



## *Turkish Studies*

*International Periodical for the Languages, Literature and History of Turkish or Turkic*  
Volume 12/11, p. 59-70

DOI Number: <http://dx.doi.org/10.7827/TurkishStudies.11902>  
ISSN: 1308-2140, ANKARA-TURKEY

---

### *Article Info/Makale Bilgisi*

*Referees/Hakemler:* Doç. Dr. Erdinç UZUN – Yrd. Doç. Dr. Heysem KAYA

---

*This article was checked by iThenticate.*

## **HAYVAN HASTALIĞI TEŞHİSİNDE NORMALİZASYON TEKNİKLERİNİN YAPAY SİNİR AĞI VE ÖZELLİK SEÇİM PERFORMANSINA ETKİSİ**

*Pınar CİHAN\* - Oya KALIPSIZ - Erhan GÖKÇE*

### **ÖZET**

Günümüzde, neredeyse her alanda yapay öğrenme yöntemleri kullanılmaktadır. Yapay öğrenme yöntemlerinin veterinerlik alanında kullanımı oldukça yenidir ve veterinerlik alanında yapılmış çalışmalarda başarılı sonuçlar elde edildiği bilinmektedir. Hayvan ve hayvansal üretimdeki azalma ülke ekonomisini kötü yönde etkilemektedir. Hayvancılığa talebi artırmak ve veteriner hekimin işini kolaylaştırmak için veterinerlik alanında yapay öğrenme yöntemlerini kullanma oldukça önem arz etmektedir. Bu çalışmada normalizasyon tekniklerinin yapay sinir ağı performansına etkisini gözlemleyebilmek için veri setindeki değerler minimum-maksimum, ondalık ölçeklendirme, z-skor ve sigmoid normalizasyon teknikleri ile normalize edilmiştir. Daha sonra kuzularda hastalık teşhisi için, ham veri seti ve elde edilen dört veri seti kullanılarak yapay sinir ağı yöntemi ile sınıflandırma yapılmıştır. Çalışma sonucunda, ham veri ile yapılan sınıflandırmadan en başarısız sonuç (f-ölçümü = 0.31) elde edilirken, sigmoid normalizasyon uygulanan veri seti ile en başarılı sonuç (f-ölçümü = 0.36) elde edilmiştir. Ayrıca özellik seçiminde normalizasyonun etkisini gözlemleyebilmek için genetik algoritma tabanlı özellik seçimi yöntemi uygulanmıştır. Elde edilen sonuçlar yine sigmoid normalizasyon tekniğinin daha başarılı olduğunu göstermiştir.

**Anahtar Kelimeler:** Makine öğrenmesi, yapay sinir ağı, veri madenciliği, normalizasyon teknikleri, özellik seçimi, veterinerlik.

---

\* Arş. Gör. Yıldız Teknik Üniversitesi Elektrik Elektronik Fakültesi Bilgisayar Mühendisliği Bölümü, El-mek: pinarkaya3@gmail.com

---

## **EFFECT OF NORMALIZATION TECHNIQUES ON ARTIFICIAL NEURAL NETWORK AND FEATURE SELECTION PERFORMANCE IN ANIMAL DISEASE DIAGNOSIS**

### **ABSTRACT**

Nowadays, machine learning methods are used in almost every field. The use of machine learning methods in the field of veterinary field is quite new and it is known that successful results have been obtained in veterinary studies. The decline in animal and animal production affects the country's economy in the worst way. It is very important to use artificial learning methods in the veterinary field in order to increase the demand for animal husbandry and to facilitate the work of the veterinarian. In this study, the values in the data set were normalized by minimum-maximum, decimal scaling, z-score and sigmoid normalization techniques in order to observe the effect of normalization techniques on artificial neural network performance. Then, for the diagnosis of disease in the lambs, the raw data set and the four obtained data sets were used to classify by artificial neural network method. As a result of the study, the most unsuccessful results were obtained without classifying the raw data set (f-measure = 0.31) and the most successful results were obtained with the sigmoid data set (f-measure = 0.36). In addition, genetic algorithm based feature selection method was applied to observe the effect of normalization in feature selection. The results also showed that the sigmoid normalization technique was more successful.

### **STRUCTURED ABSTRACT**

**Introduction and purpose of the work.** Nowadays, machine learning methods are used in almost every field. The use of machine learning methods in the field of veterinary field is quite new and it is known that successful results have been obtained in veterinary studies. The decline in animal and animal production affects the country's economy in the worst way. It is very important to use artificial learning methods in the veterinary field in order to increase the demand for animal husbandry and to facilitate the work of the veterinarian.

In the field of medicine, artificial learning methods in diagnosing and predicting human health are frequently used. In the field of veterinary medicine, artificial learning methods have recently started to be applied and publications in this area has gained momentum in recent years (Cihan, 2017). The aim of this study is to determine the optimal normalization technique for the data set by observing the effect of normalization techniques on artificial neural network performance in neonatal lamb diagnosis. For this reason, raw data set normalized by minimum-maximum, decimal scaling, z-score and sigmoid normalization techniques, and then classified by artificial neural network method. The accuracy, sensitivity, specificity and f-measure criteria were used to compare the results obtained.

**Material and Method.** Dataset used in this study was collected from two sheep flocks in Kars province, (a city in Northeastern Turkey) in

2009. Comprising 347 lambs in total, the dataset consists of biochemical and physical measurements. The lambs were registered with an individual ear tag number and for each lamb the gender, birth weight, parity and maternal ear tag number and birth type were recorded. In order to observe the effect of normalization techniques on artificial neural network performance, the minimum- maximum, decimal scaling, z-score and sigmoid normalization techniques were used in normalizing the data in this study. In addition, the study was performed using the statistical programming language R.

**Results.** The box-plot graph of the raw data set (not normalized data set) and the box-plot graph after the minimum-maximum normalization were examined. When the box-plot graph of the raw data set is examined; it is seen that the values ranges of IgG, GGT and WG28 parameters are much wider than those of LT, TP, ALB, BW and MDG28. After normalization applied, the value of this difference is reduced so that they get closer to each other has been found that the properties.

In order to investigate the effect of normalization techniques on artificial neural network performance, raw data set and data sets obtained after applying four different normalization techniques, disease classification was performed on the these five data sets. In the study, the artificial neural network classification algorithm was run 50 times.

The most successful classification with 86% accuracy was obtained from the sigmoid normalization applied dataset, and the worst with approximately 84% accuracy was obtained from the z-score dataset. Also, the most successful classification with 0.24 sensitivity, 0.99 specificity and 0.36 f-measure was obtained after sigmoid normalization.

When the results obtained are examined, it is seen that the sensitivity value is generally low. That's why 60 of 347 lambs are sick and 287 of them are healthy. Since 70% of these 60 lambs leave for training and 30% for testing, so the number of sick lambs for training of the model is quite small. In other words, since there are not enough lambs in the model, the model can not obtain very successful results in distinguishing healthy lambs from ill lambs. Thus, the number of healthy lambs is high, so the value of specificity seems to be quite high. It appears that the system has successfully diverted healthy lambs 99% in normalized data sets.

In addition, genetic algorithm based feature selection method was applied to investigate the effect of normalization techniques on feature selection. In the study, the artificial neural network classification algorithm was run 50 times. When the findings obtained after the feature selection were examined, the most successful classification results were obtained from the data set of the sigmoid normalization method. F-measure value increased from 36% to 47% after the selection of properties.

**Conclusion.** In the field of medicine, computer aided diagnosis is used extensively. In this way, both the decisions that will be taken will be supported and the information that the physician does not see is being released. On the other hand, it has been observed that the computer aided diagnosis studies in veterinary fields is very limited. Indeed, in the

Science Direct database; abstracts, titles and keywords of publications were searched for 'artificial neural network' AND 'veterinary', 'artificial neural network' AND 'animal' and 'artificial neural network' AND 'livestock' keywords searched and there was no publications about the disease diagnosis.

The aim of this study is to determine the successful normalization technique for the data set by examining the effect of normalization techniques on neural network and feature selection performance in neonatal lamb diagnostics.

First of all, since the mean and standard deviations of the properties in the data set are very different from each other, they must be subjected to normalization processing. Because normalization makes the distribution of values in the dataset regular. Thus, the performance of the network changes as the input and output of the model are scaled or normalized.

For this reason, minimum-maximum, decimal scaling, z-score and sigmoid normalization techniques have been applied to the raw data set. Then, these four normalized data sets and the raw data set were classified into neonatal lambs using artificial neural network method. The most successful classification with 0.24 sensitivity, 0.99 specificity and 0.36 f-measure was obtained from the data set with sigmoid normalization applied. The most unsuccessful results were obtained from the classification of the raw data set with a sensitivity of 0.23, specificity of 0.97 and f-measure of 0.31.

In addition, genetic algorithms were used to select features in raw and normalized data sets. According to the properties selected by the genetic algorithm, the data sets are again classified by artificial neural network. The results obtained (sensitivity = 0.37, specificity = 0.97, f-measure = 0.47) again showed that the sigmoid normalization technique was more successful. It is also seen that the f-measure value increases from 36% to 47% after the feature selection.

As a result, it has been determined that the most successful normalization technique in diagnosing disease in neonatal lambs is sigmoid. In our previous study, the performance of the k-means clustering method and the normalization techniques were compared and the most successful method was found to be sigmoid (Cihan, Kalipsiz, & Gökçe, 2016).

**Keywords:** Machine learning, artificial neural network, data mining, normalization techniques, feature selection, veterinary.

## Giriş

Yapay öğrenme, bilgisayarların geçmiş bilgilerinden elde edilen tecrübelerden yararlanarak, gelecekteki olayları tahmin etmeye ve modelleme yapmaya imkan veren yalnızca veriler üzerinde uygulanan bir teknik değil, bir yapay zeka alanıdır. Yapay öğrenme metotları sayesinde makineler, insanlığın işgücüne sağladıkları katkıyı insanlığın beyin gücüne de sağlamaya başlamışlardır. Yapay öğrenme metotları, her tür uygulama için çok miktarda verinin analiz edilerek gelecekle ilgili varsayımlar yapmamıza ve kararlar vermemize yardımcı olan yöntemlerdir. Bu nedenle yapay

öğrenme metotlarının önemi ve katkıları her gün daha fazla artmaktadır (Alpaydin, 2014; Amasyalı, 2008; Cihan, 2017; Çentik, 2013).

Günümüzde insan ve hayvan sağlığında yapay öğrenme yöntemleri; hekim kararını desteklemek için, hekimlerin ön göremediği veri içinde gizli kalmış bilgileri açığa çıkarmak için veya sınıf bilgisi bilinmeyen bir örneğin sınıfını tahmin etmek için kullanılmaktadır (Aksoy, 2009; Gülten & Doğan, 2009). Tıp alanında insan sağlığında teşhis ve tahminde yapay öğrenme yöntemlerinden sıklıkla yararlanılmasına karşın hayvan sağlığında teşhis ve tedavide yapay öğrenme yöntemlerinden yetime yararlanılmamaktadır.

Science Direct veri tabanında, 2007-2017 tarihleri arasındaki yayınların özet, başlık ve anahtar kelimelerinde ‘artificial neural network’ AND ‘veterinary’, ‘artificial neural network’ AND ‘animal’ ve ‘artificial neural network’ AND ‘livestock’ anahtar kelimeleri aranmış olup yapay sinir ağı ile hayvan hastalıkları teşhisi çalışmalarıyla karşılaşılmalıdır. Bunun yanında veterinerlik alanında yapay sinir ağları kullanılarak farklı çalışmalar yapılmış olup bunlar da sınırlı sayıdadır. Küçükönder ve ark.(Küçükönder, Üçkardeş, & Narinç, 2014), yaptıkları çalışmada sınıflandırma algoritmalarından YSA, RBF Network, Naive Bayes, KStar, ve Ridor yöntemlerini kullanarak Japon bildircını yumurtalarının döllülük üzerine etkisi olan mevsim, seleksiyon ve yerleşim sıklığı faktörlerine göre sınıflandırmayı ve bu faktörlerin etkisini belirlemeye çalışmışlardır. Ghotoorlar ve ark.(Ghotoorlar, Ghamsari, Nowrouzian, Ghotoorlar, & Ghidary, 2012), yapay sinir ağı kullanılarak öznel olarak hesaplanan topallama puanına uygun bir otomatik puan hesaplama sistemi geliştirmeye çalışmışlardır. Takma ve ark.(Takma, Atıl, & Aksakal, 2012), çoklu regresyon ve yapay sinir ağı modellerini kullanarak siyah alaca ineklerin laktasyon süt verimleri üzerine laktasyon süt verimine laktasyon süresi, buzağılama yılı ve servis periyodu etkilerini modellemeye çalışmışlardır.

Yapay sinir ağları, birçok nörondan oluşan paralel ve dağıtık bilgi işleme yapısındadır. Yani yapay sinir ağları, beyin hücreleri olan nöronların çalışma prensibini modelleyen ve öğrenebilen algoritmalarıdır. Yapay sinir ağının yapısı kurulduktan sonra sinir ağacı eğitilir. Daha sonra giriş verisine karşılık çıkış verisi alınır. Elde edilen bu değer ile gerçek değer arasındaki hata miktarına göre ağın içerisindeki nöron fonksiyonlarının ayarlanması yapılır. Bu şekilde elde edilen veriler ağına verilerek, ağın eğitilmesi sağlanır.

Yapay sinir ağlarında, modelin girdi ve çıktılarının ölçeklendirilmesi yani normalize edilmesi ağın performansını yakından etkilemektedir. Ağına girilen bütün değerlerin belirli bir aralıkta ölçeklendirilmesi, hem birbirinden farklı ortalama ve standart sapmaya sahip özelliklerin aynı ölçek üzerine indirgenmesine hem de yanlış girilen çok küçük veya çok büyük değerlerin etkisinin ortadan kalkmasını sağlar. Çünkü normalizasyon ile veri setindeki değerlerin dağılımı düzenli bir hale getirmiş olur. Literatürde birçok veri normalizasyon çeşidi bulunmaktadır. Bunlar; minimum-maksimum, ondalık ölçeklendirme, Ortanca, z-skor ve sigmoid gibi sıralanabilir (Jayalakshmi & Santhakumaran, 2011).

Veterinerlik alanına, yapay öğrenme yöntemleri yeni yeni uygulanmaya başlanmış olup, bu alanda yapılmış çalışmalar son yıllarda ivme kazanmıştır (Cihan, 2017). Bu çalışmanın amacı, neonatal kuzularda hastalık teşhisinde normalizasyon tekniklerinin yapay sinir ağı performansına etkisini gözlemlemek, veri seti için en uygun normalizasyon tekniğini belirlemektir. Bu nedenle ham veri setindeki veriler minimum-maksimum, ondalık ölçekleme, z-skor ve sigmoid normalizasyon teknikleri ile normalize edildikten sonra yapay sinir ağı yöntemi ile sınıflandırma yapılmıştır. Elde edilen sonuçları karşılaştırabilmek için ise doğruluk, duyarlılık, seçicilik ve f-ölçümü kriterleri kullanılmıştır.

### Materyal ve Metot

Normalizasyon tekniklerinin yapay sinir ağı performansına etkisini gözlemleyebilmek için bu çalışmada verilerin normalleştirilmesinde minimum-maksimum, ondalık ölçekleme, z-skor ve sigmoid normalizasyon teknikleri kullanılmıştır. Ayrıca çalışma R istatistiksel programlama dili kullanılarak gerçekleştirilmiştir.

**Minimum-maksimum metodu:** Bu yöntem, verileri doğrusal olarak normalize eder. Herhangi iki değer arasındaki fark ya da değerlerdeki büyüklüklerin orantısallığı değişmez. Minimum bir verinin alabileceği en düşük değer iken, maksimum verinin alabileceği en yüksek değeri ifade eder. Veri genellikle 0-1 aralığındadır. Minimum-maksimum modelinde değerler Eşitlik 1'den hesaplanabilir. Minimum-maksimum normalizasyonu için R'da 'mnorm' yöntemi kullanılmıştır.

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

$x'$ : Normalize edilmiş veri

$x_i$ : Girdi değeri

$x_{max}$ : Girdideki en küçük sayı

$x_{min}$ : Girdideki en büyük sayı

**Ondalık ölçekleme metodu:** Ele alınan değişkenin değerlerinin ondalık kısmı hareket ettirilerek normalleştirme gerçekleştirilir. Hareket edecek ondalık nokta sayısı, değişkenin maksimum mutlak değerine bağlıdır. Sonuçta aralık -1 ve +1 arasındadır. Ondalık ölçekleme modelinde değerler Eşitlik 2'den hesaplanabilir. Ondalık ölçekleme normalizasyonu için R'da 'dscale' yöntemi kullanılmıştır.

$$x' = \frac{x_i}{10^j} \quad (2)$$

$j$ :  $\text{Max}(|x'|) < 1$  olacak şekildeki en küçük tam sayı

**Z-skor metodu:** Z Skor normalleştirmede değişkenin her hangi bir y değeri, değişkenin ortalaması ve standart sapmasına bağlı olarak bilinen Z dönüşümü ile normalleştirilir. Z-skor, her ne kadar eksi sonsuz ile artı sonsuz arasında tanımlı olsada dağılım genellikle -1.5 ile +1.5 arasında gerçekleşir. Z-skor modelinde değerler Eşitlik 3'den hesaplanabilir. Z-skor normalizasyonu için R'da 'znorm' yöntemi kullanılmıştır.

$$x' = \frac{x_i - \mu_i}{\sigma_i} \quad (3)$$

$x'$ : Normalize edilmiş veri

$x_i$ : Girdi(gözlem) değeri

$\mu_i$ : Girdi setinin ortalaması

$\sigma_i$ : Girdi setinin standart sapması

**Sigmoid normalizasyonu metodu:** Sigmoid normalizasyonu fonksiyonu verileri 0 ile 1 veya -1 ile 1 arasında sınıflandırır. Bir kaç tane doğrusal olmayan sigmoid fonksiyon çeşidi vardır. Sigmoid modelinde değerler Eşitlik 4'den hesaplanabilir. Sigmoid normalizasyonu için R'da 'signorm' yöntemi kullanılmıştır.

$$x' = \frac{1}{1+e^{-x_i}} \quad (4)$$

$x'$ : Normalize edilmiş veri

$x_i$ : Girdi değeri

$e$ : Doğal logaritma değeridir

### Materyal

Veri 2009 yılında Kuzeydoğu Anadolu bölgesinde Kars ilinde bulunan iki koyun çiftliğinden toplanmıştır. Veri toplama çalışması 301 Akkaraman melezi koyun ve bunlardan doğan 347 kuzu üzerinde yürütülmüştür. Tüm koyun ve kuzular aynı beslenme şekli ve işletme koşullarında tutulmuştur. Kuzeydoğu Anadolu bölgesindeki sürülerde kuzulama sezonu tipik olarak kış (Aralık-Şubat) veya bahar (Mart-Mayıs)'dır. Doğumda anne ve kuzulara kulak küpesi uygulanarak hastalık ve diğer bilgiler bu numara üzerinden takip edilmiştir. Doğumda kuzuların cinsiyet, doğum tarihi, doğum ağırlığı, anne doğum şekli (güç doğum ve normal) ve doğan kuzu sayısı (ikiz/tek) kayıt altına alınmıştır. Kuzular doğumda (kolostrum almadan önce) bir baskül yardımıyla tartılmıştır [CASIA DB2-150 kg ( $\pm 30$ gr)]. Daha sonra kuzular doğal yolla emzirilmiştir. Yeni doğan kuzular ilk dört hafta boyunca anneleri ile birlikte tutulmuştur. Bu dönemden sonra üç ay boyunca günde iki kere (sabah ve akşam) emzirilmesine izin verilmiştir. Kuzuların sağlık muayeneleri neonatal periyotta (yaşamın ilk 1 ayı) günlük, sonraki 8 haftalık periyotta (post-neonatal periyot) 2 günde bir yapılan ziyaretlerle ilgili alandaki araştırmacılar tarafından yapılarak hastalıklar kayıt altına alınmıştır. Kuzuların yaşamlarının 28 (neonatal periyodun bittiği gün) ve 84. gündeki canlı ağırlıkları tartılarak kayıt altına alınmıştır. Kan örnekleri doğumdan sonra  $24 \pm 1$  saat sonra alınmıştır. Kan örnekleri 3000 rpm'de 5 dakika santrifüj edildikten sonra serum örnekleri elde edilerek  $-20^\circ$  de saklanmıştır. Serum ve kolostral IgG konsantrasyonları ELISA yöntemi kullanılarak ölçülmüştür. Aynı şekilde serum ve kolostrumda total protein, GGT ve ALB analizleri ticari spektrofotometrik kitler kullanılarak ölçülmüştür.

Veri setinde kuzuların neonatal (doğumdan sonraki 0-28. gün), post-neonatal (doğumdan sonraki 28-84. gün) ve bu iki dönemin tamamı (doğumdan sonraki 0-84.gün) için toplanmış toplamda 42 özellik bulunmaktadır. Neonatal dönemde kuzunun sağlık durumunda (sağlıklı / hasta) kan parametrelerinin önemli etkisi olup post-neonatal dönemde fiziksel koşullar devreye girmektedir(Gokce, Atakisi, Kirmizigul, Unver, & Erdogan, 2014). Bu nedenle çalışmada kuzunun neonatal dönemdeki sağlık durumunu etkileyebilecek faktörler olan IgG, GGT, LT, TP, ALB, BW, WG28, MDG28, AH, twin, gender, age, parity, farm ve kuzunun sağlık durumunun tutulduğu özellikler kullanılmıştır. Çalışmada kullanılan özellikler ve bu özelliklere ait sonuçlar Tablo 1'de verilmiştir.

### Bulgular

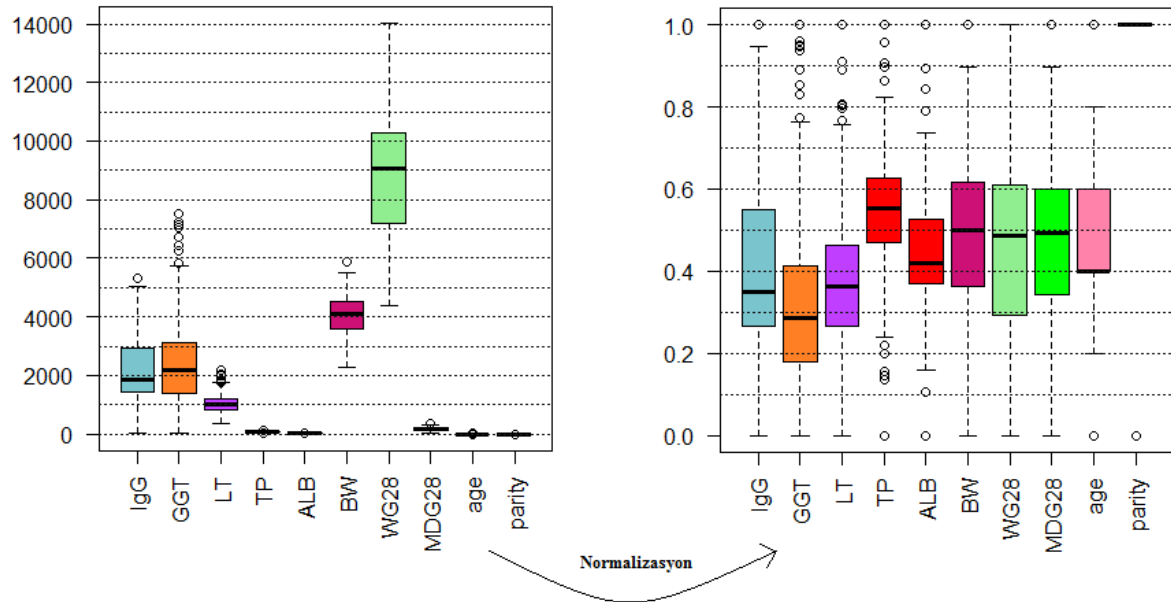
Çalışmada kullanılan veri setindeki özellikler ve bu özelliklere ait detaylı bilgiler Tablo 1'de verilmiştir.

**Tablo 1.** Veri setindeki özellikler ve bu özelliklere ait sonuçlar

<i>Nümerik değişkenler</i>				
Özellikler	Kısaltma	Min	Max	Ortalama
Immunoglobulin G	IgG	19	5302	2196
Gamma Glutamil Trasnfereraz	GGT	38	7517	2382
Laktoferrin	LT	354	2194	1052
Total Protein	TP	21	117	73
Albumin	ALB	32	51	40
Doğum ağırlığı	BW	2260	5900	4028
28. gün sonundaki vücut ağırlığı	WG28	4364	14016	8804
Ortalama günlük kilo kazanımı	MDG28	14	340	169
Anne yaşı	Age	1	6	3
Anne doğum sayısı	Parity	1	5	2
<i>Kategorik değişkenler</i>				
Özellikler	Kısaltma	0	1	
Anne sağlık durumu	AH	Hasta n=17	Sağlıklı n=330	
Doğum tipi	Twin	İkiz n=92	Tek n=255	
Cinsiyet	Gender	Erkek n=156	Dişi n=191	
Çiftlik	farm	1. işletme n=30	2. işletme n=317	

Tablo 1 incelendiğinde numerik özelliklerin aldıkları değer aralıklarının birbirinden çok farklı olduğu görülmektedir. Örneğin IgG kan seviyesinin en küçük değeri 19 iken en büyük değeri 5302, 28. gün sonundaki vücut ağırlığının en küçük değeri 4364 iken en büyük değeri 14016 veya anne yaşının en küçük değeri 1 iken en büyük değerinin 6 olduğu görülmektedir.

Şekil 2’de ise Veri setindeki özelliklerin ve bu özelliklerin normalize edildikten sonraki değerlerine ait kutu grafiği verilmiştir.

**Şekil 1.** Normalize edilmemiş ve normalize edilmiş veri kutu grafiği

Kutu grafiği ile sürekli değişkenin hem büyüklüğünü, hem de dağılımını incelememiz mümkündür. Şekil 1’de solda normalize edilmemiş yani ham veri setindeki özelliklere ait kutu grafiği, sağda ise veri setine minimum maksimum normalizasyon uygulandıktan sonraki kutu

### Turkish Studies



grafliğini göstermektedir. Ham verinin kutu grafiği incelendiğinde IgG, GGT ve WG28 parametrelerinin değer aralıklarının LT, TP, ALB, BW ve MDG28 özelliklerinin aldıkları değer aralığından çok daha geniş olduğu görülmektedir. Normalizasyon uygulandıktan sonra ise bu farkın azaldığı yani özelliklerin aldıkları değerlerin birbirlerine daha yaklaştığı görülmektedir.

Değişkenlerin ortalama ve varyansları birbirlerinden önemli ölçüde farklı olduğunda büyük ortalama ve varyansa sahip değişkenlerin diğerleri üzerindeki baskısı daha fazla olur ve onların rollerini önemli ölçüde azaltır. Ham verideki özelliklerin ortalamalarının birbirinden çok uzak değerlere sahip oldukları ancak normalizasyon uygulanmış veri setlerindeki özelliklerin ortalamalarının birbirine çok yakın değerler olduğu görülmektedir. Özellikler uzaklık ölçüsü alan yöntemlerde normalize edilmiş veri kullanımı daha doğru sonuçlar üretilmesini sağlayacaktır.

Normalizasyon tekniklerinin yapay sinir ağı performansına etkisini inceleyebilmek için ham veri seti ve dört farklı normalizasyon tekniği uygulandıktan sonra elde edilen veri setleri üzerinde hastalık sınıflandırması yapılmıştır. Çalışmada yapay sinir ağı sınıflandırma algoritması 50 kere çalıştırılmış olup elde edilen ortalama sonuçlar Tablo 2’de sunulmuştur.

**Tablo 2.** Sınıflandırma sonuçları

	<b>Doğruluk (Accuracy)</b>	<b>Duyarlılık (Recall)</b>	<b>Seçicilik (Specificity)</b>	<b>F-Ö (F-Measure)</b>	<b>Mcnemar's P-Value</b>
<i>Ham veri</i>	0,85	0,23	0,97	0,31	0,128
<i>Min max</i>	0,88	0,23	0,99	0,35	0,043
<i>Ondalık</i>	0,85	0,22	0,99	0,33	0,019
<i>Z-skor</i>	0,84	0,21	0,99	0,32	0,026
<i>Sigmoid</i>	0,86	0,24	0,99	0,36	0,024

Doğruluk, gerçekte testin hasta ve sağlam olarak toplam doğru tanı koyma oranını verir. %86 doğruluk oranı ile en başarılı sınıflandırma sigmoid normalizasyon uygulanmış veri setinden elde edilirken yaklaşık %84 doğruluk oranı ile en kötü sonuç z-skor veri setinden elde edilmiştir.

Duyarlılık, gerçekte hasta olan kuzuların ne kadarının sistem tarafından doğru tahmin edildiğini verir. Seçicilik, gerçekte sağlam olan kuzuların ne kadarının sistem tarafından doğru tahmin edildiğini verir. F-Ölçümü, tek başına kesinlik veya duyarlılığın hesaplanması sistemin başarısını değerlendirmede eksik kalabilmektedir. Bu nedenle kesinlik (precision) ve duyarlılığın harmonik ortalaması olan F-Ölçümü kullanılır.

Elde edilen sonuçlara göre 0.24 duyarlılık, 0.99 seçicilik ve 0.36 f-ölçümü ile en başarılı sınıflandırma sigmoid normalizasyon uygulandıktan sonra elde edilmiştir. En başarısız sonuçlar ise 0.23 duyarlılık, 0.97 seçicilik ve 0.31 f-ölçümü oranı ile ham veri setinin sınıflandırmasından elde edilmiştir.

Elde edilen sonuçlar incelendiğinde genel olarak duyarlılık değerinin düşük olduğu görülmektedir. Bunun sebebi 347 kuzudan 60 tanesinin hasta 287 tanesinin sağlıklı olmasıdır. Bu 60 kuzunun da %70’i eğitim, %30’u test için ayrılınca modelin eğitimi için hasta kuzu sayısı oldukça az sayıda olmaktadır. Yani yeterli sayıda hasta kuzu olmadığından model hasta kuzular ile sağlıklı kuzuları ayırmada çok başarılı sonuçlar elde edememektedir. Nitekim sağlıklı kuzu sayısı fazla olduğu için öğrenme işlemi de başarılı bir şekilde gerçekleştiği elde edilen sonuçlardan anlaşılmaktadır. Sistemin normalize edilmiş veri setlerinde sağlıklı kuzuları %99 başarılı bir şekilde ayırttığı görülmektedir.

Ayrıca genetik algoritma tabanlı özellik seçimi yöntemi uygulanarak normalizasyon tekniklerinin özellik seçimi üzerindeki etkisi incelenmiştir. Çalışmada yapay sinir ağı sınıflandırma algoritması 50 kere çalıştırılmış olup elde edilen ortalama sonuçlar Tablo 3’de sunulmuştur.

**Tablo 3.** Özellik seçimi sonrası sınıflandırma sonuçları

	Seçilen Özellik	Doğruluk (Accuracy)	Duyarlılık (Recall)	Seçicilik (Specificity)	F-Ö (F-Measure)	McNemar's P-Değeri
Ham veri	GGT, LT, ALB, WG28, MDG28, AH, gender, parity	0,85	0,23	0,98	0,32	0,058
Min max	IgG, BW, MDG28, AH, gender, age, parity	0,86	0,33	0,98	0,44	0,146
Ondalık	IgG, BW, WG28, AH, gender, age, parity, farm	0,86	0,35	0,97	0,45	0,138
Z-skor	IgG, BW, MDG28, AH, gender, age, parity, farm	0,86	0,35	0,96	0,45	0,186
Sigmoid	IgG, BW, MDG28, AH, gender, age, parity	0,87	0,36	0,97	0,47	0,043

Özellik seçimi sonrası yapılan sınıflandırma sonucunda elde edilen bulgular incelendiğinde yine en başarılı sınıflandırma sonuçlarının sigmoid normalizasyon yöntemi uygulanan veri setinden elde edilmiştir.

### Teşekkür

Bu çalışmada kullanılan veri seti TÜBİTAK projesi (Proje Kodu: TOVAG 108 O 847) kapsamında toplanmıştır.

### Sonuç

Tıp alanında, bilgisayar destekli tanı çalışmaları yoğun olarak kullanılmaktadır. Bu sayede hem hekime alacağı kararlarda destek sağlanmakta hem de hekimin ön görmediği bilgiler açığa çıkarılmaktadır. Buna karşın veterinerlik alanında hayvanlarda bilgisayar destekli tanı çalışmalarının oldukça sınırlı olduğu gözlemlenmiştir. Nitekim Science Direct veri tabanında, 2007-2017 tarihleri arasındaki yayınların özet, başlık ve anahtar kelimelerinde ‘artificial neural network’ AND ‘veterinary’, ‘artificial neural network’ AND ‘animal’ ve ‘artificial neural network’ AND ‘livestock’ kelimeleri aranmış olup hastalık teşhisi ile ilgili yapılmış çalışmayla karşılaşılmamıştır.

Bu çalışmada, neonatal kuzularda hastalık teşhisinde normalizasyon tekniklerinin yapay sinir ağı performansına etkisini gözlemlemek ve veri seti için başarılı normalizasyon tekniğini belirlemek amaçlanmıştır.

Veri setindeki özelliklerin ortalama ve standart sapmaları birbirlerinden çok farklı oldukları için normalizasyon işlemine tabi tutulmalıdırlar. Çünkü normalizasyon, veri setindeki değerlerin dağılımını düzenli hale getirmektedir. Böylelikle modelin girdi ve çıktılarının ölçeklendirilmesiyle yani normalize edilmesiyle ağın performansı değişmektedir.

### Turkish Studies

Ham veri setine minimum-maksimum, ondalık ölçeklendirme, z-skor ve sigmoid normalizasyon teknikleri uygulanmıştır. Daha sonra normalize edilmiş bu dört veri seti ve ham veri seti yapay sinir ağı yöntemi kullanılarak neonatal kuzularda hastalık sınıflandırılması yapılmıştır. Çalışma sonucunda 0.24 duyarlılık, 0.99 seçicilik ve 0.36 f-ölçümü ile en başarılı sınıflandırma sigmoid normalizasyon uygulanan veri setinden elde edilmiştir. En başarısız sonuçlar ise 0.23 duyarlılık, 0.97 seçicilik ve 0.31 f-ölçümü oranı ile ham veri setinin sınıflandırmasından elde edilmiştir.

Ayrıca ham ve normalize edilmiş veri setlerinde genetik algoritmalar ile özellik seçimi yapılmıştır. Genetik algoritmalarının seçtiği özelliklere göre, veri setleri tekrar yapay sinir ağı ile sınıflandırılmıştır. Elde edilen sonuçlar (duyarlılık = 0.37, seçicilik=0.97, f-ölçümü=0.47) yine sigmoid normalizasyon tekniğinin daha başarılı olduğunu göstermiştir.

Sonuç olarak, neonatal kuzularda hastalık teşhisinde en başarılı normalizasyon tekniğinin sigmoid olduğu tespit edilmiştir. Nitekim yapılan çalışmada k-ortalamlar kümeleme yöntemi ile normalizasyon tekniklerinin performansları kıyaslanmış olup en başarılı yöntemin sigmoid olduğu bildirilmiştir (Cihan, Kalipsiz, & Gökçe, 2016).

#### KAYNAKÇA

- Aksoy B. (2009). Cluster analysis of decompression illness. Yüksek Lisans Tezi, Galatasaray Üniversitesi, Fen Bilimleri Enstitüsü.
- Alpaydin E. (2014). Introduction to machine learning. MIT press.
- Amasyalı, F. (2008). Yeni makine öğrenmesi metodları ve ilaç tasarımına uygulamaları. Doktora Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü.
- Cihan P., Gökçe E., Kalipsiz O. (2017). A Review of Machine Learning Applications in Veterinary Field. Kafkas Üniversitesi Veteriner Fakültesi Dergisi, 23(4), 673–680. <http://doi.org/10.9775/kvfd.2016.17281>
- Cihan P., Kalipsiz O., Gökçe E. (2016). On the Application of Data Pre-processing Techniques in Data Mining to Livestocks Data. International Conference on Computer Science and Engineering, Tekirdağ, Turkey, 20-23 October, 577-585.
- Çentik G. (2013). Makine öğrenmesi yöntemlerinin polisomnografik verilere uygulanması. Doktora Tezi, Trakya Üniversitesi, Fen Bilimleri Enstitüsü.
- Ghotoorlar S.M., Ghamsari, S.M., Nowrouzian I., Ghotoorlar S.M., Ghidary S.S. (2012). Lameness scoring system for dairy cows using force plates and artificial intelligence. Veterinary Record, 170(5), 126–126. <http://doi.org/10.1136/vr.100429>
- Gokce E., Atakisi O., Kirmizigul A. H., Unver A., Erdogan H. M. (2014). Passive immunity in lambs: Serum lactoferrin concentrations as a predictor of IgG concentration and its relation to health status from birth to 12 weeks of life. Small Ruminant Research, 116(2-3), 219–228. <http://doi.org/10.1016/j.smallrumres.2013.11.006>
- Gülten A., Doğan Ş. (2009). Genetik algoritmalar yönteminin biyomedikal verileri üzerindeki uygulamaları. DAUM, 7(1), 5.
- Jayalakshmi T., Santhakumaran A. (2011). Statistical normalization and back propagation for classification. International Journal of Computer Theory and Engineering, 3(1). Retrieved from <http://www.ijcte.org/papers/288-L052.pdf>

- 
- Küçükönder H., Üçkardeş F., Nariç D. (2014). A Data Mining Application in Animal Breeding: Determination of Some Factors in Japanese Quail Eggs Affecting Fertility. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 20(6), 903–908. <http://doi.org/10.9775/kvfd.2014.11353>
- Takma Ç., Atıl H., Aksakal V. (2012). Çoklu Doğrusal Regresyon ve Yapay Sinir Ağı Modellerinin Laktasyon Süt Verimlerine Uyum Yeteneklerinin Karşılaştırılması. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 18(6), 941–944. <http://doi.org/10.9775/kvfd.2012.6764>