Citation: Özkök A, Keskin M, Tanuğur Samanci AE, Yorulmaz Önder E, Silahtaroğlu G. Discovering the chemical factors behind regional royal jelly differences via machine learning. U. Arı D. / U. Bee J. 2023,23(1):49-60. DOI: 10.31467/uluaricilik.1238027

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

DISCOVERING THE CHEMICAL FACTORS BEHIND REGIONAL ROYAL JELLY DIFFERENCES VIA MACHINE LEARNING

Makine Öğrenimi Yoluyla Bölgesel Arı Sütü Farklarının Arkasındaki Kimyasal Faktörleri Keşfetmek

Aslı ÖZKÖK[1*], Merve KESKİN[2], Aslı Elif TANUĞUR SAMANCI[3], Elif YORULMAZ ÖNDER[3], Gökhan SİLAHTAROĞLU[4]

[*1]Bee and Bee Products Application and Research Center (HARUM), Hacettepe University, Ankara, TÜRKİYE, Yazışma Yazarı / Corresponding author: E-posta: aozkok@hacettepe.edu.tr, ORCID No: 0000-0002-7336-2892.

[2]Vocational School of Health Services, Bilecik Şeyh Edebali University, Bilecik, TÜRKİYE, E-posta: merveozdemirkeskin@gmail.com, ORCID No: 0000-0001-9365-334X.

[3]SBS Scientific Bio Solutions Co Bee&You Propolis, İstanbul, TÜRKİYE, E-posta: asli@sbs-turkey.com, ORCID No: 0000-0003-1639-6495, E-posta: elif.onder@sbs-turkey.com, ORCID No: 0000-0003-1990-0693

[4]Department of Management Information Systems, School of Business, Istanbul Medipol University, İstanbul, TÜRKİYE, E-posta: gsilahtaroglu@medipol.edu.tr, ORCID No: 0000-0001-8863-8348

Geliş Tarihi / Received: 18.01.2023 Kabul Tarihi / Accepted: 16.03.2023 DOI: 10.31467/uluaricilik.1238027

ÖZ

Bu çalışmanın amacı, makine öğrenmesi yoluyla arı sütünün bölgesini belirlemek için ayırt edici kimyasal faktörleri keşfetmektir. Çalışmada, Türkiye'nin 13 farklı bölgesinden 84 numune kullanılmış ve nem, pH, asitlik ve 10-hidroksi-2-dekanoik asit (10-HDA) kimyasal parametreleri incelenmiştir. 13 yerden toplanan arı sütleri arasında dört kimyasal değer açısından farklılık olup olmadığı ANOVA testi ile incelenmiştir. İstatistiksel testlere ek olarak, arı sütlerini birbirinden neyin ayırdığını keşfetmek için bir makine öğrenimi modeli kullanılmıştır. Arı sütü, kimyasal analiz sonuçlarının tanımlayıcı istatistikleri sırasıyla, nem %63,05±2,99, pH 3,67±0,08, asitlik 45,32±3,55 ve 10-HDA 2,40±0,24 olarak bulunmuştur. Şaşırtıcı bir şekilde, makine öğrenimi modeli, 10-HDA'nın arı sütünün bölgesini belirlemek için en belirgin parametre olabileceğini öne sürmektedir. Bu bilgi, arı sütünün doğruluğunun tespitini daha kolay öğrenmemize yardımcı olacaktır.

Anahtar Kelimeler: Arı sütü, bal arısı, makine öğrenimi, 10-HDA

ABSTRACT

This study aims to discover the characteristic chemical factors for determining the region of royal jelly using machine learning. 84 samples from 13 different regions of Turkey were used for the study, and the chemical parameters of moisture, pH, acidity, and 10-hydroxy-2-decanoic acid (10-HDA) were investigated. ANOVA test was conducted to determine whether there are differences between royal jelly from 13 locations concerning the four chemical values. In addition to the statistical tests, a machine learning model was used to find out what makes royal jelly different from each other. The descriptive statistics of the chemical analysis results of royal jelly showed the following values: moisture 63.05%±2.99, pH 3.67±0.08, acidity 45.32±3.55, and 10-HDA 2.40±0.24. Surprisingly, the machine learning model suggests that 10-HDA may be the most prominent parameter for determining the region of royal jelly. This information will help us identify royal jelly's authenticity more easily.

Keywords: Royal jelly, honeybee, machine learning, 10-HDA

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

GENİŞLETİLMİŞ ÖZET

Amaç: Bu çalışmanın amacı, arı sütünün bölgesel farklılıklarının ardındaki kimyasal faktörleri makine öğrenimi kullanarak keşfetmektir. Bu amaçla Türkiye'nin farklı lokasyonlarından toplanan arı sütü ürünlerinin kimyasal analiz sonuçları karşılaştırılmıştır.

Gereç ve yöntem: 2020 yılında Türkiye'nin 13 farklı ilinden toplam 84 adet arı sütü örneği toplanmıştır. Toplanan örnekler analize kadar -18°C'de saklanmıştır. Daha sonra kimyasal analizler yapılmıştır. Bu çalışmada kullanılan veriler arı sütünün kimyasal analiz sonuçlarından oluşmaktadır. Şekil 1'deki haritada gösterilen 13 şehir ismi hedef değişken olarak yapay zeka makine öğrenmesi algoritmalarına verilmiştir. Kullanılan kimyasal değerler nem, pH, asitlik ve 10-hidroksi-2-dekanoik asittir (10-HDA). Her spesifikasyon, karşılık gelen analizle etiketlenmiştir. Diğer bir deyişle, makine analizi biliyordu ve öğrenme sırasında dikkate alıyordu. Arı sütleri arasında kimyasal analiz sonuçları açısından anlamlı istatistiksel fark olup olmadığını belirlemek için ANOVA testi yapılmıştır. Nem, 10-HDA (%), pH ve asitlik açısından bölgeler arasında anlamlı bir fark olup olmadığını görmek için tek taraflı ANOVA testi yapılmıştır. Daha sonra, 13 farklı lokasyondan toplanan arı sütlerinin özelliklerinin ardındaki gizli kuralları çıkarmak için karar ağaçlarına sahip makine öğrenme algoritması kullanılmıştır. SMOTE, makine öğrenimi çalışmalarında yaygın olarak kullanılan sentetik bir veri geliştirme tekniğidir (Silahtaroğlu 2009). Veri seti %30 / %70 oranında bölünmüştür. Bu, verilerin %70'inin makineyi eğitmek için ve %30'unun sonuçları test etmek ve değerlendirmek için kullanıldığı anlamına gelmektedir. Doğruluk, kesinlik, özgüllük, Cronbach' Kappa ve F-değeri, makinenin arı sütünün yerini tahmin etmek için ne kadar iyi eğitildiğini ölçmek için kullanılmıştır.

Bulgular ve tartışma: Arı sütü, kimyasal analiz sonuçlarının tanımlayıcı istatistikleri Tablo 1'de gösterilmiştir. Tablo 1'e göre, aşağıdaki ortalama kimyasal analiz değerleri elde edilmiştir: Nem %63,05±2,99, pH 3,67±0,08, asitlik 45,32±3,55 ve 10-HDA 2,40 ±0,24. Arı sütünün bileşiminin mevsime, ekolojik koşullara ve toplandığı bölgenin özelliklerine bağlı olduğu bildirilmiştir (Zheng ve Hu 2010). Fransız arı sütü üzerinde yapılan araştırmaya göre nem içeriği %60-70, 10-HDA değeri ise %1,4 ile %3,7 arasında değişmektedir (Wytrychowski vd. 2013). Başka bir çalışmada Anadolu arı sütünün 10-HDA değerlerinin %1,0 ile %3,9 arasında, nem içeriğinin ise %62,6 ile %73 arasında değiştiği saptanmıştır (Kolaylı vd. 2015). Ayrıca arı sütünün uluslararası standardı olan ISO/DIS 12824 (2016) uyarınca arı sütünün 10-HDA değeri minimum %1,4, nem içeriği minimum %62 ile maksimum %68,5 arasında ve asitlik 30 ila 53 m.q g/kg arasında değişmektedir. Keskin vd. (2020), arı sütü örneklerinin 10-HDA içeriğinin %2,1 ile %2,6 arasında, nem içeriğinin ise %62,6 ile %66,5 arasında değiştiğini tespit etmiştir. Tüm bu çalışmaların sonuçları ile bizim çalışmamız benzerdir (Tablo 1).

Tablo 2'de görülebileceği gibi, Levene test istatistikleri, dört parametrenin tümü için eşit varyansların varsayılabileceğini göstermektedir. Öte yandan, Tablo 3'teki test istatistikleri, 10-HDA hariç, H1 için sıfır hipotezinin reddedildiğini, yani pH, asitlik ve nem açısından araçlar arasındaki tüm farkların istatistiksel olarak anlamlı olduğunu göstermektedir. Bununla birlikte, bir p-değeri = 0,296 ile sıfır hipotezi reddedilemez, dolayısıyla 10-HDA'ya göre popülasyon araçlarının hepsinin aynı olduğu sonucuna varırız. Şekil 2, oluşturulan karar ağacını göstermektedir. Makine tarafından oluşturulan karar ağacının kökü 10-HDA'dır (%). Değerinin 2,59'dan büyük olup olmamasına göre sola veya sağa yaylanır. Bu, arı sütünün yerini tahmin etmede en önemli faktörün 10-HDA (%) olduğu anlamına gelir. Bu parametrenin eşik değeri de 2,59'dur (Şekil 2). Ağaç, arı sütünün Balıkesir'de mi yoksa Konya'da mı hasat edildiğini tek başına 10-HDA (%) değerleri ile tahmin etmenin mümkün olduğunu göstermektedir. %2,59'a bağlı olarak 10-HDA, arı sütünün hasat edildiği yeri tahmin etmede en önemli faktördür. Ağaç değerlendirme istatistikleri Tablo 4'te gösterilmektedir. İstatistiklerin minimum %75 doğrulukla tatmin edici olduğu görülmektedir. Cohen'in kappa'sı da 0,50'den büyüktür, yani 0,73'tür.

Sonuç: Klasik tek yönlü analiz ANOVA, %10-HDA açısından siteler arasında ortalama bir fark olmadığını öne sürse de, makine öğrenimi karar ağacı, %10-HDA'nın siteler arasında ayrım yapması için eşik olarak 2,59 verir. Bu durum arı sütünün Balıkesir veya Konya'da %10-HDA değeri kullanılarak üretilip hasat edildiğini gösterebilir. Bu, arı sütünün orijinalliğini daha kolay tespit etmemize yardımcı olacaktır. Burada %10-HDA değeri arasındaki ilişkiyi ortaya koyabilmek için daha fazla veri içeren ileri çalışmalara ihtiyaç vardır. Çok fazla veriye sahip bir yapay zeka modeli, bu konuda daha

fazla bilgi sağlayabilir ve gelecekte yapay zeka aracılığıyla arı ürünlerinin doğruluğunu belirlemek için kullanılabilir.

INTRODUCTION

People's interest in natural products is increasing rapidly worldwide, increasing the popularity of bee products. Bee products are rare natural products with high biological activities and therapeutic properties. Royal jelly is also one of the important bee products and is a homogeneous, white or yellow colored, sharp smelling, sour taste, water-soluble substance secreted from the hypopharyngeal and mandibular glands of worker bees (Chinou 2014, Krell 1996, Melliou and Schmidt 1997, Tamura et al. 2009, Wytrychowski et al. 2013).

It was determined that the queen bee was constantly fed with royal jelly by Huber, and for this reason, it was named "royal jelly", which means queen meal in English for the first time in Switzerland in 1793 (Crane 1997). While worker bees are fed royal jelly in the first three days of their larval stages, the only food of the queen bee is royal jelly. As a result of feeding honey bee larvae with royal jelly, a productive and long-lived adult queen develops morphological and functional characteristics that other adult individuals in the hive do not have (Krell 1996, Melliou and Chinou 2014, Schmidt 1997, Wytrychowski et al. 2013). Throughout history, human beings have always been affected by this special effect of royal jelly on bees, and for this reason, royal jelly has been used as a food supplement and therapeutic for centuries.

Fresh royal jelly contains water (60-70%), proteins (9-18%), carbohydrates (7-18%), fatty acids and lipids (3-8%), minerals (about 1.5%), and small amounts of polyphenols and vitamins (Melliou and Chinou 2014). On the other hand, lyophilized royal jelly contains <5% water, 27-41% protein, 22-31% carbohydrates, and 15-30% fat (Melliou and Chinou 2014). The composition of royal jelly is similar when compared with different bee colonies and bee breeds (Krell 1996).

The biological activities of royal jelly vary depending on the number of trace elements in its content. Anti-aging (Salazar-Olivo and Paz-Gonzalez 2005), antioxidant (Liu et al. 2008), antibacterial (Fujiwara et al. 1990; Melliou and Chinou 2005), antitumor (Salazar-Olivo and Paz-Gonzalez 2005, Tamura et al. 1987) antihypertensive (Tokunaga et al. 2004), immunomodulatory (Vucevic et al. 2007), anti-inflammatory (Kohno et al. 2004), liver protective (Kanbur et al. 2009), osteoporosis preventative (Hidaka et al. 2006) properties have attracted attention in recent years. In addition, it is known that people who use royal jelly develop a general sense of well-being in a short time and develop higher learning capacity, fatigue resistance, and better memory (Krell, 1996). In other words, royal jelly appears to be a general stimulant that improves the immune system and general body functions (Krell, 1996).

Royal jelly contains short-chain hydroxy fatty acids, which are not found in other foods, and are claimed to have anti-tumor, anti-bacterial and immunoregulatory activity and hormonal activity (Ramadan and Al-Ghamdi 2012, Terada et al. 2011). Royal jelly fatty acids are mostly short-chain fatty acids, such as dicarboxylic acids with 8-10 carbons, which are not found in most plants and animals (Ramadan and Al-Ghamdi 2012). Trans-10-hydroxy-2-decanoic acid (10-HDA), which is considered the major fatty acid of royal jelly, is considered a marker for royal jelly and is found in royal jelly at a rate of 0.5-3.5% of its dry weight (Garcia- Amoedo and Almeida-Muradian 2007, Ramadan and Al-Ghamdi 2012, Kamakura et al. 2001, Kanelis et al. 2015, Liu et al. 2008, Yukunc 2019). However, it is also stated that in a quality royal jelly, the 10-HDA should be at least 1.4 and the moisture content should be between 60% and 70% (Bogdanov and Gallmann 2008).

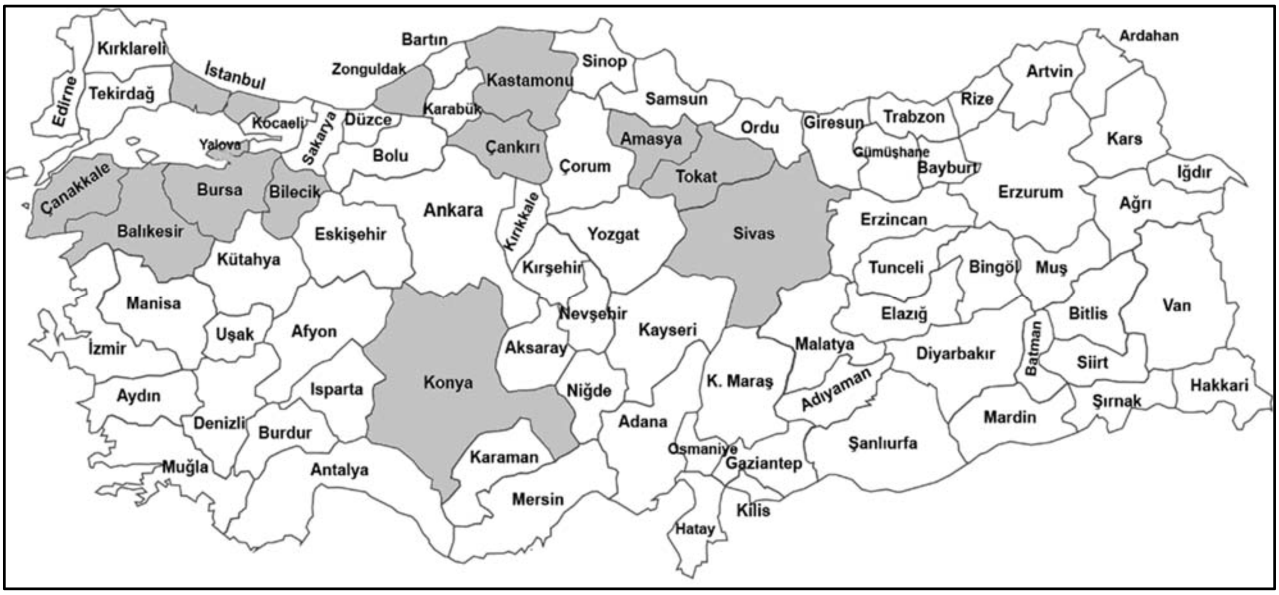
The term machine learning was coined in 1959 by American Arthur Samuel, a pioneer in computer games and artificial intelligence and an IBM employee, and machine learning or machine learning refers to the design and development processes of algorithms that enable computers to learn based on data types such as sensor data or databases. It is a science that deals with the subject. The focus of machine learning research is to give computers the ability to detect complex patterns and make rational decisions based on data. This shows that machine learning is closely related to fields such as statistics, probability theory, data mining, pattern recognition, artificial intelligence, adaptive control, and theoretical computer science (Alpaydin 2020, Arthur 1959, Kohavi and Provost 1998).

This study aims to discover the chemical factors behind the regional differences of royal jelly through machine learning. To achieve this, we collected, compared, and evaluated the results of chemical analysis of royal jelly products harvested in various locations in Turkey.

MATERIAL AND METHODS

Collection and storage of royal jelly samples

A total of 84 royal jelly samples were collected from 13 different provinces of Turkey in 2020 (Figure 1). Collected samples were stored at -18°C until analysis. Then, chemical analyses were made.



Şekil 1. Arı sütü toplanan bölgeler

Figure 1. Royal jelly collected regions

Chemical analysis of samples and data collection

The data used in this study consists of the results of the chemical analysis of royal jelly. The 13 city names shown on the map in Figure 1, are given as target variables to artificial intelligence machine learning algorithms. The chemical values used are moisture, pH, acidity, and 10-hydroxy-2-decanoic acid (10-HDA). Each data was marked with the corresponding analysis.

Moisture analysis

For moisture determination, 1 g of each sample was weighed. It was then incubated at 105°C for 3 hours. At the end of the period, the royal jelly samples were cooled in the desiccator until they reached a constant weight. This process was repeated three times for each sample (Horwitz and Latimer 2000).

pH analysis

The pH and acidity of the royal jelly samples were determined according to ISO/DIS 12824 (2016). 1 g of royal jelly sample was weighed and placed in a 100 mL beaker and 75 mL of boiled and cooled distilled water was added. It was titrated with sodium hydroxide standard solution (c = 0.1 mol/L). The endpoint was obtained when the acidometer showed a pH of 8.3.

Acidity analysis

To obtain the sample's acidity, the milliliter amount of sodium hydroxide standard solution consumed in the titration is multiplied by the concentration value (mol/L), divided by the sample's mass, and multiplied by 100. The acidity of royal jelly is calculated with the following equation:

Acidity [(1 mol/l NaOH) mL/100g] = (V x c x 100) / m

V is the volume of 0.1 mol/L NaOH standard solution consumed in the titration in milliliters. c is the concentration of the NaOH standard solution in mol/L; m is the mass of the sample in grams.

10-Hydroxy-2-decanoic acid (10-HDA) analysis

The amount of 10-HDA was determined by ultraviolet absorbing detector high-performance liquid chromatography (HPLC, VWR Hitachi Chromaster HPLC-UV). Measurements were made at 215 nm to obtain the highest absorbance value (Kim and Lee 2010). The mobile phase of the C-18 (150 x 4.6 mm) column was optimized for methanol, water, and phosphoric acid (55:45:2.2) and the flow rate was 1.0 mL min-1 and the injection volume was 3 µL. Liquid methyl 4-hydroxybenzoate (MHB) was used as a standard. Calibration solutions were obtained by diluting 160 µg mL^{-1} of stock 10-HDA solution (0.1, 0.5, 1.0, 5.0, 10, 20, 40, 80, and 160 µg mL-1). The R^2 value of the obtained calibration graph was determined as 0.9998.

Statistical analysis and machine learning

ANOVA test was performed to see if there was a statistically significant difference between royal jellies in terms of chemical analysis results. One-way ANOVA was performed to test whether there was a significant difference between locations in terms of humidity, 10-HDA (%), pH, and acidity.

H0: There is a significant difference between locations in terms of humidity, 10-HDA (%), pH, and acidity averages.

H1: There is no statistically significant difference between the means.

Then the hidden rules behind the properties of the royal jelly harvested from 13 different locations were extracted using the Decision Tree Machine Learning algorithm.

The purpose of the study analysis is to find hidden rules that will allow the machine to predict the location of the royal jelly using the four chemical analysis values. For this purpose, Gini and Gain (entropy-based) decision tree algorithms were used. Decision trees are widely used in machine learning to enable machine learning and predict class value. They are preferred to "deep learning" and "support vector machines" because they create rules that humans can understand. Decision Trees are among the most common and powerful algorithms used especially in classification and regression tasks. They are used to build models that can predict the value of the target variable based on input properties. Decision trees are created by dividing the data using a hierarchical structure according to different feature values and making predictions based on the resulting datasets. Decision Trees are used in two main types: “Classification Trees” and “Regression Trees”. Classification Trees are used for categorical or discrete target variables, while Regression Trees are used for continuous target variables. Both types of trees are created using a similar process. This process is done by separating the data into smaller subsets based on different property values. There are many pollution calculations for decision trees. Gini and entropy are two of them. Their ability to handle missing data and their ability to be used for feature selection is a great advantage (Schug et al. 2005, Shafer et al. 1996, Silahtaroğlu 2013). For this reason, these algorithms were preferred in the study.

Since the number of royal jellies was not evenly distributed across the 13 locations, SMOTE was used to eliminate the class imbalance. SMOTE is a synthetic data augmentation technique that is widely used in Machine Learning studies (Silahtaroğlu 2009). The dataset is partitioned at 30% / 70%. This means that 70% of the data is used to train the machine and 30% is used to test and evaluate the results. Accuracy, Sensitivity, Specificity, Cronbach Kappa, and F-Value were used to measure how well the machine was trained to estimate the location of royal jelly.

RESULTS

Descriptive statistics of the chemical analysis results of royal jelly are given in Table 1. According to Table 1, the mean chemical analysis results were found as follows, respectively: Humidity 63.05% ±2.99, pH 3.67±0.08, acidity 45.32 ±3.55 and 10-HDA 2.40±0.24.

Table 1. Descripitve statistics of the royal jelly samples

Tablo 1. Arı sütü örneklerinin tanımlayıcı istatistikleri

Parameter	Min.	Mean	Max.	Std. Dev.	Skewness	Kurtosis	No. Missing
Nem (%)	53.29	63.05	69.58	2.99	-0.59	1.18	0
pH	3.50	3.67	3.8	0.08	-0.52	-0.05	0
Asitlik (m.q. g/kg)	38.0	45.32	58	3.55	0.76	1.06	0
10-HDA (%)	1.46	2.40	3.06	0.24	-0.49	2.16	0

As can be seen in Table 2, Levene's test statistics show that equal variances can be assumed for all four parameters. On the other hand, the test statistics in Table 3 show that the null hypothesis is rejected on behalf of H1, excluding 10-HDA, that is, any of the differences between the means are statistically significant in terms of pH, acidity, and humidity. However, with a p-value = 0.296, we fail to reject the null hypothesis and conclude that the population means are all equal in terms of 10-HDA.

Table 2. Test of homogeneity of variances

Tablo 2. Varyansların homojenliği testi

		Levene Statistic	df1	df2	Sig.
Moisture	**Based on Mean**	1.244	9	70	.283
	Based on Median	.825	9	70	.595
	Based on the Median and with adjusted df	.825	9	54.611	.596
	Based on trimmed mean	1.191	9	70	.314
pH	**Based on Mean**	1.050	9	70	.410
	Based on Median	.467	9	70	.892
	Based on the Median and with adjusted df	.467	9	57.491	.891
	Based on trimmed mean	.980	9	70	.464
Acidity	**Based on Mean**	.639	9	70	.760
	Based on Median	.604	9	70	.789
	Based on the Median and with adjusted df	.604	9	53.631	.788
	Based on trimmed mean	.626	9	70	.771
10-HDA	**Based on Mean**	1.871	9	70	.071
	Based on Median	1.141	9	70	.346
	Based on the Median and with adjusted df	1.141	9	39.333	.358
	Based on trimmed mean	1.840	9	70	.076

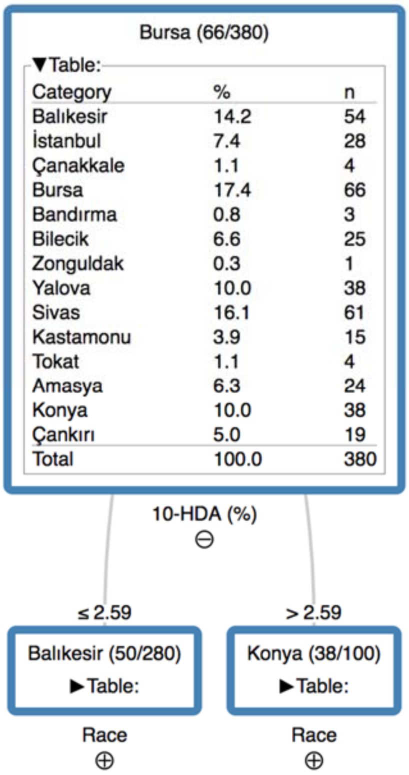
Table 3. ANOVA results
Tablo 3. ANOVA sonuçları

		Sum of Squares	df	Mean Square	F	Sig.
Moisture	**Between Groups**	218.312	13	16.793	2.237	.016
	Within Groups	525.389	70	7.506		
	Total	743.701	83			
pH	**Between Groups**	.179	13	.014	2.550	.006
	Within Groups	.379	70	.005		
	Total	.558	83			
Acidity	**Between Groups**	289.463	13	22.266	2.065	.027
	Within Groups	754.859	70	10.784		
	Total	1044.321	83			
10-HDA	**Between Groups**	.908	13	.070	1.203	.296
	Within Groups	4.065	70	.058		
	Total	4.973	83			

Table 4. Decision tree results
Tablo 4 Karar ağacı sonuçları

Region	True Pos	False Pos	True Neg	False Neg	Recall	Precision	Sensitivity	Specifity	F- Measure
Balıkesir	7.00	0.00	84.00	2.00	0.78	1.00	0.78	1.00	0.88
Tokat	3.00	4.00	81.00	5.00	0.38	0.43	0.38	0.95	0.40
Bursa	6.00	0.00	82.00	5.00	0.55	1.00	0.55	1.00	0.71
Bandırma	1.00	5.00	84.00	3.00	0.25	0.17	0.25	0.94	0.20
Zonguldak	0.00	6.00	83.00	0.00	0.00	0.00	0.00	0.93	NaN
Sivas	6.00	0.00	87.00	0.00	1.00	1.00	1.00	1.00	1.00
Kastamonu	6.00	1.00	86.00	0.00	1.00	0.86	1.00	0.99	0.92
Çanakkale	3.00	4.00	85.00	1.00	0.75	0.43	0.75	0.96	0.55
Konya	7.00	0.00	85.00	1.00	0.88	1.00	0.88	1.00	0.93
Çankırı	5.00	1.00	86.00	1.00	0.83	0.83	0.83	0.99	0.83
Amasya	5.00	2.00	86.00	0.00	1.00	0.71	1.00	0.98	0.83
Bilecik	7.00	0.00	86.00	0.00	1.00	1.00	1.00	1.00	1.00
Yalova	7.00	0.00	85.00	1.00	0.88	1.00	0.88	1.00	0.93
İstanbul	7.00	0.00	86.00	0.00	1.00	1.00	1.00	1.00	1.00

Table 4 and Figure 2 show the machine-generated decision tree results. The root of the machine-generated decision tree is 10-HDA. The tree arcs to the right or left depending on whether the percentage value is greater than 2.59 or not. This means that the most important factor for estimating the location of royal jelly is 10-HDA, and the cut-off point of this parameter is 2.59 (Figure 2). Tree suggests that with 10-HDA values alone it is possible to predict whether royal jelly was harvested in Balikesir or Konya. Based on 2.59%, 10-HDA is the most important factor in estimating the location of royal jelly.



Şekil 2. Karar ağacının kökü

Figure 2. The root of the decision tree

Tree evaluation statistics are shown in Table 5. The statistics appear to be satisfactory with a minimum accuracy of 75%. Cohen's kappa is also greater than 0.50, meaning 0.73.

Table 5. Decision tree evaluation statistics

Tablo 5. Karar ağacı değerlendirme istatistikleri

Precision	Sensivity	Specifity
0.74	0.73	0.98
F-measure	Accuracy	Cohen's kappa
0.78	0.75	0.73

DISCUSSION

It is stated that the composition of royal jelly depends on the season, ecological conditions, and the characteristics of the region where it is collected (Zheng and Hu 2010). In a study conducted with French royal jelly, it was reported that the moisture content varies between 60% and 70%, and the 10-HDA value varies between 1.4% and 3.7% (Wytrychowski et al. 2013). In another study, it was stated that the 10-HDA values of Anatolian royal jelly varied between 1.0% and 3.9%, and the moisture content between 62.6% and 73% (Kolaylı et al. 2015). In addition, according to the international royal jelly standard ISO/DIS 12824 (2016), the 10-HDA value of royal jelly should be between at least 1.4%, its moisture value between at least 62% and a maximum of 68.5%, and its acidity value between 30 and 53 m.q. It is stated that it should be between g / kg. Keskin et al. (2020), it was determined that the 10-HDA content of royal jelly samples varied between 2.1% and 2.6%, and the moisture content varied between 62.6% and 66.5%. The results of all these studies and ours are similar (Table 1).

A study by Yavuz and Gürel (2017) determined that the 10-HDA content of commercial royal jelly ranged between 0.75% and 3.11%, and the moisture content between 63.1% and 73.5%. However, according to the international royal jelly standard (ISO/DIS 12824, 2016), the 10-HDA value of royal jelly should be at least 1.4%. It has been stated that royal jelly containing 10-HDA below this value is likely to be a product with additives or adulteration (Yavuz and Gürel 2017). In this case, the machine learning method can be a guide in determining the accuracy of royal jelly.

As the success of data mining and machine learning algorithms has increased, these data analysis methods and tools have been used in many fields such as agriculture, beekeeping, and quality and yield estimation in recent years. These studies cover both supervised learning (clustering) and unsupervised learning (classification and prediction) models of machine learning and data mining. In one of the studies conducted in the field of beekeeping in recent years, a face-to-face survey was conducted with 62 beekeepers and the decision tree model, which is one of the data mining methods, was used to determine the factors affecting the cooperative membership of beekeepers. As a result of the statistical analysis, of the cooperative members of the beekeepers; It was concluded that credit utilization status, education level, and beekeeping support have a significant effect on the status (Çukur and Çukur 2022). In another study, the decision tree method was used to monitor honey bee health (Edwards-Murphy et al. 2016). Aksoy et al. (2018) conducted a study on the estimation of honey production in beekeeping enterprises located in eastern Turkey. In this study, they aimed to determine the main factors contributing to honey production and to develop a model that can accurately predict honey production in the region by using different data such as colony number, cultivated land area, and plant species. In the study using Multiple Adaptive Regression Splines (MARS), Classification and Regression Tree (CART), and (Chi-squared Automatic Interaction Detector) CHAID algorithms in honey yield per hive, it was determined that MARS (r=0.920) outperformed other approaches. In a study by Karadas and Kadirhanogullari (2017), data mining and artificial neural network algorithms were used to determine the amount of honey production. In another study, a method based on clustering and classification techniques was developed to develop seasonal honey bee models that can be customized and integrated into the computer system for remote monitoring of beehives. In the model, a clustering technique was applied that monitors the brood temperature, relative humidity, and weight of the beehives. Naive Bayes, k-NN, and Random Forest algorithms were compared to propose a high-accuracy classification model (99.67%) that recommends seasonal honey bee patterns for the remote monitoring system (Rafael Braga et al. 2020).

Beekeepers perform inspection procedures that include identifying potential problems or overall colony health in the colonies. If the inspections are not done properly, it may disrupt the microclimate balance in the hive and the work of the worker bees. Classification algorithms were used in the study, which states that these controls can be made and the health status of honey bee colonies can be predicted with high precision (Rafael Braga et al. 2020, Robles-Guerrero et al. 2017).

Data mining and machine learning algorithms are also used to detect fatal diseases of honey bees. In a study, artificial neural networks and Support Vector Machines (SVM) algorithms were used in the detection of Nosema, which is known as a common fatal disease, and disease detection was successfully performed with artificial neural networks with an accuracy of 96.25% (Dghim et al. 2021).

In our study, it has been shown that the location of the harvest and other ecological conditions can be determined by examining the composition of royal jelly. For this purpose, moisture, pH, acidity, and 10-HDA values, which are among the components of royal jelly, were used. The study has developed and introduced a smart system that predicts the harvested location with high accuracy using the composition values of the collected royal jelly samples with the decision tree method, which is one of the machine learning methods. In the results, it was seen that the 10-HDA value from the royal jelly components is effective in determining the geographical location and the harvested location can be determined with more than 75% accuracy by using the 10-HDA value. Thanks to the developed smart system, the location of the received royal jelly sample could be determined within seconds.

Conclusion

Although classical one-way ANOVA suggests no mean difference between locations in terms of 10%-HDA, the machine learning decision tree gives 2.59 as a cut-off point for 10%-HDA in distinguishing locations. This situation can reveal that royal jelly was produced and harvested in Balıkesir or Konya using a 10%-HDA value. Thus, it will contribute to our learning of the originality of royal jelly more easily. Here, more studies with more data are needed to reveal the correlation between the 10%-HDA value. This preliminary study, which we have done on the adaptation of bee products to the machine learning path, is the first study conducted in Turkey. An artificial intelligence model with a larger amount of data can provide more information on this issue and can be used in the future to determine the accuracy of artificial intelligence bee products.

Author contributions: Idea-A.Ö. Analysis and/or interpretation- M.K.; E.Y.Ö.; G.S. Literature review - A.Ö; M.K; G.S. Article writing - A.Ö; M.K; G.S. Critical review -A.Ö.; M.K.; A.E.T.S.; E.Y.Ö.; G.S.

Financial Source: This work was supported by SBS Scientific Bio Solutions Inc. and Bee&You Propolis R&D Center.

Conflict of Interest: There is no conflict of interest between the authors.

Ethics Certificate: Ethics certificate is not required for this study.

Data access: Raw data can be provided upon request.

REFERENCES

Aksoy A, Ertürk YE, Erdoğan S, Eyduran E, Tariq MM. Estimation of honey production in beekeeping enterprises from the eastern part of Turkey through some data mining algorithms. Pak. J. Zool. 2018;50(6):2199-2207, http://dx.doi.org/10.17582/journal.pjz/2018.50.6.2199.2207.

Alpaydin E. Introduction to machine learning (Fourth bas.) MIT. Ss.xix, 2020, 1-3, 13-18.

Arthur S. Some studies in machine learning using the game of checkers. IBM J. Res. Devep. 1959;3(3):210-229, doi:10.1147/rd.33.0210.

Bogdanov S, Gallmann P. Authenticity of Honey and Other Bee Products State of the Art, ALP Science Switzerland, 2008, No.520,p.1-12.

Crane E. The Past and Present Importance of Bee Products to Man, A. Mizrahi, Y. Lensky. (eds.), Bee Products. Springer Science & Business Media New York. 1997,p.1-15.

Çukur T, Çukur F. Determining Factors Affecting Cooperative Membership of the Beekeepers Using Decision Tree Algorithms. Tarim Bilim. Derg. 2022;28(1):25-32, https://doi.org/10.15832/ankutbd.739230

Dghim S, Travieso-González CM, Burget R. Analysis of the Nosema cells identification for microscopic images. Sensors, 2021;21(9):3068, https://doi.org/10.3390/s21093068.

Edwards-Murphy F, Magno M, Whelan PM, O'Halloran J, Popovici EM. b+ WSN: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring. Comput. Electron. Agric. 2016;124:211-219, https://doi.org/10.1016/j.compag.2016.04.008.

Fujiwara S, Imai J, Fujiwara M, Yaeshima T, Kawashima T, Kobayashi K. A potent antibacterial protein in royal jelly. Purification and determination of the primary structure of royalisin. J. Biol. Chem. 1990;265(19):11333-11337.

Garcia-Amoedo LH, Almeida-Muradian LB. Physicochemical composition of pure and adulterated royal jelly. Quím. Nova, 2007;30(2):257-259, https://doi.org/10.1590/S0100-40422007000200002.

Hidaka S, Okamoto Y, Uchiyama S, Nakatsuma A, Hashimoto K, Ohnishi ST, Yamaguchi M. Royal jelly prevents osteoporosis in rats: beneficial effects in ovariectomy model and in bone tissue culture model. Evid. Based Complementary and Altern. Med. 2006;3(3):339-48, doi: 10.1093/ecam/nel019

Horwitz W, Latimer G. Official methods of analysis of AOAC International. Gaithersburg MA, USA: Association of Official Analytical Chemist. 2000.

Kamakura M, Mitani N, Fukuda T, Fukushima M. Antifatigue effect of fresh royal jelly in mice. J. Nutr. Sci. Vitaminol. 2001;47(6):394-401, doi: 10.3177/jnsv.47.394.

Kanbur M, Eraslan G, Beyaz L, Silici S, Liman BC, Altinordulu S, Atasever A. The effects of royal jelly on liver damage induced by paracetamol in mice. Exp. Toxicol. Pathol. 2009;61(2):123-132, doi: 10.1016/j.etp.2008.06.003.

Kanelis D, Tananaki C, Liolios V, Dimou M, Goras G, Rodopoulou MA, Karazafiris E, Thrasyvoulou A. A suggestion for royal jelly specifications. Arh. Hig. Rada. Toksikol. 2015;66(4):275-284, doi: 10.1515/aiht-2015-66-2651.

Karadas K, Kadirhanogullari IH. Predicting honey production using data mining and artificial neural network algorithms in apiculture. Pak. J. Zool. 2017;49(5):1611-1619, DOI:10.17582/journal.pjz/2017.49.5.1611.1619.

Keskin M, Özkök A, Karahalil F, Kolaylı S. What should be the amount of 10-Hydroxi-2-Decanoic Acid (10-HDA) in royal jelly? Mediterr. Agric. Sci. 2020;33(3):347-350, https://doi.org/10.29136/mediterranean.698926.

Kim J, Lee J. Quantitative analysis of trans-10-hydroxy decanoic acid in royal jelly products purchased in the USA by high-performance liquid chromatography. J. Agric. Food Chem. 2010; 54:77-86.

Kohavi R, Provost F. Glossary of terms. Mach. Learn. 1998;20(2-3):271-274.

Kohno K, Okamoto I, Sano O, Arai N, Iwaki K, Ikeda M, Kurimoto M. Royal jelly inhibits the production of proinflammatory cytokines by activated macrophages. J. Biosci. Biotechnol. Biochem. 2004;68(1):138-145, doi: 10.1271/bbb.68.138.

Kolaylı S, Sahin H, Can Z, Yildiz O, Malkoc M, Asadov A. Member of complementary medicinal food: Anatolian royal jellies, their chemical compositions, and antioxidant properties. Evid. Based Complementary and Altern. Med. 2015;21(4):43-48, doi: 10.1177/2156587215618832.

Krell, R. Value-Added Products from Beekeeping, Fao Agricultural Services Bulletin No. 124, Food and Agriculture Organization of the United Nations Rome. 1996.

Liu JR, Yang YC, Shi L, Peng C. Antioxidant properties of royal jelly associated with larval age and time of harvest. J. Agric. Food Chem. 2008;56(23):11447-11452, doi: 10.1021/jf802494e.

Melliou E, Chinou I. Chemistry and bioactivities of royal jelly. Stud. Nat. Prod. Chem., 2014;43: 261-290, https://doi.org/10.1016/B978-0-444-63430-6.00008-4.

Melliou E, Chinou I. Chemistry and bioactivity of royal jelly from Greece. J. Agric. Food Chem. 2005;53(23):8987-8992, doi: 10.1021/jf051550p.

Rafael Braga A, Gomes DG, Freitas BM, Cazier JA. A cluster-classification method for accurate mining of seasonal honey bee patterns. Ecol. Inform. 2020;59:101107, DOI:10.1016/j.ecoinf.2020.101107

Ramadan MF, Al-Ghamdi A. Bioactive compounds and health-promoting properties of royal jelly: A Review. J. Funct. Foods. 2012;4(1):39-52, https://doi.org/10.1016/j.jff.2011.12.007

Robles-Guerrero A, Saucedo-Anaya T, González-Ramérez E, Galván-Tejada CE. Frequency Analysis of Honey Bee Buzz for Automatic Recognition of Health Status: A Preliminary Study. Res. Comput. Sci. 2017;142:89-98, DOI:10.13053/RCS-412-1-9.

Salazar-Olivo LA, Paz-Gonzalez V.Screening of biological activities present in honeybee (*Apis mellifera*) royal jelly. Toxicol. In Vitro, 2005;19(5):645-651, doi: 10.1016/j.tiv.2005.03.001.

Schmidt JO. Bee Products: Chemical Composition and Application, A. Mizrahi&Y. Lensky. (eds.), Bee Products. Springer Science&Business Media New York; 1997, p.15-25.

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. 2005;6(4):1-24.

Shafer J, Agrawal R, Mehta M. SPRINT: A scalable parallel classifier for data mining. VLDB, 1996;96: 544-555.

Silahtaroğlu G. An attribute-center-based decision tree classification algorithm. World Academy of Science, J. Eng. Technol. 2009;56:302-306, https://doi.org/10.5281/zenodo.1084171

Silahtaroğlu G. Veri madenciliği: Kavram ve algoritmaları. Papatya Bilim, 2013, s.300.

Tamura S, Amano S, Kono T, Kondoh J, Yamaguchi K, Kobayashi S. Molecular characteristics and physiological functions of major royal jelly protein 1 oligomer. *Proteomics,* 2009;9:5534-5543.

Tamura T, Fujii A, Kuboyama N. Antitumor effects of royal jelly (RJ). Nihon Yakurigaku Zasshi, 1987; 89(2):73-80.

Terada Y, Narukawa M, Watanabe T. Specific hydroxy fatty acids in royal jelly activate TRPA1. J. Agric. Food Chem. 2011;59(6):2627-2635, DOI: 10.1021/jf1041646

Tokunaga KH, Yoshida C, Suzuki KM, Maruyama H, Futamura Y, Araki Y, Mishima S. Antihypertensive effect of peptides from royal jelly in spontaneously hypertensive rats. Biol. Pharm. Bull. 2004;27(2):189-192, DOI: 10.1248/bpb.27.189

Vucevic D, Melliou E, Vasilijic S, Gasic S, Ivanovski P, Chinou I, Colic M. Fatty acids isolated from royal jelly modulate the dendritic cell-mediated immune response in vitro. Int. Immunopharmacol. 2007;7(9):1211-1220, DOI: 10.1016/j.intimp.2007.05.005

Wytrychowski M, Chenavas S, Daniele G, Casabianca H, Batteau M, Guibert S, Brion B. Physicochemical characterization of French royal jelly: Comparison with commercial royal jellies and royal jellies produced through artificial bee-feeding. J. Food Compos. Anal. 2013;29(2):126-133, DOI: 10.1016/j.jfca.2012.12.002

Yavuz İ, Gürel F. Chemical properties of the royal jellies in Turkish markets. Mediterr. Agric. Sci. 2017.30(3): 281-285.

Yukunc GO. Royal jelly: Proteins and peptides. J. Apit. Nat. 2019;2(2):59-70.

Zheng HQ, Hu FL, Dietemann, V. Changes in the composition of royal jelly harvested at different times: Consequences for quality standards. Apidologie, 2010;2(1):39-47.