in 2015 (Sozuer, 2019). After this development, it was reported that the phylogenetic similarity of Honamli and Hair goats to each other was over 85% and that these breeds could not be distinguished with the help of microsatellite markers (Bulut, et al., 2016; Karadag, 2016; Gumus, 2018; Karsli, et al., 2020; Gul, et al., 2020).

The distinction of nine domestic goat races raised in Turkey is made with microsatellite markers, but only the Honamli goat distinction has not been successful. It is seen that individuals belonging to other races are more clearly and genetically separated from each other (Gumus, 2018). In a study, which determined the genetic differences of the breeds with 20 different microsatellite markers in five different domestic goat breeds, it was reported that Honamli and Hair goats had the lowest genetic distance (0.0587) (Agaoglu and Ertugrul, 2012). The low genetic distance might be related to proximity of the breeding areas of Honamli and Hair goats (Agaoglu and Ertugrul, 2012; Bulut, et al., 2016).

Considering the genotypic characterization studies, the phenotypic characterization has great importance because Hair and Honamli goats cannot be separated. Although Honamli and Hair goats are morphologically similar to each other, they could be separated subjectively by experienced breeders (Akbas and Saatci, 2016; Sozuer, 2019). It is predicted that this distinction will reveal reliably with robust quantitative methods, and will fill an important gap in literature. Considering both the protection of gene resources and economical aspects, dissemination of the Honamli breed will be beneficial for goat breeders. The Honamli goat is under threat of extinction and its breeding is of great importance, both as a new breed registration and as a gene source. (Anonim, 2020). Due to their very different morphological and physiological characteristics, in recent years, goat breeders have preferred Honamli goats, raised within the scope of improvement project in public conditions, to hair goats (Aytekin, 2016).

Comparisons of some local goat genotypes raised in the world in terms of yield and morphology traits (Student t, Paired t, F and  $\chi^2$  test) and breed discrimination were made with the help of multivariate statistical methods (MANOVA), principal components analysis (PCA), canonical analysis, hierarchical, k-means clusters and step-wise, linear and nonlinear discriminate analysis (Jordana et al., 1993; Herrera, et al. 1996; Capote et al., 1998; Crepaldi et al., 2001; Macciotta et al., 2002; Herrera, et al. 2003; Lanari et al., 2003; Rodero et al., 2003; Nsoso et al., 2004; Zaitoun et al. 2005; Traoré et al., 2006; Dossa et al., 2007; Vargas et al., 2007; Traoré et al., 2008; Chacón, et al. 2011; Tölü and Savaş, 2012; El Moutchou, et al. 2014; El Moutchou, 2017; Çelik and Oflaz, 2018). Although classification trees and MARS algorithm applications, which are data mining algorithms that have become popular in the field of animal science in recent years,

are applied, there is no study in which racial discrimination is made by using morphological features in goats.

In this study, it was aimed to determine the phenotypic characterization of Hair and Honamli goats by using data mining algorithms and some of the morphological characteristics belonging to Honamli and Hair goats such as LW, WH, BH, RH, CD, BL, CG, LG, HL, FH, EL, and TL.

## **Material And Methods**

### **Animals**

The animal material of the study consists of 65 Hair goats (45 females, 20 males) and 83 Honamli goats (73 females, 10 males) of different ages (1, 2, 3, 4, 5 and 6 years and over), which are grown extensively in a private farm in Çalca district of Kütahya province. Honamli pedigree goats were brought to the enterprise in 2015 from an enterprise which is a member of the sheep and goat breeding association in Antalya. Hair goats in the enterprise were bred in previous years. Pedigree records of the breeds were checked by the business owner through the TürkVet system. In the busy winter season or in adverse weather conditions, goats have been fed in the pen. Concentrated feed is given for 1 month before the breeding season for flushing, and in addition to pasture, straw, alfalfa, and fescue grass are given as roughage sources in addition to pasture.

### **Measurement of Morphological Characteristics of Breeds**

A special scale designed for weighing small ruminants is used in determining the live weight of animals. All body characteristics of the goats were measured due to the suggestions of (Ertugrul, 1996), and LW was taken with 0.1 precision. All body measurements were made after the animals had adapted to the environment for a certain period of time on a flat platform and the stress factors were minimized. WH, BH, RH, BL, and CD, which are the height measurements of goats, were taken using a measuring stick and body circumference measurements (CG, and LG) were taken using a measuring tape. HL, FH, EL, and TL values were measured using calipers.

### **Statistical analysis**

The structure created by using all arguments and dividing the data into subgroups is termed a classification tree. The root node, which does not contain any fragmentation and contains only the dependent variable, is at the top of the classification trees. First, this root node is divided into two or more parts. While these separated parts are called parent branches, similarly, breaking up of parent branches is called daughter node or subset (Eyduvan et al. 2016). When fragmentation is over in the daughter nodes, and in the node where there is no branching anymore, the terminal becomes node (Orucoglu, 2011). By testing the independent variables taken into the model, the cut values of the explanatory variable are determined in a way to provide the specified category in the new node to be formed (Aksahan and Keskin, 2015).

Classification performances of CART (Breiman et al., 1984), CHAID (Kass, 1980), Exhaustive CHAID (Biggs et al., 1991), QUEST (Loh and Shih, 1997) and MARS (Friedman, 1991) data mining algorithms were utilized comparatively in terms of accuracy, sensitivity, specificity and area under ROC curve. While CHAID and Exhaustive CHAID classification tree algorithms are divided into nodes in the form of multiway (Akin *et al.*, 2018), CART and QUEST algorithms are divided according to the binary node structure rule (Kovalchuk *et al.*, 2017). MARS algorithm, which is a modified version of the CART algorithm, makes better predictions than binary logistic regression thanks to the hinges function in its structure. In the classification trees, the maximum tree depth was used as CHAID (3), Exhaustive CHAID (3), CART (5), and QUEST (3), respectively. After the whole data set (148 records) was randomly divided into 10 parts, nine parts of the training set of the models were created, while the model was estimated 5 times in the remaining part. In the formation of classification trees, the minimum number of parent and child nodes was taken as 10 and 5. Accuracy is the proportion at which a classification algorithm correctly separates Honamli and Hair goat. Sensitivity is the proportion at which the algorithm correctly classifies Honamli goats, while specificity is the proportion at which the algorithm correctly classifies Hair goats. Table 1 presents the confusion table for classifying algorithms.

Table 1

Confusion table for the classifier algorithms

		Predicted as Breeds	
		Honamli	Hair
Observed	Honamli	A	B
	Hair	C	D

The expressions  $T^+$ ,  $T^-$ ,  $F^+$  and  $F^-$  used in the accuracy, sensitivity and specificity equations represent numbers of true positive, true negative, false positive, and false negative, respectively. The formula used below to determine AUC (AUCse) was developed by (Hanley and McNei, 1982).

$$\text{Accuracy} = (T^+ + D) / (T^+ + F^+ + F^- + T^-)$$

$$\text{Sensitivity} = T^+ / (T^+ + F^+)$$

$$\text{Specificity} = T^- / (F^- + T^-)$$

$$\text{Error proportion} = 1 - \text{Accuracy}$$

$$se_{AUC} = \sqrt{\frac{AUC(1 - AUC) + (n_A - 1)(q1 - AUC^2) + (n_B - 1)(q2 - AUC^2)}{n_A n_B}}$$

$$n_A = T^+ + F^- \text{ and } n_B = F^+ + T^-$$

$$q1 = \frac{AUC}{2 - AUC} \text{ and } q2 = \frac{2AUC^2}{1 + AUC}$$

Statistical analyses of the classification trees were performed in IBM SPSS 23 package program (IBM Corp. Released, 2015). Earth (v5.1.2; Eyduran et al, 2019) and caret (v60.86; Kuhn, 2020) packages of R studio programmes were used for MARS analysis (R Core Team, 2020). To determine the area under the ROC curve and comparison (AUC) and to compare the algorithm pairs in the area under the ROC curve, the trial version of MedCalc package program was carried out with the help of version 19.5.1. Also, the z test was used to compare algorithms in terms of area under ROC curve.

## Results

Categorical variables belonging to Honamli and Hair goats in the study are presented in Table 2. Descriptive statistics of continuous variables obtained from Honamli and Hair goats are given in Table 3. Honamli females had higher values compared to males because females are older (Table 2). In other words, young billy goats are preferred in the herd to shorten the time between generations.

Table 2

Categorical variables belonging to Honamli and Hair goats

Factors	Levels	N	Percentage (%)
Sex	Female	118	79.73%
	Male	30	20.27%
Breeds	Honamli	83	56.08%
	Hair	65	43.92%
Age	1	48	32.43%
	2	8	5.41%
	3	18	12.16%
	4	10	6.76%
	5	4	2.70%
	6	60	40.54%
Ear	Comuk (native terms)	48	32.43%
	Lop	100	67.57%

Table 3

Descriptive statistics on live weight and some body measurements in Honamli and Hair goats of different age

Traits	Breed	Sex	N	Minimum	Maximum	Mean±SE	StDev	CoefVar
LW	Honamli	Female	73	27.10	84.70	60.76±1.54	13.13	21.60
		Male	10	37.20	63.10	48.02±2.77	8.76	18.24
	Hair	Female	45	27.30	72.20	48.40±1.76	11.77	24.33
		Male	20	29.00	43.60	36.15±0.95	4.25	11.74
WH	Honamli	Female	73	51.90	89.50	75.78±1.04	8.90	11.75
		Male	10	66.00	87.00	76.25±2.37	7.51	9.84
	Hair	Female	45	51.70	85.00	70.63±1.20	8.05	11.39
		Male	20	63.50	77.50	68.70±0.94	4.22	6.15
BH	Honamli	Female	73	64.50	90.50	79.14±0.71	6.05	7.65
		Male	10	66.50	86.00	76.60±1.97	6.23	8.13
	Hair	Female	45	60.00	79.50	71.09±0.69	4.66	6.55
		Male	20	61.00	75.50	68.40±1.01	4.50	6.58
RH	Honamli	Female	73	65.50	89.00	79.14±0.66	5.63	7.12
		Male	10	70.00	86.50	77.45±1.69	5.36	6.92
	Hair	Female	45	60.50	81.00	71.82±0.72	4.80	6.68
		Male	20	61.00	77.00	68.08±0.98	4.39	6.45
CD	Honamli	Female	73	16.50	29.50	24.66±0.29	2.47	10.00
		Male	10	20.50	27.50	24.45±0.76	2.41	9.85
	Hair	Female	45	18.50	27.50	22.36±0.28	1.86	8.31
		Male	20	18.50	23.50	21.13±0.32	1.44	6.82
BL	Honamli	Female	73	46.00	91.50	80.57±0.97	8.25	10.24
		Male	10	62.50	88.00	75.65±2.70	8.54	11.29
	Hair	Female	45	60.00	92.00	74.63±1.12	7.50	10.05
		Male	20	58.00	76.50	68.30±0.83	3.69	5.41
CG	Honamli	Female	73	70.00	104.00	91.04±0.82	6.96	7.65
		Male	10	81.00	95.00	86.25±1.36	4.30	4.99
	Hair	Female	45	73.00	102.50	86.64±1.05	7.02	8.11

		Male	20	72.50	90.50	81.50±0.98	4.38	5.37
LG	Honamli	Female	73	35.00	67.00	51.21±0.74	6.29	12.28
		Male	10	46.00	65.00	55.90±2.05	6.48	11.59
	Hair	Female	45	39.00	77.50	50.87±1.13	7.58	14.90
		Male	20	46.50	61.00	55.00±0.75	3.35	6.09
HL	Honamli	Female	73	17.00	24.00	21.23±0.18	1.49	7.03
		Male	10	19.50	22.50	20.80±0.31	0.98	4.70
	Hair	Female	45	17.00	23.00	19.67±0.22	1.45	7.39
		Male	20	17.50	22.50	20.13±0.29	1.30	6.44
FH	Honamli	Female	73	11.00	21.00	14.26±0.25	2.13	14.95
		Male	10	12.00	16.00	13.95±0.46	1.46	10.48
	Hair	Female	45	11.00	17.00	13.37±0.24	1.58	11.79
		Male	20	11.50	23.00	14.03±0.53	2.37	16.86
EL	Honamli	Female	73	8.00	22.50	16.97±0.44	3.74	22.02
		Male	10	9.50	21.00	15.50±1.34	4.22	27.24
	Hair	Female	45	13.00	28.00	17.79±0.45	3.00	16.88
		Male	20	7.00	20.50	14.68±0.86	3.86	26.31
TL	Honamli	Female	73	13.00	27.50	19.11±0.39	3.37	17.61
		Male	10	15.00	29.00	20.35±1.35	4.28	21.04
	Hair	Female	45	11.00	22.00	15.97±0.35	2.31	14.50
		Male	20	14.00	20.00	16.43±0.42	1.88	11.44
Live weight (LW), Withers height (WH), Back height (BH), Rump height (RH), Chest Depth (CD), Body length (BL), Chest girth (CG), Leg girth (LG), Head length (HL), Fore head (FH), Ear length (EL), and Tail length (TL)								

The MARS algorithm, which provides the best classification of Honamli and Hair goats, takes the form of body features "LW", "BH", "CD", "HG" and "HL" as independent variables in the prediction model. In addition, the model also includes "Age" and "Sex Male" variables that do not have body features.

$$GLM_{HONAMLI} = -0.5232799 - 3.033782 * SexMale + 1.5192 * \max(0, 4 - Age) - 1.068315 * \max(0, 35.4 - LW) + 0.4609831 * \max(0, BH - 72) + 25.86152 * \max(0, BH - 82) - 0.6795643 * \max(0, 25 - CD) + 1.559002 * \max(0, 77.5 - HG) - 0.5741605 * \max(0, 21.5 - HL)$$

The probability of any goat being Honamli can be estimated with the help of the  $P_{\text{HONAMLI}} = \frac{\exp^{\text{GLMHONAMLI}}}{(1 + \exp^{\text{GLMHONAMLI}})}$  equation. The "exp" value used in the equation refers to the base of the natural logarithm of 2.718. Using the basic MARS model, it is possible to derive a new prediction equation used in the classification of females. If the goats used in breed discrimination estimation are female animals older than 4 years old, the following equation can be used.

$$\text{GLM}_{\text{HONAMLI}} = -0.5232799 - 1.068315 * \max(0, 35.4 - \text{LW}) + 0.4609831 * \max(0, \text{BH} - 72) + 25.86152 * \max(0, \text{BH} - 82) - 0.6795643 * \max(0, 25 - \text{CD}) + 1.559002 * \max(0, 77.5 - \text{HG}) - 0.5741605 * \max(0, 21.5 - \text{HL})$$

For example, if the body characteristics of a 4-year-old female goat are LW = 40 kg, BH = 78 cm, CD = 25 cm, HG = 75 cm and HL = 20 cm, the breed could be estimated using the equation as follows;

$$\text{GLM}_{\text{HONAMLI}} = -0.5232799 - 3.033782 * \text{SexMale (Female=0)} + 1.5192 * \max(0, 4 - 4) - 1.068315 * \max(0, 35.4 - 40) + 0.4609831 * \max(0, 78 - 72) + 25.86152 * \max(0, 78 - 82) - 0.6795643 * \max(0, 25 - 25) + 1.559002 * \max(0, 77.5 - 75) - 0.5741605 * \max(0, 21.5 - 20)$$

$$\text{GLM}_{\text{HONAMLI}} = -0.5232799 + 0.4609831 * \max(0, 78 - 72) + 1.559002 * \max(0, 77.5 - 75) - 0.5741605 * \max(0, 21.5 - 20)$$

$$\text{GLM}_{\text{HONAMLI}} = -0.5232799 + 0.4609831 * 6 + 1.559002 * 2.5 - 0.5741605 * 1.5$$

$$\text{GLM}_{\text{HONAMLI}} = 5.27888295$$

$$P_{\text{HONAMLI}} = \frac{\exp^{\text{GLMHONAMLI}}}{(1 + \exp^{\text{GLMHONAMLI}})}$$

$$P_{\text{HONAMLI}} = \frac{2.718^{5.27888295}}{(1 + 2.718^{5.27888295})}$$

$$P_{\text{HONAMLI}} = 0.994924974$$

probably the goat belongs to the Honamli breed. Classification performances of data mining algorithms used for race discrimination are shown in Table 4. The areas under ROC (AUC) were statistically significant for all algorithms that made breed discrimination ( $P < 0.01$ ). Among the algorithms used for breed discrimination, the best classification performance is the MARS algorithm in terms of sensitivity (0.916), specificity (0.846), and general accuracy rate (0.937). The MARS algorithm was able to correctly classify 75 of 83 Honamli goats, 55 of 65 Hair goats and 88.50 (%) of all goats. In addition, the MARS algorithm was found to have the largest area in the breed discrimination diagnostic test with the area under the ROC curve of 0.942. Using the morphological characteristics of Honamli and Hair goats, the best performing algorithms for breed discrimination are MARS and CART. By using the morphological characteristics of Honamli and Hair goats, it was determined that the performances of the discrimination made by the CHAID algorithm had the values of sensitivity, specificity and accuracy respectively as 0.911, 0.841, and 0.878. The CHAID algorithm, which is used to classify goat breeds correctly, allocated 11 of 83 Honamli incorrectly and 72 correctly, while it separated 58 of 65 Hair goats correctly. CHAID has the area under the largest ROC with a value of 0.880 after the MARS algorithm. In the study, performance results of CART were determined as 0,849, 0,927, and 0,848 for sensitivity, specificity, and accuracy rate respectively. The CART algorithm estimated 79 of 83 Honamli goats, 51 of 65 Hair goats and 87.80 (%) of all goats by classifying them correctly. In addition, the CART algorithm was found to have the third largest AUC (0.868) value in the diagnostic test used for breed discrimination. Although the results are similar to the CHAID algorithm, the performance value of the Exhaustive CHAID algorithm is lower. The Exhaustive CHAID algorithm used to distinguish Honamli and Hair goats has performance values such as sensitivity (0.861),

specificity (0.855), and accuracy rate (0.858). While Exhaustive CHAID algorithm classified 74 of 83 Honamli goats correctly, 9 of them were classified incorrectly. While this algorithm allocated 12 of 65 Hair goats incorrectly, 53 of them were classified correctly. The area under ROC, which is one of the diagnostic test performance criteria in racial discrimination, was determined as (0.853). In this study, in which the breed discrimination was made by using the morphological characteristics of Honamli and Hair goats, performance values of the QUEST algorithm, which is one of the data mining algorithms, were determined as sensitivity (0.889), specificity (0.682), and accuracy (0.770) rate, respectively. The QUEST algorithm, which has the worst performance among the algorithms, classified only 56 of the 83 Honamli goats correctly, while 27 of them were incorrectly classified. Although the QUEST algorithm failed to classify Honamli goats, it classified 58 of 65 Hair goats correctly and seven of them incorrectly. The fact that the sensitivity and specificity values of the model performance value criteria are close to each other and that the AUC value is close to one is an indication of the correct classification. Models compared in terms of AUC can be mathematically expressed as MARS = CHAID = CART > Exhaustive CHAID > QUEST. When all data mining algorithms are compared among themselves in terms of performance criteria, it was determined that the most successful algorithm used in breed discrimination is MARS. Among the classification tree algorithms, it was determined that the CHAID algorithm has the best diagnostic test performance. Although the CART algorithm correctly classified Honamli goats with a high rate (95.20%), the percentage of correctly classifying Hair goats (78.50%) remained low. Although the Exhaustive CHAID algorithm correctly separated both breeds in close percentages, their performance values were insufficient compared to other algorithms (MARS, CHAID, and CART). The QUEST algorithm made the worst classification according to the model performance criteria. While QUEST algorithm correctly separated Hair goats with a high rate (89.20%), the separation percentage of Honamli goats (67.50%) remained quite low.

Table 4

Classification performances of the data mining algorithms for each diagnosis test

Algorithm	Sensitivity	Specificity	AUC±SE	Accuracy of Model	Correctly Classify of Honamli breed	Correctly Classify of Hair breed	P-value
MARS	0.916	0.846	0.942±0.028 <sup>a</sup>	0.885	0.894	0.892	<0.001
CHAID	0.911	0.841	0.880±0.027 <sup>a</sup>	0.878	0.867	0.892	<0.001
CART	0.849	0.927	0.868±0.023 <sup>a</sup>	0.878	0.952	0.785	<0.001
Exhaustive CHAID	0.861	0.855	0.853±0.030 <sup>ab</sup>	0.858	0.892	0.815	<0.001
QUEST	0.889	0.682	0.784±0.032 <sup>b</sup>	0.770	0.675	0.892	<0.001
<b>a, ab, b</b> The difference between AUC with letter in any data mining algorithm column is significant (P<0.05).							

CHAID was chosen as the best classifier among the classification trees for the distinction of Honamli Hair goats (Table 4). In the root node of the CHAID diagram, 83 (56.10 %) of the 148 goats were classified as Honamli 65 (43.90 %) and as Hair (Figure 2). When the CHAID diagram is examined, it was determined that the first order effective independent variable on breed discrimination is RH (Adj. P-value = 0.000,  $\chi^2 = 59.332$ ), second order is Age (Adj. P-value = 0.014,  $\chi^2 = 9.981$ ), and BH (Adj. P-value = 0.036,  $\chi^2 = 6.313$ ), and third-order independent variables were LG (Adj. P-value = 0.045,  $\chi^2 = 13.362$ ) and CD (Adj. P-value = 0.003,  $\chi^2 = 12.577$ ). Branches generated by independent variables in the entire tree structure are statistically significant ( $P < 0.05$ ).

All goats (Node 0) considered in the study were divided into 3 sub-groups (nodes) in terms of RH variable. In the first node, 39 (83%) of the goats with RH =  $< 71.00$  cm shorter were Hair and 8 (17%) of them were Honamli. In the second node, 25 Hair (43.1%) and 33 Honamli (56.9%) of 58 goats were classified in a range of  $71.0 < RH = < 79.0$ . In the third node, it was determined that 42 of the goats ( $71.9 < RH$ ) with RH trait more than 79 cm were Honamli (97.7%) and only one of them was Hair goat.

Goats (Node 3rd) with RH characteristics greater than 79 cm formed nodes 6th and 7th in terms of GH characteristics. In the 6th Node, 83.30% of the goats with the BH trait less than 79.50 cm or equal value are classified as Honamli and 16.70% as Hair goat. All of the goats with the BH trait values greater than 79.50 belong to the Honamli breed (Node 7th).

While the 3rd and 4th nodes of the CHAID algorithm diagram showed a division according to the age variable, it did not have a direct but indirect effect on breed discrimination. Accordingly, it was determined that 13 of the goats aged 3 and under are Honamli (92.9%) and one of them is Hair goat (7.1%) ( $LG = < 53.50$ ) in terms of LG (Node 8th). In Node 9, the goats with a value between  $53.50 < LG = < 59.50$  were classified as 2 Honamli (28.6%) and 5 as Hair (71.4%). At node 10, 100% of all goats with the LG trait larger than 79 cm ( $71.90 < LG$ ) belong to the Honamli breed. In node 11, when CD characteristics of goats older than 3 years are less than and equal to 23.50 cm, ( $CD = < 23.50$ ), 93.30% of goats are classified as Hair and 6.7% as Honamli. If the CD trait is greater than 23.50 cm ( $23.50 < CD$ ), the probability of finding Honamli goat is 68.80% and 31.20% is Hair goat (Node 12th).

## Discussion

The most successful data mining algorithms used in the phenotypic characterization of Honamli and Hair goats are MARS and CHAID. While the MARS algorithm uses "LW", "BH", "CD", "CG", "Sex", "Age", and "HL" properties as independent variables in breed discrimination, CHAID algorithm uses "RH", "Age", "BH", "LG", and "CD". Essentially the same key variables can be used to describe closely related animal species (FAO, 2012). Nsoso et al. (2004) reported that the effect of age is important in the phenotypic characterization of indigenous Tswana goats raised in Botswana. This finding seems to coincide with the results of the CHAID and MARS models used in our study.

Body length (BL) and chest girth (CG) characteristics were reported to differ significantly in the distinction between brown and gray Bengal goats (Mukeherjee et al., 1979). It was emphasized that front shark circumference (FSC), chest girth (CG), chest depth (CD), rump height (RH) and rump width (RW), and

shoulder height (SH) traits are important for the distinction of five different indigenous goat breeds in Spain (Herrera, et al. 1996). Martínez et al. (2014) reported that the chest depth and rump length characteristics of the Murciano-Granadina and Malagueña dairy goat breeds raised in Spain are important in breed discrimination. Since the goat breeds in these literatures are in different environmental conditions and their genetic structure is different, they are not consistent with our study results.

Dossa et al. (2007) reported that the rate of correct assignment of goats in Benin to the appropriate grazing system is 76.60%. In Jordan, two indigenous goat breeds were separated by simple discriminant analysis by using the morphological characteristics, and it was stated that the Desert and crossbred goats were assigned to the correct class as 65.60% and 79.80%, respectively (Zaitoun, 2005). In the current study, the data mining algorithms used to separate Honamli and Hair breeds are MARS, CHAID, CART, Exhaustive CHAID, and QUEST. The assignment rates of Honamli goats to the correct class were 89.40%, 86.70%, 95.20%, 89.20%, and 67.50%, respectively, while the rates of assigning the Hair goats to the correct class were 89.20%, 89.20%, 78.50%, 81.50%, and 89.20%, respectively. All data mining algorithms also performed better than Desert goats' results in terms of the rate of assignment of Honamli and Hair goats to the correct class. However, the assignment rate of Benin and crossbred goats to the correct class performed better than the Honamli goat QUEST algorithm results. The QUEST algorithm appeared as less successful than other data mining algorithms.

Orhan et al. (2018) reported that Honamli and Hair goats could not be distinguished between breeds only because of genotypic characterization, and morphological examinations on the hair of goats do not have important traits in terms of breed discrimination considering the types of bulbus. They reported that although it was determined that the bulbus and scapus pili diameters of the Honamli goat hairs were larger than the bristle goat hairs, they could not be used in racial discrimination due to the heterogeneous structure of the hair samples (Orhan et al. 2018). In the present study, even if the data set has a very heterogeneous structure, it enables successful segregation with data mining algorithms.

## Conclusion

Breeding of hair goats has become essential due to the economic and good adaptation of goats in the environmental conditions, where they are raised. In this context, Honamli goats, which have high adaptability and yield capability, should be crossbred with Hair goats after pure breeding. As a first step to pure breeding, it is necessary to distinguish between goat breeds correctly.

Data mining algorithms used in racial discrimination are among the most reliable (robust) methods. Considering the successful performances of four different classification trees and MARS analysis in racial discrimination according to the current study results, it was determined that more accurate classification could be made with the CHAID and MARS methods. Especially in the breeding season of Hair goats, the preference of Honamli billy goats as breeding breeders due to their superior characteristics will provide success in crossbreeding. These crossbreeding studies could provide employment for breeders as well as obtaining higher productivity animals and will play a triggering role in increasing red meat production nationally.

The data mining algorithms evaluated in the present study might shed a light on further studies. In addition, these algorithms can distinguish breeds or crossbred that cannot be distinguished between races by genetic proximity, and they should not be ignored especially in statistical models that can be applied in gene resources conservation programs.

## **Declarations**

## **Acknowledgements**

I would like to thank Mehmet Tostcu goat farm in Kütahya for obtaining body characteristics and providing the records. In addition, the authors acknowledge that this work is original and that no part of or complete content of this has been published in any other journal and congress.

## **Author contribution**

The design of the study, data collection, statistical analysis, writing of the manuscript, conversion to the journal format and submission to the journal were done by author. The author has approved the submitted version.

## **Funding**

This study was not funded by any organization.

## **Data availability**

Not applicable.

## **Code availability**

Not applicable.

## **Ethics approval**

Ethical rules were followed by following all applicable international, national and institutional guidelines for the care and use of animals. In the study, there is no need for an ethical approval due to the lack of blood sampling from the animals and the absence of any surgical procedures. All data were collected with the approval of the breeder.

## **Conflict of interest**

The author declare no competing interests.

## References

1. Akbas, A.A., Saatci, M., 2016. Growth, slaughter, and carcass characteristics of Honamlı, Hair, and Honamlı x Hair (F1) male goat kids bred under extensive conditions. *Turkish Journal of Veterinary and Animal Sciences*, 40(4), 459-467. <https://doi.org/10.3906/vet-1511-5>.
2. Akin, M., Hand, C., Eyduran, E., Reed, B.M., 2018. Predicting minor nutrient requirements of hazelnut shoot cultures using regression trees. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 132(3), 545–559. <https://doi.org/10.1007/s11240-017-1353-x>.
3. Aksahan, R., Keskin, I., 2015. Determination of the some body measurements effecting fattening final live weight of cattle by the regression tree analysis. *Selcuk Journal of Agriculture and Food Sciences*, 2(1), 53-59.
4. Aktas, A.H., Gok, B., Tekin, M.E., Halici, I., Bas, Demirci, H., Dursun, S., 2013. Comparison of intensive fattening performance of Hair Goat and Honamli type of Hair kids, 8th National Animal Science Congress, Çanakkale, Turkey.
5. Anonymous, 2020. <https://www.esk.gov.tr/tr/11129/Honamli-Kecisi>. Date of access: 09/12/2020.
6. Agaoglu, O.K., Ertugrul, O., 2012. Assessment of genetic diversity, genetic relationship and bottleneck using microsatellites in some native Turkish goat breeds. *Small Ruminant Research*, 105(1-3), 53-60. <https://doi.org/10.1016/j.smallrumres.2011.12.005>.
7. Aytekin, 2016. Effects of Two Different Rearing Systems on Kid Growth in Honamlı Goat. pp: 487-494. *International Human and Nature Sciences: Problems and Solution Seeking Congress October 7 – 9, 2016 Sarajevo*.
8. Biggs, D., De Ville, B., Suen, E., 1991. A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, 18(1), 49–62. <https://doi.org/10.1080/02664769100000005>.
9. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and regression trees*. Chapman & Hall/CRC
10. Bulut, Z., Kurar, E., Ozsensoy, Y., Altunok, V., Nizamlioglu, M., 2016. Genetic diversity of eight domestic goat populations raised in Turkey. *BioMed Res. Int.* <https://doi.org/10.1155/2016/2830394>.
11. Celik, H.T., Olfaz, M., 2018. Investigation on survival rate and growth characteristics of pure hair goat and Saanen x hair goat (F<sub>1</sub>, B<sub>1</sub>, B<sub>2</sub>) crossbreds in breeder conditions. *Mediterranean Agricultural Sciences*, 31(1), 77-85.
12. Chacón, E., Macedo, F., Velázquez, F., Paiva, S.R., Pineda, E., McManus, C., 2011. Morphological measurements and body indices for Cuban Creole goats and their crossbreds. *Revista Brasileira de Zootecnia*, 40, 1671-1679. <https://doi.org/10.1590/S1516-35982011000800007>.
13. Capote, J., Delgado, J.V., Fresno, M., Camacho, M.E., Molina, A., 1998. Morphological variability in the Canary goat population. *Small Ruminant Research*, 27: 162-167. [https://doi.org/10.1016/S0921-4488\(97\)00047-3](https://doi.org/10.1016/S0921-4488(97)00047-3).

14. Crepaldi, P., Negrini, R., Milanese, E., Gorni, C., Cicogna, M., Ajmone-Marsan, P., 2001. Diversity in five goat populations of the Lombardy Alps: comparison of estimates obtained from morphometric traits and molecular markers. *Journal of Animal Breeding and Genetics*, 118, 173-180.  
<https://doi.org/10.1046/j.1439-0388.2001.00289.x>.
15. Daskiran, I., Savas, T., Koyuncu, M., Koluman, N., Keskin, M., Esenbuga, N., Konyali, A., Cemal, I., Gul, S., Elmaz, O., Kosum, N., Dellal, G., Bingöl, M., 2018. Goat production systems of Turkey: Nomadic to industrial. *Small Ruminant Research*, 163, 15-20. <https://doi.org/10.1016/j.smallrumres.2017.10.001>.
16. Dossa, L.H., Wollny, C., Gauly, M., 2007. Spatial variation in goat populations from Benin as revealed by multivariate analysis of morphological traits. *Small Ruminant Research*, 73, 150-159.  
<https://doi.org/10.1016/j.smallrumres.2007.01.003>.
17. El Moutchou, N., González, A.M., Chentouf, M., Lairini, K., Rodero, E., 2017. Morphological differentiation of Northern Morocco goat. *Journal of Livestock Science and Technologies*, 5(1), 33-41.  
<https://doi.org/10.22103/jlst.2017.1622>.
18. El Moutchou, N., González Martínez, A.M., Chentouf, M., Lairini, K., Rodero, E., 2014. Approach to morphological characterization of northern Morocco goat population. *Options Méditerranéennes*, 108, 427-432.
19. Ertugrul, M. 1996. *Ovine Breeding Practices*. Second Edition, Ankara University Faculty of Agriculture Publications Number: 1446, Ankara, Turkey.
20. Eyduran, E., Keskin, I., Erturk, Y.E., Dag, B., Tatliyer, A., Tirink, C., Aksahan, R., Tariq, M.M., 2016. Prediction of fleece weight from wool characteristics of sheep using regression tree method (Chaid Algorithm). *Pakistan Journal of Zoology*, 48(4), 957-960.
21. Eyduran, E., Akin, M., Eyduran, S.P., 2019. *Application of multivariate adaptive regression splines in agricultural sciences through R software*. Ankara: Nobel Academic Publishing.
22. FAO, 2012. *Phenotypic Characterization of Animal Genetic Resources*. FAO Animal Production and Health Guidelines No. 11. Rome.
23. Friedman, J.H., 1991. Multivariate adaptive regression splines. *Annals of Statistics*, 19: 1-141.
24. Gok, B., Aktas, A.H., Dursun, S., 2011. Honamli goat: Rising star of the Taurus Mountains, RBI 8th Global Conference on the Conservation of Animal Genetic Resources. Tekirdağ, Turkey, 65-72.
25. Gul, S., Yilmaz, O., Gunduz, Z., Keskin, M., Cemal, I., Ata, N., Onel, S.E., 2020. The genetic structure of the goat breeds belonging to northwest part of fertile crescent. *Small Ruminant Research*, 182, 22–28.  
<https://doi.org/10.1016/j.smallrumres.2019.09.009>.
26. Gumus, N., 2018. *The Identification of genetic diversity in Turkish native goat breeds by using microsatellite DNA markers*. Namik Kemal University Graduate School of Natural and Applied Sciences (Master Thesis).
27. Hassen, H., Baum, M., Rischkowsky, B., Tibbo, M., 2012. Phenotypic characterization of Ethiopian indigenous goat populations. *African Journal of Biotechnology*, 11(73). 13838-13846.  
<https://doi.org/10.5897/AJB12.2092>.
28. Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>.

29. Herrera, M., Rodero, E., Gutierrez, M.J., Peña, F., Rodero, J.M., 1996. Application of multifactorial discriminant analysis in the morphostructural differentiation of Andalusian caprine breeds. *Small Ruminant Research*, 22, 39-47. [https://doi.org/10.1016/0921-4488\(96\)00863-2](https://doi.org/10.1016/0921-4488(96)00863-2).
30. Herrera, P.S.J.I., Rodero, E., Sánchez, M.D., Luque, M., 2003. Raza caprina Moncaina. 1.- caracteres cuantitativos morfoestructurales. SEOC.
31. IBM Corp. Released, 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.
32. Jordana, J., Ribo, O., Pelegrin, M., 1993. Analysis of genetic relationships from morphological characters in Spanish goat breeds. *Small Ruminant Research*, 12, 301-314. [https://doi.org/10.1016/0921-4488\(93\)90065-P](https://doi.org/10.1016/0921-4488(93)90065-P).
33. Karadag, O., Soysal, M.İ., 2018. The determination of some, reproduction, growth and morphological traits in Honamli goats breeds. *Journal of Tekirdag Agricultural Faculty*, 15(1), 135-142.
34. Karadag, O. 2016. A Study on the investigation of Honamli through some morphological characteristics fertility and casein genes polymorphism. Namık Kemal University Graduate School of Natural and Applied Sciences (PhD Thesis).
35. Karsli, T., Demir, E., Fidan, H. G., Aslan, M., Karsli, B. A., Arik, I. Z., Semerci, E.S., Karabag, K., Balcioglu, M.S., 2020. Determination of genetic variability, population structure and genetic differentiation of indigenous Turkish goat breeds based on SSR loci. *Small Ruminant Research*, 190, 106147. <https://doi.org/10.1016/j.smallrumres.2020.106147>.
36. Kass, G.V., 1980. An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2), 119-127. <https://doi.org/10.2307/2986296>.
37. Kaymakci, M., Eliçin, A., Tuncel, E., Pekel, E., Karaca, O., Işın, F., Taşkın, T., Aşkın, Y., Emsen, H., Özder, M., 2000. Small ruminant production in Turkey, Turkey V. Technical Agricultural Engineering Congress, 17-21 January, 765-793, Ankara, Turkey.
38. Kovalchuk, I.Y., Mukhitdinova, Z., Turdiyev, T., Madiyeva, G., Akin, M., Eyduran, E., Reed, B.M., 2017. Modeling some mineral nutrient requirements for micropropagated wild apricot shoot cultures. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 129(2), 325–335. <https://doi.org/10.1007/s11240-017-1180-0>.
39. Kuhn, M., 2020. Caret: Classification and regression training. Retrieved from <https://CRAN.R-project.org/package=caret>
40. Lanari, M., Taddeo, H., Domingo, E., Pérez-Centeno, M., Gallo, L., 2003. Phenotypic differentiation of Criollo goat population in Patagonia (Argentina). *Archives Animal Breeding*, 46, 347-356.
41. Loh, W.Y., Shih, Y.S., 1997. Split selection methods for classification trees. *Statistica Sinica*, 815-840.
42. Macciotta, N.P.P., Cappio-Borlino, A., Steri, R., Pulina, G., Brandano, P., 2002. Somatic variability of Sarda goat breed analysed by multivariate methods. *Livestock Production Science*, 75: 51-58. [https://doi.org/10.1016/S0301-6226\(01\)00309-8](https://doi.org/10.1016/S0301-6226(01)00309-8).
43. Martínez, A.G., García, M.H., Cuesta, M.L., Serrano, E.R., 2014. Influence of farming system and production purpose on the morphostructure of Spanish goat breeds. *Spanish Journal of Agricultural Research*, (1), 117-124.

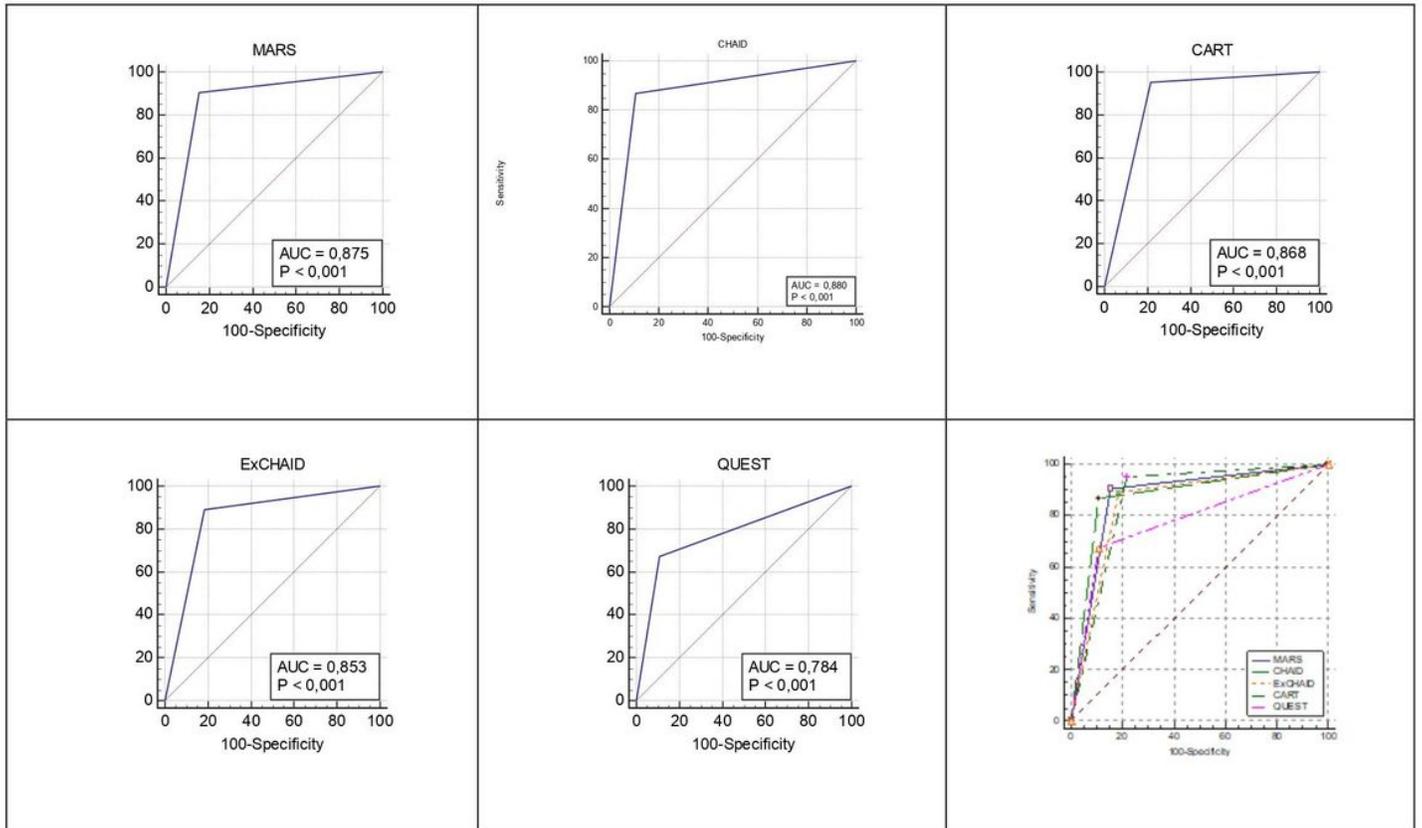
44. Mukeherjee, D.K., Singh, C.S.P., Mishra, H.R., 1979. A note on some phenotypic parameters in grey and brown Bengal goats. *Indian Journal Animal Science*, 49, 671–671.
45. Nsoso, J.J., Podisi, B., Otsogie, E., Mokhutshwane, B.S., Ahmadu, B., 2004. Phenotypic characterization of indigenous Tswana goats and sheep in Botswana: continuous traits. *Tropical Animal Health and Production*, 36, 789-800. <https://doi.org/10.1023/B:TROP.0000045979.52357.61>.
46. Orhan, I., Duzler, A., Alan, A., Elmaz, O., Ozgel, O., 2018. Morphological investigations on the hairs of the Honamli and the Hair goat (Black Goat). *Kocatepe Veterinary Journal*, 11(2), 173-179. <https://doi.org/10.30607/kvj.407473>.
47. Orucoglu, O., 2011. Determination of environmental factors affecting 305-day milk yield of Holstein cows by regression tree method. Suleyman Demirel University Graduate School of Natural and Applied Sciences (Master Thesis).
48. R Core Team, 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
49. Rodero, E., Herrera, M., Peña, F., Molina, A., Valera, M., Sepúlveda, N., 2003. Morpho-structural model for Florida and Payoya Spanish dairy goats in extensive (grazing) systems. *Revista Científica de la Facultad de Ciencias Veterinarias de la Universidad Del Zulia*, 13(5), 403-412.
50. Sozuer, O. 2019. Race of Honamli goat, as a biocultural diversity item. Mehmet Akif Ersoy University, Health Sciences Institute (Master Thesis).
51. Vargas, S., Larbi, A., Sánchez, M., 2007. Analysis of size and conformation of native Creole goat breeds and crossbreds used in smallholder agrosilvopastoral systems in Puebla, Mexico. *Tropical Animal Health and Production*, 39: 279-286. <https://doi.org/10.1007/s11250-007-9012-6>.
52. Taskaya, H., Kale, M., 2020. Investigation of Caprine Arthritis Encephalitis Virus (CAEV) infection in Honamli goat breed. *Mehmet Akif Ersoy University Journal of the Faculty of Veterinary Medicine*, 5(2), 58-63. <https://doi.org/10.24880/maevfd.682590>.
53. Tekin, M.E., Arli, M., 2019. The growth and survival rate of hair goat kids raised by public in Karaman region. *Journal of the Turkish Veterinary Medical Society*, 90(2), 152-157. <https://doi.org/10.33188/vetheder.539444>.
54. Tolu, C., Savas, T., 2012. Comparison of Gökçeada, Maltese and Turkish Saanen Goat Genotypes for Parturition and Kid Growth. *Journal of Animal Production*, 53(2), 17-25.
55. Traoré, A., Tamboura, H.H., Kaboré, A., Yaméogo, N., Bayala, B., Zaré, I., 2006. Caractérisation morphologique des petits ruminants de race locale Mossi au Burkina Faso. *Agri*, 39: 39-50.
56. Traoré, A., Tamboura, H.H., Kaboré, A., Royo, L.J., Fernández, I., Álvarez, I., Sangaré, M., Bouchel, D., Poivey, J.P., Francois, D., Toguyeni, A., Sawadogo, L., Goyache, F., 2008. Multivariate characterization of morphological traits in Burkina Faso sheep. *Small Ruminant Research*, 80: 62-67. <https://doi.org/10.1016/j.smallrumres.2008.09.011>.
57. Varol, M. 2014. Definition of Morphological Traits of Hair Goats in Denizli Province. Adnan Menderes University Graduate School of Natural and Applied Sciences (Master Thesis).
58. Yakubu, A., Salako, A.E., Imumorin, I.G., 2011. Comparative multivariate analysis of biometric traits of West African Dwarf and Red Sokoto goats. *Tropical Animal Health and Production*, 43(3): 561-566.

<https://doi.org/10.1007/s11250-010-9731-y>.

59. Zaitoun, I.S., Tabbaa, M.J., Bdour, S., 2005. Differentiation of native goat breeds of Jordan on the basis of morphostructural characteristics. *Small Ruminant Research*, 56: 173-182.

<https://doi.org/10.1016/j.smallrumres.2004.06.011>.

## Figures



**Figure 1**

All and individual ROC curves of classifying algorithms for diagnosis test of breed discrimination

