

KAN VERMEYE ELVERİŞLİ DONÖRLERİN BİLGİSAYAR DESTEKLİ TESPİTİ

Kerim KARADAĞ^{1*}

¹Harran Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 63000, Şanlıurfa, Türkiye
Geliş Tarihi/Received Date: 10.09.2021 Kabul Tarihi/Accepted Date: 24.10.2021 DOI: 10.54365/adyumbd.993772

ÖZET

İnsanlar hayatları boyunca farklı zamanlarda çeşitli sebeplerden dolayı kan ve kan ürünlerine gerek duymaktadırlar. Kan ve kandan elde edilen ürünler laboratuvar koşullarında elde edilemeyip yalnızca insandan elde edildiğinden, hayati önem taşıyan kanın sağlıklı bireylerden sağlanması gerekmektedir. Karaciğer rahatsızlıkları ile ilgili, Hepatit çeşitleri, karaciğer kanseri, karaciğer büyümesi, karaciğer yetmezliği, siroz ve benzeri hastalıklar kanın durumunu belirleyen etmenlerin başında gelir. Bu hastalıklar ilerledikçe telafisi mümkün olmayan zararlara yol açabileceği için erken teşhis edilip tedavi edilmesi ve kan verilecek kişilere de bulaş riskini ortadan kaldırmak için hayati önem arz etmektedir. Son yıllarda birçok alanda kullanılan bilgisayar destekli öğrenme yöntemleri, bu alanda da geleneksel yöntemlere ilaveten hekimlere fikir verici olarak kullanılmaktadır. Bu çalışmada amaç, makine öğrenme yöntemleri ile karaciğer hastalıklarının tespiti ve donörlerin kan bağıışı yapmaya elverişli olup olmadıklarını kan değerlerinden tespit etmektir. Sınıflandırma için, Karar Ağaçları, Destek Vektör Makinesi ve k-en yakın komşuluk algoritmaları kullanılıp doğruluk ve gerçek oran performans ölçütleri ile değerlendirilmiştir.

***Anahtar Kelimeler:** Kan Bağıışı, Karaciğer Hastalıkları, Makine Öğrenme Algoritmaları, Sınıflandırma, Performans Ölçütü*

COMPUTER AIDED DETERMINATION OF SUITABLE BLOOD DONORS

ABSTRACT

People need blood and blood products for various reasons at different times throughout their lives. Since blood and products obtained from blood cannot be obtained in laboratory conditions, but only from humans, vital blood must be obtained from healthy individuals. Hepatitis types, liver cancer, liver enlargement, liver failure, cirrhosis and similar diseases are among the factors that determine the state of the blood. Since these diseases can cause irreparable harm as they progress, it is vital to diagnose and treat early and to eliminate the risk of transmission to those who will be given blood. Computer-aided learning methods, which have been used in many fields in recent years, are also used in this field as an idea for doctors in addition to traditional methods. The aim of this study is to detect liver diseases with machine learning methods and to determine whether donors are suitable for blood donation from blood values. For classification, Decision Trees, Support Vector Machine and k-nearest neighbor algorithms were used and evaluated with accuracy and true rate performance measures.

***Keywords:** Blood Donation, Liver Diseases, Machine Learning Algorithms, Performance Measure, Classification*

1. Giriş

Kan bağıışı, kişileri ruhen ve bedenen olumlu etkiler. Kan bağıışı sonrası doku ve organlar uyarılarak, kan hücreleri yenilenmiş olur ve kandaki yüksek yağ oranı düşer. Ayrıca bağıışçılar, her bağıış esnasında Hepatit, HIV gibi farklı enfeksiyon etkenlerinin de taramasından geçmiş olur [1]. İlk kan nakli

*1 e-mail: k.karadag@harran.edu.tr ORCID ID: <https://orcid.org/0000-0001-5167-4054>

hayvandan hayvana 17. yy da gerçekleştirilmiş, insandan insana kan nakli ise 19. yy da gerçekleştirilmiştir. Kan transferine; kazalar, lösemi, anemi, cerrahi müdahaleler gibi rahatsızlıklarda tedavi edici başka seçeneğin olmadığı durumlarda ihtiyaç duyulmaktadır [2]. İnsanlar hayatları boyunca farklı zamanlarda çeşitli sebeplerden dolayı kan ve kan ürünlerine gerek duymaktadırlar. Kan ve kandan elde edilen ürünler laboratuvar koşullarında elde edilemeyip yalnızca insandan elde edildiğinden, hayati önem taşıyan kanın sağlıklı bireylerden sağlanması gerekmektedir [2]. İnsan vücudundaki birçok organ karmaşık ve bir o kadarda farklı görevlere sahiptir. Bu organlardan biri olan karaciğer; kanı, birçok yabancı ve toksik maddelerden arındırma, vücuttaki yağları sindirme, atıkları vücuttan uzaklaştırma ve safra üretme gibi hayati fonksiyonları yerine getirme gibi görevlere sahiptir [3]. Kan bağışını gerekliliği, uygun donörün seçimi ve bilgisayar destekli karar verme yöntemlerinin kullanıldığı çalışmalar incelendiğinde, Masser ve ark., yaptıkları çalışmalarında kan bağışçılarının karar vermesi üzerindeki etkilerin ele alınmasını, bağışçı toplama ve elde tutma yaklaşımına nasıl fayda sağlayabileceğini araştırmışlar [4]. Altındiş ve ark., yaptıkları çalışmada kan verme önündeki engelleri ve bu konudaki tutum-davranışlarını belirlemek ve kan bağışı ile ilgili farkındalığı artırmak için hangi aşamalarda eksiklikler olduğunu saptamayı amaçlamışlar [1]. Czeizler ve Garbarino yaptıkları çalışmalarında, kan bağışı kararının zamansal kurgusal çerçevesi ile kan bağışı talebi arasındaki eşleşmenin bir uyumsuzluktan daha yüksek bağış niyetlerine yol açıp açmadığını test ederek yorum teorisini genişletmişler [5]. Khan ve ark., Hepatit C hastalığını makine öğrenme yöntemleri kullanarak tahmin etmeye çalışmışlar [6]. Mueller-Breckenridge ve ark., yaptıkları çalışmada, HBeAg durumunu doğru bir şekilde sınıflandıran viral varyantları belirlemek için makine öğrenimi yaklaşımı kullanarak bir model geliştirmişler [7]. Keleş ve ark., elde ettikleri modellerle karaciğer hastalıklarının erken evrede yakalanmasını sağlayacak akıllı bir sistemin alt yapısını oluşturmaya çalışmışlar [3]. Tian ve ark., çalışmalarında her hastanın HBsAg seroklerans durumunu laboratuvar ve demografik değişkenlere göre regresyonlayarak, makine öğrenimi algoritmasının değerlendirmesini ve karşılaştırmasını yapmışlar [8]. Nilashi ve ark., yaptıkları çalışmalarında, çeşitli parametreler kullanılarak hepatit hastalığının tahmin edilmesi için topluluk öğrenme tekniklerinin etkinliğini araştırmışlar [9]. Lichthagen ve ark., çalışmalarında, hastaların fibrozis evrelerini histolojik olarak tanımlayıp, belirteç geçerliliğini araştırmak için ROC analizleri yapmışlar [10]. Jekarl ve ark., kronik hepatit B hastalarında WFAM2BP ile karaciğer fibrozisi tanısı ile ilgili bir çalışma yapmışlar [11]. Omran ve ark., kronik HCV hastalarında karaciğer fibrozunun ileri evresini ayırt etmede geliştirilmiş karaciğer fibrozisi (ELF) skorunun geçerliliğini değerlendirmişler [12]. Guo ve ark., hepatit E insidansını ve vaka sayısını tahmin etmek için makine öğrenme yöntemleri ve bu yöntemlerin etkinliğini doğrulamak ve her bir modelin performansını değerlendirmek için farklı performans metriklerini uygulamışlar [13]. Salazar-Concha ve Ramírez-Correa anket verilerinden, donörlerin kan vermeye eğilimlerini karar ağacı yöntemi kullanarak yapmaya çalışmışlar [14]. Hoffmann ve ark., Hepatit C hasta verilerini kullanarak, makine öğrenmesi yöntemlerinin uygulanabilirliğini göstermişler [15].

Yapılan çalışmada, UCI HCV veri seti kullanılarak, kan bağışında bulunacak donörlerin kan değerlerinden sağlıklı ve hastalıklı oluşları makine öğrenmesi yöntemleri ile teşhis edilip kan verip veremeyeceklerine karar verilmeye çalışılmıştır. Sınıflandırma için, Karar Ağaçları (KA), Destek vektör Makineleri (DVM) ve k en yakın komşu (KNN) yöntemleri kullanılmış, performans ölçütü olarak ise doğruluk ve gerçek oran (GO) kullanılmıştır.

2. Materyal ve Metod

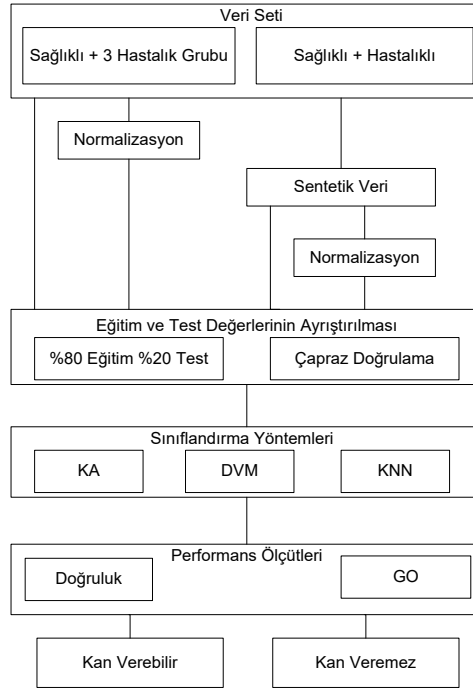
2.1. Veriseti

Bu çalışmada UCI HCV dataset isimli veri kümesi kullanılmıştır [16]. Normalde bu veri kümesi 615 örnek 13 özellik ve 1 çıkıştan oluşmuştur. Bu örneklerden 533 tanesi sağlıklı, 7 tanesi şüpheli, 24 tanesi hepatit, 21 tanesi fibrosis ve 30 tanesi cirrhosis hastalığından oluşmuştur. Fakat bunların 32 tanesinde eksik bilgiler olduğu için ve 7 tanede şüpheli olduğu için sonuçta sağlıklı ve hasta olmak üzere

toplam 583 örnek işleme alınmıştır. Özellik olarak donörlerin yaşı, cinsiyeti, albümin (ALB), bilirubin (BIL), choline (CHE), γ -lutamyl-transferase (GGT), aspartate amino-transferase (AST), alanine amino-transferase (ALT), ALP, CHOL, CREA ve PROT değerleri kullanılmıştır. Sonuç olarak çalışmada 583 örnek, 13 özellikten oluşan bir veri kümesi değerlendirilmiştir.

2.2. Paradigma

Elde edilen veri setinde 583 örneğin; 526 tanesi sağlıklı, 20 tanesi hepatit, 12 tanesi fibrosis ve 24 tanesi cirrhosis hastalığından oluşmuştur. İlk olarak sağlıklı ve 3 farklı hastalığın örnekleri alınarak sınıflandırma işlemi yapılmıştır. Daha sonra sağlıklı (526 örnek) ve hastalıklı (56 örnek) olacak şekilde iki grup işleme alınmıştır. Sağlıklı ve hastalıklı örnekler arasında dengesizlik olduğu için synthetic minority oversampling technique yöntemi ile sentetik veriler oluşturularak aynı şekilde işleme alınmıştır. Son aşamada normal ve sentetikli verisetleri normalize edilerek sınıflandırma işlemine alınmıştır. Sınıflandırmada kullanılacak veriseti eğitim ve test verileri olarak ayrıştırılıp makine öğrenme yöntemlerinden KA, DVM ve KNN kullanılarak kan verecek donörlerin tahmini doğruluk ve GO performans ölçütleri ile yapılmıştır. Çalışma akış diagramı şekil 1 de verilmiştir.



Şekil 1. Çalışma akış diagramı

2.3. Sınıflandırma ve Başarı Ölçütü

Çalışmada, makine öğrenme algoritmalarından yararlanarak sağlıklı ve hasta kişilerin kan değerlerinden kan bağıışı yapıp yapamayacaklarına karar verilmeye çalışılmıştır. Bu algoritmaların temel amacı hesaplama yöntemlerini kullanarak veriden bilgi elde etmeyi sağlamaktır. Kullanılan birçok farklı yöntem vardır, fakat doğru algoritmaların seçimi önemlidir. Bu çalışmada; KA, DVM ve KNN yöntemleri kullanılmıştır. KA, bir kararın her olası sonucunu göstermek için dallanma tekniğini kullanan bir grafik veya ağaç oluşturur. Bir karar ağacı temsiline, her dahili düğüm bir özelliği test eder, her dal, ana düğümün sonucuna karşılık gelir ve her yaprak sonunda sınıf etiketini atar. Bir örneği sınıflandırmak için, ağacın kökünden başlayarak yukarıdan aşağıya bir yaklaşım uygulanır. Belirli bir özellik veya

düğüm için, bir yaprağa ulaşıncaya veya bir etikete karar verilene kadar, o özneliğin veri noktasının değeriyle uyuşan dal dikkate alınır [17]. DVM yöntemi, farklı sınıflara ait veri setini birbirinden en uzak olacak şekilde, aynı sınıfa ait veri kümelerini ise en yakın olacak şekilde ayırmaya çalışan bir yöntemdir. KNN, sınıflandırma işlemi için en yakın komşu sayısı k 'yı seçer ve k 'nın grup üyeliğine göre verileri sınıflandırır. [18]. Veri setleri sınıflandırma işlemine tutulmadan önce normalizasyon işlemine tabi tutulmuştur. Normalizasyon için eş.1 kullanılmıştır [19].

$$N = 0,8 * \frac{(X_i - X_{\min})}{(X_{\max} - X_{\min})} + 0,1 \quad (1)$$

Literatürde sıklıkla karşılaştığımız doğruluk performans ölçütüne ilave olarak bu çalışmada özellikle vurgulamak istediğimiz GO ölçütü kullanılmıştır. Doğruluk performans ölçütü için eş.2 kullanılmıştır [20].

$$\text{Doğruluk} = \frac{\text{Doğru Tahmin Edilen Örnekler}}{\text{Tüm Örnekler}} \quad (2)$$

GO ölçütleri için iki grubun sınıflandırılmasın da eş.3, dört grubun sınıflandırılmasın da ise eş.4 kullanılmıştır.

$$\text{TR} = \frac{\frac{T_{c1} + T_{c2}}{c_1 + c_2}}{2} \quad (3)$$

$$\text{TR} = \frac{\frac{T_{c1} + T_{c2} + T_{c3} + T_{c4}}{c_1 + c_2 + c_3 + c_4}}{4} \quad (4)$$

3. Uygulama ve Başarımlar

Çalışmada kullanılan veri kümesi, sağlıklı ve üç grup hastalıklı olacak şekilde toplam dört grup direkt olarak işleme alınmıştır. Eğitim ve test verileri harmanlanarak 10 çapraz doğrulama ve veri setinin %80 eğitim, %20 test işlemine kullanılacak şekilde iki farklı aşamada sınıflandırılmıştır. Sağlıklı ve diğer hastalıklı gruplar arasında dengesiz bir dağılım olduğu için sentetik veriler oluşturularak tekrardan sınıflandırma işlemi yapılmış ve elde edilen sonuçların doğruluk ve GO değerleri tablo 1'de verilmiştir.

Çizelge 1. Dört grup'a ait başarı oranları.

Yöntem	Gerçek Değerler				Sentetik Değerler			
	GO		Doğruluk		GO		Doğruluk	
	%80-%20	CV	%80-%20	CV	%80-%20	CV	%80-%20	CV
KA	0,536	0,570	0,94	0,933	0,967	0,969	0,967	0,96
DVM	0,900	0,743	0,983	0,967	0,995	0,989	0,995	0,989
KNN	0,725	0,589	0,966	0,952	0,978	0,979	0,978	0,979

Elde edilen sonuçlar karşılaştırıldığında, tüm gruplar direkt işleme alındığında doğruluk performans ölçütü kullanılarak elde edilen sonuçlar GO ölçütü ile elde edilen sonuçlardan daha yüksek olmuştur. Sentetik olarak oluşturulan veri seti sınıflandırma işlemine alındığında doğruluk ve GO performans ölçütleri kullanılarak elde edilen sonuçların yaklaşık ve yüksek değerler olduğu

gözlemlenmiştir. Sınıflandırma yöntemleri karşılaştırıldığında ise en yüksek sonuçların DVM yöntemi kullanılarak elde edildiği görülmüştür.

İlk uygulamada kullanılan veri seti normalizasyon işlemine tabi tutularak aynı aşamalardan geçirilmiştir. Bu şekilde yapılan sınıflandırma işleminde elde edilen sonuçların doğruluk ve GO değerleri çizelge 2’de verilmiştir.

Çizelge 2. Normalizasyonlu dört grup’a ait başarı oranları.

Yöntem	Gerçek Değerler				Sentetik Değerler			
	GO		Doğruluk		GO		Doğruluk	
	%80-%20	CV	%80-%20	CV	%80-%20	CV	%80-%20	CV
KA	0,437	0,494	0,931	0,926	0,947	0,958	0,947	0,958
DVM	0,729	0,741	0,966	0,967	0,988	0,993	0,988	0,993
KNN	0,521	0,601	0,94	0,954	0,983	0,978	0,983	0,978

Normalizasyon işlemine tabi tutularak oluşturulmuş veri setinin sınıflandırma sonuçları karşılaştırıldığında direk işleme alınan veri setinde doğruluk performans ölçütü kullanılarak elde edilen sonuçlar GO ölçütü ile elde edilen sonuçlardan daha yüksek olmuştur. Sentetik olarak oluşturulan veri seti sınıflandırma işlemine alındığında doğruluk ve GO performans ölçütleri kullanılarak elde edilen sonuçların yaklaşık ve yüksek değerler olduğu gözlemlenmiştir. Sınıflandırma yöntemleri karşılaştırıldığında ise en yüksek sonuçların DVM yöntemi kullanılarak elde edildiği görülmüştür.

Bir sonraki uygulamada veri seti sağlıklı ve hastalıklı olacak şekilde iki grup direkt olarak işleme alınmıştır. Eğitim ve test verileri harmanlanarak 10 çapraz doğrulama ve veri setinin %80 eğitim, %20 test olacak şekilde sınıflandırılmıştır. Sağlıklı ve hastalıklı gruplar arasında dengesiz bir dağılım olduğu için sentetik veriler oluşturularak tekrardan sınıflandırma işlemi yapılmış ve elde edilen sonuçların doğruluk ve GO değerleri çizelge 3’de verilmiştir.

Çizelge 3. İki grup’a ait başarı oranları.

Yöntem	Gerçek Değerler				Sentetik Değerler			
	GO		Doğruluk		GO		Doğruluk	
	%80-%20	CV	%80-%20	CV	%80-%20	CV	%80-%20	CV
KA	0,713	0,889	0,922	0,973	0,976	0,978	0,976	0,978
DVM	0,954	0,954	0,991	0,990	0,990	0,995	0,99	0,995
KNN	0,818	0,845	0,966	0,966	0,980	0,985	0,981	0,986

İki grubun sınıflandırılması sonucu elde edilen başarı değerleri karşılaştırıldığında, veri seti direk olarak işleme alındığında doğruluk performans ölçütü kullanılarak elde edilen sonuçlar GO ölçütü ile elde edilen sonuçlardan daha yüksek olmuştur. Sentetik olarak oluşturulan veri seti sınıflandırma işlemine alındığında doğruluk ve GO performans ölçütleri kullanılarak elde edilen sonuçların yaklaşık ve yüksek değerler olduğu gözlemlenmiştir. Sınıflandırma yöntemleri karşılaştırıldığında ise en yüksek sonuçların yine DVM yöntemi kullanılarak elde edildiği görülmüştür.

Aynı şekilde iki grubun sınıflandırılmasında kullanılan veri seti, normalizasyon işlemine tabi tutulup direkt ve sentetik veriler oluşturulmuş şekilde sınıflandırılmıştır. Eğitim ve test verileri bir önceki uygulama ile aynı şekilde ayrıştırılmıştır. Bu şekilde yapılan sınıflandırma işleminde elde edilen sonuçların doğruluk ve GO değerleri çizelge 4’de verilmiştir.

Çizelge 4. Normalizasyonlu iki grup'a ait başarı oranları.

Yöntem	Gerçek Değerler				Sentetik Değerler			
	GO		Doğruluk		GO		Doğruluk	
	%80-%20	CV	%80-%20	CV	%80-%20	CV	%80-%20	CV
KA	0,773	0,869	0,957	0,966	0,986	0,975	0,986	0,975
DVM	0,909	0,945	0,983	0,988	0,995	0,989	0,995	0,989
KNN	0,818	0,845	0,966	0,966	0,990	0,988	0,99	0,988

Normalizasyon işlemine tabi tutularak iki grup için oluşturulmuş veri setinin sınıflandırma sonuçları karşılaştırıldığında direk işleme alınan veri setinde doğruluk performans ölçütü kullanılarak elde edilen sonuçlar GO ölçütü ile elde edilen sonuçlardan daha yüksek olmuştur. Sentetik olarak oluşturulan veri seti sınıflandırma işlemine alındığında doğruluk ve GO performans ölçütleri kullanılarak elde edilen sonuçların yaklaşık ve yüksek değerler olduğu gözlemlenmiştir. Sınıflandırma yöntemleri karşılaştırıldığında ise tüm denemelerde de görüldüğü gibi en yüksek sonuçlar DVM yöntemi kullanılarak elde edilmiştir.

Veri setinde sağlıklı ve hasta örnek sayılarının dağılımı dengesiz olduğundan değerlendirmenin daha hassas olabilmesi için sentetik veriler oluşturulmuş ve ayrıca performans ölçütü olarak GO tercih edilmiştir. İlk olarak sentetik veriler oluşturulmadan yapılan çalışmada doğruluk performans ölçütünün daha yüksek olduğu görülmüştür. Fakat veri setlerinin dengesiz dağılımında doğruluk ölçütünün yanıltıcı olabileceği düşünüldüğünden GO performans ölçütünün daha belirleyici olabileceği görülmüştür. Veri setinin dengeli olduğu durumlarda doğruluk performans ölçütünün belirleyici olabileceği bilindiği için sentetik veriler üretilerek tekrardan sonuçlar karşılaştırılmış ve her iki performans ölçütünün de başarılı olduğu gözlemlenmiştir.

4. Sonuç

Bu çalışmada UCI HCV veri seti kullanılarak, sağlıklı ve hasta kişilerin kan değerlerinden kan bağışı yapıp yapamayacakları ele alınmış olup KA, DVM ve KNN olmak üzere üç farklı sınıflandırma algoritması ile sınıflandırılmıştır. Elde edilen sonuçlar değerlendirildiğinde en başarılı sınıflandırma yönteminin DVM olduğu sonucuna varılmıştır. Performans ölçütlerinin karşılaştırılmasında ise veri setlerinin dengesiz dağılımlarında GO ölçütünün, dengeli dağılımda ise doğruluk ve GO performans ölçütünün belirleyici olduğu söylenebilir. Hastalık teşhislerinde uzmanlar belirli kriterler ve kendi tecrübelerini kullanarak karar verirler. Karar aşamasında tecrübe eksikliği veya o anki konsantrasyon eksikliği farklı bir teşhis olasılığına sebep olabilir. Ayrıca bazı teşhis durumlarında laboratuvar ortmalarına ihtiyaç duyulur. Bu ve benzeri durumlar zamansal ve ekonomik olarak da maliyet anlamına gelir. Bu çalışmada sunulan makine öğrenimi algoritmaları ile uzmanlara teşhis koyma aşamasında yardımcı olacak fikirler vererek çalışmaya disiplinler arası bir nitelik verilmiş ve sağlıkla ilgili çalışmalarda kapsam dahada genişletilerek farklı çalışmalar yapılabileceği görülmüştür.

Kaynaklar

- [1] Altındış, S., Aşıcı, N., Şenol, U., Bakır, A. B., Yücedağ, Ç. Y., Demir, C., ... & Altındış, A. Transfüzyon Amaçlı Kan Verme Önündeki Engellerin Belirlenmesi. *Sakarya Tıp Dergisi*, 9(2), 237-244.
- [2] Yıldız, C., Emekdas, G., & Kanık, A. (2006). Why don't we donate blood. A general view of blood donation by people living Mersin: a public survey. *Turkish Journal of Infection*, 20(1), 41-55.

- [3] Keleş, A., Kaslı, Ö.B., Keleş, A. (2020). Makine öğrenme algoritmaları ile karaciğer hastalığının teşhisi, *Turkish Studies-Applied Sciences*, 15(1), 75-83.
- [4] Masser, B., Ferguson, E., Merz, E. M., & Williams, L. (2020). Beyond description: The predictive role of affect, memory, and context in the decision to donate or not donate blood. *Transfusion Medicine and Hemotherapy*, 47(2), 175-185.
- [5] Czeizler, A., & Garbarino, E. (2017). Give blood today or save lives tomorrow: matching decision and message construal level to maximize blood donation intentions. *Health marketing quarterly*, 34(3), 175-186.
- [6] Khan, M. A., Soh, J. E., Maenner, M., Thompson, W. W., & Nelson, N. P. (2019). A machine-learning algorithm to identify hepatitis C in health insurance claims data. *Online Journal of Public Health Informatics*, 11(1).
- [7] Mueller-Breckenridge, A. J., Garcia-Alcalde, F., Wildum, S., Smits, S. L., Robert, A., van Campenhout, M. J., ... & Haagmans, B. L. (2019). Machine-learning based patient classification using Hepatitis B virus full-length genome quasispecies from Asian and European cohorts. *Scientific reports*, 9(1), 1-12.
- [8] Tian, X., Chong, Y., Huang, Y., Guo, P., Li, M., Zhang, W., ... & Hao, Y. (2019). Using machine learning algorithms to predict hepatitis B surface antigen seroclearance. *Computational and mathematical methods in medicine*, 2019.
- [9] Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O., & Akbari, E. (2019). A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique. *Journal of infection and public health*, 12(1), 13-20.
- [10] Lichtinghagen, R., Pietsch, D., Bantel, H., Manns, M. P., Brand, K., & Bahr, M. J. (2013). The Enhanced Liver Fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. *Journal of hepatology*, 59(2), 236-242.
- [11] Jekarl, D. W., Choi, H., Lee, S., Kwon, J. H., Lee, S. W., Yu, H., ... & Yoon, S. K. (2018). Diagnosis of liver fibrosis with Wisteria floribunda agglutinin-positive Mac-2 binding protein (WFA-M2BP) among chronic hepatitis B patients. *Annals of laboratory medicine*, 38(4), 348.
- [12] Omran, D., Yosry, A., Darweesh, S. K., Nabeel, M. M., El-Beshlawey, M., Saif, S., ... & Zayed, R. A. (2018). Enhanced liver fibrosis test using ELISA assay accurately discriminates advanced stage of liver fibrosis as determined by transient elastography fibroscan in treatment naïve chronic HCV patients. *Clinical and experimental medicine*, 18(1), 45-50.
- [13] Guo, Y., Feng, Y., Qu, F., Zhang, L., Yan, B., & Lv, J. (2020). Prediction of hepatitis E using machine learning models. *Plos one*, 15(9), e0237750.
- [14] Salazar-Concha, C., & Ramírez-Correa, P. (2021). Predicting the Intention to Donate Blood among Blood Donors Using a Decision Tree Algorithm. *Symmetry*, 13(8), 1460.
- [15] Hoffmann, G., Bietenbeck, A., Lichtinghagen, R., & Klawonn, F. (2018). Using machine learning techniques to generate laboratory diagnostic pathways—a case study. *J Lab Precis Med*, 3, 58.
- [16] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- [17] Das, K., & Behera, R. N. (2017). A survey on machine learning: concept, algorithms and applications. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(2), 1301-1309.
- [18] Karadağ, K. (2020). Makine Öğrenme Yöntemleri ile Semen Kalitesi Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (18), 306-311.
- [19] Gökhan, A. K. S. U., Güzeller, C. O., & Eser, M. T. (2019). The effect of the normalization method used in different sample sizes on the success of artificial neural network model. *International Journal of Assessment Tools in Education*, 6(2), 170-192.
- [20] Jiao, Y., & Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantitative Biology*, 4(4), 320-330.