

# Miyeloproliferatif Hastalık Ön Tanılı Hastalarda Makine Öğrenmesi Yöntemi ile Genetik Test Seçimine İlişkin Metodolojik Bir Modelleme Çalışması

## A Methodological Modeling Study on the Selection of Genetic Testing with Machine Learning Method in Patients with Pre-Diagnosis of Myeloproliferative Disease

• Gözde KUBAT<sup>a,b</sup>, • Feride İffet ŞAHİN<sup>c</sup>, • Bülent ÇELİK<sup>d</sup>

<sup>a</sup>Başkent Üniversitesi Kahramankazan Meslek Yüksekokulu, Yönetim ve Organizasyon Bölümü, Ankara, Türkiye

<sup>b</sup>Gazi Üniversitesi Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye

<sup>c</sup>Başkent Üniversitesi Tıp Fakültesi, Tıbbi Genetik ABD, Ankara, Türkiye

<sup>d</sup>Gazi Üniversitesi Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye

Bu çalışma, 1. Ulusal HematoOnkoGenetik Kongresi'nde (25-28 Kasım 2021, Antalya) sözlü bildiri olarak sunulmuştur.

Bu çalışma, Gözde Kubat'ın "Miyeloproliferatif Hastalık Ön Tanılı Hastalarda Makine Öğrenmesi Yöntemi ile Genetik Test Seçimine İlişkin Metodolojik Bir Modelleme Çalışması" başlıklı doktora tezinden üretilmiştir (Ankara: Gazi Üniversitesi; 2022).

**ÖZET Amaç:** Miyeloproliferatif hastalıkların tanısının konulabilmesi için Dünya Sağlık Örgütü'nün belirlediği hastalık tanı şeması doğrultusunda, klinisyen tarafından genetik mutasyonların varlığı incelenmektedir. Bu çalışmada; kliniğe başvuru yapmış hastaların kemik iliği ve tam kan sayımı bulgularından yola çıkılarak oluşturulabilecek model ile uygun tanı testinin tahmin edilebilmesi hedeflenmiştir. Bu doğrultuda, kliniğe başvuran hastaların zaman ve maddi açıdan tasarruf etmesi amaçlanmaktadır. **Gereç ve Yöntemler:** Başkent Üniversitesi Ankara Hastanesi Tıbbi Genetik Ana Bilim Dalı Genetik Hastalıklar Değerlendirme Merkezine yönlendirilmiş hastaların bulguları ele alınarak, makine öğrenmesi algoritmaları kullanılarak tahminleme yapılmıştır. Çalışma verilerine ait tanımlayıcı istatistikler medyan olarak verilmiş olup, Kruskal-Wallis test istatistiği kullanılarak istatistiksel olarak anlamlı farklılıklar araştırılmıştır. Tahminlemede Naive Bayes, K-En Yakın Komşuluk, Doğrusal Diskriminant Analizi, Destek Vektör Makineleri, Entropi Tabanlı Sınıflandırma ve Karar Ağacı gibi sınıflandırma algoritmaları kullanılmıştır. Algoritmalar ile doğruluk, özgüllük, duyarlılık gibi belirleyici değerler elde edilmiştir. Yapılan tahminlemeler elde edilen doğruluk oranlarına göre incelenmiş ve en iyi model seçilmeye çalışılmıştır. **Bulgular:** İncelenen tam kan sayım değerleri ile mutasyon varlığı arasında istatistiksel olarak anlamlı farklılıklar tespit edilmiştir. Ele alınan Naive Bayes, K-En Yakın Komşuluk, Doğrusal Diskriminant Analizi, Destek Vektör Makineleri, Entropi Tabanlı Sınıflandırma ve Karar Ağacı Algoritmaları ile oluşturulan modellemelerde doğruluk oranları %60 olarak saptanmıştır. **Sonuç:** Kullanılan makine öğrenmesi algoritmalarından elde edilen doğruluk oranı orta seviyede olmasına rağmen benzer çalışmaların literatürde yer almaması sebebiyle çalışma sonuçlarının alana önemli katkısının olacağı sonucuna varılmıştır.

**Anahtar kelimeler:** Miyeloproliferatif hastalık; makine öğrenmesi; karar ağacı; K-en yakın komşuluk; doğrusal diskriminant analizi

**ABSTRACT Objective:** In order to diagnose myeloproliferative diseases, the presence of genetic mutations is examined by the clinician in line with the disease diagnosis scheme determined by the World Health Organization. In this study, it is aimed to predict the appropriate diagnostic screening test with the model created based on bone marrow and complete blood count findings of patients who applied to the clinic. Accordingly, it is aimed to save time and financial for patients who apply to the clinic. **Material and Methods:** Prediction was made using the machine learning algorithms by considering the findings of patients referred to the Başkent University Ankara Hospital, Department of Medical Genetics Genetic Diseases Diagnosis Center. Descriptive statistics of the study data were given as median, and statistically significant differences were investigated using the Kruskal-Wallis test statistic. Classification algorithms such as Naive Bayes, K-Nearest Neighbor, Linear Discriminant Analysis, Support Vector Machines, Entropy Based Classification and Decision Tree were used in the estimation. With the algorithms, determinative values such as accuracy, specificity and sensitivity were obtained. The estimations made were examined according to the accuracy rates obtained and the best model was tried to be selected. **Results:** Statistically significant differences were found between the examined complete blood count values and the presence of mutation. The accuracy rates were found to be around 60% in the models created with Naive Bayes, K-Nearest Neighbor, Linear Discriminant Analysis, Support Vector Machines, Entropy Based Classification and Decision Tree Algorithms. **Conclusion:** Although the accuracy rate obtained from the machine learning algorithms used is at a moderate level, it was concluded that the results of the study would make a significant contribution to the field, since similar studies have not been included in the literature.

**Keywords:** Myeloproliferative disease; machine learning; decision tree; K-nearest neighbor; linear discriminant analysis

**Correspondence:** Gözde KUBAT

Başkent Üniversitesi Kahramankazan Meslek Yüksekokulu, Yönetim ve Organizasyon Bölümü, Ankara, Türkiye

**E-mail:** ozer.gozde@gmail.com

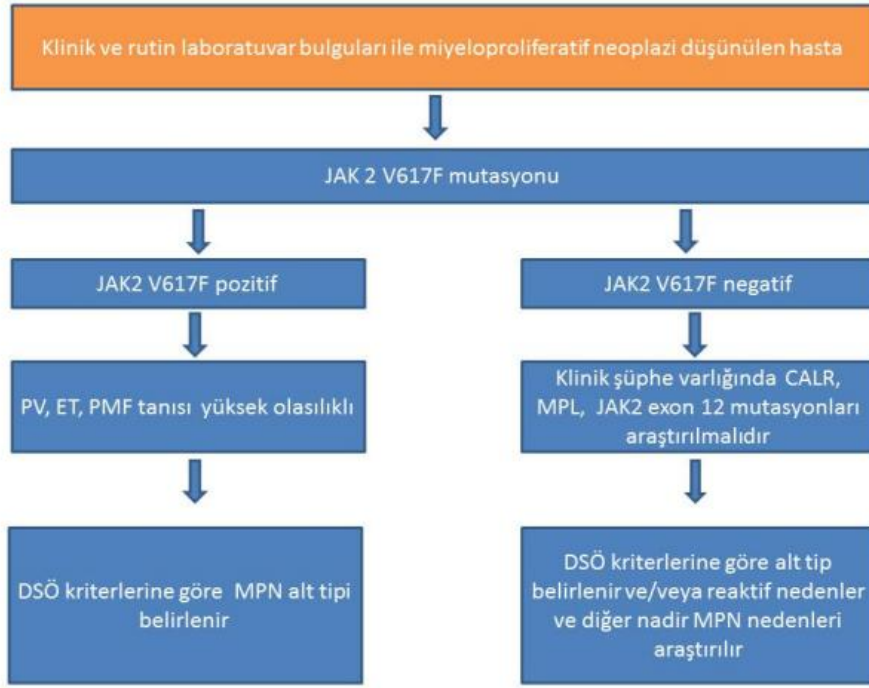


Peer review under responsibility of Türkiye Klinikleri Journal of Biostatistics.

**Received:** 10 Nov 2021 **Received in revised form:** 22 Mar 2022 **Accepted:** 23 Mar 2022 **Available online:** 20 Apr 2022

2146-8877 / Copyright © 2022 by Türkiye Klinikleri. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Miyeloproliferatif hastalıklar; kırmızı kan hücrelerinin, trombositlerin veya beyaz kan hücrelerinin kontrolsüz çoğalması sonucu oluşmaktadır.<sup>1</sup> Kronik miyelositer lösemi, polisitemia vera, esansiyel trombositopeni ve primer miyelofibrozis olmak üzere 4 grupta değerlendirilmektedir.<sup>2</sup> Bu hastalıklar için klinisyene başvuran hastalara, Dünya Sağlık Örgütü'nün (DSÖ) belirlediği tanı kriterlerine göre ön tanı konulmaktadır.<sup>3</sup> Bu kapsamda, ön tanısı konulan hastalara genetik mutasyon taraması uygulanarak kesin teşhis için karar verilmektedir. Kesin teşhis için uygulanabilecek mutasyon taramalarına ilişkin çok sayıda yöntem olmasına karşın, belirlenen ön tanı hastalar için sıklıkla *JAK2 V617F*, *JAK2 Exon 12*, *MPL* ve *CALR* mutasyon taramaları tercih edilmektedir.<sup>4,5</sup> [Şekil 1](#)'de *JAK 2 V617F*, *JAK2 Exon 12*, *MPL* ve *CALR* mutasyon taramalarına ilişkin uygulama şeması yer almaktadır.



DSÖ: Dünya Sağlık Örgütü  
 PV: Polisitemia vera, ET: Esansiyel trombositemi, PMF: Primer miyelofibrozis  
 MPN: Miyeloproliferatif neoplazi

**ŞEKİL 1:** Miyeloproliferatif neoplazi ön tanı hastalara ilişkin tanı testlerinin uygulama yöntemi.<sup>3</sup>

Klinisyen tarafından tarama testleri [Şekil 1](#)'de verilen akış şemasına göre uygulanmaktadır. Kliniğe gelen miyeloproliferatif neoplazi ön tanı hastalarda, öncelikle tarama testi yapılarak *JAK2 V617F* mutasyonunun varlığı saptanmaya çalışılmaktadır. Bu mutasyon saptandığı durumda, hastalık riski yüksek olarak tanı konulabilmekteyken, mutasyonun saptanamaması durumunda *MPL*, *JAK2 Exon 12* veya *CALR* mutasyonları araştırılmaktadır.

Günümüzde yapılmakta olan bu uygulamaların, süre ve maliyet kapsamında hasta kazanımlarının düşük seviyede olduğu düşüncesi ortaya çıkmaktadır. Oluşturulacak istatistik bir model ile kliniğe gelen hastaların süre ve maliyet kazanımlarının artırılabilirliği öngörülmüştür. DSÖ'nün tanımladığı akış şemasına göre hastaya tarama testi yapıldığı bilgisi ışığında, çalışma sonucunda oluşturulan istatistik model ile klinisyenin uygulaması gereken mutasyon taramasına ilişkin yönlendirme yapabileceği düşünülmüştür.

Çalışmada hedeflenen model, makine öğrenmesi yöntemi kullanılarak elde edilmiştir. Makine öğrenmesi, günümüzde yaygın olarak kullanılan bir analiz yöntemidir ve mevcut verilerden doğru ve tutarlı tahminler yapabilmek amacıyla sıklıkla kullanılmaktadır.<sup>6</sup> Makine öğrenmesi yönteminde, gerçekte karşılaşılmış durumlara ilişkin verilerin belirli bir oranı kullanılarak ele alınan algoritmalara eğitim verisi olarak verilmekte, geri kalan kısmı ise test verisi olarak tahmin edilmeye çalışılmaktadır.<sup>7</sup> Algoritmalar sonucunda tahmin edilen değerlerin gerçek durumlarına göre tahminlemenin güven düzeyleri saptanmakta ve saptanan bu değerlere göre klinisyene tanı testi önerilebilmektedir.<sup>8</sup>

Yapılan bu çalışmanın amacı, miyeloproliferatif hastalık ön tanısı konulmuş hastalara ait kemik iliği ve kan sonuçları ile demografik özelliklerine ilişkin veriler ele alınarak, makine öğrenmesi yöntemleri yardımıyla hastalara uygulanacak olan mutasyon testinin tespit edilebileceği modelin oluşturulmasıdır.

## GEREÇ VE YÖNTEMLER

Metodolojik olarak yapılan bu çalışma kapsamında, 03 Ocak 2012-11 Mart 2019 tarihleri arasında Başkent Üniversitesi Ankara Hastanesi Tıbbi Genetik Ana Bilim Dalı Genetik Hastalıklar Merkezine yönlendirilmiş miyeloproliferatif hastalık ön tanılı hastaların kemik iliği ve tam kan sayım sonuçları ele alınmıştır. Bu çalışma, 24 Temmuz 2019 tarihli, 19/80 sayılı proje ile Başkent Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurul tarafından onaylanmış ve Helsinki Deklarasyon Prensipleri'ne uygun olarak gerçekleştirilmiştir.

### MAKİNE ÖĞRENMESİ YÖNTEMİ

Makine öğrenmesi, günümüzde yaygın olarak kullanılan bir analiz yöntemidir. Bu yöntem ile çalışmada kullanılacak olan veriler modellenerek tahminleme yapılabilmektedir. Makine öğrenmesi yöntemi, yapay zekâ algoritmalarının bir alt boyutudur.<sup>9</sup> Makine öğrenmesi teknikleri, denetimli ve denetimsiz öğrenme olmak üzere 2 gruba ayrılmaktadır.<sup>10</sup> Denetimli öğrenme, önceden gözlemlenmiş ve sonuçları bilinen (etiketlenmiş) verileri kullanarak yeni bir model oluşturmayı, denetimsiz öğrenme ise etiketlenmemiş veriler arasındaki bilinmeyen bir yapı bulmayı hedeflemektedir.<sup>11</sup>

### MAKİNE ÖĞRENMESİ YÖNTEMİNDE KULLANILAN ALGORİTMALAR

**Naive Bayes Sınıflandırma Algoritması:** Basit olasılıksal sınıflandırıcı olarak bilinen Bayes sınıflandırıcısı, istatistiksel sınıflandırma teknikleri arasında yer alır. Bu algoritma göre bir sonucun çıkma olasılığı, o sonucu etkileyen tüm faktörlerin sonucu sağlama olasılıklarının çarpımı ile elde edilmektedir. Yapılan bu işlem sonrasında elde edilen çarpım değeri üzerinden yorumlama yapılabilmektedir.<sup>12</sup> Bu sınıflandırıcının avantajı, eğitim için az sayıda niteliğe ihtiyaç duymasındadır.<sup>13</sup>

Naive Bayes (NB) sınıflandırma algoritmasında sınıf değişkeni (C) 'ye ilişkin bir koşul oluşturularak Eşitlik (1)'de verilen olasılık değerleri hesaplanır.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

Olasılıklar hesaplandıktan sonra veriye ilişkin sınıflandırma için  $P(X|C_i)P(C_i)$  çarpım değerinin en yüksek olduğu sınıf belirlenerek atama yapılır.<sup>14</sup>

**K-En Yakın Komşuluk Sınıflandırma Algoritması:** K-En Yakın Komşuluk [K-Nearest Neighbors (KNN)] Sınıflandırma Algoritması, denetimli ve örnek tabanlı bir sınıflandırma algoritmasıdır.<sup>15</sup> Bu algoritma, uygulanabilirlik yönüyle daha kolay olmakla birlikte, diğer algoritmalarından farklı olarak eğitim verisine ihtiyaç duymayabilmektedir. İşleyiş aşamasında birbirine yakın olan nesnelere aynı kategoride toplayarak sınıflandırma yapmaktadır.<sup>16</sup> KNN Algoritması'nda sınıfı bilinen örnekler kullanılarak sınıflama yapılmaktadır. Tahmin edilen özellikler tek tek işleme alınarak, sınıflandırma yapılabilmesi için eğitim kümesindeki o örneğe en yakın K tane örnek seçilir. Seçilen örneklerden oluşan küme içerisinde en fazla

hangi sınıfa ait örnek varsa tahmin edilecek olan bu sınıfa aittir denilmektedir.<sup>17</sup> Yapılacak çalışmalarda, örnekler arası uzaklıklar Eşitlik (2)'de verilen Öklid (Euclidean) uzaklığı ile hesaplanmaktadır.<sup>18</sup>

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (dizi_r(x_i) - dizi_r(x_j))^2} \quad (2)$$

**Doğrusal Diskriminant Analizi Sınıflandırma Algoritması:** Doğrusal Diskriminant Analizi [Linear Discriminant Analysis (LDA)], temelleri 1930'lu yıllarda Fisher tarafından atılan çok değişkenli istatistiksel yöntemlerden biridir.<sup>19</sup> Bu algoritma ile sınıflandırma, kümenin daha küçük boyuta indirgenmesi ile gerçekleştirilmektedir. Yöntemin temel amacı, verilerin sınıflandırılması için gerekli olan ayırt edici özellikleri belirleyip, ayırt edici olmayan özellikleri elemektir. Böylece sınıflandırma işlemi daha kolay yapılabilmektedir. LDA ölçüsü, sınıf merkezleri arası uzaklığı en büyük yapmayı, sınıf içindeki dağılımı ise en az düzeye indirmeyi sağlamaktadır.<sup>20</sup> LDA, doğrudan bir sınıflandırıcı olarak kullanılması yanı sıra daha karmaşık doğrusal olmayan sınıflandırmalarda bir ön işlemci olarak da kullanılabilir.

**Destek Vektör Makineleri:** Bu algoritma, 2. dereceden bir optimizasyon problemini çözerek, yüksek boyutlu doğrusal olmayan bir sınıflandırma yapabilme yeteneğine sahiptir.<sup>21</sup> Temel Destek Vektör Makineleri [Support Vector Machine (SVM)] teorisi doğrultusunda 2 sınıfta yer alan örneklerin doğrusal olmayan şekilde ayrılabilirliği bilinmektedir. Bu kapsamda, belirtilen algoritma ile seçilmiş bir veri kümesi için sınıfları ayırabilen sonsuz sayıda doğru olduğu bilinmektedir. Sınıflar oluşturulurken en iyi ayırıcı çizginin seçimi, destek vektörleri olarak tanımlanan eğitim örneklerinin alt kümesi kullanılarak gerçekleştirilmektedir. Destek vektörleri arasındaki en büyük sınır uzaklığı, en iyi karar sınırı olarak tanımlanmaktadır. Sınıflandırmanın doğrusal olmadığı ayrıştırma problemleri için SVM, kernel işlevinden yola çıkarak giriş değişkenlerinde örtülü bir dönüşümü kullanmaktadır. Bu algoritmada, Kernel işlevleri doğrusal bir düzlem kullanarak, doğrusal olmayan ayrılabilir bir destek vektörü yapısı ile sınıflandırmaya izin vermektedir.<sup>22</sup> Birçok uygulamada, performansı en iyilemek için probleme uygun bir kernel genişliğinin seçilmesi gerekir.<sup>21</sup>

**Karar Ağacı Algoritması:** Bu algoritmada bir ağaç yapısı oluşturulmaktadır. Ağacın yaprakları seviyesinde oluşturulan sınıf etiketleri ile başlangıçtan çıkan ve bu yapraklara giden özellikler üzerindeki işlemler ifade edilmektedir.<sup>23</sup> Karar Ağacı Algoritması, temel olarak tanımlanan bir algoritma olup, Random Forest ve Gradient Boosted Tree şeklinde farklı karar ağacı algoritmaları türetilmiştir. Algoritma, anlaşılması ve yorumlanması açısından hem sayısal hem de sınıflandırılmış verilerin işlenmesi şeklinde kullanılabilir.

**Entropi Tabanlı Sınıflandırma:** Bu sınıflandırma yönteminde, C4.5 ve C5.0 algoritmaları kullanılarak tahmin yapılmaktadır. C5.0 algoritmasında tahmin edilen özelliğin kategorik olması gerekmektedir.<sup>24</sup> Öncelikle veri kümesinin beklenen değeri Eşitlik (3) ile hesaplanmaktadır.

$$Info(S) = - \sum_{i=1}^k \left( \frac{f(C_i|S)}{|S|} \right) \log_2 \left( \frac{f(C_i|S)}{|S|} \right) \quad (3)$$

Daha sonra gerçekleştirilen bölümlenme sonrası beklenen değer, Eşitlik (4) yardımıyla hesaplanır.

$$Info_x(T) = - \sum_{i=1}^m \left( \frac{|T_i|}{|T|} \right) Info(T_i) \quad (4)$$

En yüksek kazanımı elde edebilecek değişkenin belirlenebilmesi için Eşitlik (5)'te verilen hesaplamalar yapılmaktadır.<sup>24</sup>

$$Gain(x) = Info(T) - Info_x(T) \quad (5)$$

## MAKİNE ÖĞRENMESİ YÖNTEMİNDE ALGORİTMALARIN KARŞILAŞTIRILMASI

Yapılan algoritmalar sonucunda elde edilen tahmin değerlerinin karşılaştırması karşıtlık matrisi yardımı ile yapılmaktadır. Bu matriste verilen "Doğru Pozitif (DP)", "Doğru Negatif (DN)", "Yanlış Negatif (YN)" ve "Yanlış Pozitif (YP)" değerleri kullanılarak Eşitlik (6-10)'da verilen "Doğruluk", "Duyarlılık", "Özgüllük",

“Keskinlik” ve “F-Ölçütü” değerleri hesaplanmaktadır. Elde edilen değerler sonucunda en iyi tahminleme yapılabilmektedir.<sup>25</sup>

**Doğruluk (Accuracy):** Makine öğrenmesinde kullanılan eğitim modeline göre test verisi için yapılan tahminlemelerin gerçek değerlerle karşılaştırılması sonucu elde edilen doğru sınıflandırma oranıdır ve Eşitlik (6) yardımıyla hesaplanmaktadır.<sup>26</sup>

$$\text{Doğruluk} = \frac{DP+DN}{DP+YP+YN+DN} \quad (6)$$

**Duyarlılık (Sensitivity):** DP oranı olarak da tanımlanan duyarlılık değeri, yapılan tahminleme sonucu gerçekte pozitif sınıfa ait olan verilerin doğru olarak tahmin edilme oranını vermektedir ve Eşitlik (7) yardımıyla hesaplanmaktadır.<sup>26</sup>

$$\text{Duyarlılık} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Negatif}} \quad (7)$$

**Özgüllük (Specificity):** Negatif belirleyicilik değeri olarak da tanımlanan özgüllük değeri, tahmin edilen test verisindeki doğru tahminlenen negatif sınıf verilerinin oranını vermektedir ve Eşitlik (8) yardımıyla hesaplanmaktadır.<sup>27</sup>

$$\text{Özgüllük} = \frac{\text{Doğru Negatif}}{\text{Yanlış Pozitif} + \text{Doğru Negatif}} \quad (8)$$

**Keskinlik (Precision):** Çalışmada ele alınan test verisinden ve eğitim verisinden elde edilen pozitif değerlerin ne kadarının doğru olarak tahmin edildiğini belirtmektedir ve Eşitlik (9) ile hesaplanmaktadır.<sup>28</sup>

$$\text{Keskinlik} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Pozitif}} \quad (9)$$

**F Ölçütü:** F ölçütü, elde edilen tahminleme için daha kesin bir ifade kullanılabilmesi amacıyla hesaplanan bir değerdir ve Eşitlik (10)'da verildiği gibi kesinlik ve duyarlılık ölçülerinin harmonik ortalaması alınarak hesaplanmaktadır.<sup>29</sup>

$$F - \text{Ölçütü} = \frac{2 * \text{Keskinlik} * \text{Duyarlılık}}{\text{Keskinlik} + \text{Duyarlılık}} \quad (10)$$

## BULGULAR

Çalışmada ele alınan hastaların demografik dağılımı [Tablo 1](#)'de verilmiştir. Yapılan tetkiklerin %66,3'ünde *JAK2 V617F* mutasyonu olduğu saptanmıştır.

**TABLO 1:** Demografik özelliklerin dağılımı.

Değişkenler		Frekans	%
Cinsiyet	Kadın	353	33,9
	Erkek	688	66,1
Materyal	PV	968	93,0
	Kemik iliği	73	7,0
Tetkik	<i>JAK2 V617F</i>	690	66,3
	<i>JAK2 EKZON12</i>	220	21,1
	<i>CALR</i>	46	4,4
	<i>MPL</i>	85	8,2

Çalışma kapsamında kullanılan tahmin yöntemlerinde, verinin %70, %80 ve %90'lık bölümü makine öğrenmesi yöntemlerinde kullanılmak üzere eğitim verisi olarak tanıtılmış, geri kalan kısmı ise tahmin edilmeye çalışılmıştır. Belirtilen oranlara ilişkin eğitim ve test verisi sayıları [Tablo 2](#)'de verilmiştir.

**TABLO 2:** Makine öğrenmesi algoritmalarında kullanılan veri sayılarının dağılımı.

Doğru sonucun verildiği oran	Eğitim verisi	Test verisi
0,70	730	311
0,80	833	208
0,90	938	103

Makine öğrenmesi algoritmalarından LDA, NB Sınıflandırması, KNN, SVM, Karar Ağacı Gini Algoritması ve Entropi Tabanlı Sınıflandırma C5.0 Algoritması kullanılarak analiz sonuçları elde edilmiştir.

Çalışma kapsamında, %70, %80 ve %90 olmak üzere 3 farklı oranda eğitim verisi kullanılmıştır. [Tablo 3](#)'te, ele alınan makine öğrenmesi algoritmalarından elde edilen en iyi doğruluk oranlarına göre tarama testleri için hesaplanan duyarlılık, özgüllük, kesinlik ve F-ölçütü değerleri verilmiştir.

**TABLO 3:** Tarama testleri için hesaplanan doğruluk, duyarlılık, özgüllük, kesinlik ve F-ölçütü değerleri.

	Doğruluk	Tarama testi	Duyarlılık	Özgüllük	Kesinlik	F ölçütü
LDA*	0,660	JAK 2 V617F	0,957	0,118	0,941	0,949
		JAK 2 Exon 12	0,000	0,988	-	-
		MPL	0,000	1,000	-	-
		CALR	0,250	0,958	0,999	0,400
Naive Bayes*	0,592	JAK 2 V617F	0,870	0,277	0,985	0,924
		JAK 2 Exon 12	0,000	0,951	0,000	
		MPL	0,000	0,949	0,000	
		CALR	0,125	0,947	0,997	0,222
SVM*	0,670	JAK 2 V617F	1,000	0,000		
		JAK 2 Exon 12	0,000	1,000		
		MPL	0,000	0,000		
		CALR	0,000	0,000		
KNN*	0,670	JAK 2 V617F	0,986	0,029	0,773	0,866
		JAK 2 Exon 12	0,045	0,988	1,000	0,087
		MPL	0,000	1,000		
		CALR	0,000	1,000		
Gini*	0,670	JAK 2 V617F	1,000	0,000	1,000	1,000
		JAK 2 Exon 12	0,000	1,000		
		MPL	0,000	1,000		
		CALR	0,000	1,000		
C5.0**	0,601	JAK 2 V617F	0,874	0,231	0,977	0,923
		JAK 2 Exon 12	0,061	0,947	0,998	0,114
		MPL	0,000	0,970	0,000	
		CALR	0,080	0,923	0,991	0,148

\*Eğitim verisi oranı=0,90; \*\*Eğitim verisi oranı=0,70; LDA: Doğrusal Diskriminant Analizi; SVM: Destek Vektör Makineleri; KNN: K-En Yakın Komşuluk.

LDA sonucunda, makine öğrenmesi yönteminde ele alınan modelin tahminleme yapabilmesi için verilerin %90'ının eğitim verisi olarak alınması sonucunda elde edilen tahminlerin doğruluğu %66 ile en iyi sonuç olarak tespit edilmiştir. Tarama testlerinin duyarlılığı JAK2 V6127F için %95,7, CALR için %25 ve diğer 2 mutasyon taraması için %0 olarak elde edilmiştir.

NB sınıflandırmasına göre yapılan analiz sonuçlarında, verinin %10'unun tahmin edildiği durumda tahminlerin doğruluğu %59,2 olarak tespit edilmiştir. Tarama testlerinin duyarlılığı JAK2 V6127F için %87,0, CALR için %12,5 ve diğer 2 mutasyon taraması için %0 olarak hesaplanmıştır.



KNN algoritması ile yapılan tahminleme modelinde, en yüksek doğruluk %67 ile verilerin %90'unun eğitim verisi olarak alındığı uygulamadan saptanmıştır. Tarama testlerinin duyarlılığı *JAK2 V6127F* için %98,6, *JAK2 Exon 12* için %4,5 ve diğer 2 mutasyon taramaları için %0 olarak elde edilmiştir.

SVM kullanılarak yapılan analiz kapsamında, doğrusal fonksiyon, polinomiyal fonksiyon ve radyal tabanlı fonksiyon kullanılarak sınıflandırma yapılmıştır. Belirtilen 3 fonksiyon kullanılarak yapılan sınıflandırma sonuçları aynıdır. En yüksek doğruluk değeri, verinin %10'unun tahmin edildiği modelden elde edilmiştir. %67 doğruluk ile *JAK2 V617F* mutasyon taramasının duyarlılık değeri %100, diğer tarama testlerinin duyarlılık değerleri %0 olarak tespit edilmiştir.

Gini algoritması ile ele edilen analiz sonuçlarına göre en yüksek doğruluk değeri örneklem verisinin %10'unun tahmin edildiği durumda saptanmıştır. %67 olan bu doğruluk değeri ile *JAK2 V617F* mutasyon tarama testinin duyarlılık değeri %100, diğer tarama testlerinin duyarlılık değerleri %0 olarak tespit edilmiştir.

C5.0 algoritması ile ele edilen analiz sonuçlarına göre en yüksek doğruluk değeri, verinin %30'unun tahmin edildiği durumda ortaya çıkmıştır. %60,1 olan bu doğruluk değeri ile hesaplanan duyarlılık değerleri *JAK2 V617F* mutasyon taraması için %87,4, *CALR* mutasyon taraması için %8 ve *JAK2 Exon 12* mutasyon taraması için %6,1, diğer tarama testi için %0 olarak hesaplanmıştır.

[Tablo 3](#)'te görüldüğü gibi C5.0 haricindeki tüm algoritmalarda en yüksek doğruluk oranları %90 eğitim verisinin kullanıldığı durumda elde edilmiştir.

Çalışma kapsamında oluşturulan model ile birlikte incelenen mutasyonlara ilişkin tam kan sayım değerlerine dair bir karşılaştırma yapılmıştır. Yapılan karşılaştırma, verilerin dağılımı normal dağılım göstermediği için parametrik olmayan testlerden Kruskal Wallis-H testi ile gerçekleştirilirken, gruplar arasındaki farklılığa LSD ile bakılmıştır. Analize ilişkin sonuçlar [Tablo 4](#)'te verilmiştir.

**TABLO 4:** Çalışmada ele alınan hasta bulgularının yapılan tarama testine göre dağılımı.

	<i>JAK2 V617F</i>		<i>JAK2 EKZON12</i>		<i>CALR</i>		<i>MPL</i>		p değeri	Farklılık gösteren mutasyonlar
	Medyan	Ortalama Rank	Medyan	Ortalama Rank	Medyan	Ortalama Rank	Medyan	Ortalama Rank		
Eritropoietin	7,3	229,3	6,0	296,7	21,5	351,3	10,1	267,5	0,059	
Ferritin	74,2	362,2	65,3	335,9	71,5	327,7	76,4	386,6	0,382	
Hemoglobin	15,8	507,9	16,6	599,7	14,4	410,4	13,4	32,5	0,000*	<i>JAK2</i> mutasyonları
Hematokrit	4,7	511,7	49,9	605,4	42,7	408,6	40,7	321,7	0,000*	<i>JAK2</i> mutasyonları
Eritrosit	5,4	507,7	5,7	607,6	5,0	409,9	4,7	342,0	0,000*	<i>JAK2</i> mutasyonları
Lökosit	7,8	508,9	8,0	527,8	7,8	514,9	7,2	483,0	0,688	
Trombosit	247,0	512,0	236,0	491,5	287,5	554,0	278,0	522,4	0,567	
MCH	9,3	505,9	9,3	512,8	9,9	548,9	9,8	527,0	0,749	
MCV	87,2	509,7	87,6	514,5	88,2	526,8	87,8	504,1	0,974	
MCHC	33,0	508,3	33,0	522,9	33,1	530,2	33,0	492,4	0,819	
RDW	14,7	502,2	14,6	477,1	15,6	603,5	15,8	618,2	0,000*	<i>CALR</i> ve <i>MPL</i>
PDW	19,3	337,7	19,2	344,3	20,1	417,6	19,7	400,6	0,243	
MPV	8,5	507,0	8,6	504,8	7,9	418,9	8,0	438,3	0,053	
Nötrofil	4,6	507,9	4,6	517,3	4,8	538,4	4,8	505,3	0,896	
Lenfosit	2,2	506,3	2,3	554,0	2,0	466,8	2,0	465,0	0,052	
Monosit	0,6	512,1	0,6	525,8	0,6	503,6	0,5	444,0	0,184	
Eozinofil	0,2	517,1	0,2	534,6	0,2	448,2	0,1	411,9	0,004*	<i>JAK2</i> mutasyonları
Bazofil	0,1	492,7	0,1	545,1	0,1	550,7	0,1	527,2	0,089	

\*p<0,05; MCH: Ortalama eritrosit hemoglobin; MCV: Ortalama hücresel hacim; MCHC: Ortalama eritrosit hemoglobin konsantrasyonu; RDW: kırmızı kan hücresi dağılım genişliği; PDW: Trombosit dağılım genişliği; MPV: Ortalama trombosit hacmi.

Yapılan istatistiksel karşılaştırmalar sonucunda; hemoglobin, hematokrit, eritrosit, kırmızı kan hücresi dağılım genişliği ve eozinofil değerleri için yapılan mutasyon taramaları arasında istatistiksel olarak anlamlı fark bulunurken ( $p=0,004<0,05$ ), diğer klinik değerleri bakımından istatistiksel anlamlı fark saptanmamıştır ( $p=0,089>0,05$ ). Analiz sonucunda çıkan istatistiksel anlamlı farklılıklar ışığında; eozinofil, hemoglobin, hematokrit ve eritrosit değerleri için *JAK2* mutasyonlarda yüksek olarak saptandığından, bu hastalarda belirtilen mutasyonların incelenmesi gerektiği, eritrosit dağılım genişliğinde saptanan anlamlılık doğrultusunda da hastaların *CALR* ve *MPL* mutasyonlarına ilişkin taramaya yönlendirilmesi gerektiği düşünülmektedir. Böylece, [Şekil 1](#)'de verilen DSÖ'nün hastalık tanı şemasını destekleyen sonuçların saptandığı değerlendirilmiştir.

## TARTIŞMA

Miyeloproliferatif hastalıklar ile ilgili yapılan literatür taramalarında, Uysal ve ark. tarafından yapılan çalışmada, miyeloproliferatif hastalıkların tanısının saptanmasında kullanılan mutasyon taramaları ile konulan tanının karşılaştırılması yapılmıştır. Bu kapsamda, *JAK2 V617F*, *CALR* ve *MPL* olmak üzere 3 mutasyon incelenerek, bu 3 mutasyonun miyeloproliferatif hastalıklarda en sık tekrarlanan mutasyon olduğu değerlendirilmiştir.<sup>30</sup> Zulkeflee ve ark.nın yapmış olduğu benzer çalışmada, yine aynı 3 mutasyon incelenmiş olup, mutasyonun varlığında konulan tanılara ilişkin değerlendirme yapılmıştır. Bu çalışmada, bir öncekinden farklı olarak, *JAK2 V617F* ve *CALR* mutasyonlarının hastalık tanısında önemli rolü olduğu saptanmıştır.<sup>31</sup> Li ve ark. tarafından yapılan çalışma kapsamında ele alınan *JAK2 V617F*, *JAK2 Exon 12*, *CALR* ve *MPL* mutasyonları incelenmiştir. Yapılan bu çalışmada, diğer çalışmalara benzer olarak, mutasyonların tanı ile karşılaştırması yapılarak sonuçlandırılmıştır.<sup>32</sup> “*JAK2* ve *MPL* mutasyonu negatif esansiyel trombositoz ve primer miyelofibroz hastalarında kalretikulin (*CALR*) mutasyon durumunun belirlenmesi” isimli tez çalışmasında, *JAK2 V617F*, *CALR* ve *MPL* mutasyonlarının varlığı incelenerek *JAK2 V617F* ve *MPL* mutasyon varlığının negatif çıktığı durumlarda *CALR* mutasyon varlığının incelenmesinin hastaya zaman kazandırdığı sonucuna varılmıştır (Öte, 2015). Bahsi ve Yiğenoğlu'nun yapmış olduğu araştırmada, *JAK2 V617F*, *CALR* ve *MPL* mutasyonlarının varlığı incelenerek mevcut hasta gruplarındaki durum analizi yorumlanmıştır.<sup>1</sup> Lang ve ark., (2018) *JAK2 V617F*, *CALR* ve *MPL* mutasyonları ile Çin'deki Uygur ve Han uyruklu hastaların klinik özellikleri arasında ilişkilendirme yaptıkları çalışmalarında, 2 uyruk arasında saptanan farkları değerlendirmişlerdir.<sup>33</sup>

Daha önceki çalışmalarda, hastalık ile mutasyon varlığı araştırılırken, ele alınan bu çalışmada hastanın klinik bilgileri ile mutasyonu tespit edebilecek bir model oluşturulması amaçlanmıştır.

İncelenen modelde, *JAK2 V617F*, *JAK2 Exon 12*, *CALR* ve *MPL* mutasyonları ele alınmıştır. Çalışma kapsamında oluşturulan model ile klinisyene doğruluk oranlarıyla değerlendirilen 4 mutasyondan 1'i önerilecektir. Önerilen modelden elde edilen doğruluk oranlarının %60 civarında olduğu göz önüne alınarak, daha sonra yapılacak çalışmalarda örneklem büyüklüğü daha büyük alınarak doğruluk değerinin yükseltilebileceği ön görülmektedir.

## SONUÇ

Miyeloproliferatif hastalıklar, bireyin kemik iliğinde oluşan bozukluklar olarak tanımlanmaktadır. Bu bozukluklarda hücre öncüllerinin aşırı miktarlarda üretimi, olgun hücre sayısında artışa ve buna bağlı olarak diğer kan hücrelerinin sayısında da artış veya azalışlara yol açmaktadır.

Miyeloproliferatif hastalık tanısı ile kliniğe başvuru yapmış hastalar için DSÖ'nün belirlemiş olduğu kriterler doğrultusunda, klinisyen tarafından genetik tarama testlerine yönlendirilerek hastalığın tanısı konulmaktadır. *JAK2 V617F* mutasyonu ile hastalık şiddeti arasında ilişki olduğu ve birçok çalışmada bu mutasyonun pıhtılaşma oluşumunda etken olduğu değerlendirilmiştir.<sup>34</sup> Bu kapsamda, kliniğe başvuran hastalarda *JAK2 V617F* mutasyonunun varlığı araştırılarak tanı konulmaktadır. Mutasyonun durumu tespit



edildikten sonra sonuç pozitif ise tanı konulmakta, sonuç negatif ise diğer mutasyonların varlığı araştırılmaktadır.

Çalışma kapsamında, miyeloproliferatif hastalıklarında gözlenen *JAK2 V617F*, *JAK2 Exon 12*, *CALR* ve *MPL 4* mutasyon ele alınmıştır. Hastaların kemik iliği ve tam kan sayımı sonuçlarından oluşturulan bir model ile hangi mutasyonun mevcut olduğunun araştırılması gerektiği belirlenmeye çalışılmıştır. Gerçekleştirilen bu prosedür, başvuru yapan hastalar için hem maliyetli hem de uzun zaman alan bir süreç olarak ortaya çıkmaktadır. Bu doğrultuda, çalışma kapsamında bir modelleme kurgulanmıştır. Analizler sonucu yapılan istatistiksel değerlendirmeler, hastalık tanı sürecinde izlenen yöntemi desteklenmektedir. Sonuç olarak çalışmanın klinisyene hızlı bir yönlendirme sağlanmasının yanı sıra kliniğe başvuran hastalara hem zaman hem maddi yükümlülükler bakımından kolaylık sağlaması ön görülmektedir.

### Finansal Kaynak

*Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyebilecek maddi ve/veya manevi herhangi bir destek alınmamıştır.*

### Çıkar Çatışması

*Bu çalışma ile ilgili olarak yazarların ve/veya aile bireylerinin çıkar çatışması potansiyeli olabilecek bilimsel ve tıbbi komite üyeliği veya üyeleri ile ilişkisi, danışmanlık, bilirkişilik, herhangi bir firmada çalışma durumu, hissedarlık ve benzer durumları yoktur.*

### Yazar Katkıları

**Fikir/Kavram:** Feride İffet Şahin, Bülent Çelik; **Tasarım:** Gözde Kubat; **Denetleme/Danışmanlık:** Feride İffet Şahin, Bülent Çelik; **Veri Toplama ve/veya İşleme:** Feride İffet Şahin, Gözde Kubat; **Analiz ve/veya Yorum:** Gözde Kubat, Feride İffet Şahin, Bülent Çelik; **Kaynak Taraması:** Gözde Kubat, Feride İffet Şahin, Bülent Çelik; **Makalenin Yazımı:** Gözde Kubat, Feride İffet Şahin, Bülent Çelik; **Eleştirel İnceleme:** Feride İffet Şahin, Bülent Çelik.

## KAYNAKLAR

- Bahsi T, Yiğenoğlu TN. Myeloproliferatif neoplazilerde CALR, JAK2 ve MPL gen mutasyonlarının sıklığının ve birlikteliğinin değerlendirilmesi; Tek merkez deneyimi [CALR, JAK2 and MPL genes mutations in myeloproliferative neoplasms, Single center experience]. Acta Oncol Tur. 2019;52(3):388-92. [\[Crossref\]](#)
- Rumi E, Cazzola M. Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms. Blood. 2017;129(6):680-92. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
- Türkiye Hematoloji Derneği [Internet]. Türk Hematoloji Derneği İktisadi İşletmesi © 2008. [Erişim tarihi: 04 Kasım 2021]. Miyeloproliferatif hastalıklarda tanı ve tedavi kılavuzu. Erişim linki: [\[Link\]](#)
- Saeidi K. Myeloproliferative neoplasms: current molecular biology and genetics. Crit Rev Oncol Hematol. 2016;98:375-89. [\[Crossref\]](#) [\[PubMed\]](#)
- Michiels JJ, Tevet M, Trifa A, Niculescu-Mizil E, Lupu A, Vladareanu AM, et al. 2016 WHO Clinical Molecular and Pathological Criteria for Classification and Staging of Myeloproliferative Neoplasms (MPN) Caused by MPN Driver Mutations in the JAK2, MPL and CALR Genes in the Context of New 2016 WHO Classification: prognostic and Therapeutic Implications. Maedica (Bucur). 2016;11(1):5-25. [\[PubMed\]](#) [\[PMC\]](#)
- Filiz E, Karaboğa HA, Akoğul S. BIST-50 Endeksi değişim değerlerinin sınıflandırılmasında makine öğrenmesi yöntemleri ve yapay sinir ağları kullanımı [BIST-50 Index change values classification using machine learning methods and artificial neural networks]. Çukurova University Institute of Social Sciences. 2017;26(1):231-41. [\[Link\]](#)
- Çağlayan Akar E. Ekonometride yeni bir ufuk: Büyük veri ve makine öğrenmesi [A new horizon in econometrics: Big data and machine learning]. Social Sciences Research Journal. 2018;7(2):41-53. [\[Link\]](#)
- Aksu G, Doğan N. Veri madenciliğinde kullanılan öğrenme yöntemlerinin farklı koşullar altında karşılaştırılması [Comparison of learning methods used in data mining under different conditions]. Ankara University Journal of Faculty of Educational Sciences. 2018;51(3):71-100. [\[Crossref\]](#)
- Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. J Med Imaging Radiat Sci. 2019;50(4):477-87. [\[Crossref\]](#) [\[PubMed\]](#)
- Murphy KP. Machine Learning: a Probabilistic Perspective. 1st ed. Massachusetts: MIT Press; 2012.
- Frank SM, Qi A, Ravasio D, Sasaki Y, Rosen EL, Watanabe T. Supervised learning occurs in visual perceptual learning of complex natural images. Curr Biol. 2020;30(15):2995-3000.e3. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
- Vembandasamy K, Sasipriya R, Deepa E. Heart diseases detection using naive bayes algorithm. IJSET. 2015;2(9):441-4. [\[Link\]](#)
- Zhang Z. Naïve Bayes classification in R. Ann Transl Med. 2016;4(12):241. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
- Güldal H, Çakıcı Y. Ders yönetim sistemi yazılımı kullanıcı etkileşimlerinin sınıflandırma algoritmaları ile analizi [Analysis of course management system software users' interactions using classification algorithms]. Journal of Graduate School of Social Sciences. 2017;21(4):1355-67. [\[Link\]](#)

15. Konieczny J, Stojek J. Use of the K-Nearest neighbour classifier in wear condition classification of a positive displacement pump. *Sensors (Basel)*. 2021;21(18):6247. [[Crossref](#)] [[PubMed](#)] [[PMC](#)]
16. Türk Ö. Determination of emotional status from EEG time series by using EMD based local binary pattern method. *European Journal of Technique*. 2020;10(2):313-21. [[Crossref](#)]
17. Karakoyun M, Hacıbeyoğlu M. Biyomedikal veri kümeleri ile makine öğrenmesi sınıflandırma algoritmalarının istatistiksel olarak karşılaştırılması [Statistical comparison of machine learning classification algorithms using biomedical data sets]. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*. 2014;16(48):30-42. [[Link](#)]
18. Hu LY, Huang MW, Ke SW, Tsai CF. The distance function effect on k-nearest neighbor classification for medical datasets. *Springerplus*. 2016;5(1):1304. [[Crossref](#)] [[PubMed](#)] [[PMC](#)]
19. Fisher R. The Use of Multiple measurements in taxonomic problems. In *Annals of Eugenics*. 1936;7(2):179-88. [[Crossref](#)]
20. Ye Q, Fu L, Zhang Z, Zhao H, Naiem M. Lp- and Ls-norm distance based robust linear discriminant analysis. *Neural Netw*. 2018;105:393-404. [[Crossref](#)] [[PubMed](#)]
21. Hsu C, Chen MC, Chen LS. Intelligent ICA-SVM fault detector for non-Gaussian multivariate process monitoring. *Expert Systems with Applications*. 2010;37(4):3264-73. [[Crossref](#)]
22. Yu H, Kim S. SVM Tutorial - Classification, Regression and Ranking, Rozenberg G, Back T, Kok JN, editor. *Handbook of Natural Computing*. 1st ed. Berlin: Springer; 2012. p.479-506. [[Crossref](#)]
23. Chern CC, Chen YJ, Hsiao B. Decision tree-based classifier in providing telehealth service. *BMC Med Inform Decis Mak*. 2019;19(1):104. [[Crossref](#)] [[PubMed](#)] [[PMC](#)]
24. Başbağ M, Çağan E, Sayar M, Fırat M. Çayır-mera ve doğal alanlardan toplanan koca fiğ (*Vicia narbonensis* L.) türüne ait ot kalite değerlerinin belirlenmesi ve C5.0 türüne ait algoritmasına göre lokasyon sınıflandırmasının yapılması. Çelik Ş, editör. *Veri Madenciliği Yöntemleri: Tarım Alanında Uygulamaları*. 1. Baskı. Rating Academy Ar-Ge Yazılım Yayıncılık Eğitim Danışmanlık ve Organizasyon Tic. Ltd. Şti.; 2020. p.115-23.
25. Wu MT. Confusion matrix and minimum cross-entropy metrics based motion recognition system in the classroom. *Sci Rep*. 2022;12(1):3095. [[Crossref](#)] [[PubMed](#)] [[PMC](#)]
26. Alan A, Karabatak M. Veri seti-sınıflandırma ilişkisinde performansa etki eden faktörlerin değerlendirilmesi [Evaluation of the factors affecting performance on the data set-classification relationship]. *Firat University Journal of Engineering Science*. 2020;32(2):531-40. [[Crossref](#)]
27. Cesur S, Kınıklı S, Hatipoğlu ÇA, Arslan K, Karakök T, Demircan ŞA. Vankomisine dirençli enterokokların saptanmasında iki farklı kromojenik besiyerinin karşılaştırılması [Comparison of two different chromogenic agar for isolation of vancomycin-resistant enterococci]. *Turkish Journal of Clinics and Laboratory*. 2019;10(3):319-23. [[Crossref](#)]
28. Erdaş ÇB, Sümer E. A Machine learning-based approach to detect survival of heart failure patients. 2020 Medical Technologies Congress (TIPTEKNO); 2020 Nov 19-20; Online:IEEE; 2020. [[Crossref](#)]
29. Dilki G, Deniz Başar Ö. İşletmelerin iflas tahmininde k-en yakın komşu algoritması üzerinden uzaklık ölçütlerinin karşılaştırılması [Comparison study of distance measures using k- nearest neighbor algorithm on bankruptcy prediction]. *Istanbul Commerce University Journal of Science*. 2020;19(38):224-33. [[Link](#)]
30. Uysal A, Altınar Ş, Çelik S, Uysal S, Çebi AH. Genetic analysis of BCR-ABL negative chronic myeloproliferative diseases at initial diagnosis and their clinical effects. *Cukurova Medical Journal*. 2020;45(3):933-6. [[Crossref](#)]
31. Zulkeflee RH, Zulkafii Z, Johan MF, Husin A, Islam MA, Hassan R. Clinical and laboratory features of JAK2 V617F, CALR, and MPL mutations in Malaysian patients with classical myeloproliferative neoplasm (MPN). *Int J Environ Res Public Health*. 2021 ;18(14):7582. [[Crossref](#)] [[PubMed](#)] [[PMC](#)]
32. Li MY, Chao HY, Sun AN, Qiu HY, Jin ZM, Tang XW, et al. [Clinical significance of JAK2, CALR and MPL gene mutations in 1 648 Philadelphia chromosome negative myeloproliferative neoplasms patients from a single center]. *Zhonghua Xue Ye Xue Za Zhi*. 2017;38(4):295-300. [[PubMed](#)] [[PMC](#)]
33. Lang T, Nie Y, Wang Z, Huang Q, An L, Wang Y, et al. Correlation analysis between JAK2, MPL, and CALR mutations in patients with myeloproliferative neoplasms of Chinese Uyghur and Han nationality and their clinical characteristics. *J Int Med Res*. 2018;46(11):4650-9. [[Crossref](#)] [[PubMed](#)] [[PMC](#)]
34. Akgun Cagliyan G, Keskir A. JAK2V617F and platelet functions. *Clin Lymphoma Myeloma Leuk*. 2015;15(2):56-7. [[Crossref](#)]