

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**HEMATOLOJİK HASTALIKLARIN TEŞHİSİNDE
YAPAY ZEKA TEKNİKLERİNİN PERFORMANS
KARŞILAŞTIRMASI**

DOKTORA TEZİ

Tuba KARAGÜL YILDIZ

Enstitü Anabilim Dalı	:	BİLGİSAYAR VE BİLİŞİM MÜHENDİSLİĞİ
Tez Danışmanı	:	Doç. Dr. Nilüfer YURTAY
Ortak Danışman	:	Doç. Dr. Birgül ÖNEÇ

Mart 2021

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**HEMATOLOJİK HASTALIKLARIN TEŞHİSİNDE
YAPAY ZEKA TEKNİKLERİNİN PERFORMANS
KARŞILAŞTIRMASI**

DOKTORA TEZİ

Tuba KARAGÜL YILDIZ

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ**

Bu tez 05/03/2021 tarihinde aşağıdaki jüri tarafından oybirliği ile kabul edilmiştir.

Jüri Başkanı

Üye

Üye

Üye

Üye

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Tuba KARAGÜL YILDIZ

02.02.2021

TEŞEKKÜR

Doktora eğitimim boyunca yardımlarını esirgemeyen danışman hocalarım Doç. Dr. Nilüfer YURTAY'a ve Doç. Dr. Birgül ÖNEÇ'e teşekkürlerimi sunarım. Tez çalışmam boyunca bana rehberlik eden tez izleme komite üyeleri Prof. Dr. Emin GÜNDOĞAR ve Prof. Dr. Kürşat AYAN hocalarıma teşekkür ederim. Ayrıca anlayış ve yardımlarını esirgemeyen tüm çalışma arkadaşlarıma teşekkür ederim.

Hayatım boyunca desteklerini hissettiğim ve sahip olduğuma şükrettiğim çok değerli babam Refik KARAGÜL'e, annem Emel KARAGÜL'e, kardeşlerim Aslı ve Talha KARAGÜL'e ve eşim Bayram YILDIZ'a maddi ve manevi destekleri için minnettarım.

Bu çalışma Sakarya Üniversitesi Bilimsel Araştırma Projeleri (BAP) Komisyon Başkanlığınca (Proje No: 2014-50-01-027) desteklenmiştir. Çalışmada kullanılan veriler Düzce Üniversitesi Tıp Fakültesi İnvaziv Olmayan Klinik Araştırmalar Etik Kurulu'nun 2014/82 numaralı kararı ile insan sağlığına ait etik standartlara uygun olarak toplanmıştır.

İÇİNDEKİLER

TEŞEKKÜR	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ	iv
ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ	viii
ÖZET	x
SUMMARY	xi
BÖLÜM 1.	
GİRİŞ	1
BÖLÜM 2.	
KAYNAK ARAŞTIRMASI	6
BÖLÜM 3.	
HEMATOLOJİK HASTALIKLARIN TEŞHİSİ	35
3.1. Kanın Yapısı	36
3.1.1. Kırmızı küreler (Eritrositler)	37
3.1.2. Beyaz küreler (Lenfositler)	38
3.1.3. Kan pulcukları (Plateletler / Trombositler)	38
3.2. Kanın Üretimi	39
3.3. Hematolojik Hastalıkların Belirlenmesinde Kullanılan Laboratuvar Yöntemleri	40
3.3.1. Tam kan sayımı	41
3.3.2. Çevre (periferik) kan yayması	42

3.3.3. Kemik iliği incelemesi	43
3.4. Kan Hastalıkları	43
3.4.1. Anemi çeşitleri ve tanı kriterleri	45
 BÖLÜM 4.	
YAPAY ZEKA TEKNİKLERİ	51
4.1. Öznitelik Seçme Yöntemleri	54
4.1.1. Bilgi kazancı	55
4.1.2. Bilgi kazanç oranı	58
4.1.3. Temel bileşen analizi	61
4.1.4. Korelasyon tabanlı öznitelik altküme seçimi	63
4.2. Sınıflandırıcılar	64
4.2.1. Yapay sinir ağları	65
4.2.2. Destek vektör makineleri	67
4.2.3. Karar ağaçları	69
4.2.4. Naive Bayes	71
4.3. Değerlendirme	73
 BÖLÜM 5.	
ARAŞTIRMA BULGULARI VE TARTIŞMA	76
5.1. Veri Seti	76
5.2. Bulgular ve Tartışma	92
 BÖLÜM 6.	
SONUÇ VE ÖNERİLER	114
KAYNAKLAR	118
EKLER	132
ÖZGEÇMİŞ	134

SİMGELER VE KISALTMALAR LİSTESİ

AUC	: Eğri Altında Kalan Alan (Area Under Curve)
CFS	: Korelasyon Tabanlı Öznitelik Altkümesi Seçimi (Correlation-Based Feature Subset Selection)
CRP	: C-Reaktif Protein
D.BİLİRUBİN	: Direkt Bilirubin
DSÖ	: Dünya Sağlık Örgütü
DVM	: Destek Vektör Makineleri
FN	: Yanlış Negatif (False Negative)
FP	: Yanlış Pozitif (False Positive)
HCT	: Hematokrit
HGB	: Hemoglobin
I. BİLİRUBİN	: İndirekt Bilirubin
ICD	: Hastalıkların Uluslararası Sınıflaması (International Classification Of Diseases)
InfoGain	: Bilgi Kazancı (Information Gain)
InfoGainRatio	: Bilgi Kazanç Oranı (Information Gain Ratio)
KA	: Karar Ağaçları
MCH	: Ortalama Eritrosit Hemoglobini (Mean Corpuscular Hemoglobin)
MCHC	: Ortalama Eritrosit Hemoglobin Konsantrasyonu (Mean Corpuscular Hemoglobin Concentration)
MCV	: Ortalama Eritrosit Hacmi (Mean Corpuscular Volume)
NB	: Naive Bayes
NEUT	: Nötrofil Sayısı
P	: Kesinlik, Güven (Precision)
PCA	: Temel Bileşen Analizi (Principal Component Analysis)

PLT	: Trombosit Sayısı
R	: Geri Çağırma, Hassasiyet (Recall)
RBC	: Kırmızı Kan Hücresi Sayısı (Red Blood Cells)
RDW	: Kırmızı Hücre Dağılım Genişliği (Red Cell Distribution Width)
ROC	: Alıcı İşletim Karakteristiği (Receiver Operating Characteristic)
SQL	: Biçimlendirilmiş Sorgu Dili (Structured Query Language)
T. BİLİRUBİN	: Toplam Bilirubin
TIBC	: Toplam Demir Bağlama Kapasitesi (Total Iron Binding Capacity)
TN	: Doğru Negatif (True Negative)
TNR	: Doğru Negatif Oranı (True Negative Rate)
TP	: Doğru Pozitif (True Positive)
TPR	: Doğru Pozitif Oranı (True Positive Rate)
UIBC	: Bağlanmamış Demir Bağlama Kapasitesi (Unsaturated Iron Binding Capacity)
WBC	: Beyaz Kan Hücresi Sayısı (White Blood Cells)
YSA	: Yapay Sinir Ağları

ŞEKİLLER LİSTESİ

Şekil 3.1. Kanın bileşenleri	36
Şekil 3.2. Eritrositlerin yapısı	37
Şekil 3.3. Beyaz kan hücreleri ve kan pulcukları	38
Şekil 3.4. Kanın üretimi	40
Şekil 3.5. Çevre kan yayması	42
Şekil 3.6. Hematolojik hastalıkların sınıflandırılması	44
Şekil 3.7. Anemi belirtileri	46
Şekil 3.8. Aneminin işlevsel açıdan sınıflandırılması	48
Şekil 3.9. Aneminin eritrosit morfolojiksine göre sınıflandırılması	49
Şekil 4.1. Çalışmanın akış diyagramı	53
Şekil 4.2. Bilgi kazancına göre öznitelik ağırlıkları	58
Şekil 4.3. Bilgi kazanç oranına göre öznitelik ağırlıkları	60
Şekil 4.4. Temel bileşen analizine göre öznitelik ağırlıkları	63
Şekil 4.5. CFS yöntemine göre seçilen özniteliklerin ağırlıkları	64
Şekil 4.6. Önerilen yapay sinir ağının yapısı	65
Şekil 4.7. Yapay sinir ağının eğitim durumu performans grafiği.....	67
Şekil 4.8. Destek vektör makineleri yönteminin yapısı	68
Şekil 4.9. DVM eğitim süreci minimum sınıflandırma hatası	69
Şekil 4.10. Örnek bir karar ağacı yapısı	70
Şekil 4.11. Karar ağaçları eğitim süreci minimum sınıflandırma hatası	71
Şekil 4.12. Naive Bayes yöntemi eğitim süreci minimum sınıflandırma hatası ..	72
Şekil 5.1. Veri tablolarının oluşturulması ve içeriğinin düzenlenmesi	82
Şekil 5.2. Tabloların düzenlenmesi ve ilişkilerin kurulması.....	82
Şekil 5.3. Sorgu işlemleri ve sorgu sonucu elde edilen veri seti	83
Şekil 5.4. Açılış formu	84
Şekil 5.5. Veri giriş ve kayıt formu	84

Şekil 5.6. Veri seti hazırlama formu	85
Şekil 5.7. Çalışmanın aşamaları	91
Şekil 5.8. CFS-7 veri seti ile doğrusal DVM yönteminin karmaşıklık matrisi ...	99
Şekil 5.9. CFS-7 veri seti ile doğrusal DVM yönteminin ROC eğrisi grafiği	100
Şekil 5.10. CFS-7 veri seti ile karesel DVM yönteminin karmaşıklık matrisi ...	101
Şekil 5.11. CFS-7 veri seti ile karesel DVM yönteminin ROC eğrisi grafiği	102
Şekil 5.12. InfoGain-5 veri seti ile kübik DVM yönteminin karmaşıklık matrisi	103
Şekil 5.13. InfoGain-5 veri seti ile kübik DVM yönteminin ROC eğrisi grafiği.	104
Şekil 5.14. PCA-16 veri seti ile artırılmış KA yönteminin karmaşıklık matrisi ..	105
Şekil 5.15. PCA-16 veri seti ile artırılmış KA yönteminin ROC eğrisi grafiği ...	106
Şekil 5.16. CFS-7 veri seti ile torbalanmış KA yönteminin karmaşıklık matrisi .	107
Şekil 5.17. CFS-7 veri seti ile torbalanmış KA yönteminin ROC eğrisi grafiği ..	108
Şekil 5.18. CFS-7 veri seti ile Naive Bayes yönteminin karmaşıklık matrisi	109
Şekil 5.19. CFS-7 veri seti ile Naive Bayes yönteminin karmaşıklık matrisi	110
Şekil 5.20. Orijinal veri seti ile YSA yönteminin karmaşıklık matrisi	111
Şekil 5.21. Orijinal veri seti ile YSA yönteminin ROC eğrisi grafiği	112

TABLolar LİSTESİ

Tablo 4.1. Öznitelik seçimi sonrası oluşturulan veri setleri	55
Tablo 4.2. Bilgi kazancı yöntemi ile bulunan öznitelik ağırlıkları	57
Tablo 4.3. Bilgi kazanç oranı yöntemi ile bulunan öznitelik ağırlıkları	59
Tablo 4.4. Temel bileşen analizi yöntemi ile bulunan öznitelik ağırlıkları	62
Tablo 4.5. ROC analizi için karmaşıklık matrisi	73
Tablo 5.1. Veri setinin ham halinde yer alan anemi ilişkili tanılar ve ICD kodları	86
Tablo 5.2. Veri setinde bulunan öznitelikler ve özellikleri	87
Tablo 5.3. Veri setinde bulunan verilerin tanıya göre dağılımı	89
Tablo 5.4. Modellerin 25 öznitelik içeren orijinal veri seti için elde edilen performans değerleri	92
Tablo 5.5. Modellerin 7 öznitelik içeren CFS veri seti için elde edilen performans değerleri	93
Tablo 5.6. Modellerin 3 öznitelik içeren bilgi kazancı veri seti için elde edilen performans değerleri	93
Tablo 5.7. Modellerin 5 öznitelik içeren bilgi kazancı veri seti için elde edilen performans değerleri	94
Tablo 5.8. Modellerin 13 öznitelik içeren bilgi kazancı veri seti için elde edilen performans değerleri	94
Tablo 5.9. Modellerin 7 öznitelik içeren bilgi kazanç oranı veri seti için elde edilen performans değerleri	95
Tablo 5.10. Modellerin 11 öznitelik içeren bilgi kazanç oranı veri seti için elde edilen performans değerleri	95
Tablo 5.11. Modellerin 16 öznitelik içeren bilgi kazanç oranı veri seti için elde edilen performans değerleri	96
Tablo 5.12. Modellerin 16 öznitelik içeren temel bileşen analizi veri seti için elde edilen performans değerleri	97

Tablo 5.13. Modellerin veri setlerine göre doğruluk deęerleri	97
Tablo 5.14. Veri setlerinin modellere göre doğruluk deęerleri	98



ÖZET

Anahtar kelimeler: Anemi, kan hastalıkları, sınıflandırma, veri madenciliği, yapay öğrenme

Hematolojik hastalıkların en yaygın görülen grubu anemi ve anemi ilişkili hastalıklardır. Dünya Sağlık Örgütü'nün kırmızı kan hücrelerinin veya oksijen taşıma kapasitelerinin yetersiz kaldığı bir durum olarak tanımladığı anemi, özellikle kadınları ve okul öncesi çağıdaki çocukları etkilemektedir. Hayat kalitesini önemli oranda düşüren aneminin hem bir hastalık hem de ciddi hastalıklara eşlik eden bir bulgu olması nedeniyle tedavisi çok önemlidir. Etkin bir tedavi için hatasız bir tanı koyma süreci de gerekmektedir. Artan hasta sayısı ve hastanelerdeki yoğunluk, hastaların uzman hekime ulaşmalarını güçleştirmektedir. Bu güçlükler nedeniyle genel sağlık hizmetleri koşullarında anemi tanımayı sağlayacak bir sistem kullanışlı olacaktır. Bu sayede birinci basamak sağlık hizmeti sunulan merkezlerde tanı için istenen ilk tetkiklerle beraber geliştirilen sistemin işletilmesi ile bu sağlık merkezlerinde çalışan uzman olmayan personel daha etkin yönlendirme yapacaktır.

Bu çalışmada, anemi ve anemi ilişkili hastalıklara tanı koyma sürecinin yapay öğrenme yöntemleri ile gerçekleştirilmesi hedeflenmiştir. Bu çalışma için 12 farklı anemi çeşidine tanı koyabilen bir yapay öğrenme mimarisi geliştirilmiştir. İlk olarak geliştirilen bir ara yüz yardımıyla üzerinde herhangi bir sayısal işlem yapılmamış tamamen orijinal veriler elde edilmiştir. Veriler Düzce Üniversitesi Araştırma ve Uygulama Hastanesi'nden etik kurul izni ile alınmıştır. Sonraki aşamada, bu verilerin ağırlıkları, Bilgi Kazancı, Bilgi Kazancı Oranı, Korelasyon ve Temel Bileşen Analizi teknikleri kullanılarak hesaplanmıştır. Bu hesaplamaların ardından hem orijinal veri seti ile hem de elde edilen öznitelikler kullanılarak oluşturulan yeni veri setleri ile yapay öğrenme tekniklerinden Destek Vektör Makinaları, Karar Ağaçları, Naive Bayes ve Yapay Sinir Ağları yöntemleri uygulanarak sonuçlar değerlendirilmiştir. Tüm modellerde çapraz doğrulama yöntemi kullanılmış ve sonuçlar doğruluk, sınıflandırma hatası, duyarlılık, hassasiyet ve f1-skoru performans göstergeleri ile değerlendirilmiştir.

Araştırmada elde edilen bulgulara göre, anemi çeşitlerine tanı koymak için yapay öğrenme yöntemlerinin başarılı sonuçlar verdiği görülmüştür. Ayrıca öznitelik seçme yöntemlerinin sınıflandırma yöntemlerinde başarıyı arttırdığı fakat bu çalışma özelinde yapay sinir ağlarının en başarılı sonucu orijinal veri seti ile verdiği sonucuna varılmıştır.

PERFORMANCE COMPARISON OF ARTIFICIAL INTELLIGENCE TECHNIQUES IN THE DIAGNOSIS OF HEMATOLOGICAL DISEASES

SUMMARY

Keywords: Anemia, artificial learning, classification, data mining, hematological diseases.

The most common hematological diseases are anemia and anemia-related diseases. Anemia, defined by the World Health Organization as a condition in which red blood cells or their oxygen-carrying capacity are insufficient, especially affects women and preschool children. Since anemia, significantly reduces the quality of life, is both a disease and a symptom accompanying serious diseases, its treatment can be vital. An accurate diagnosis process is also required for effective treatment. The increasing number of patients and the congestion in hospitals make it difficult for patients to reach expert medical doctors. Because of these difficulties, a system to recognize anemia in general practice conditions would be useful. Using this system with the first examinations requested for diagnosis in the primary health care services are provided, non-specialist personnel working in these health centers will provide more effective guidance.

This study aims to perform the diagnosis of anemia and anemia-related diseases with artificial learning methods. For this study, an artificial learning architecture that can diagnose 12 different types of anemia developed. With the help of an interface first, completely original data without any numerical intervention obtained. The data are taken from Düzce University Research and Application Hospital with the permission of the ethics committee. At the next step, the weights of these data calculated using correlation, information gain, information gain ratio, and principal component analysis techniques. After these feature selection calculations, the results are evaluated by applying the methods of Support Vector Machines, Decision Trees, Naive Bayes, and Artificial Neural Networks from artificial learning techniques with the data sets created using both the original data set and the obtained features. The cross-validation method is used in all models and the results are evaluated with accuracy, classification error, AUC, recall, precision, and f-score performance indicators.

According to the results obtained in this study, it is seen that artificial learning methods give successful results to diagnose types of anemia. Besides, feature selection methods increase the success in classification methods, but specific to this study, the artificial neural networks gave the most successful result with the original dataset.

BÖLÜM 1. GİRİŞ

Teknolojinin hızla geliştiği günümüzde bilgisayarlar ve bilgisayarlı sistemler hayatımızın her alanında bir ihtiyaç haline gelmiştir. Özellikle insan hayatının söz konusu olduğu sağlık alanındaki hızlı değişimler, bu alanda çeşitli bilgisayar sistemlerinin kullanılmasını zorunlu hale getirmiştir. Günümüzde hasta sayısının yoğunluğu, takip gerektiren hastalıkların çoğalması, hastanede tedavi oranındaki ve sağlık giderlerindeki artış, hastanede kalma süresinin uzunluğu gibi sebeplerle tıp alanında bilgisayarlı sistemlere duyulan ihtiyaç giderek artmaktadır [1]. Bütün bu gelişmeler göz önünde bulundurulduğunda disiplinler arası bir konu olan tıp bilişimi, üzerinde çok çalışılan, dikkat çekici bir konu haline gelmiştir.

İnsanlara tıpta bilgisayar kullanımı dendiğinde genellikle doktorun yaptığı işleri yapabilecek bir cihaz veya bilgisayar programı akla gelir. Günümüzde sağlık alanında sağlık personelinin yükünü hafifletecek hasta kayıt, hasta yatış, raporlama gibi işlemler için geliştirilen hastane bilgi sistemleri ve hastaya verilen hizmetin kalitesini arttırma, tıbbi hataları önleme, uygun tedavi süreci ve tedavi maliyetlerinin azaltılması gibi işlemler için geliştirilen karar destek sistemleri gibi çeşitli bilgisayar yazılımları kullanılmaktadır. Ülkemizde hâlihazırda kullanılan yazılımlar sağlık personeline hasta girişi, kaydı ve yönetiminde önemli kolaylıklar sağlamaktadır. Bununla birlikte literatürde tanı ve tedavide doktorlara karar desteği sağlamaya yönelik çalışmalara sıklıkla rastlanmaktadır. Fakat böyle sistemlerin aktif olarak kullanımı ülkemizde henüz yaygınlaşmamıştır.

Sağlık alanında en önemli işlemlerden biri hasta bireye tanı koymak ve doğru bir tedavi yolu izlemektir. Türk Dil Kurumu tanı kelimesini “hastalığın ne olduğunu araştırıp ortaya koyma, tanılama, teşhis” ve “belirtiler ve bulgulara göre hastalığın

belirlenmesi” şekillerinde tarif etmiştir [2]. Hangi hastalık söz konusu olursa olsun doğru tanı koymak ve tedavi sürecine erken başlamak hastaların iyileşme ve tedavinin başarılı gerçekleşme oranını arttıracaktır. Erken tanı hemen hemen bütün hastalıklarda önemli olmakla birlikte örneğin kanser çeşitleri gibi ölümcül ve otizm gibi bütün yaşam tarzını etkileyecek hastalıklarda özellikle hayat kurtarmaktadır. Ülkemizde doktorların yoğun çalışma saatleri göz önünde bulundurulduğunda tanı koymak için gerekli zamanı kısaltmak ve hatalı tanı oranını en düşük seviyeye getirmek için ilerleyen teknolojiden faydalanmak gerekmektedir. İnsan hayatının söz konusu olduğu sağlık sistemlerinde bu yönde çalışmalar yapılmakta ve birçok farklı yöntem denenmektedir.

Tanı süreci hangi soruların sorulacağı, hangi testlerin isteneceği veya hangi prosedürlerin yerine getirileceği ve sonuçların riskleri ve finansal maliyetleri ile ilişkili değerlerini belirlemeye karar vermedir. Bununla birlikte tanı koyma yalnızca bir hasta hakkında neyin doğru olduğuna karar vermeyi değil aynı zamanda neyin doğru olduğunu belirlemek için hangi verilere ihtiyaç duyulduğunu belirlemeyi de içerir. Tanı bilinirken bile doktorun bilgi ve tecrübesini ölçen çeşitli yönetim kararları da vardır. Hastayı tedavi etmeli mi, kendi kendine iyileşmesini mi beklemeli? Tedavi uygulanacaksa nasıl olmalı? Tedavi süresince hastanın tepkisi normal mi, alternatif yaklaşımlar denemeye ihtiyaç var mı? Hatta bazı durumlarda, verdiğim ilk tanı kararı yanlış mıydı? Tıp alanında özellikle hastaları ve hastalıkları içermeyen karar görevleriyle de oldukça sık karşılaşılmaktadır [3].

Bir doktor tanı koyma sürecinde hastaya yaptığı fiziksel muayenenin yanı sıra hastanın şikayetlerini ve öyküsünü öğrenmeye yönelik birtakım sorular sorar ve gerekli gördüğü takdirde çeşitli tahlillerin (kan, idrar vb) tıbbi laboratuvar sonuçlarından ve MR, röntgen, ultrason, tomografi vb. tıbbi ölçüm cihazlarının yardımından faydalanmaktadır. Buna rağmen farklı doktorların aynı hasta hakkında verdikleri kararlar farklı olabilmekte, aynı sonucu farklı doktorlar farklı şekillerde yorumlayabilmektedirler. Bu, büyük ölçüde tecrübe faktöründen kaynaklanmaktadır. Bu noktada doktorlar ve doktor adayları için karar verme sürecinde, uzman doktorların

meslekteki tecrübelerinden faydalanmaları, daha erken bir zamanda doğru tanı koymaları için yardımcı olacaktır.

Bu tez çalışmasında uzman olmayan sağlık personeline, tıp öğrencilerine, yeni mezun ve uzmanlık yapmakta olan doktorlara yardımcı olacak bir bilgisayar sistemi geliştirilmesi düşünülmüştür. Geliştirilecek sistemin bir hastane bilgi sistemi şeklinde bütün birimlerde uygulanabilir olması düşünülse de böyle kapsamlı bir çalışma için ekip çalışması yapmak gerekir. Bu çalışma bireysel bir tez çalışması olduğu için ekip kurmak mümkün olamadı, ancak ileride bu düşünce gerçekleştirilebilir. Ayrıca bu sistem için hastanelerden ve uzman doktorlardan destek almak çok gereklidir, fakat maalesef sağlık çalışanlarının böyle akademik çalışmalara vakit ayırmaları mümkün olmamaktadır. Bu nedenlerden dolayı çalışma pilot olarak Düzce Üniversitesi Eğitim ve Araştırma Hastanesi'nde görev yapan Sayın Doç. Dr. Birgül ÖNEÇ'in yönlendirmeleri ve yardımları ile anemi hastalıkları özelinde gerçekleştirilmiştir.

Çalışmamızın amacı hematolojik hastalıklardan anemi çeşitlerine tanı koyulması sürecini önerilen sisteme aktararak sistemi eğitmek ve uzman doktorun karar verme sürecini kopyalamaktır. Böylelikle yeni hastalar geldiğinde sistemin uzman doktor gibi karar vermesinin de sağlanması ve bu yeteneğinin ölçülmesi düşünülmüştür. Bunun yanında doktora karar desteği sağlamak da bir diğer çıktı olmuştur. Bu çalışmanın uzman olmayan sağlık çalışanlarına faydalı olmasının yanında, literatüre önemli katkılar da sağlayacağına inanıyoruz.

Önerilen çalışmanın ilk aşamasında, tasarlanan sisteme uzman doktorun tanı koyma sürecinde kullandığı veriler ve tanıları girilerek sistem eğitilmiştir. Eğitim aşamasının sonrasında sistem yeni verilerle test edilmiştir. Test aşamasından sonra sistemin doktorun verdiği kararlar gibi karar verme yeteneği kazanma durumu kontrol edilerek başarısı analiz edilmiştir.

Çalışmanın gerçekleştirilmesi için kullanılan yazılımsal yöntemler şu şekilde tasarlanmıştır: Veri tabanı işlemleri için ücretsiz bir yazılım olan PostgreSQL [4]

kullanılmıştır. Verilerin hazırlandığı arayüz tasarımı Microsoft Visual Studio [5] programı ile gerçekleştirilmiştir. Kullanılan öznelik seçme ve yapay öğrenme yöntemleri ise MATLAB [6], RapidMiner [7] ve WEKA [8] programları üzerinde tasarlanmış ve uygulanmıştır.

Anemi dünyada en yaygın görülen kan hastalığıdır [9]. Dünya Sağlık Örgütü'ne (DSÖ) göre anemi kırmızı kan hücrelerinin sayılarının ve buna bağlı olarak oksijen taşıma kapasitelerinin vücudun ihtiyacını karşılamada yetersiz olduğu bir tıbbi durumdur [10]. Anemi, kandaki eritrosit kütesinin veya hemoglobin ve hematokrit konsantrasyonunun azalması şeklinde de tanımlanabilir. Kanda hemoglobin ve hematokrit değerlerinin normallığı yaş, cinsiyet gibi faktörlere göre değişiklik göstermekle birlikte belli sınırların altına düştüğünde anemi varlığından söz edilmektedir. Kiassebaum ve ark.'nın 2010 DSÖ küresel hastalık yükü çalışmasının verilerini kullanarak yaptığı çalışmada 189 ülke, 20 farklı yaş grubu incelenmiş ve küresel anemi sıklığı %32,9 olarak hesaplanmıştır. Anemi en fazla okul çağı öncesi çocuklarda ve kadınlarda görülmektedir. En yaygın çeşidi ise demir eksikliği anemisidir [11]. Anemi hem başlı başına bir hastalık hem de ciddi hastalıklara eşlik eden bir bulgudur. Bu nedenle hayat kalitesini fazlaca etkilemekte ve tedavisi hayati önem taşımaktadır. Doğru tedavi için ise doğru tanı koymak elzemdir. Bu tez çalışmasında çok sınıflı bu problem yapay öğrenme mimarileri kullanılarak çözülmüştür. Amacımız anemi tanısını genel sağlık hizmetleri şartlarında koyabilecek bir sistem tasarımı gerçekleştirmektir. Bu sistemin anemi tanısında kullanılan ilk testler ile birlikte kullanımı, uzman olmayan sağlık personeline anemi çeşitlerine tanı koyma konusunda yardımcı olacaktır. Artan hasta yükü ve uzmanlara erişimin güçlüğü anemi tanısının gecikmesine ve buna bağlı olarak yaşam kalitesinin ciddi oranda düşmesine neden olmaktadır. Böylelikle bu soruna çözüm bulunması hedeflenmiştir. Bu sistemin kullanılması ile ileri tetkikler gerektiren ve tedavi gerektiren hastaların daha hızlı ve daha kesin tanı ile tespit edilmesi sağlanacaktır.

Türkiye'nin Düzce ilinde en sık görülen 12 farklı anemi çeşidi, kullanılan yapay öğrenme yöntemleri ile sınıflandırılmıştır. Gerçek hasta verilerinin kullanılması, veriler üzerinde hiçbir sayısal müdahale olmaması ve verilerin tamamen orijinal

olması önemli bir noktadır. Ayrıca bir doktorun ilk etapta hastalık tanısına karar vermek için kullandığı niteliklerin kullanılması da bu çalışmayı önemli kılmaktadır. Yapay öğrenme yöntemleri hem orijinal veri seti üzerinde hem de çeşitli nitelik azaltma/seçme yöntemleri kullanılarak azaltılmış nitelikler ile oluşturulan yeni veri setleri üzerinde çalıştırılmış ve sonuçlar karşılaştırılmıştır. En yüksek başarı Karar Ağacı tabanlı yöntemler ile elde edilmiştir. Yapay sinir ağı yönteminin en başarılı sonucu orijinal veri seti ile vermesi dikkat çekici bir sonuç olarak karşımıza çıkmıştır.

Bu tez çalışması 6 bölümden oluşmaktadır. İlk bölümde tezin konusuna giriş yapılmış ve çalışmanın ana hatları verilmiştir. İkinci bölümde tıp bilişimi, hematoloji, anemi ve tez çalışmasında kullanılan yapay öğrenme yöntemleri ile ilgili yapılmış çalışmalar özetlenmiştir. Üçüncü bölümde hayati sıvımız olan kan ve kan hastalıkları, hastalıkların teşhisi için kullanılması gereken laboratuvar yöntemlerine değinilmiş, özellikle anemi çeşitleri ve tanı kriterleri hakkında bilgi verilmiştir. Ayrıca kan hastalıklarının sınıflandırılması ve anemilerin sınıflandırılması yapılmış, bu hastalıklara tanı koyma süreci anlatılmıştır. Dördüncü bölümde tez çalışmasında kullanılan yöntemler hakkında bilgi verilmiş ve önerilen sistem tanıtılmıştır. Beşinci bölümde çalışmada kullanılan verilerin nasıl elde edildiği anlatılmış, veri setinin içeriği detaylıca açıklanmıştır. Ayrıca yapılan çalışmalar sonucu elde edilen bulgulara yer verilmiş, alınan sonuçlar tartışılmıştır. Son olarak altıncı bölümde tez çalışmasının sonuçları, gelecekte yapılması planlanan çalışmalar ve öneriler sunulmuştur.

BÖLÜM 2. KAYNAK ARAŞTIRMASI

Tıp bilişimi literatürde “Tıbbi bilgi ve becerinin yönetimi, kontrolü, karar verme süreçleri ile analizi için sistem analizi araçlarını kullanarak işlem ve algoritmalar geliştiren bir bilim dalıdır.” [12] ve “Medikal enformatik veri işleme ve iletişim yöntemlerini tıp ve sağlık alanlarında elde edilen bilgi birikimi ile beraber kullanarak yapılan her tür teorik ve pratik uygulamadır.” [13] şeklinde tanımlanmıştır.

Literatürde tıp bilişimi biyomedikal bilişim olarak da geçmektedir. Biyomedikal veri, enformasyon ve bilginin bilimsel araştırma, problem çözme ve karar verme işlemleri için etkili kullanımını takip eden, insan sağlığını geliştirme çabalarıyla teşvik edilen disiplinler arası bir çalışma alanıdır. Bu konuda çalışanlar biyomedikal veriler, yöntemler, teoriler geliştirmekle kalmazlar aynı zamanda bütün bunların sağlık alanında etkili ve faydalı bir şekilde nasıl modelleneceğini ve kullanılacağını sorgularlar [14].

Bugün tıp alanında biyomedikal mühendisliği tıbbi görüntüleme, tıbbi tanı, radyasyon terapisi (kanser tedavisinde), organ desteği (örneğin periton diyalizi), tedavi edici fonksiyon (insülin üreten pankreatik adacık hücrelerinin diyabet tedavisi için kapsüllenmesi) gibi işlevlere sahiptir. Biyomedikal son zamanlarda popüler olmaya başlamış ve yeni bir alan zannedilse de esasında oldukça eski bir alandır. Ghista yaptığı çalışmada 1500’lerde Da Vinci’nin tıp alanındaki çalışmalarından, 1800’lü yıllarda sinir iletim dağılımının hızını tanımlayan Hermann Von Helmholtz’dan ve daha nicelerinden bahsetmiştir. Bir mühendislik alanı olarak 1960larda olgunlaşmaya başlayan biyomedikal konusunda bilgiler vermiş, üniversitelerde bu alanda verilen dersleri incelemiştir [15].

Hassanien ve ark. yaptıkları bir çalışmada hesaplamalı zekâ teknikleri hakkında ayrıntılı bilgiler vermişler ve biyobilişim alanındaki kullanım alanlarından bahsetmişlerdir. Yapay sinir ağları, Boltzmann makineleri, derin inanç ağları, destek vektör makineleri, bulanık mantık, geçici küme teorisi, tahminsel algoritmalar, genetik algoritmalar, yapay bağışıklık sistemleri ve sürü zekâsı (karınca kolonisi, parçacık sürü vb.) çalışmada bahsi geçen yöntemlerdir. Bu yöntemler ve bunların birlikte kullanılmalarıyla elde edilen hibrit yöntemlerin gen ifadesi kümelemesi ve sınıflandırılması, protein dizilerinin sınıflandırılması, protein yapı tahmini gibi alanlarda başarıyla kullanılmakta olduğu ifade edilmiş ve örnek çalışmalar sunulmuştur [16].

Sağlık profesyonelleri bilgisayarların ilk günlerinden beri tanı sürecinde makinelerin kendilerine yardımcı olacakları zamanı beklemişlerdir. Bu ihtimalle ilgili olarak ilk makaleler 1950lerin sonunda görülmüştür ve deneysel prototipler birkaç yıl içinde ortaya çıkmaya başlamıştır [17, 18]. 1970'lerdeki 3 danışma sistemi, LEEDS, MYCIN ve HELP, klinik karar destek sistemlerinin kökenindeki iş bakış açısına fayda sağlamışlardır [3]. Özata ve Aslan çalışmalarında klinik karar destek sistemleri hakkında bilgi vermişler, dünyada aktif olarak kullanılan bazı sistemlerden bahsetmişlerdir [19]. Tıp eğitiminde eğitsel araştırma için mikrobilgisayar tabanlı, olay güdümlü bir sistem tasarlayan Beck ve ark. çalışmalarına PlanAlyzer adını vermişlerdir. 1988'de kalp hastalıkları tanısı [20], 1989'da anemi tanısı [21] için yayınlanan PlanAlyzer sonraki yıllarda öğrencilerin yaygın bir tıbbi bozukluğa teşhis yaklaşımını ortaya çıkarmayı ve eleştirmeyi amaçlamıştır. 1993 yılında yayınlanan çalışmada Lyon ve ark. bu programın 7 yıl boyunca çeşitli geliştirmeler testler ve değerlendirmelerden sonra Dartmouth tıp okulunun kardiyoloji ve hematoloji birimlerinde anemi ve göğüs ağrısı teşhislerini öğretmek için kullandıklarını belirtmişlerdir [22].

Biyomedikal bilişimi araştırmalarında bilgisayar tabanlı karar araçlarının klinik karar vermeyi desteklemede kullanılmasıyla çok fazla karşılaşılmaktadır. Son 50 yıldaki araştırmalar değişik alanlardaki tıbbi karar problemlerinde çok çeşitli modelleme tekniklerinin kullanılmasıyla tıbbi karar desteği uygulamalarındaki gelişmelere sahne

olmuştur. Wagholikar ve ark. tanısal karar desteği için modelleme tekniklerinin karar doğruluğu odaklı olarak literatürünü araştırmışlardır [23]. Yazarlar bu konuda şunları önermişlerdir.

- a. Tıbbi karar destek uygulamalarının doğruluğunun geliştirilmesi, belirsiz ve geçici verileri modelleme, çıkarım algoritmaları üzerinde araştırma, hasta bilgisinin farklı kaynaklarla birleştirilmesi ve gen profili algoritmalarının geliştirilmesiyle mümkün olabilir.
- b. Tıbbi karar desteği araştırmaları tanımlanmamış tıbbi veri setlerinin ve açık kaynaklı veri madenciliği araçlarının geliştirilmesi ile kolaylaştırılabilir.
- c. Geliştiricilerin tıbbi problemlerdeki teknikleri seçebilmeleri için özelliklerini anlamalarında kılavuz olabilecek farklı modelleme tekniklerinin karşılaştırmalı sonuçlarına ihtiyaçları vardır.
- d. Klinik karar desteklerinin doktorların tıbbi karar desteği uygulamalarındaki değer tahminlerinden faydalanması ile beslenmesi ihtiyacı vardır.

Yarım yüzyıllık araştırmaya rağmen tıbbi karar desteği uygulamalarının geniş kabul görmeme ve sağlık alanında faydalanılmama sebeplerini ise şöyle sıralamışlardır;

- Karar doğruluğu: Teşhis performansı insan uzmanlar kadar iyi olan çok az uygulama mevcuttur.
- İş akışı ile bütünleşme: Bu araçlar doktorların ihtiyacı olan tüm farklı bilgi tiplerini sağlayamamakta ve bu durum doktorun kendi iş akışını da aksatabilmektedir.
- Kullanılabilirlik: Çoğu sistemin ara yüzü kullanıcı dostu değildir ve kullanımları için önce eğitilmeleri gerekmektedir.
- Doğal dil işleme (NLP): Çoğu hasta bilgisi yapılandırılmamış serbest metin biçimindedir. Böyle veriler NLP algoritmaları kullanılarak incelenmektedir. Fakat NLP yöntemleri hala klinik uygulamalar için yeterli değildir.
- Açıklama: Karar önerisinin temeli bilimsel kanıt sağlanarak açıklanırsa doktorlar sistemi daha istekli kullanırlar. Çoğu sistemde bu tarz açıklama modülleri eksik bırakılmıştır.

Yazarlar PUBMED veritabanında “bilgisayar yardımcı teşhis” olarak sorgu yapmışlar ve 2880 makale incelemişlerdir. Bu çalışmadaki eksiklik sadece basılı yayınların incelenmiş olmasıdır. Literatürdeki modelleme tekniklerinden Naive Bayes, Ledley ve Lusted’in çalışmalarından [17] incelenmiş, bu yaklaşımla yapılmış bir diğer çalışma olan Warner ve ark.’nın tasarımı [18] ve ilerlemiş haliyle karar destek modülüne sahip ilk hastane bilgi sistemi olan HELP [24] anlatılmıştır. DeDombal’in İngiltere’de abdominal hastalıkların belirti sürecini incelemek ve karar desteği sağlamak amacıyla gerçekleştirdiği karın ağrısı sistemi LEEDS [25], 1987’de dâhili tıp için geliştirilmiş ve halen web tabanlı bir sistem olarak kullanılmakta olan ve bilgi tabanında 2200 hastalık ve 5000 belirti yer alan DXPLAIN [26] sistemleri incelenmiştir. HEME hematolojik bozuklukların teşhisi için Bayes yaklaşımı ile geliştirilmiş bir sistemdir [27]. Kural tabanlı sistemlere örnek olarak Stanford Üniversitesinde enfeksiyon hastalıklarının teşhis ve tedavi yönetiminde kullanılmak üzere geliştirilmiş MYCIN [28], Weiss tarafından geliştirilmiş EXPERT [29] ve Politakis tarafından geliştirilmiş SEEK [30] sistemleri incelenmiştir. Kural tabanlı sistemlerin dezavantajı veri sayısı artınca kontrolden çıkmalarıdır. Sezgisel yöntemlerin çoğu yapay zekâ tabanlı programlardır. Bunlara CASENET [31], CHRONOS [32], Internist [33], CADUCEUS [34] gibi çalışmalar örnek olarak verilmiştir. CADIAG [35] bulanık küme teorisiyle gerçekleştirilen bir diğer örnektir. Tartışma kısmında karar doğruluğunun eksikliği, modelleme tekniği seçimi, klinik ve genomik hasta bilgisinin birleştirilmesi, geçici bilginin modellenmesi, bulanıklığın modellenmesi, değer tahmini ölçümleri, klinik ayarlamalarda değer tahmini, karar problemlerinin karmaşıklığı, yüksek başarılı hesaplamadan faydalanma, açık kaynaklı olma, serbest metin verisinin kullanılması gibi konular açıklanmıştır [23].

Fizyolojik işaretler tıbbi görüntüler ve biyosistemler bir hastanın sağlık durumuna erişmek için kullanılabilir ve doktorlar tedavi için bunlarla tanıyı desteklerler. Bilgisayar destekli tanı sağlık uygulamalarında otomatik karar verme görüntüleme ve saklı karmaşık özelliklerde klinik tanıyı desteklemeye yardım edebilir. Bu sistemler hastayla ilgilenmenin kalitesini arttırmaya ve cihaz hataları dâhil tüm hataları minimuma indirmeye odaklanırlar. Faust ve ark. [36] makalelerinde biçimsel ve model sürücülü tasarım yöntemlerinin bu gereklilikle karşılaşmış sistemlere öncülük

edebilme durumunu tartışmışlardır. Modelleme bilgisayar destekli tanı için yeni değildir ancak sistem tasarımının modellenmesi daha az araştırılmıştır. Bu nedenle seçilen sistem tasarım teknikleri, bilgisayar destekli tanı ve otomatik karar verme işlemlerinde daha somut örnekler incelenmiştir. Bilgisayar destekli tanının temel kavramı biyomedikal veri analizinin desteklenmesinde ikinci fikir olarak hizmet verecek bir bilgisayar çıktısı sağlamaktır. Bu açıdan bilgisayar destekli tanı sistemleri geliştirilmesinde doktor, radyolog, bilgisayar uzmanı, mühendisler, fizyolog ve istatistikçiler gibi farklı birikimleri olan kişilerin bir takım çalışması gereklidir. Ortaya çıkan geniş ölçekli performansı gözlemek için güvenilir bir yöntem kullanılması gerekir. ROC analizi yapmak bilgisayar destekli tanı uygulamalarında algoritma geliştirmek kadar önemlidir. Bilgisayar destekli tanıda kullanılan algoritmalar; yapay sinir ağları, destek vektör makinesi, olasılıksal sinir ağı gibi çeşitli algoritmalarla gerçekleştirilen çeşitli hastalık tanıları ve güvenilirlik hesaplamaları hakkında yapılan birçok çalışmayı incelemişler ve bu inceleme makalesinin ana fikrini biçimsel modelleme ve sistem mühendisliği kombinasyonunun tanı sistemlerinin güvenilirliğini geliştirecek kapasiteye sahip olduğu şeklinde açıklamışlardır [36].

Sepulveda ve Young 2013'te yayınlanan bir makalelerinde [37] geleceğin laboratuvar bilgi sisteminde olması istenen, klinik laboratuvarlara ve daha önemlisi sağlık hizmeti sağlayıcıları ile klinik laboratuvarlar arasındaki ara yüz işlemlerini optimize ederek hasta bakımındaki kaliteyi ve maliyet verimi arttırmayı amaçlayan önemli sayıda özelliği listelemişlerdir. İnsan hatalarından kaynaklanan problemleri ortadan kaldırmak amacıyla laboratuvar bilgi sistemlerinin kullanılmasını önermişlerdir. İnsanlar verimliliği optimize etmek ve üretimi en üst düzeye çıkarmak için yalın bir yaklaşımla tasarlanmış kullanıcı dostu ara yüzleri aracılığıyla laboratuvar bilgi sistemleriyle etkileşimli olacaklardır. Bu çalışmada belirtilen fikirler şu an kullanılan laboratuvar bilgi sistemlerinde kısmen mevcut fakat yapay zekâ, uzman sistemler, gelişmiş veri tabanları, veri madenciliği ve diğer bilgi teknolojilerinin birleşimini içeren kapsamlı, işlevsel, kullanıcı dostu ve klinik olarak kullanışlı ideal bir laboratuvar bilgi sistemi için hatırı sayılır bir çaba gerekmektedir. Bu çalışmada bilgi güvenliği, test istemi, örnek toplama, katılım ve işleme, analitik aşama, sonuç girişi ve geçerliliği, sonuç raporlama, duyuru yönetimi, veri madenciliği, çapraz raporlar,

yöntem geçerliliği, kalite yönetimi, yönetsel ve finansal konular ve diğer işlevsel konular başlıklarında açıklamalar mevcuttur [37].

Gerçekleştirilen tez çalışmasında yoğun olarak veri tabanı işlemleri gerekmektedir. Wang ve Ling 2012’de tıbbi kayıtları çoklu veri tabanından çekerken karşılaşılan veri eşleme probleminin çözümüne yönelik bir çalışma yapmışlardır. Yöntem olarak bulanık mantık kullanmışlar ve iyi sonuçlar almışlardır [38].

Stantica ve ark. 2012’de periyodik olarak tekrarlanan tıbbi verilerle uğraşmak için üstü kapalı bir yaklaşımda bulunmuşlardır. Tıpta geçici bilgiler çok önemli bir rol oynamaktadır ve tıp bilişiminde bunu desteklemek ve saklamak için uygun veri tabanı yaklaşımları konusunda artan bir farkındalık vardır. Yazarlar periyodik bağlantılı tıbbi veri ile üstü kapalı bir şekilde başa çıkmak için yenilikçi bir yaklaşım önermişlerdir. Bu, TSQ2 fikir birliği yaklaşımının istikrarlı bir uzantısı olan yeni bir veri modelidir. Yapılan deneylerde mevcut açık yöntemlerden daha başarılı olduğunu kanıtlamışlardır [39].

Seera ve Lim 2014’te tıbbi veri sınıflandırmak için bulanık sinir ağı sınıflandırma (fuzzy min-max) ve regresyon ağacı ve rastgele orman (random forest) modellerini içeren hibrit bir zeki sistem önermişlerdir. Tıbbi veri sınıflandırılması için karar destek aracı olarak etkinliği test edilmiş ve hibrit sistemler model bileşenlerinin avantajlarından istifade ederken sınırlılıklarını azalttıklarından yüksek performans elde edilmiştir. UCI makine öğrenmesi veri tabanından test için Wisconsin Göğüs Kanseri, Pima Hint Diyabeti ve Karaciğer Bozuklukları olmak üzere 3 farklı tescilli tıbbi veri seti kullanılmış ve sonuçlar değerlendirilmiştir. Tıbbi uygulamalarda doğruluk, hassasiyet, kesinlik ve eğri altında kalan alan gibi kullanışlı performans metrikleri olan ROC değerleri hesaplaması yapılmıştır. Deneysel sonuçlar hibrit zeki sistemin sınıflandırma veriminin iyi olduğunu göstermiştir [40].

Trindadea ve ark. 2012’deki bir çalışmalarında veri madenciliği ve grup analiz uygulamaları için bir tıbbi görüntü veri tabanı geliştirmişlerdir. Günlük olarak tıbbi

görüntüleme merkezlerinde onlarca gigabayt tıbbi görüntüleme bilgisi üretilir. Uygun bilgiyi etkili olarak kullanarak toplanan bilgiyi olabildiğince çabuk ve kolayca saklamak, işlemek ve analiz etmek gereklidir. Yazarlar bu çalışmalarında tıbbi görüntü bilgisini depolayan, birleştiren ve analiz eden bir platform geliştirmişlerdir. Çalışmaya gerekirse diğer yetkili kullanıcılar ile güvenli olarak paylaşma seçeneği de eklenmiştir [41].

Polat ve Güneş 2007’de yaptıkları bir çalışmada çok yaygın ve önemli bir hastalık olan hepatit tanısını bir makine öğrenme sistemi ile gerçekleştirmişlerdir. Önerilen yaklaşımın 3 aşaması vardır. İlki hepatit hastalığı veri setinin özelliklerinin, C4.5 karar ağacı algoritması kullanan bir alt program ile 19’dan 10’a indirilmesi, sonra veri setinin 0-1 arasına normalize edilmesi ve bulanık (fuzzy) ön işleme ile ağırlıklandırılması ve en son elde edilen bu değerlerin yapay bağışıklık tanıma sistemi (AIRS) ile sınıflandırılmasıdır. %94,12 doğruluk elde edilen bu sınıflandırma işlemi bu hastalık için literatürdeki diğer yöntemlere göre başarılıdır [42]. Yine başka bir çalışmalarında aynı hibrit yöntemi kullanarak UCI veri tabanındaki kalp ve hepatit hastalıklarının veri setleriyle bir tıbbi karar destek sistemi gerçekleştirmişlerdir. Sistemin elde edilen sınıflandırma doğrulukları, kalp hastalığı için 92.59% ve hepatit hastalığı için 81.82% olarak bulunmuştur [43].

Shouzhong ve Fanglu on yıldan uzun sürmüş bir çalışmalarında çok büyük bir bilgi tabanlı sistem (EBME- Büyük elektronik beyin bilgeliği) geliştirilmesini hedeflemişlerdir. Örnek olarak anemi alt sisteminin çalışma prensibi ve detayları verilmiştir. Amaçlarını; bilgi mühendisliği işlemlerini otomatikleştirerek verimi arttırmak, tekrarlanan iş gücünü mümkün olabildiğince önlemek, çok büyük bir bilgi tabanlı sistemde sadece farklı kullanıcıların farklı ihtiyaçlarıyla karşılaşmayı değil aynı zamanda kombinasyon patlamasının oluşumunu önlemede faydalı olan farklı alt sistemleri birleştirmek, tıbbi bilgi işleme standardizasyonu araştırmalarını geliştirmek şeklinde açıklamışlardır [44].

Sağlıkta bakım kalitesini arttırmak için gösterdikleri yeteneklere rağmen klinik karar destek sistemleri yaygın olarak kullanılmamaktadır. Bu sınırlı kullanım tıbbi bilginin makine tarafından yürütülebilir biçimde paylaşımının zorluğundan kaynaklanmaktadır. Kawamoto ve Lobach bu sorunu çözmek için SEBASTIAN olarak bilinen karar destek web servisini geliştirmişlerdir [45]. Bu sistemde bireysel bilgi modülleri bir hastayı değerlendirmek için veri gereksinimini, o veriyi kullanarak çıkarılabilecek sonuçları ve bu sonuçların nasıl üretileceğine dair talimatları tanımlar. Bu makalede, anlatılan sistemin tasarım hedefleri, çerçevenin tanımı, çalışmanın mimarisi açıklanmış, güncel kullanımdan çerçevenin değerlendirilmesi yapılmıştır. Tasarım hedefleri “bir kere yaz her yerde çalıştır” taşınabilirliğini tıbbi bilgi için sağlamak hem basit hem karmaşık bilgi modüllerinin basit bir şekilde yazılmasına izin vermek ve bu çerçeveyi kullanmak ve anlamak için gereken çabayı minimize etmek. Yazarlar bu hedeflere ek olarak ikincil hedeflerini ise şöyle sıralamışlardır: Gerekli yerlerde mevcut standartları kullanmak, uygun araçlardan yararlanmak, bilgi bakımını kolaylaştırmak, kapsamlı testleri desteklemek, çalışma performansını optimize etmek ve klinik açıdan önemli olan karar destek sistemlerini etkinleştirmek [45].

İnternet büyük ölçekli tıbbi uzman sistemler oluşturmak için benzersiz bir fırsat sunmaktadır. Bu sistemlerde farklı tıbbi alanlarda ve farklı coğrafi bölgelerde olan tıp uzmanlarından toplanan tıbbi bilginin büyük miktarda ve genelde karmaşık halde olmasıyla başa çıkabilmek için düşük maliyetli bilgi toplama ve yüksek ölçüde yönetim düzeni olması istenecektir. Yan ve ark. makalelerinde bir tıbbi bilgi toplama ve yönetme sisteminin üç katmanlı dağıtık istemci-sunucu mimarisi üzerine inşa edilebilir olduğunu göstermişlerdir. Alınan sonuçlar yöntemin internet üzerinde dağıtık tıbbi uzman sistemi oluşturmak, değiştirmek ve genişletmek için kalıcı bir çözüm olduğunu göstermiştir [46].

Dragusin ve ark. 2012’de nadir hastalıkların tanısı için web tabanlı bir arama motoru tasarlamışlardır. Sistemde 56 adet gerçek hayattaki tanısal vaka, performans ölçümleri, bilgi kaynakları ve Google arama motorunu bu görev için kişiselleştirme rehberi bulunmaktadır. Ek olarak FindZebra olarak adlandırdıkları açık kaynaklı arama teknolojisi ile güçlendirilmiş ve ücretsiz erişilebilen çevrimiçi tıbbi bilgi kullanan bir

nadir hastalık arama motoru gerçekleştirilmiştir. Sonuçlar FindZebra'nın Google arama motorunu varsayılan düzende ve FindZebra tarafından özelleştirilmiş kaynaklarda geçtiğini göstermektedir [47].

Tradigo ve ark. 2012'de sundukları bir çalışmada geniş bir nüfus hakkında sağlık verileri içeren klinik veriyi analiz etmek için Coğrafi bilgi sistemi- GIS teknolojisi kullanma deneyimlerinden bahsetmişlerdir. Geomedica adını verdikleri platform sağlık bilgilerine dair veri setlerini etkin olarak görselleştirmek ve analiz etmek için geliştirilmiş prototip bir sistemdir. Sağlıkta hem epidemiyolojik çalışmalarda hem kara ve uzay-zamansal analizde kullanılabilir. Alkolizm ve hipertiroid olmak üzere iki yeni geniş veri seti ile yeni testler gerçekleştirme çalışmaları sürmektedir [48].

2011 de DeMaio ve ark.'nın yaptığı çalışmada ODINO (ontolojik hastalık bilgisi) adı verilen bir sistem tanıtılmıştır. Bu sistemde tıbbi hastalık ve belirtileri arasındaki ilişkiler bulanık etiketler kullanılarak semantik tabanlı modellenmiştir. Tıbbi bilgi INMP (ulusal sağlık göçü ve yoksulluk enstitüsü) üyelerinden oluşan uzman hekimler tarafından tanımlanmıştır. ODINO'nun temel amacı hastalıkların aranması için ontolojileri ve kontrollü kelimelerin kullanıldığı etkili bir kullanıcı ara yüzü sağlamaktır. Çalışmada dermatolojik hastalıklar alanında bazı deneysel sonuçlar verilmiştir [49].

Bu alanda çalışılan bir diğer konu da klinik kılavuzlardır. Klinik kılavuzlar sağlık alanında geliştirme ve standart oluşturma için faydalı araçlardır. Isern ve Moreno yaptıkları bir çalışmada 2000-2007 yılları arasında, tıbbi veri tabanları ve uluslararası konferanslardan toparlanmış tıbbi bilişim alanındaki literatürü incelemişlerdir [50].

Bir bilgisayar ne kadar gelişmiş teknolojiye sahip olursa olsun bir kullanıcının yönlendirmesi olmadan kendi başına işlem yapma yeteneğine sahip değildir. Gelişen teknoloji bize bilgisayarlara istediğimiz işlemleri yaptırma olanağı sunmaktadır. Akıllı telefonlar tabletler derken teknolojiye her yerden erişebilme günlük hayatta işlerimizi kolaylaştırmaktadır. Daha zeki sistemler tasarlama, bilgisayarlara yaptıracağımız

işlemler için bir ihtiyaç haline gelmiştir. Önceleri makine diline yakın bir dille programlar yazılırken günümüzde programlama dilleri insan diline daha benzer hale gelmiştir. Otomatik pilotlar gibi hayati işlerde dahi bilgisayar sistemleri yoğun olarak kullanılmaktadır. Bütün bu sebeplerden ötürü artık bizim yerimize hareket edebilecek, çıkarlarımızı gözeterek diğer sistemlerle de etkileşerek hayatımızı kolaylaştıracak türden sistemler tasarlanmaya başlanmıştır [51].

Çeşitli problemleri çözmek, hesaplamalı modeller, teknolojiler ve gereçler için yeni yöntemler geliştirilmiştir. 2007’de Mago ve Devi tarafından yayınlanan bir makalede Hindistan’ın kırsal ve uzak bölgelerinde yaşayan yeni doğan ve 5 yaşına kadar olan çocuklarla ilgilenen sağlık görevlilerine karar desteği sağlayacak çok ajanlı bir sistem sunulmuştur. Devletin yeterli çocuk uzmanı olmadığından bu bölgelere gönderilemediği ve bu durumun yüksek çocuk ölümlerine neden olduğu belirtilmiştir. Böyle bir durumda yazılım ajanları gerçekçi çözümler sunabildiğinden IPA-Zeki Pediatri Asistanı adı verilen bilgi tamamlayıcı da içeren ajan tabanlı bir prototip geliştirilmiştir. İçinde kullanıcı ajanı (user agent-UA) ve grafiksel kullanıcı ara yüzü (graphical user interface-GUI) olan bu sistem bilgi paylaşımı için sunucu istemci mimarisinde çalışmaktadır. Sistem sağlık çalışanlarına sunduğu ara yüz yardımıyla belirtileri istemekte ve uzman görüşü sunmaktadır [52].

Bennett ve Hauser tarafından 2013’te yapılan çalışmada yazarlar modern sağlık sisteminde hızla genişleyen maliyet ve karmaşık, büyüyen sayısız tedavi seçeneği ve patlayan bilgi akışı, en uygun tedavi kararını seçme yeteneğine etkin ulaşılması gerekliliğine neden olduğunu belirtmişler ve bunları ortadan kaldırmaya yönelik genel amaçlı (hastalığa özel olmayan) bir sistem tasarlamışlardır [53]. Bu sistemin 2 potansiyel işlevi; (1) çeşitli sağlık politikaları, ödeme yöntemleri gibi şeyleri keşfedecek bir simülasyon ortamı oluşturmak ve (2) klinik yapay zekâ tabanlı-doktor gibi düşünebilecek bir yapay zekâ geliştirmektir. Yöntem olarak Markov karar işlemi ve dinamik karar ağları kombine edilerek klinik veri ile öğrenen ve sağlık sisteminin çeşitli bileşenlerindeki bazen karmaşık bazen sinerjik etkileşimleri yakalayıp alternatif dizilmiş karar yollarını simüle ederek karmaşık planlar yapabilecek bir yaklaşım seçilmiştir. Ayrıca hastanın sağlık durumu ve işlevlerinde eksik bir veri olursa, online

ajan şeklinde hareketlerini yeniden planlayan bir sistem de mevcuttur. Bu sistem gerçek bir elektronik sağlık kaydından alınan hasta verileriyle denenmiş ve başarılı sonuçlar elde edilmiştir. Zaman ve maliyetten tasarruf edilmiş, hasta sonuçlarında %30-35 artış elde edilmiştir [53].

Aynı yazarın yaptığı bir diğer çalışmada hastaların bireysel tedavi cevapları tahmin algoritmaları ile üretilmiştir. Elektronik sağlık kayıtları (EHR), sağlıkla ilgili verilerden faydalanmak için sadece ilk adımdır. Asıl konu bu verileri faydalı bilgiye dönüştürmektir. Dahası elektronik sağlık kayıtları hastaların sonuçları ile ilgili bilgi klinik karar desteği gibi işlevsellik ve genetik bilgi ayrıca hastanın sağlık durumu ve tedaviye zamanla verdiği cevap gibi değerli bilgilerin raporları gibi verileri içermeye başlamıştır. Çalışmada 423 hastanın Centerstone'da geliştirilen EHR'den faydalanılarak bireysel hasta tedavi cevabının tahminsel algoritmalar ile üretilmesi durum çalışması anlatılmıştır. Naive Bayes, gizli Naive Bayes, AODE (Aggregating One-Depence Estimation), Bayes ağları, çok katmanlı algılayıcı sinir ağları, rastgele orman (random forest), J48 karar ağaçları, Log Regresyonu ve K-en yakın komşu yöntemleri ile optimize edilen modeller ROC analizi ve AUC (eğri altında kalan alan) yöntemi ile test edilmiştir. En yüksek doğruluklar %70-72 arası çıkmıştır, AUC değerleri ise 0,75-0,79 arasındadır [54].

Yine Bennett bir çalışmasında gerçek dünya klinik düzeninde karar desteği işlevselliğinin verimliliğini sağlamak için EHR verilerini baskılayan bir klinik verimlilik sisteminin etkilerini değerlendirmiştir. Sistem yılda 75.000'in üzerinde farklı müşteri gören geniş davranışsal bir sağlık bakımı sağlayıcısı için gerçekleştirilmiştir. Bulgular 3 ayda sisteme entegre edilen her bir değişkenin önemli oranda geliştiğini göstermektedir. Gelirde %30, kliniklerde %10, tedavi planı tanımlamada %25, vaka oranı niteliğinde %20 artış ve diğer konularda da benzer gelişmeler görülmüştür [55].

Son yıllarda yapılan çalışmalar göstermektedir ki dünyada hem bilgisayar sistemleri hem de tıp alanındaki gelişmelerle birlikte bilgisayarlı sistemlerin tıpta kullanımı

artmaya başlamıştır. Bu nedenle disiplinler arası bir çalışma olan tıp bilişimi alanında yeni sistemler geliştirmek amacıyla yapılan çalışmalar dikkat çekici hale gelmiştir. Sağlayacağı kolaylıklar ve faydalar göz önünde bulundurulduğunda ülkemizde de bu sistemlerin kullanımının yaygınlaştırılması gerekmektedir.

Tıp çok geniş bir alan olduğundan çok çeşitli uzmanlık alanlarına ayrılmıştır. Hematoloji vücudumuz için hayati önem taşıyan kanı inceleyen bir bilim dalıdır. Hematoloji dâhili tıp bilimlerinin en önemli dallarından biridir. Kanın vücutta oksijeni, besinleri ve hormonları dokulara taşımak, dokularda oluşan hücre metabolizması artıklarını toplamak, enfeksiyonlara ve çevresel faktörlere karşı vücudu korumak gibi önemli görevleri vardır. Vücudumuzun yaklaşık %7-8'ini oluşturan kanın, %40-45'i hücresel elementlerden oluşur. Koagüle edilmiş kanın hücresel elementlerden arındırılmasından sonra kalan sıvı kısım ise kanın %55-60'ını oluşturur. Bunun %90'ı sudur ve elektrolit, protein gibi maddeleri içerir Hücresel elementler kırmızı kan hücreleri (eritrositler), beyaz kan hücreleri (lökositler) ve kan pulcukları (trombositler) olarak açıklanabilir. Kan hücreleri, kemik iliğinde bulunan öncül kan hücrelerinin farklılaşması ve olgunlaşması sonucu ortaya çıkmakta ve işlev kazanmaktadırlar [56, 57].

Kemik iliğinde bulunan öncül kök hücreler, kendilerini yenileyebilir ve diğer hücrelere dönüşebilir. Bu hücreler bazı büyüme faktörleri ile önce miyeloid ve lenfoid öncül hücrelere yönelmekte, daha sonra olgun kan hücrelerini oluşturmaktadırlar. Kemik iliği, erişkin bir insanda kan yapımının gerçekleştiği en önemli yerdir. Normal bir erişkinde kemik iliği günde yaklaşık 500 milyar hücre üretir, gerektiğinde bu üretimi artırır. Doğum öncesi, yeni doğan ve erken çocukluk dönemlerinde ise kan yapımı ve kan değerleri erişkin bir kişinin durumundan farklı bir şekilde gerçekleşmektedir [56, 57]. Bu konuda detaylı bilgi Bölüm 3'te verilmiştir.

Hematoloji alanında yapılmış bilgisayarlı sistemlere baktığımızda ağırlıklı olarak görüntü işleme ve veri madenciliği ile yapılmış çalışmalar görmekteyiz.

Hematoloji otomasyonu Wallace Coulter kırmızı ve beyaz kan hücrelerini saymak için elektrik impedansı teknolojisini uyguladığından beri sürekli gelişmektedir [58]. 1980’lerde çoğu hematoloji laboratuvarı 7 parametrelili tam kan sayımı raporu yapmaktaydı. Zamanla bunun hızı önemli olmaya başladı. Sonunda manuel yöntemler ilk aspirasyonda daha hızlı raporlanabilir sonuçlar sağlayan net, güvenilir otomatikleştirilmiş hematoloji sistemleriyle yer değiştirmiştir. Hematoloji laboratuvarı otomasyonunun en üst seviyesi ölçeklenebilir ve düzenlenebilir tam kan sayımı raporu, altı parça beyaz kan hücre sayımı farkı, çekirdekli kırmızı kan hücresi, retikülosit sayımı ve olgunlaşmamış retik bölümü, otomatik olgunlaşmamış plaket bölümü ve otomatik yayma çözeltisi ve boyama analitik tespitlerini hizmete sunan sistemlerdir. Bu isteğe bağlı testler uzmanlar için standart hale gelmiştir ve performanslı olmaları gerekmektedir. Bu test işlemleri hematolojiye özgü bir ürünle desteklendiğinde klinik laboratuvarlar otomatik olarak test hacimlerinin %85’ini hiçbir operatör müdahalesi olmaksızın raporlamaktadırlar [58]. Patel 2014’te yayınlanan bir yazısında ön analitik, klinik karar verme, standardizasyon, son analitik, personel açığı, çapraz eğitim gibi alanlardaki otomasyon detaylarını anlatmıştır [58]. Bu hematoloji otomasyonu araştırmasında hem analitik hem destekleyici yazılım yetenekleri tarif edilmiştir. Bu açıdan bakıldığında hematoloji laboratuvarının üç hücre hattının (beyaz kan hücreleri, kırmızı kan hücreleri ve plateletler) olgunlaşmamış şekillerini analiz ve rapor etme yeteneği, doktorların tedavi edici karar verme işlemlerinde tam otomatik parametrelerle desteklenmesini sağlamaktadır. Yeni yazılımlar hematoloji sonuçlarını ve laboratuvar bilgi sisteminde anında raporlanmasının otomatik doğrulanmasını gerçekleştirmektedirler. Bu işlemler testlerin de entegre edilmesiyle daha verimli hale gelecektir [58].

Hematoloji hücre morfolojisinin değerlendirilmesine en bağlı olan klinik laboratuvar bilimlerinden biridir. 100 yıldan uzun bir süredir uzmanlar kan filmi lekelerini manuel olarak görüntülemişler ve beyaz veya kırmızı kan hücrelerinin sayımını ve anormalliklerin belirlenmesini de manuel olarak gerçekleştirmişlerdir. İlerleyen teknoloji sayesinde görüntüler dijital olarak elde edilmeye başlanmış ve analiz, görüntüleme, basım ve işleme daha kolay hale gelmiştir. 2012’de Linder ve Zahniser

bir yazılarında geçmişten günümüze bu konudaki gelişmeleri ve yeni yöntemleri incelemişler ve gelecekte yapılması muhtemel çalışmalardan bahsetmişlerdir [59].

Hematoloji alanında gerçekleştirilen yapay zekânın ilk uygulamaları laboratuvar veri yönetimi konusunda olmuştur. Anemi, talasemi, lösemi gibi özel hastalıkların çevresel kan analizi verileri ile yapay zekâyâ dayalı yöntemler geliştirilmiştir. Kanser teşhisinde, hematolojik hastalıklarda, DNA mikro dizileri kullanarak binlerce genin simültane dizilimlerini görüntülemeye dayalı sistematik yaklaşımlarda, yapay zekâ araçları kullanılarak analiz yapılmıştır. Zini'nin yaptığı çalışmada yapay sinir ağları hakkında bilgi verilmiş, tıbbi uygulamalarda kullanımına örnekler verilmiştir [60]. Hematoloji alanında ilk kullanımı genellikle laboratuvar veri yönetimi araçları şeklinde başlamıştır. Bu konuda yapılan çalışmalar anlatılmış. Bir diğer kullanım yeri çevresel kan analizidir. Kalıtsal kan hastalıklarının tahmini hemoglobin değerlerine bakılarak gerçekleştirilmiştir. 4 çevresel kan hematoloji parametreleri (kırmızı kan hücresi sayısı, hemoglobin, hematokrit ve ortalama eritrosit hacmi) hemakromositometrik analizlerinin sonuçlarının test edildiği ve sınıflandırıldığı (ileri beslemeli yapay sinir ağı mimarisi ile) doğruluğunun alfa ve beta taşıyıcılığının %94 olduğu çalışmalar mevcuttur. Polisitemi vera teşhisini geliştirmek için bir veri madenciliği yaklaşımının kullanıldığı, kararsal kuralların laboratuvar parametrelerinin optimizasyonu ile çıkarıldığı çalışmalar da mevcuttur. Bu yöntem gen profili analizi çalışmalarında da kullanılmıştır. Sonuçta hematolojide yapay zekâ uygulamaları çevresel kan analizinin tanısal tahmininden gen profili analizine göre bireye özel tedavilere kadar hasta yönetimini her açıdan ilgilendirmektedir [60].

Lipkin 1960 yılında yaptığı çalışmasında dijital bir bilgisayar yardımıyla hematolojik hastalıkların veri karakteristiğini ve hastane verilerini karşılaştırmıştır. 49 hasta ve 20 hastalık seçilmiş ve hastane verileri bilgisayar programı ile bağlantı kurulmuştur. Hastane vakalarının diferansiyel tanısı yazılı formda basılmıştır [61].

1976'da yayınlanan bir yayınlarda Engle ve ark. hekimlere hematolojik hastalıkların tanıya yönelik analizini sağlayan HEME adını verdikleri bir bilgisayar programını

tanıtmışlardır. Sistemde 40 hematolojik hastalık vardır. Kural tabanlı bir sistem tasarlanmış, Bayes yöntemi kullanılmıştır [27].

Slonko ve ark. 2013 yılında yayınlanan bir çalışmalarında, kliniklerinde immunohematoloji raporlarını geleneksel bir yöntem olan dikte ile yazdırmak şeklinde oluşturduklarından ve 2008'e kadar hiçbir değişiklik yapılmadığından bahsetmişlerdir. Bu raporların doğruluğunu ve kalitesini geliştirmek için 'SpezLab' dedikleri yeni bir bilgisayar yazılım uygulaması geliştirmişlerdir. Böylece çok kısa bir sürede laboratuvar raporlarını daha düşük hata oranıyla yazdırmaktadırlar. Bu yazılımın diğer enstitülere de faydalı olabilmesi için genel düşüncelerin tartışılması gerektiğini belirtmişlerdir [62].

Venkatalakshmi ve Thilagavathi yaptıkları bir çalışmada etkili ve maliyet verimli bir bilgisayarlı görüntü sistemi tanıtmışlardır. Anemi, lösemi gibi hastalıkların belirlenmesi ve tedavi takibinde kırmızı kan hücrelerinin sayısı çok önemlidir. Mikroskop altında kırmızı kan hücrelerini saymak eski ve geleneksel yöntemi klinik laboratuvar teknisyeninin yeteneğine bağlı, güvensiz ve eksik sonuç verir. Bir diğer yöntem olan hematoloji analizörü denen maliyetli bir cihazdır ki her laboratuvar böyle pahalı bir makineyi koyma imkânı bulamaz. Yazarlar otomatik kırmızı kan hücrelerinin sayımı için görüntü tabanlı analiz yapan bir yöntem sunmuşlardır. Mikroskopa bağlı dijital kameradan gerçek zamanlı mikroskopik görüntü yakalanmakta ve kırmızı kan hücrelerinin sayımı gerçekleştirilmekte, sonuçlar kullanılan hastane bilgi sistemi yoluyla ilgili doktora gönderilmektedir [63].

Haun ve ark. bir çalışmalarında beyaz kan hücreleri tanımlama yeteneğine sahip bilgisayar tabanlı bir eğitim aracı geliştirdiler. Amaçları öğrencilerin klinik laboratuvarında hematoloji eğitimlerine katkı sağlamaktır. Sistem web tabanlıdır ve kişisel bilgisayar veya mobil cihazlarla erişilebilirdir [64].

Wicht ve ark.'nın yayınlanan bir makalelerinde web tabanlı klinik karar destek sistemi ve ciddi enfeksiyonlar ve hayat tehdit edici komplikasyonlar geliştirme riski yüksek

olan hemoto-ontolojik hastaların ciddi eş tanısını (komorbiditesini) tespit etmek için kural tabanlı bilgi bakımı sistemi tanıtılmıştır. Sistemin tanısal performansını ölçmek için çok aşamalı bir değerlendirme işlemi uygulanmış, toplam başarıımı %84 bulunmuştur [65].

Maity ve ark. çalışmalarında hasta kan patoloji bilgisi edinimi ve bilgisayar görmesi yaklaşımı kullanarak otomatik anemi tanı raporu oluşturulması ve bunların yönetimi için bir uygulama tasarlayıp geliştirmişlerdir. Geliştirilmiş görüntü işleme algoritması ve veri madenciliği yaklaşımı hastanın tıbbi bilgisini analiz etmek için anormal eritrositleri tanımlayabilmektedir. Danışmanlı karar ağacı sınıflandırıcısı C4.5, görüntü örneklerini %98,1 hassasiyet ve %99,6 kesinlik değerleriyle sınıflandırmıştır [66].

Rupp ve ark. bilgisayar destekli hematoloji için kan yaymalarındaki sayma alanının tam otomatikleştirilmiş belirlenmesi için bir çalışma yapmışlardır. Algoritmanın performansının değerlendirilmesi için bir dizi kan yayması, uzman bir hematolog tarafından tanımlanarak karşılaştırma yapılmıştır. Sonuçlar göstermiştir ki gösterge fonksiyonu güvenilir ve güçlü bir şekilde sayma alanını belirlemektedir [67].

Mustafee ve ark. yaptıkları bir çalışmada hematoloji servislerinde ayakta tedavi edilebilen hastaların kapasitelerini ve uzman talebini soruşturmuşlar, servislerin birleştirilmesi için bir strateji geliştirmişlerdir. Yaptıkları durum çalışmasında 4 farklı hastanenin hematoloji servislerini 2'ye düşürmeye ve bunun için tekrar kullanılabilir bir simülasyon modeli geliştirmeye çalışmışlardır [68].

Hematolojik hastalıkların tam teşhisi ve sınıflandırılmaları için çevresel kan analizi, kemik iliği muayenesi ve imünofenotipleme (immünofenotipleme hücreler tarafından ifade edilen proteini incelemek için kullanılan bir tekniktir) yöntemi de dâhil çok parametrelili bir yaklaşım gerektiği uzun zamandır bilinmektedir. Diamond ve ark. hemogram ve çevresel kan yaymalarının yorumlanması, akış sitometrik imünofenotipleme panellerinin analizi ve kemik iliği numunelerinin morfolojik

değerlendirilmeleri için bilgi tabanlı bir bilgisayar sistemi tasarlamışlardır. Bu 3 modül, ilgilinin klinik geçmişi ve laboratuvar sonuçlarını barındıran ortak bir veri tabanını paylaşmaktadır. Sistem 3 modülün etkileşimi ve yorumlu raporların kaliteli olması yeteneğine göre 100 lösemi hastasında test edilmiştir. 100 vakanın 94'ünde kemik iliği sisteminin yaptığı son tanı hastane tanısı ile eşleşmiş ve 99'u ise yazarların yorumu ile örtüşmüştür [69].

Kırmızı kan hücreleri, kan hücrelerinin en yaygın türüdür ve vücut dokularına oksijen taşımaktan sorumludur. Kırmızı kan hücresi anormallikleri kırmızı hücrenin fiziksel özelliklerini değiştirir ya da yaşam süresini kısaltır ve felç veya anemiye neden olabilir. Lee ve Chen 2014'te yaptıkları bir çalışmada, kırmızı hücrelerin normal veya anormal olup olmadığını belirlemek için kırmızı kan hücresi görüntülerinden çıkarılan görsel bilgiyi kullanan sinir ağı tabanlı melez bir bir sınıflandırıcı önermişlerdir. Deneysel sonuçlar önerilen sistemdeki sınıflandırma doğruluğunun tek giriş katmanlı sınıflandırıcıya göre önemli ölçüde geliştirdiğini göstermektedir [70].

Türkiye'de hematoloji alanın yapılan veri madenciliği çalışmalarından Prof. Dr. Yahya BÜYÜKİŞİK tarafından gerçekleştirilen 3 tanesi Çekirdek, Kür ve Rapor programlarıdır. Çekirdek bütün hematolojik hastalıkların kaydını sağlamak üzere MS Access Tabanlı bir program geliştirmiştir. Böyle bir programa olan ihtiyaç hematoloji alanında çok geniş bir hastalık spektrumu olmasından ve bunların hatırlanmasının zorluğundan doğmuştur. Program hasta verilerini kaydetme ve gerektiğinde bu verileri form, rapor ve grafik şeklinde çağırma özelliklerine sahiptir. Son sürümü Şubat 2011'de çıkan Kür adını verdiği programda ise hastalıkların tedavi yöntemleri için önerilerde bulunmaktadır. Rapor programı ise takip edilen hasta verilerini raporlayan, aynı tanıya sahip hastaların listesini çıkaran bir programdır [71-73].

Türkiye'de ve dünyada en sık görülen kan hastalığı anemiler grubudur [9]. Anemi kandaki eritrosit kütlelerinin veya hemoglobin ve hematokrit konsantrasyonunun azalması olarak tanımlanır. Normal hemoglobin ve hematokrit değerleri yaş ve cinsiyete göre değişir. Hemoglobin ve hematokrit değerleri ilgili yaş ve cinsiyet için normal

değerlerin alt sınırından düşükse anemi vardır. Aneminin temel nedenleri arasında eritrositlerin morfolojik (yapısal özellikleri) veya fizyolojik fonksiyonlarında bozulma yer alır. Anemili hastada ilk istenecek laboratuvar tetkikleri tam kan sayımı ve eritrosit indeksleri (MCV-ortalama eritrosit hacmi, MCH-ortalama eritrosit hemoglobini, MCHC-ortalama eritrosit hemoglobin konsantrasyonu, RDW-eritrosit dağılım genişliği) olacaktır. Ayrıca periferik yayma, formül lökosit ve retikülosit sayımı da önemli tetkiklerdir. Öykü, fizik muayene ve bahsi geçen incelemeler ile anemi nedeni hakkında ön fikir elde edilir. Daha sonra kesin tanı için ileri tetkikler istenir [56, 57, 74].

Bilgisayar destekli anemi tanısı için yapılmış en erken çalışmalardan birisi Beck ve ark.'nın 1989 yılında yaptıkları çalışmadır [21]. İleriki zamanlarda bahsedilen çalışma geliştirilerek tıp eğitiminde de kullanılmıştır [22].

Quagliini ve ark.'nın tasarladığı ANEMIA, anemik durumlar için bilgi-tabanlı bir konsültasyon programıdır. Bu, Rutgers Üniversitesi'nde geliştirilen EXPERT denilen bir yapay zekâ programlama düzeni kullanılarak inşa edilmiştir. 1986'da ANEMIA, demir eksikliği anemisi, kronik hastalık anemisi, talasemiler, hemolitik anemiler ve diğer çeşitli şartlar içeren 64 hasta kişinin tanı ve tedavisine destek verebilmektedir. ANEMIA 220 retrospektif veri tabanına karşın 100 yeni vaka ile test edilmiş, teşhis performansının genel doğruluğu oldukça iyi bulunmuştur. Sonuçta ANEMIA'yı hematolog olmayan kullanıcı hekimin hematolog vekili olarak kullanmak amaçlanmıştır [75]. Aynı yazarların 1988'de yaptıkları diğer bir çalışmada ANEMIA sisteminin mevcut performans seviyesini değerlendirme sonuçlarını bildirmişlerdir [76].

Birndorf ve ark. anemi tanısı ve rapor sonuçlarının doğrudan hematoloji formlarında olması için bir uzman sistem geliştirmişlerdir. Çalışmada 473 adet veride %96,5 doğru sınıflandırma yapılmıştır [77].

1990 yılında Ortiz ve ark.'nın yaptığı çalışmada bir bilgisayar veri tabanı kullanılarak bazı kurumlardaki hastaların demografik bilgiler ve hemoglobin test sonuçlarının

analiz edildiği ve özetlendiği bir program yazılmıştır. Sonuçlar göstermektedir ki çalışılan 3 büyük kurumdaki anemi yaygınlığı yüksektir. Bu yüksek yaygınlığın nedenleri araştırılmış ve ileride yapılacak bir çalışma için veri tabanına eklenmiştir [78].

Tam kan sayımı anemi ve diğer hematolojik bozuklukları tanımlamak için kullanılır. Ancak, demir eksikliği anemisi ve talasemi tutarsız ve belirsiz olan ayırt edici özellik olarak ortalama hücre hacminin 80 fL (sıvı ons)'den az olmasına bağlıdır. Yeh ve Cheng 2005'te yaptıkları bir çalışmada, kural tabanlı bir yazılım yöntemi olan hiyerarşik yazılımsal hesaplama kullanarak bu soruna çözüm önermişlerdir. 50 örnek üzerinde %96 doğruluk bulmuşlar ve geleneksel yöntemle göre daha doğru olduğunu belirtmişlerdir [79].

Doğan ve Türkoğlu bir çalışmalarında kan biyokimya parametreleri ile demir eksikliği anemisi teşhisinde, hekime yardımcı olacak ve kolaylık sağlayabilecek bir karar destek sistemi oluşturmuşlardır. Örüntü tanıma süreci esas alınmış olup, sistemin işleyişi veri madenciliği tekniklerinden olan karar ağaçları yapısı ile sağlanmaktadır. Sisteme giriş olarak, biyokimya parametrelerinden demir eksikliği anemisi hastalığı için temel belirleyiciler olan Serum demiri, Serum demir bağlama kapasitesi ve Ferritin enzimleri verilmiş, çıkış olarak Anemi (+) ve Anemi (-) değerlendirmelerinde bulunulmuştur. Tasarlanan sistemde 96 hasta verisi değerlendirilmiştir. Karar destek sisteminin sonuçları, doktorun verdiği kararlarla tamamen örtüşmüştür [80].

Setsirichok ve ark., bir makalelerinde C4.5 karar ağacı, Naive Bayes sınıflandırıcı ve çok katmanlı algılayıcı ağ yöntemleri ile talasemi taraması için kan özelliklerinin sınıflandırılmasını sunmuşlardır. Seçilen tam kan sayımı özellikleri bir hemoglobin konsantrasyonu ve ortalama eritrosit hacmidir. Bayes sınıflandırıcı ve çok katmanlı algılayıcı uygulandığında sınıflandırma performansının ortalama doğruluğu, %93,23 ve %92,60 bulunmuştur. Bu sonuçlar, Naive bayes sınıflandırıcı ya da çok katmanlı algılayıcı ile tam kan sayımı ve hemoglobinin bir arada analizinin otomatik talasemi taraması için son derece uygun olduğunu göstermektedir [81].

Demir eksikliği anemisi ve talasemi, tanı konurken birbiri ile karıştırılma riski taşıyan iki tür mikrositik anemidir [82]. Bir araştırma makalesinde 200 beta talasemi vakası, 65 alfa talasemi vakası, 170 demir eksikliği anemisi vakası ve 45 hem demir eksikliği anemisi hem beta-talasemi vakasından oluşan bir eğitim seti kullanılarak diskriminant analizi yöntemi ile mikrositik aneminin ayırıcı tanısı yapılmıştır [83].

2016'da Jamei ve Talarposhti demir eksikliği anemisi ve β -talasemi taşıyıcılığı ayrımı için örüntü tabanlı giriş seçimli yapay sinir ağı dedikleri bir model geliştirmişlerdir. Bu yöntem bir insan uzmanın karar verme yetisi ile yapay sinir ağıları kombinasyonundan oluşmaktadır. Tam kan sayımı testinin sonuçları kural tabanı benzeri bir yöntem ile benzerlik katsayısı hesaplamışlar ve bu hesaba göre çok katmanlı perceptron yapay sinir ağı girişlerini belirlemişlerdir. Yazarlar kullandıkları yöntemin başarımını ANFIS, ANN, MLP, SVM, KNN, RBF, PNN ve Math yöntemleri kullanan çeşitli yazarların yayınları ile karşılaştırmışlar ve %99,5 ile en yüksek doğruluk oranına ulaştıklarını belirtmişlerdir [84].

Talasemilerin sınıflandırılması ve teşhisi [85] için çeşitli karar ağaçları ve yapay sinir ağıları algoritmaları denenmiş başarılı bulunmuştur. Kalıtsal bir hastalık olan orak hücre anemisi için gömülü bir sistem uygulaması yapılmış, bilirubin testi, kan oksijen testi, tam kan sayımı, hemoglobin elektroforezi, serum kreatin testi, serum potasyum ve orak hücre testlerinden alınan çeşitli parametreler işleme için mikroişlemciye gönderilmiştir. Hastaların test sonuçlarının işlenmesi ve değerlendirilmesi sonucunda mikroişlemci tarafından tanı gerçekleştirilmiştir. Gerçekleştirilen tanı hem LCD ekranda gösterilmiş hem de verilen GSM numarasına SMS olarak iletilmiştir [86].

Yılmaz ve ark. demir eksikliği anemisi seviyesinin belirlenmesi için bulanık uzman bir sistem tasarlamışlardır. Sistem tasarımının gerçekleştirilmesi sırasında gerçek hastalardan alınan laboratuvar kayıtları kullanılmış, uygun parametreler seçilmiş, giriş ve çıkış parametreleri bulanıklaştırılmış ve sonuçta uzman hekimle birlikte bir kural tabanı inşa edilmiştir. Geliştirilen bulanık çıkarım sistemi 100 hasta verisi üzerinde

uygulanmış sistemin başarı seviyesi %90-95,8 arasında belirlenmiştir. Önceki çalışmalarla karşılaştırıldığında sistemin doğruluğu onaylanmıştır [87].

Wongseree ve ark. 2007 de sinir ağları ve genetik programlama yöntemlerini kullanarak Talasemi sınıflandırması yapmışlardır. Hastaların kan parametreleri kullanılarak, hastalar Talasemi taşıyıcısı, Talasemi hastası ve normal olmak üzere 3 farklı sınıf şeklinde tanımlanmıştır. Sınıflandırma başarıları %82 dir [88].

Amendolia ve ark. 2003'te örüntü tanımanın 2 tipik yönteminden DVM (destek vektör makineleri) ve ÇKP (çok katmanlı perseptron) ile birlikte KNN (k en yakın komşu) sınıflandırıcısını karşılaştırmışlardır. ÇKP yönteminin DVM yönteminden daha iyi sonuçlar verdiği belirtilmiştir. Her iki teknik de sağlıklıları talasemililerden %95 özgüllük ile ayırmış, fakat ÇKP yönteminin hassasiyeti %92 iken DVM yönteminin hassasiyeti %83 olmuştur [89].

2009'da yapılan bir çalışmada olası talasemi, sıtma ve diğer anormal kırmızı hücre bozukluklarını belirlemede hızlı ve doğru olan, otomatik bir kan görüntü analiz sistemi geliştirilmiştir. 2 geri yayımlı YSA-yapay sinir ağı modeli (3 veri katmanlı) görüntü analiz tekniğiyle birlikte kullanılarak kandaki eritrositlerin morfolojik özellikleriyle birlikte tıbbi görüntü örüntülerinin tanınmasında doğru bir sınıflandırma gerçekleştirilmiştir. En iyi performans $2,74545 \times 10^{-5}$ hata ve %86,54 doğru tanıma oranı ile 3 katmanlı YSA vermiştir [90].

2011'de Sanap ve ark. tam kan sayımı raporlarında anemi şiddetini ölçen bir sınıflandırma yapmışlardır. Weka veri madenciliği aracı ile C4.5 karar ağacı algoritması ve destek vektör makinesi algoritmalarını kullanmışlardır. Yaş, WBC, HGB, RBC, HCT, MCV, MCH, MCHC, RDW, PLT olmak üzere 10 nümerik değer kullanmışlar ve 4 sınıfı normositik (kronik hastalık anemisi), mikrositik (demir eksikliği ve talasemi), makrositik (Vitamin B12 ve folat eksikliği) ve mikrositik (renal anemi) anemi olarak belirlemişlerdir. C4.5 karar ağacı algoritmasının başarıları %99,42 ile %88,13 başarı bulunan destek vektör makinesini geçmiştir [91].

2011’de Allahverdi ve ark. çocuk anemisini belirlemek için Takagi-Sugeno tipi nörofuzzy Ağ yöntemini kullandıkları bir çalışmalarını yayınlamışlardır. İstatistiksel hesaplarına göre sistemdeki hataları MPE (Mean percentage error)-0,0018, MAE (mean absolute error) 0,2090, MAPE (mean absolute percentage error) 0,0511, RMSE (root mean square error) 0,2743 ve R2 (coefficient of determination) 0,9957 olarak bulmuşlardır. Tahmin edilen anemilerin ölçülen değerleri çok yakın olduğunu göstermişler ve pratikte kullanılabilir bir sistem olduğunu söylemişlerdir [92].

2012’de WEKA ve Orange veri madenciliği araçları ile sınıflandırma, regresyon, bölütleme, birliktelik kuralları ve dizilim analizi yöntemleri kullanılarak hamile kadınlarda anemi ve depresyon arasında bir korelasyon bulunduğu belirtilmiştir. Bu bilgiye istinaden hamilelerde depresyon tedavisinde anemi de göz ardı edilmemelidir [93].

2014’te Solanki WEKA veri madenciliği programı ile orak hücre hastalığını kan grubuna göre sınıflandırabilecek bir tahmin modeli geliştirmiştir. Yöntem olarak C4.5 Wekadaki adıyla J48 karar ağacı algoritmasını kullanmış, 15 yaş ve altında 0+ ve A+ kan grubuna sahip kişilerin risk grubunda olduğu, B+, AB+, A- kan grubuna sahip kişilerin risk grubunda olmadığı iddiasında bulunmuştur [94].

Kishore ve ark. 2015’te giriş olarak tam kan sayımı formlarından yaş, cinsiyet, HGB, MCV, MCH, HCT değerlerini, çıkış olarak ise demir eksikliği anemisi ve Vitamin B12 eksikliği anemisini belirleyerek bir çalışma yayınlamışlardır. Yöntem olarak ID3 ve non-threaded ID3 karar ağacı algoritmalarını inceleyerek bir threaded ID3 yaklaşımı geliştirmişlerdir. 480 veri seti ile sistemi hem threaded hem de non-threaded ID3 ve gini algoritmaları ile test etmişler ve buldukları yöntemin kullanılabilir olduğunu söylemişlerdir [95].

Amin ve Habib’in 2015’te yaptıkları çalışmada yine WEKA veri madenciliği aracı kullanılarak sınıflandırma yöntemlerini karşılaştırmışlardır. Veri girişi olarak yaş, tam

kan sayımı parametrelerinden WBC, RBC, HB, HCT, MCV, MCHC, PLT, NEU, LYMP, MONO, EO, BO ve yorum değerlerini kullanmışlardır. Sınıf olarak ise kronik hastalık anemisi, eozinofili, mikrositik hipokromik anemi, normositik anemi, nötrofil lökositozu, nötrofili, bilinmeyen bulgular ve yüksek ESR tanılarını kullanmışlardır. Sınıflandırıcı olarak kullandıkları J48 karar ağacı %97,16 başarı ile %86,55 başarılı çok katmanlı perseptron ve %70,28 başarılı naive bayes yöntemlerini geçmiştir [96].

Kimura ve ark. aplastik anemi ile miyelodisplastik sendrom hastalıklarının ayırıcı tanısını yapmak üzere derin öğrenme tabanlı görüntü analiz sistemi geliştirmişlerdir. Veri setini 17 hücre tipi ve 97 farklı morfolojik özellik için bölmüşler ve sınıflandırma yapmak için evrimsel sinir ağı modülü oluşturmuşlardır. Performans ölçümü için ROC eğrisi yöntemini kullanmışlar ve AUC değerini 0,99 olarak bulmuşlardır [97].

Alsheref ve Gomaa makine öğrenmesi yöntemleriyle kan hastalıkları tahmini yaptıkları bir çalışmada 668 hastanın kan testi sonuçlarından bir veri seti oluşturmuşlardır. Veri setinde 28 parametre ve 4 sınıf vardır. WEKA programını kullanarak uyguladıkları makine öğrenmesi yöntemlerinden en yüksek başarı gösteren yöntemin %98,16 ile logit boost yöntemi olduğunu belirtmişlerdir [98].

Meena ve ark. çocuklarda görülen aneminin sınıflandırılması ile ilgili bir çalışma yapmışlardır. Çalışmalarında veri madenciliği yöntemleri kullanarak bir karar destek sistemi önermişlerdir. Geliştirdikleri modelde karar ağaçları yöntemini ve birliktelik kuralları yöntemini kullanmışlar ve başarılı sonuçlar aldıklarını belirtmişlerdir [99].

Çil, Ayyıldız ve Tuncer demir eksikliği anemisi ile beta-talasemi hastalıklarının ayırıcı tanısını yapmak için bir karar destek sistemi geliştirmişlerdir. 342 hastaya ait HGB, RBC, HCT, MCV, MCH, MCHC ve RDW parametrelerini kullanmışlardır. Yöntem olarak ise lojistik regresyon, KNN, DVM, ELM ve RELM algoritmalarını kullanmışlar ve %95,59 oranında başarılı bir performans elde etmişlerdir [100].

Varghese makine öğrenmesi yöntemlerinin kan hücrelerinin ve kan hastalıklarının sınıflandırılmasında ne kadar önemli olduğunu belirten bir çalışma yapmıştır.

İncelediği birçok makaleyi analiz ederek CNN ve SVM yöntemlerinin en başarılı sonuçları verdiğini iddia etmiştir [101].

Alzubaidi ve ark. RBC hücrelerinin şekillerinden orak hücre anemisi teşhisi için bir çalışma yapmışlardır. RBC hücrelerini normal, orak hücre ve kan içeriği olmak üzere üç sınıf olarak tanıtmışlardır. Çalışmalarında üç farklı derin öğrenme modeli önermişlerdir. Eğitim verileri olmadığı için transfer öğrenimi ve veri artırımı kullanmışlardır. Modellerinin sınıflandırmada yüksek doğruluk verdiğini belirtmişlerdir [102].

Shaik ve Subashini, anemi teşhisi için bulanık bir mantık yaklaşımı sundular. Girdi parametreleri olarak HGB, HCT, MCV, MCHC, WBC, retikülosit, toplam demir bağlama kapasitesi (TIBC), serum demiri ve hiper segmentli beyaz küre (HSWC) laboratuvar test sonuçlarını kullanmışlardır. Çıktı olarak, aplastik, sideroblastik, megaloblastik, kronik, miyelopitizik ve demir eksikliği anemilerini içeren altı anemi türü kullanmışlardır [103].

Dalvi ve Vernekar, yaptıkları bir çalışmada anemi teşhisi için kırmızı kan hücrelerini sınıflandırmanın en uygun yöntemini belirlemek istemişlerdir. Beş toplu öğrenme yöntemi (AdaBoost, torbalama, istifleme, oylama ve Bayes artırma) ve dört sınıflandırıcı (k-en yakın komşu, Naive Bayes, karar ağacı ve YSA) kullanmışlardır [104].

Belginova ve ark. demir eksikliği anemisinin teşhisine yönelik kuralla dayalı bir yaklaşım sunmuşlardır. Uzman tıbbi danışmanlar için hasta verilerini içeren bir karar destek sistemi önermişlerdir (örneğin, kimlik, sosyo-ekonomik durum, tıbbi geçmiş, şikayetler veya duyumlar, tıbbi göstergeler ve hastalıkla ilgili istatistiksel bilgiler). Bu verileri kullanmanın danışmanın hastalıkla ilgili daha doğru kararlar vermesini sağladığını belirtmişlerdir [105].

Dimauro ve ark. konjonktival bölgenin bir görüntüsünü analiz edebilen non-invaziv bir cihaz kullanarak hastaların hemoglobin değerini tahmin etme üzerine bir çalışma

yapmışlardır. Bu verileri KNN sınıflandırıcısı ile 113 birey üzerinde test etmişler ve iyi sonuçlar elde ettiklerini belirtmişlerdir [106].

Anemi başlı başına bir hastalık iken, bazı başka hastalıklara eşlik ettiği durumlar sık görülmektedir. Kanser hastalarında anemi yaygınlığı ve sıklığı yüksektir. Avrupa kanser anemisi anketi performans durumunu hafif orta ve şiddetli anemi arasındaki ilişki de dahil olmak üzere Avrupa'daki kanser hastalarında yaygınlık, sıklık ve anemi tedavisini ileriye dönük olarak değerlendirmek amacıyla yapılmıştır. Hastalar 6 ay değerlendirildi. Nüfus, tümör tipi, performans durumu, hemogloblin seviyeleri, kanser tedavileri ve anemi tedavilerini içeren 15367 veri toplandı. Kayıtlardaki anemi yaygınlığı (HGB<10) %39,3 ve %67 arasında gezindi. Düşük HGB seviyeleri ile düşük performans durumu anlamlı şekilde ilişkilidir. Anemi görülme sıklığı (HGB<10) %53,7'dir. Anemili hastaların %38,9'u (Epotein 17,4, transfüzyon 14,9 ve demir %6,5) tedavi edilmiştir. Anemi tedavisi başlatmak için ortalama HGB 9,7 g/dL'dir [107].

Repetto ve ark. 2006'da bu anketle ilgili açıklamaların yer aldığı bir çalışma yapmışlardır. Miyelosupresif kemoterapi alan kanserli hastalarda sık sık anemi gelişmektedir; platin-bazlı kemoterapi, özellikle, kemik iliği uyarıcı bir hormon olan eritropoietin indirgenmiş üretimine yol açar. Avrupa Kanser Anemi Anketi birçok hastanın eritropoez uyarıcı madde (ESA) tedavisi almadığını gösterdi ve kanser hastalarında anemi tanısı ve tedavisi için net kurallar olması ihtiyacının önemine değindi. ESA ile anemi tedavisi maliyet verimli ve uzun vadede kaliteli düzeltilmiş yaşam yılı kazancı sağlar [108].

2006 EORTC yönergeleri kanserle ilgili anemi eritropoietik proteinlerin kanıt tabanı en güncel değerlendirmesini sağlar. Kanıta dayalı rehberlerin klinisyenlerin kabul konusunda tıpta genel endişeleri göz önüne alındığında, bu bakım noktasına yönergeleri getirme yollarını bulmak için çok önemlidir. Aapro ve ark. yaptıkları çalışmada RESPOND ile EORTC yönergelerine dayanan bir web tabanlı klinik

rehberlik sistemi ve sistemi doğrulamak için yapılan iki çalışmanın metodolojileri ardındaki mantığı açıklamışlardır [109].

2007’de Bokemeyer ve ark. kanserli hastalarda aneminin yaşam kalitesini düşürdüğü ve durumun sık gerçekleştiğinden bahsedilen bir çalışma yapılmıştır [110].

2013’te Ostrovsky ve ark. tarafından yapılan çalışmada çevresel kan plazmasının fourier dönüşümlü kızılötesi spektroskopisi kullanılarak biyokimyasal analizine dayalı kanser tespiti için yeni bir yaklaşım geliştirmişlerdir. Bu yaklaşım hızlı güvenli minimal invaziv ve etkili olduğu kanıtlanmıştır. Çalışmalarında bir spektrum ölçerek toplam moleküler kompozisyonun ve çevresel kan örneklerini yapısına bakılarak ve vücuttaki yerine veya kanser tipine bakılmaksızın bilgi veren bir herhangi bir katı tümör varlığının işaretlerini tanır. Yöntemlerini çapraz doğrulama ile değerlendirmişler ve %93,33 hassasiyet, %87,8 özgüllük ve toplamda %90,7 doğruluk bulmuşlardır. Bu yöntemi kullanarak daha az gereksiz invaziv işlemler ve katı tümörlerin daha hızlı tespitinin sağlanacağını söylemişlerdir [111].

Anemi, son dönem böbrek yetmezliği (SDBY) hastalarında yaygındır. Böbrek hastalarında tedavi edilmeyen anemi, kalp hastalıklarını da içeren bir dizi duruma öncülük edebilir, düşük hayat kalitesine ve ölümlere neden olabilir. Bir doktor böyle bir durumda hastayı tedavi etmek için ilaçların uygun dozlarını kullanmalı ve yan etkilerini gözlemlemelidir. Gaweda ve ark. yayınladıkları bir makalede kronik durumların farmakolojik tedavisinde deneme ve hata işlemini resmileştirmek için takviye öğrenme temelli verimli bir hesaplamalı çerçeve sağlamak ve özelleştirmeyi amaçlamışlardır [112].

30 yıl önce rekombinant insan eritropoietinin (rHuEPO) keşfi ile diyaliz hastaları için anemi tedavisi kan naklinden rHuEPO tedavisine kaymıştır. Klinisyenler tarafından kullanılan birçok anemi yönetim protokolleri rHuEPO dozlarında haftalıktan aylığa titrasyon için hemoglobin (HGB) ölçümlerine dayalı deneyim-temelli bir dizi kural içermektedir. Anemi yönetim protokolü tasarımını kolaylaştırmak için, model tabanlı

geribildirim kontrol teorisini kullanarak fizyolojik eritropoiz modeli sunmak ve klinik verilerini kullanarak uygulanabilirliğini göstermişlerdir [113].

Rekombinant insan eritropoietinin harici yönetimi, periyodik hemodiyalize giren kronik böbrek yetmezliği nedeni ile ikincil anemili hastaların tedavisinde seçilen bir tedavi yöntemidir. Bu çalışmanın amacı uygulanacak eritropoetin dozajının bireyselleştirilmiş tahminini gerçekleştirmektir. İlacın yüksek maliyeti ikincil etkileri ve bazı bireylerin muzdarip olduğu potansiyel direnç olgusu nedeniyle çalışılmıştır. Çalışmaya 110 hasta dâhil edilmiş ve çeşitli etkenler sinir ağı modellerini geliştirmek amacıyla toplanmıştır. Elde edilen sonuçlar oldukça iyi olduğundan kolay kullanılan karar yardımcı bir bilgisayar uygulaması hayata geçirilmiştir. Bilinen sinir ağlarından çok katmanlı perseptron, elman tekrarlayan ağı, sonlu darbe cevabı ağı kullanılmıştır ki bunlar tahmin problemlerinde iyi yöntemlerdir. En iyi sonuçlar çok katmanlı perseptron yöntemi ile alınmıştır. %97'den fazla tahminde 0,5'ten küçük hatalar olmuştur [114].

Böbrek yetmezliğinin neden olduğu aneminin yönetimi konusunda hekimlere yardımcı olmak amacıyla eritropoez uyarıcı ajanlar (ESA) ile tedavi edilebilir olan hemodiyaliz hastalarındaki anemi için ilaç dozlarının ayarlamasını gerçekleştiren bir diğer çalışma Montero ve ark. tarafından 2014'te gerçekleştirilmiştir. Önerilen yöntem Markov karar süreçlerinde veri yükleme yaklaşımı ile ikincil karar verme problemlerinde kullanılan R1 algoritmasıdır. Performansı bilinen Q öğrenme yöntemiyle karşılaştırılarak etkili sonuçlar verdiği söylenmektedir [115].

Kronik böbrek yetmezliği ilişkili anemiden muzdarip hastalar için hemoglobin seviyesini bilmek esastır. Martinez ve ark. 2014'te Avrupa hemodiyaliz hastalarının bir veri tabanını kullanarak hemoglobin değerlerini tahmin etmek amaçlı bir çalışma yapmışlardır. Yöntem olarak Lineer modeller, Yapay sinir ağları, Destek vektör makineleri ve Regresyon ağaçları gibi farklı küresel modeller veri setine uygulanmıştır [116].

Kronik böbrek yetmezliği ilişkili anemi hastaları için Akabua ve ark.'nın yaptığı bir çalışmada, yarı-kör sağlam tanımlama denilen bir teknik ile bireysel anemi hastalarının gerçek hemoglobin yanıtları ve eritropoietin (EPO) yönetimine dayalı benzersiz bir model tanımlanmıştır [117].

Anemi konusunda 2014'te yayınlanmış bir çalışmada anemi tanısı için farklı yapay zekâ yöntemleri birlikte kullanılmıştır. Bu çalışmanın amacı kan değerleri parametrelerine bakarak kadınlarda demir eksikliğine bağlı aneminin teşhisinde hekimlere yardımcı olan bir sistem geliştirmektir. Önerilen sistemde verilerin eğitim aşamasında Klonal seçim algoritması (KSA) kullanılmaktadır. Test aşamasında ise bellek tabanlı sınıflandırma tekniklerinden k-en yakın komşu algoritması (KNN) ve regresyon ağacı sınıflandırma tekniklerinden Gini algoritması olmak üzere iki farklı sınıflandırıcı kullanılmaktadır. Karşılaştırma yapmak için yapay sinir ağı yöntemlerinden İleri beslemeli sinir ağı ve Olasılıksal sinir ağı ile de sınıflandırmalar yapılmıştır. Test işlemi sonucunda, başarı oranlarının mukayesesi Alıcı İşletim Karakteristiği (Receiver operating characteristics- ROC) analizi yöntemi ile gerçekleştirilmiştir. Bellek tabanlı sınıflandırmada doğruluk %96 iken regresyon ağacı yöntemi ile yapılan sınıflandırmada doğruluk %98,73 olarak bulunmuştur. Böylece KSA'nın standart sınıflandırıcısı olan KNN yerine Gini algoritmasının kullanımı ile daha yüksek bir başarı oranı elde edildiği gösterilmiştir [118].

Anemi dışında diğer hastalıklar için yapılmış benzer çalışmalar da literatürde mevcuttur. 2011'de Uzoka ve ark.'nın klinik karar destek sistemi olarak tasarladıkları çalışmada sıtma teşhisi için bulanık mantık ve analitik hiyerarşi işlemi (AHP) yöntemleri karşılaştırılmış, istatistiksel olarak yöntemler arasında fark olmadığı fakat uzman doktorun fikrine yakın olan sonucu bulmada bulanık mantık yönteminin daha iyi olduğu belirtilmiştir [119].

2016'da Arslan ve ark. iskemik inme tahmini için 80 hasta 112 sağlıklı bireye ait verilerde 17 değişken ve 1 sonuç değişkeni belirleyerek bir çalışma yapmışlardır. Destek vektör makineleri, stokastik eğim artırıcı ve cezalı lojistik regresyon

yöntemlerini 10 fold çapraz doğrulama örnekleme metodu ile uygulamışlar ve en iyi sonucu %97 ile destek vektör makineleri yöntemiyle aldıklarını belirtmişlerdir [120].

2020’de yapılan bir çalışmada kronik böbrek hastalığının tanısı için hibrit bir yöntem tasarımı gerçekleştirilmiştir. Yapay sinir ağları, karar ağaçları ve k-en yakın komşu algoritması kullanılarak yapılan sınıflandırma iyi sonuçlar vermiştir. Sınıflandırma öncesi verileri klonal seçim algoritması ile eğiterek veri sayısını arttırmışlar ve bu yöntemin yetersiz veri sayısı olan durumlar için kullanılabilir olduğunu göstermişlerdir [121].

Görüldüğü üzere veri madenciliği yöntemleri ile yapay öğrenme yöntemleri tıbbi veriler üzerinde sıklıkla kullanılmakta ve başarılı sonuçlar elde edilmektedir. Tüm bu bilgiler ışığında gerçekleştirilmesi düşünülen çalışma oldukça anlamlıdır ve literatüre katkı sağlayacağı düşünülmektedir.

BÖLÜM 3. HEMATOLOJİK HASTALIKLARIN TEŞHİSİ

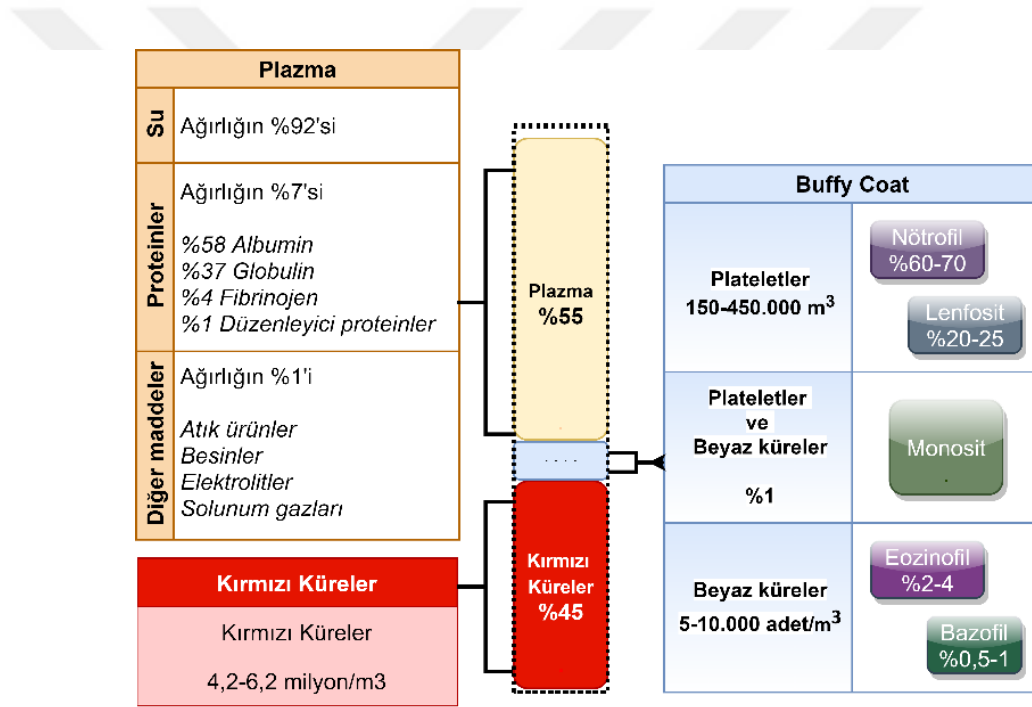
Bu bölümde öncelikle vücudumuz için hayati önem taşıyan kanı inceleyen bir bilim dalı olan hematoloji ve hematolojik hastalıklar hakkında bilgilendirme yapılacak, daha sonra çalışmada kullanılan veriler hakkında detaylı bilgiler verilecektir.

Hematoloji normal ve hastalıklı durumlarda kanın yapısını inceleyen ve işlevi ile ilgilenen bir bilim dalıdır. Kelimenin kökeni latince'den gelmektedir; haima kan logia bilim demektir. Hematoloji dâhili tıp bilimlerinin en önemli dallarından biridir. Hematoloji uzmanları ülkemizde tıp eğitiminin ardından yaptıkları 5 yıllık iç hastalıkları veya pediatri uzmanlığından sonra 3 yıl daha yan dal eğitimi almaktadırlar [122]. Hematoloji tıptaki en geniş hastalık spektrumuna sahip alanlardan biridir. Durum böyle olunca hematolojik hastalıkları anlayabilmek için kan ile ilgili birçok parametrenin bilinmesi gerekmektedir [123].

Kanın besinleri, hormonları ve oksijeni dokulara taşımak, hücre metabolizması sonucu oluşan artıkları toplamak, enfeksiyonlar ve diğer çevre faktörlerine karşı vücudu korumak gibi önemli görevleri vardır. Vücudumuzun yaklaşık %7-8'ini oluşturan kanın %40-45'i hücresel elementler ve kalanı plazmadan oluşur. Koagüle kanın hücresel elemanlarından ayrılmasından sonra kalan sıvı kısmı; kanın %55-60'ını oluşturur. Bunun %90'ı sudur, elektrolit ve protein gibi organik olmayan ve organik maddeleri içerir Hücresel elementleri kırmızı kan hücreleri (eritrositler), beyaz kan hücreleri (lökositler) ve kan pulcukları (trombositler) olarak açıklayabiliriz. Kandaki hücreler kemik iliğinde bulunan öncül kan hücrelerinin farklılaşması ve olgunlaşması sonucu ortaya çıkar ve işlev kazanırlar [56, 57].

3.1. Kanın Yapısı

Kan plazma, kırmızı küreler (eritrosit), beyaz küreler (lökosit) ve kan pulcukları (platelet/trombosit) denen elementlerden oluşan yaşamsal bir sıvıdır. Vücutta oksijen, besin maddeleri, hormonlar, vitaminler ve antikorlar gibi önemli bileşenleri dokulara taşır ve oluşan karbondioksit ve diğer atık maddeleri vücuttan uzaklaştırır. Kan oksijenle temas ettiğinde pıhtılaşır ve böylece kan kaybı engellenir. Metabolizmada oluşan ısı, kan sayesinde vücutta dolaşır ve vücut ısısı dengelenir. Vücudumuz için olmazsa olmaz bir sıvıdır. Şekil 3.1.'de kanın bileşenleri görülmektedir.

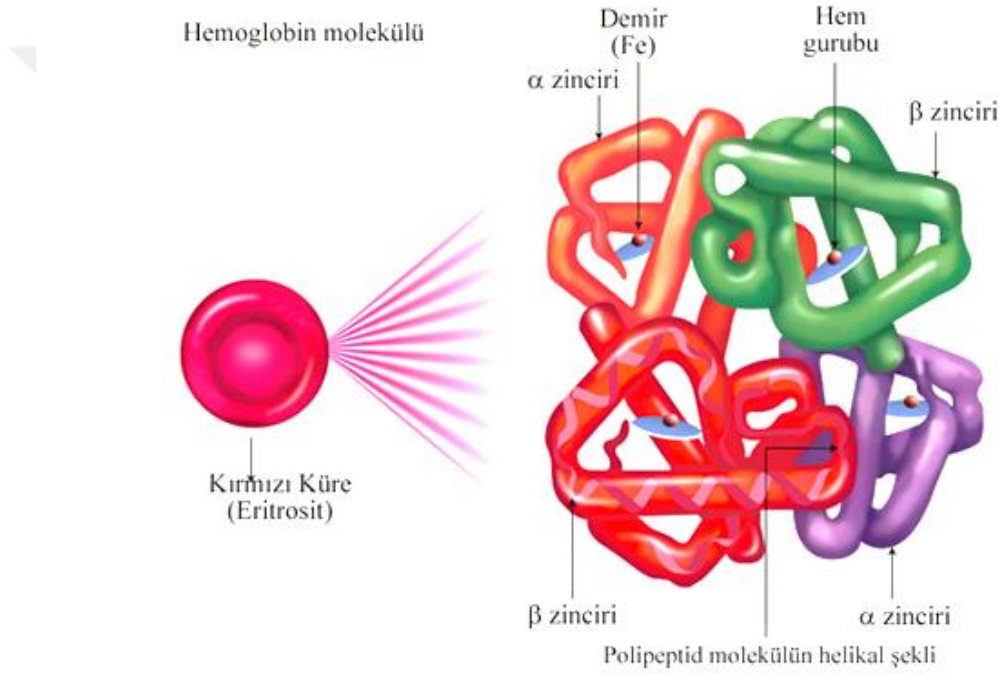


Şekil 3.1. Kanın Bileşenleri [122].

Koagüle edilmiş olan bir tüp kan santrifüj edilirse, altta kırmızı küreler ve üstte plazma olmak üzere iki ana kısma ayrılır. Birleştikleri yerde ise, “buffy coat” (ince beyaz tabaka) adı verilen, kan pulcukları ve beyaz kürelerden oluşan çok ince bir hat bulunur. Plazma, kanın yaklaşık %55’ini oluşturur. Plazma içerisinde; su, proteinler ve suda çözünen diğer maddeler bulunur [122].

3.1.1. Kırmızı küreler (Eritrositler)

Kırmızı küreler, aynı zamanda alyuvarlar veya eritrositler olarak da bilinir. Testlerde kısaca İngilizcesinin (Red Blood Cells) kısaltması olan “RBC” şeklinde yazılırlar. Kanın yarısına yakın kısmını eritrositler oluşturmaktadır. Eritrositlerin büyüklükleri 8 mm kadar olup mm^3 'te 4-6 milyon adet bulunurlar. Şekil 3.2.'de eritrositlerin yapısı görülmektedir.

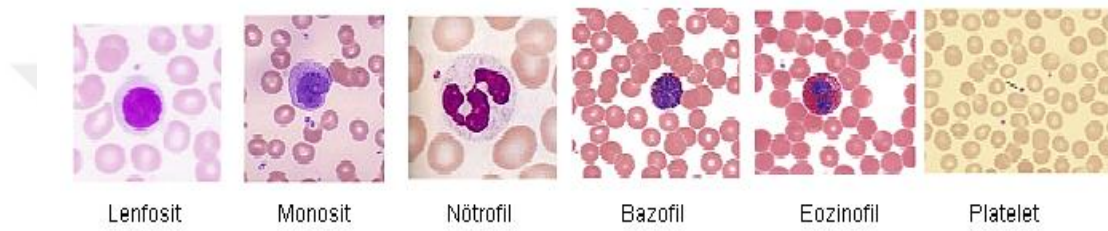


Şekil 3.2. Eritrositlerin Yapısı [122].

Eritrositlerin içleri “Hemoglobin” adı verilen bir yapıyı barındırır. Hemoglobin ortasında demir içeren bir proteindir. Bu protein yapısı, eritrositlerin akciğerlerden dokulara oksijen taşımalarını ve dokularda oluşan karbondioksiti tekrar akciğere taşımalarını sağlar. Embriyonik dönemde ve erişkinlerde hemoglobin oranları değişmektedir.

3.1.2. Beyaz küreler (Lökositler)

Beyaz küreler, aynı zamanda akyuvarlar veya lökositler olarak bilinir. Testlerde ingilizcesinin (White Blood Cells) kısaltması olan “WBC” şeklinde yazılırlar. En önemli görevleri vücudu enfeksiyonlara karşı savunmaktır. Düşük lökosit sayısı lökopeni olarak bilinir. Değişik türlerde lökositler mevcuttur. Şekil 3.3.’te çeşitli lökositler görülmektedir.



Şekil 3.3. Beyaz Kan Hücreleri ve Kan Pulcukları [122].

Her bir lökosit çeşidi vücudu farklı bir enfeksiyondan korumaktadır. En çok görülen lökosit çeşidi “nötrofiller”dir. Bakterilere en çok saldıran ve öldüren lökositler bunlardır. Nötrofil azlığına “nötropeni” denir. Her hücre değişik bir şekilde savaşır.

Lökositlerin birkaç sınıflaması vardır. Tek çekirdekli (mononükleer) hücreler; “monosit” ve “lenfosit” (B ve T hücreleri), parçalı çekirdekli hücreler; “nötrofil”, “bazofil” ve “eozinofil” olarak adlandırılır.

3.1.3. Kan pulcukları (Plateletler/Trombositler)

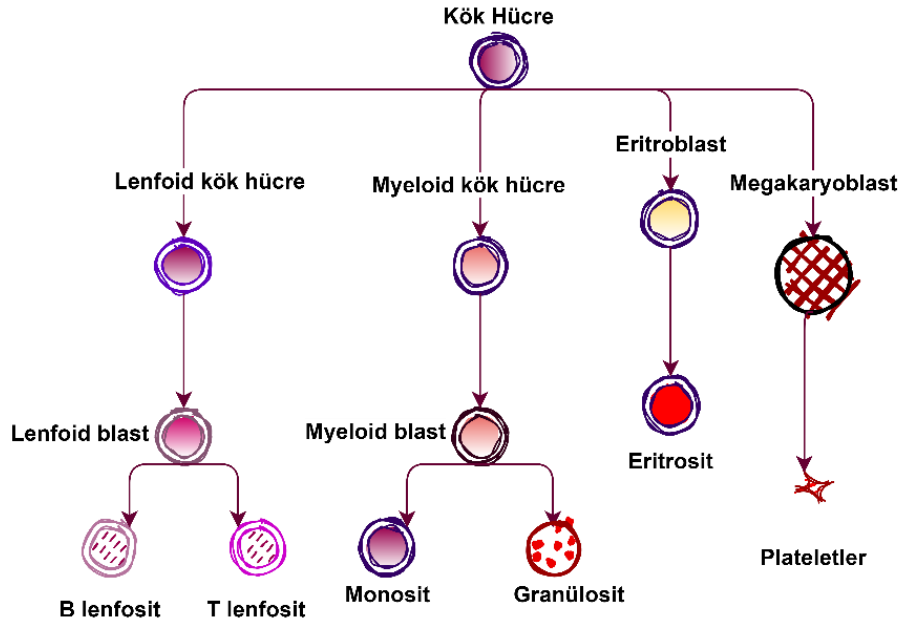
Trombositler kanda pıhtılaşmadan sorumlu hücrelerdir. Yapıları Şekil 3.3.’te görülmektedir. Genelde sayıları mm^3 ’te 150.000-450.000 arasında bulunur. Düşük trombosit sayısı “trombositopeni” olarak bilinir. Trombositopeni varlığında küçük sıyrık ve travmalarda bile kolayca kanama ya da morarma olabilir. Diğer bulgular arasında burun ve diş eti kanamaları ve özellikle alt bacaklarda oluşan peteşeni denen nokta şeklinde kanamalar bulunur [122].

3.2. Kanın Üretimi

Kanın üretimi daha embriyo döneminde sarı kesede başlar. Bu üretim bu şekilde 3 ay kadar devam eder ve sonrasında kan üretimi karaciğer ve dalakta yapılmaya başlar. Doğumdan itibaren kan üretimi uzun ve yassı kemiklerde gerçekleşir. Erişkin bir kişide üretim tamamen yassı kemiklerde yapılmaktadır. Kan üretimine “Hematopoez” adı verilir. Kanın yassı kemiklerde üretilmesine “medüller hematopoez” adı verilir. Bazı durumlarda kan, kemik iliği dışında karaciğer ve dalak gibi organlarda üretilir. Bu duruma da “ekstra medüller hematopoez” adı verilir [122].

Kan hücrelerinin üretimini sağlayan kemik iliği, bazı kemiklerin içinde bulunan süngerimsi bir yapıdır. Hepsi “kök hücre (stem cell)” denilen bir hücreden köken alırlar. Kemik iliğinde hematolojik kök hücreler bulunmaktadır. Bu hücreler çeşitli beslenme faktörleriyle vücudun ihtiyaçlarına göre kırmızı hücreler, beyaz hücreler ve kan pulcukları üretirler [122]. Şekil 3.4.’te kan üretiminin görseli verilmiştir.

Kemik iliğinde kendini yenileyebilen ve diğer hücrelere farklılaşabilen öncül kök hücreler, bazı büyüme faktörleri ile önce miyeloid ve lenfoid öncül hücrelere yönelir, daha sonra olgun kan hücrelerini oluştururlar. Kemik iliği erişkin bir insanda kan yapımının en önemli yeri ve en büyük organlardan biridir. Normal bir erişkinde kemik iliği günde yaklaşık 500 milyar hücre üretir, gerektiğinde bu üretimi artırabilir. Doğum öncesi, yeni doğan ve erken çocukluk dönemlerinde kan yapımı ve kan değerleri erişkin bir kişinin kan yapımı ve kan değerlerinden farklılık gösterir [56, 57].



Şekil 3.4. Kan Üretimi [122].

Şekil 3.4.’te de görüldüğü gibi kök hücreler her tip hücreyi üretebilirler. Ayrıca kendilerini çoğaltabilir ya da kopyalayabilirler. Bu kopya hücreler zaman içinde olgun kan hücrelerine dönüşürler. Olgunlaşan kan hücreleri kemik iliğini terk ederler. Sağlıklı bir vücut rahatlıkla kendine yetecek kadar kan hücresi üretebilir. Bu hücre türlerinden herhangi birinin eksikliğine “sitopeni” ismi verilir. İki tip hücrenin eksik olduğu durumlarda “bisitopeni” tanımı kullanılır. Eğer her üç seri de eksikse o zaman “pansitopeni” varlığından bahsedilir [122].

3.3. Hematolojik Hastalıkların Belirlenmesinde Kullanılan Laboratuvar Yöntemleri

Hematolojik hastalıkların belirlenmesinde çok çeşitli yöntemler kullanılmaktadır. 1642’de Leuweenhook’un kan hücrelerini keşfetmesinden bu yana birçok yeni laboratuvar yöntemi geliştirilmiştir [123]. Günümüzde hastanelerde çeşitli hematolojik hastalıklara tanı koyabilmek için Kan grubu ve Rh, Tam kan sayımı (Hemogram), Periferik yayma boyama (Çevresel kan yayması), Retükülosit sayımı, Sedimentasyon, Trombin ve Protrombin zamanı, İnvitro kanama zamanı, Faktörler (II, V, VII, VIII, IX), Direk ve İndirek coombs, Protein (C, S) aktiviteleri, Akım sitometri,

Kemik iliği incelemesi ve daha birçok farklı laboratuvar yöntemleri kullanılmaktadır [124, 125]. Burada hematoloji alanında en yaygın kullanılan laboratuvar yöntemlerinden olan tam kan sayımı, çevresel kan yayması ve kemik iliği incelemesi yöntemlerinden kısaca bahsedilecektir. Bu laboratuvar testleri aynı zamanda anemi tanısı koyabilmek için mutlaka bakılması gereken değerleri belirlemede çok önemlidir.

3.3.1. Tam kan sayımı

Tam kan sayımı kanda hangi hücreden ne kadar olduğunu ve dağılımını belirleyen bir laboratuvar yöntemidir. Tam kan sayımı ile genelde eritrosit sayısı, lökosit sayısı, trombosit sayısı, standart eritrosit indeksleri (MCV, MCHC, MCH, RDW), Hemoglobin (HGB), ve Hematokrit (HCT) değerleri belirlenmektedir [126].

Eritrosit indekslerinden MCV (Mean Corpuscular Volume) ortalama eritrosit hacmini ifade eder. Normal eritrositlerin hacmi 80-100 (fL) femtolitredir. Anemilerin sınıflandırılmasında etkin bir parametredir. Anemilerin sınıflandırılması MCV değerine göre mikrositik (<80 fL), makrositik (>100 fL) ve normositik olarak yapılır.

MCH (Mean Corpuscular Hemoglobin) ortalama eritrosit hemoglobini ifade eder. Eritrosit içindeki ortalama HGB miktarını gösterir. Normal değeri 30-34 (pg) pikogramdır.

MCHC (Mean Corpuscular Hemoglobin Concentration) ortalama eritrosit hemoglobininin konsantrasyonunu ifade eder. Eritrositlerin içerdiği HGB değerinin yüzde olarak ifadesidir. Normal değeri 30-36 g/dL dir.

RDW (Red Cell Distribution Width) kırmızı hücrelerin dağılımının genişliğini ifade eder. Eritrosit histogramlarından elde edilen istatistiksel bir değerdir. Eritrositlerin büyüklüklerinin farkını gösterir.

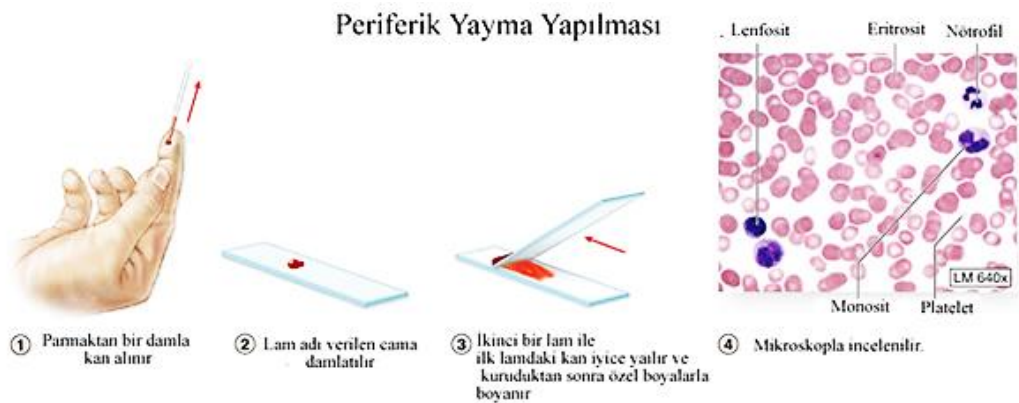
HCT (Hematokrit) kan sayımı cihazlarında bu değer ölçülmemekte, MCV ve eritrosit sayımından faydalanılarak (Denklem 3.1) formülü ile hesaplanmaktadır [126].

$$HCT = \frac{MCV (fL) \times Eritrosit Sayısı(\frac{milyon}{\mu L})}{10} \quad (3.1)$$

Burada “MCV (fL)” ortalama eritrosit hacmi ve “Eritrosit Sayısı” çarpımı 10’a bölünerek elde edilen değer hematokrit olarak hesaplanır.

3.3.2. Çevre (periferik) kan yayması

Periferik yayma tam kan sayımını tamamlayan bir yöntemdir. Periferik yayma yönteminde çevresel kan örneğinden alınan kan, lamaların arasına yerleştirilerek mikroskop ile incelenir. İyice kurumuş olan kan yayması plazmadaki ve çekirdekteki ayrıntıların ortaya çıkarılabilmesi için Wright boyasıyla boyanır. Eritrosit morfolojisinin düzgün yorumlanabilmesi için yaymada eritrositlerin tipik bikonkav bir disk şeklinde göründükleri yerlerin seçilmesi gerekir [126,127]. Şekil 3.5. çevresel kan yayma yapımını göstermektedir.



Şekil 3.5. Çevre Kan Yayması [122].

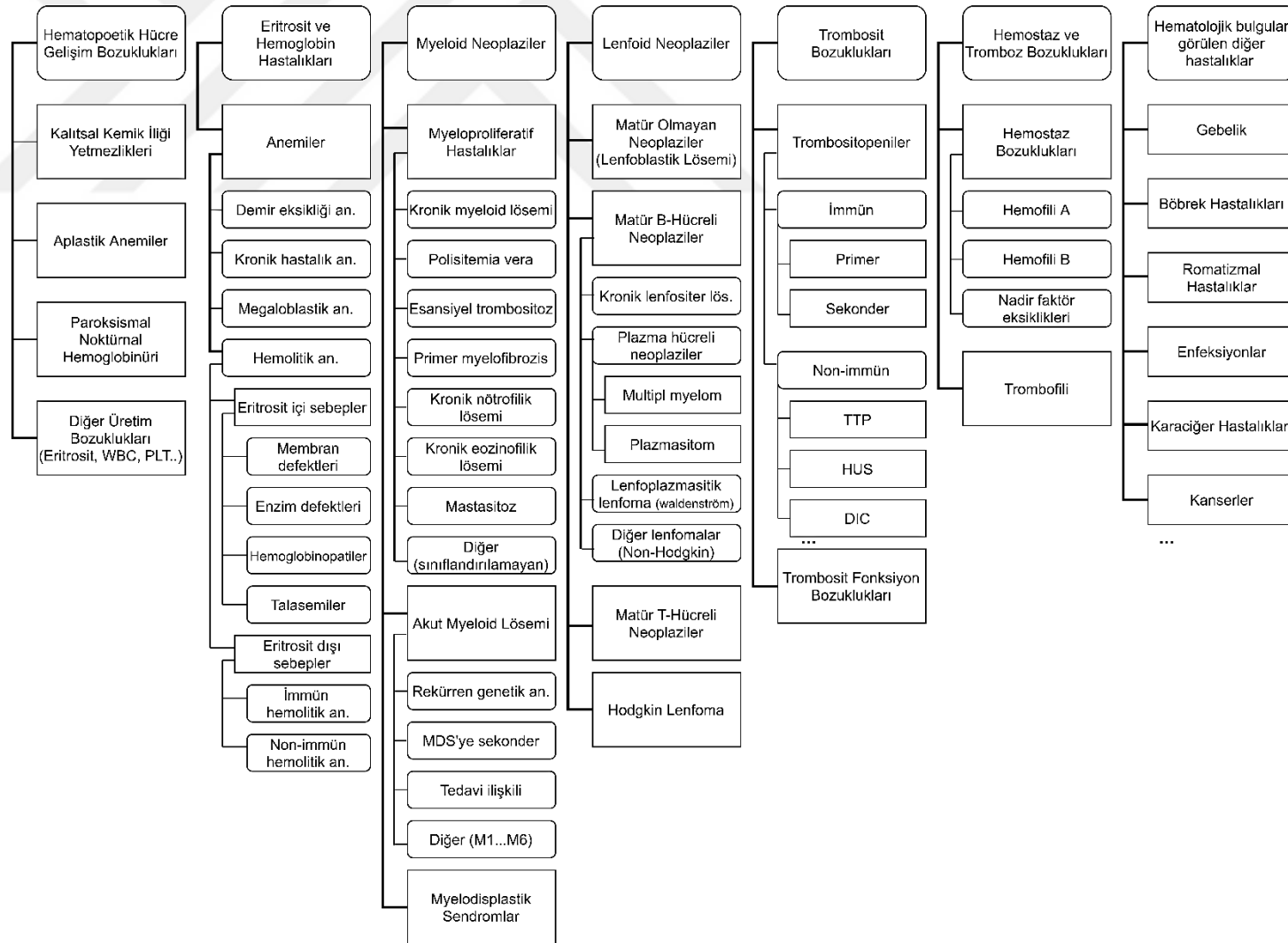
3.3.3. Kemik iliği incelenmesi

İğne aspirasyonu veya biyopsi ile kemik iliği örneği alınarak kemik iliğinin laboratuvarında incelenmesidir. Kemik iliği aspirat örneği hücre morfolojisinin değerlendirilmesi için kullanılır. İğne biyopsisi ise genel hücresellüğün ve kemik iliği yapısıyla hematopoetik hücreler arasındaki ilişkinin değerlendirilmesi için en uygun yöntemdir [127].

3.4. Kan Hastalıkları

Kan vücudumuzdaki en hayati sıvıdır ve bütün doku ve organlarda bulunmaktadır. Bu nedenle kan hastalıklarının tanısı ve tedavisi çoğu zaman hayati önem taşımaktadır. Bu bölümde kan hastalıklarının sınıflandırılmasından kısaca bahsedilecektir. Kan hastalıklarının genel sınıflandırılması Şekil 3.6.'da verilmiştir.

Bu sınıflandırma hematolojinin temel kitaplarından biri olan Hoffman ve ark.'nın yaklaşımına göre yapılmıştır. Buna göre moleküler-genetik temellerden sonra hematolojik hastalıklar önce hematopoetik hücre gelişim bozuklukları, sonra eritrosit hastalıkları, myeloid neoplaziler, lenfoid neoplaziler, trombosit bozuklukları, kanama/pıhtılaşma bozuklukları ve diğer hastalıklardan kaynaklanan hematolojik bozukluklar olarak sınıflandırma yapılmıştır [128].



Şekil 3.6. Hematolojik Hastalıkların Sınıflandırılması.

3.4.1. Anemi Çeşitleri ve Tanı Kriterleri

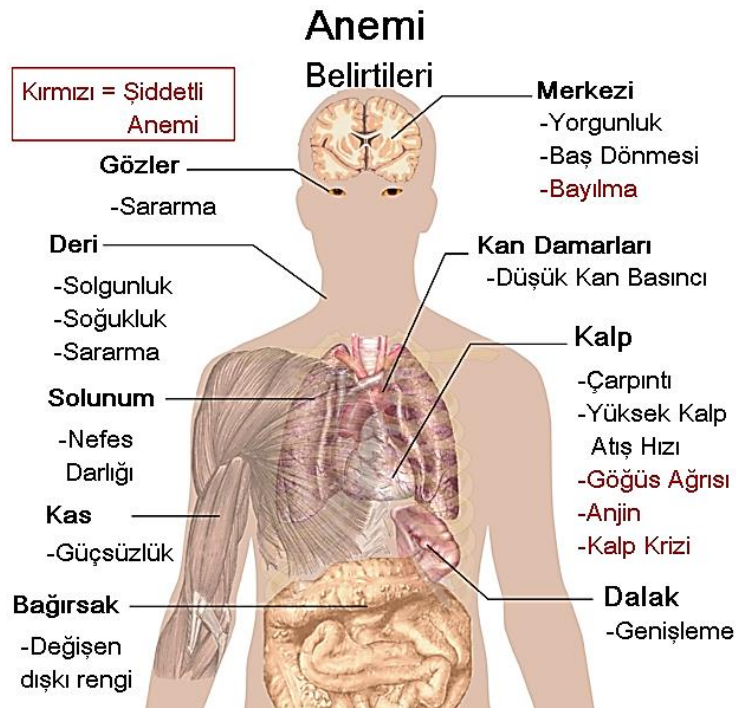
Türkiye’de ve dünyada en sık görülen kan hastalığı anemiler grubudur [9]. Dünya sağlık örgütüne göre anemi, kırmızı kan hücrelerinin sayılarının (ve buna bağlı olarak oksijen taşıma kapasitelerinin) vücudun fizyolojik ihtiyaçlarını karşılamada yetersiz kaldığı bir durumdur [10]. Anemi eritrosit kütesinin veya kan hemoglobinin ve hematokrit konsantrasyonunun azalması olarak da tanımlanır. Normal hemoglobin ve hematokrit değerleri yaş ve cinsiyete göre değişir. Hemoglobin ve hematokrit değerleri o yaş ve cins için normal değerlerin alt sınırından düşükse anemi vardır.

Aneminin başlıca nedenleri arasında kan hücrelerindeki bozukluklar yer alır. Eritrositlerin morfolojik veya fizyolojik işlevlerinde meydana gelen hasarlar anemilerin ortaya çıkmalarında önemli etkenlerdir. Anemiler temelde dört sebeple ortaya çıkmaktadır. Bu sebepler şöyle sıralanabilir:

1. Eritrositlerin yapım aşamasındaki bozukluklar (kemik iliğinden yeterli miktarda eritrosit üretilmemesi durumu).
 - a. Kemik iliğinin yeterli çalışmaması, kemik iliği yetmezliği (örnek: aplastik anemi, enfeksiyon, ilaç veya kansere bağlı kemik iliği yetmezliği).
 - b. Eritrositlerin olgunlaşmasında çok önemli bir rolü olan ve %90’ı böbrekten salınan eritropoetin sentezinde meydana gelen bozukluklar (örnek: kronik böbrek yetmezliği, hipotiroidi, romatizmal hastalıklar).
2. Eritrositlerin olgunlaşma aşamasındaki yapısal ve işlevsel bozukluklar (örnek: demir eksikliği, hemoglobin yapı ve işlev bozuklukları, kurşun zehirlenmesi, vitamin B12 eksikliği, folik asit eksikliği).
 - a. Eritrositlerin metabolizmada zamanından önce yıkılması ve yok edilmesi (örnek: hemolitik anemiler).

- b. Eritrosite ait diğer nedenler (örnek: eritrosit zarında oluşan bozukluklar, eritrositte enzim eksiklikleri, hemoglobinopatiler).
3. Eritrosit dışı nedenler (örnek: immün ve immün olmayan nedenler).
4. Kan kayıpları (kanamalar).

Aneminin yaygın görülen klinik bulguları halsizlik, yorgunluk ve solukluktur. Bu belirtilere ek olarak yaygın kemik ve eklem ağrıları ile karaciğer, dalak ve lenf bezelerinde büyüme lösemi ve diğer bazı hastalıklarda izlenebilir. Diğer belirtiler arasında baş ağrısı, çarpıntı, sık sık enfeksiyon geçirme, iştahsızlık, yenmemesi gereken toprak, sıva, buz gibi maddeleri yeme isteği, tad alma duyusunda kayıp, dilde ağrı ve ağız kenarlarında yaralar, tırnaklarda bozukluklar yer alabilir. Uzun süredir anemisi olan hastalar anemiye daha rahat tolere edebilirler ve anlamlı şikayetleri bulunmayabilir [10, 74, 127, 129]. Şekil 3.7.'de anemi belirtileri gösterilmektedir.



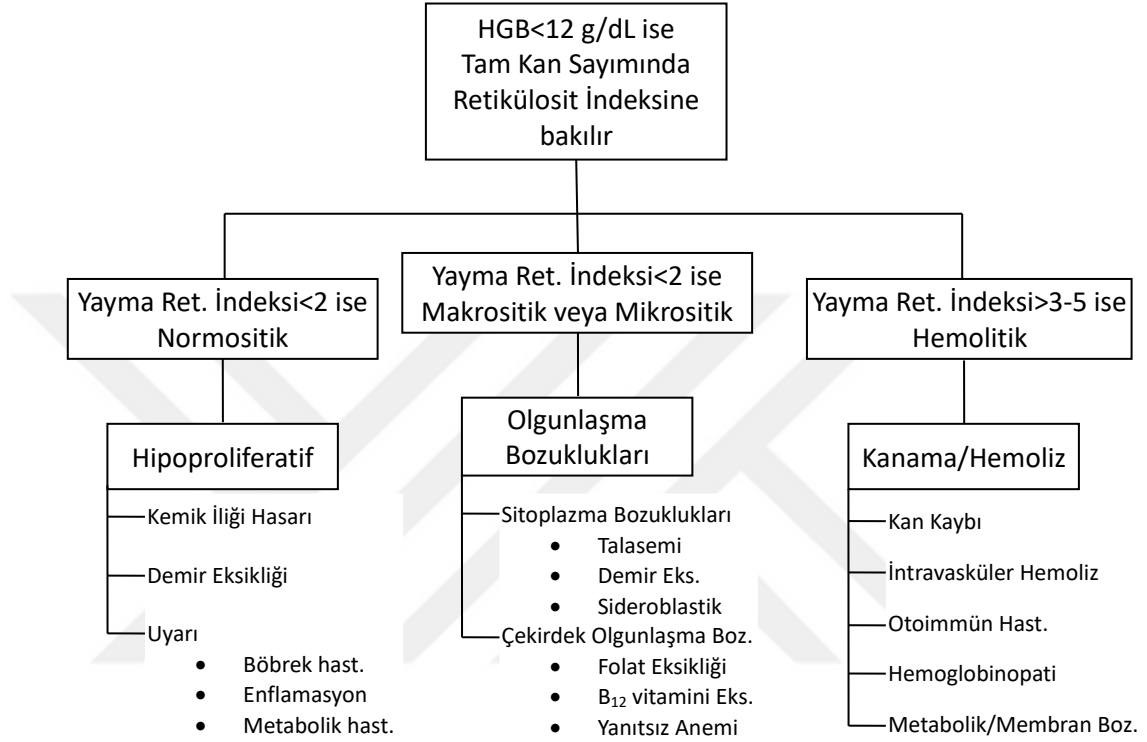
Şekil 3.7. Anemi Belirtileri [121].

Anemili hastada fizik muayenede önemli olan bulgular şu şekilde sıralanabilir: [10, 74, 127, 129, 130].

- Solukluk; en iyi avuç içi, tırnak yatakları, gözler ve kulak kepçesine bakarak anlaşılır.
- Yüksek nabız; ağır bir anemi durumu varsa kalp yetmezliği açısından dikkat edilmelidir.
- Anemiye bağlı olarak kalpte üfürüm duyulabilir.
- Demir ve vitamin B12 eksikliği anemilerinde dil kızarıklık ve ağrılı, tırnaklarda çizgilenme ve içe doğru çökme, dudak kenarlarında çatlama görülebilir.
- Enfeksiyon bulguları araştırılmalıdır.
- Deride koyu lekeler, kırmızı benekler, morluklar, renk değişikliği durumlarına dikkat edilir.
- Kemik ve eklemlerde ağrı olup olmadığına bakılır (özellikle lösemi açısından önemlidir).
- Kemiklerde genişleme ve kemik değişiklikleri varsa kronik kalıtsal hemolitik anemiler düşünülebilir.
- Başparmak anomalisi, kalpte, kemiklerde, deride, böbreklerde ve tırnaklarda anormal bulgulara bazı kalıtsal kemik iliği yetmezliklerinde rastlanabilir.
- Yaşa göre motor fonksiyonlarda ve zekâda gerilik, büyüme ve gelişme geriliği, yutma güçlüğü, halsizlik, huzursuzluk gibi bulgular değerlendirilmelidir.

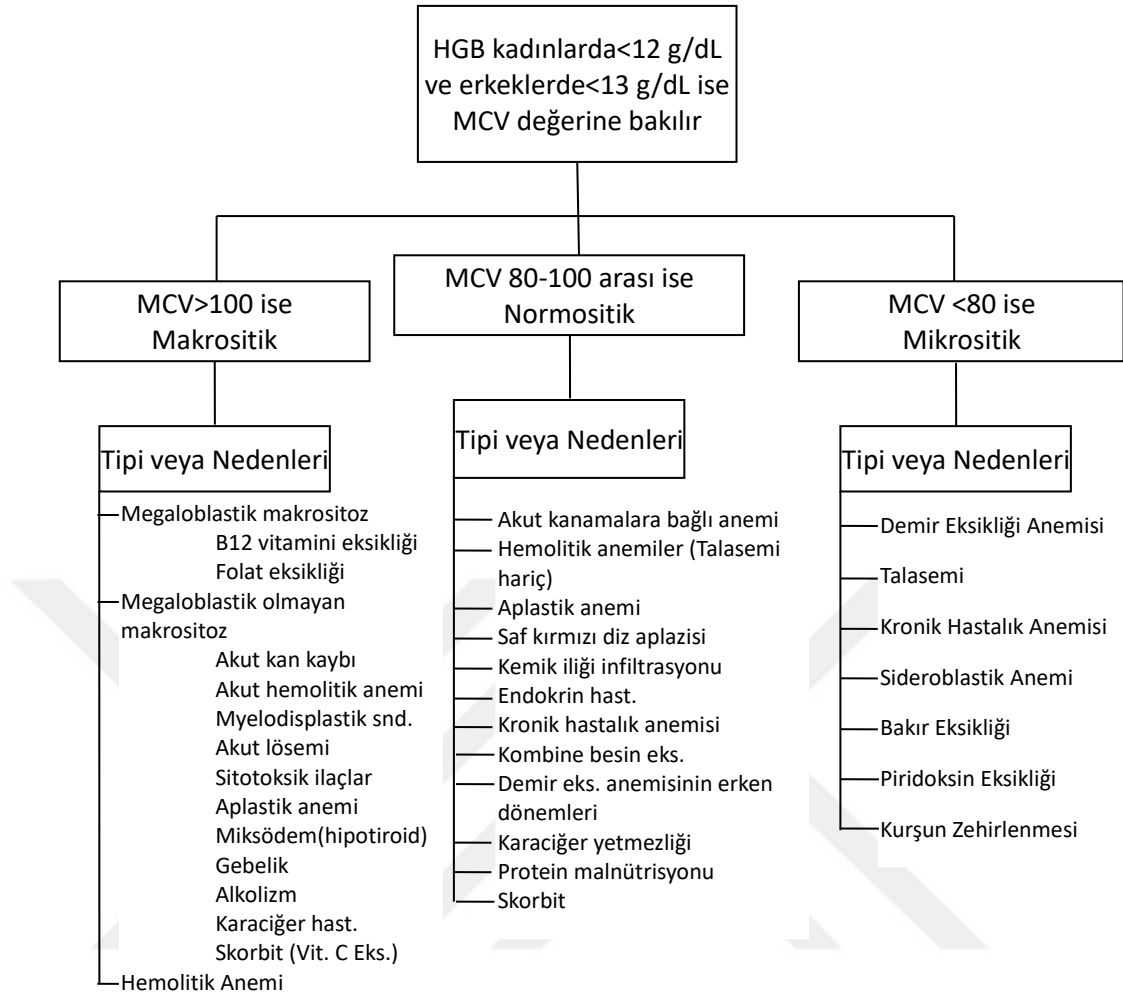
Anemili hastada ilk istenecek laboratuvar tetkikleri tam kan sayımı eritrosit indeksleridir (MCH, MCHC, MCV, RDW). Ayrıca duruma göre ileri tetkikler olan periferik yayma, formül lökosit, retikülosit sayımı değerleri istenebilir. Hastanın öyküsü, fizik muayenesi ve bu tetkikler ile anemi nedeni hakkında bir ön fikir elde edilir. Sonraki aşamada kesin tanı için ileri tetkikler istenebilir [74].

Aneminin sınıflandırılması rutin laboratuvar test sonuçlarına dayandırılabilir. Aneminin işlevsel açıdan sınıflandırılması Şekil 3.8.'de ve eritrosit morfolojisine göre sınıflandırılması Şekil 3.9.'da gösterildiği gibidir.



Şekil 3.8. Aneminin İşlevsel Açıdan Sınıflandırılması [127].

Üretim bozukluğu/hipoproliferatif anemileri düşük retikülosit üretim indeksi ve eritrosit yapısında hiç veya çok az değişiklik olması ile nitelenir. Kemik iliği yapısının öncül kök hücre havuzunun hasar görmesi, eritropoetin uyarısının olmaması veya demir eksikliği sonucunda meydana gelebilir. Eritropoetin, eritrositler için sitokin görevi gören bir glikoprotein hormondur. Burada bahsi geçen sitokin hücrelerin birbirleriyle iletişimini sağlayan bir protein grubudur. Eritropoetin böbreklerde üretilir ve eritrosit üretiminin kontrolünden sorumlu hormondur [130].



Şekil 3.9. Aneminin Eritrosit Morfolojisine Göre Sınıflandırılması [74].

Olgunlaşma bozukluklarında düşük retikülosit üretim indeksine makrositik veya mikrositik bir eritrosit yapısı eşlik eder. Eritrosit öncül hücre olgunlaşma sırasının bozulması folik asit ve B12 vitamini eksikliğine, kemoterapiye veya miyelodisplastik veya prelösemik bir duruma bağlı olabilir. Bunların tümü çekirdek olgunlaşma bozukluğu ile ilişkili olduğundan hastalarda makrositik anemiler, megaloblastik kemik iliği yapısı ve değişen derecelerde inefektif eritropoez mevcut olabilir.

Hemolize bağlı artmış eritrosit yıkımının söz konusu olduğu hastalarda normalin 3 katını aşacak düzeyde bir dengeleyici retikülosit indeks artışı ve hastalık süreci için ayırt ettirici olan veya olmayan bir eritrosit yapısı görülür.

Anemi sınıflandırılırken atılan ilk adım hem tanı hem tedavi için önemlidir. İşlevsel bozukluğa göre hastalığın tedavisi de değişecektir [127, 130].

Bu tez çalışması için incelenen anemi türleri Bölüm 5.1.'de detaylı bir şekilde incelenmiş ve açıklanmıştır.



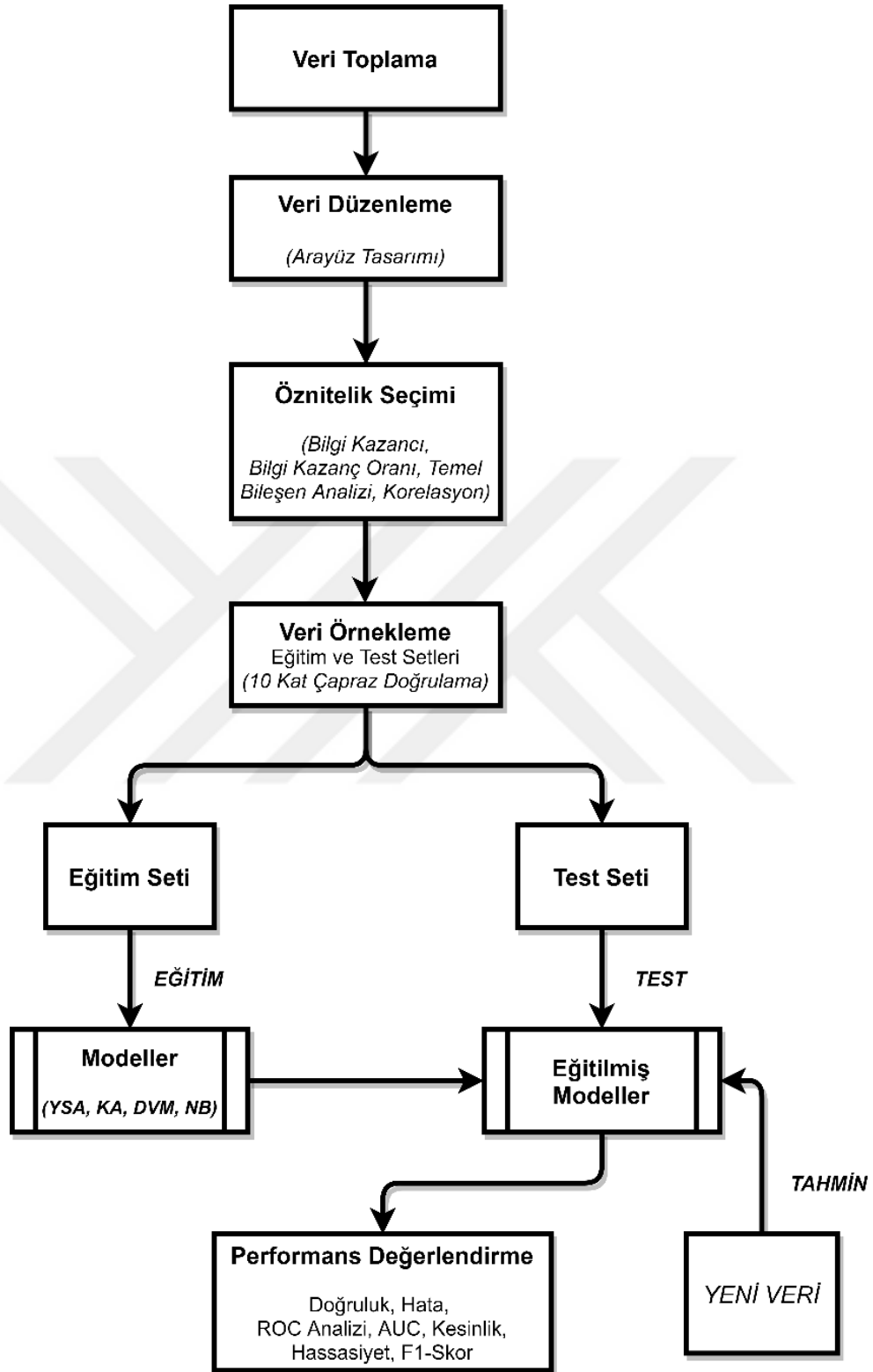
BÖLÜM 4. YAPAY ZEKA TEKNİKLERİ

Çalışmamızda hematolojik hastalıklardan en yaygını olan anemi tiplerine tanı koyulması süreci tasarlanan bir bilgisayar programına aktararak uzman doktorun karar verme sürecinin programa öğretilmesi amaçlanmaktadır. Artan hasta sayısı ve hastanelerdeki yoğunluk, bunun yanında uzman hekime ulaşmadaki güçlükler nedeniyle genel sağlık merkezi koşullarında anemiyi tanımayı sağlayacak bir sistem geliştirilmesi hedeflenmiştir. Bu sistemin birinci basamak sağlık hizmeti sunulan merkezlerde tanı için istenen tetkiklerle beraber işletilmesi ile bu sağlık sunucularında çalışan uzman olmayan personele yardımcı olması düşünülmektedir. En azından buna dayanılarak ileri merkeze sevk edilmesi gereken hastalar daha hızlı ve doğru seçilebilir. Bu amaca yönelik olarak hematoloji polikliniğine anemi öntanısı ile başvuran hastaların verileri ele alınarak bilgisayar programına anemi ilişkili tanımlar alan hastaların tetkiklerinin değerlendirilmesi öğretilmiştir. Öğrenme işleminden sonra sistem yeni verilerle test edilerek uzman doktor gibi karar verme yeteneği analiz edilmiştir. Çalışmanın asıl çıktısının doktorun karar vermede nasıl bir yöntem izlediğinin programa aktarılması olması planlanmaktadır. Diğer bir çıktı olarak uzman olmayan personele, doktor ve doktor adaylarına karar desteği sağlanması düşünülmektedir.

Veriler, program arayüzü kullanılarak elde edildikten sonra, sınıflandırma işlemi için dört temel model geliştirilmiştir. Yöntem olarak destek vektör makineleri (DVM), karar ağaçları (KA), yapay sinir ağları (YSA) ve Naive Bayes (NB) seçilmiştir. Bu modeller, güncel ve kanıtlanmış sınıflandırma yöntemleri oldukları ve literatürde oldukça başarılı sonuçlar verdikleri görüldüğü için seçilmiştir. Her bir model 9 farklı veri seti üzerinde çalıştırılmıştır. Bu veri setleri öznelitliklerin seçilmesi ile elde edilmiştir. Çeşitli öznelitlik seçme algoritmaları ile her bir öznelitlik için bir ağırlık hesabı yapılmış ve en yüksek ağırlığa sahip olan öznelitliklerden oluşan yeni veri setleri

hazırlanmıştır. Bu aşamada kullanılan öznitelik seçme yöntemleri bilgi kazancı, bilgi kazanç oranı, korelasyon ve temel bileşen analizidir. Son olarak, geliştirilen model kaydedilmiş ve yeni veriler üzerinde test edilmiştir. Bu değerlendirme işlemi k-kat çapraz doğrulama yöntemi ile gerçekleştirilmiştir. Alıcı işletim karakteristiği (ROC-Receiver Operating Characteristic) analizi kullanılarak modellerin performans değerlendirmesi yapılmıştır [131]. Her modelin başarısı doğruluk değerlerine ek olarak sınıflandırma hatası, eğri altında kalan alan (AUC-Area under the curve), kesinlik, geri çağırma ve F1-skoru ölçütleri kullanılarak ölçülmüştür. Ayrıca her bir model için ROC eğrileri ve karmaşıklık matrisleri de verilmiştir. Gerçekleştirilen çalışmanın akış diyagramı Şekil 4.1.'de görülmektedir.





Şekil 4.1. Çalışmanın akış diyagramı.

4.1. Öznitelik Seçme Yöntemleri

Öznitelik basitçe veri setinde yer alan sütunlara verilen isimdir. Yapay öğrenme yöntemleri kullanılırken, veri setinde birbiri ile ilişkili olan ve sayıca fazla öznitelik olması istenen bir durum değildir. Seçilen yapay öğrenme yöntemine bağlı olarak öznitelik sayısının artması algoritmanın çalışma süresini uzatabilir, başarısını olumsuz etkileyebilir. Veri setinde bağımsız özniteliklere sahip olmak, sınıflandırma sürecinde daha güçlü sonuçlar elde edilmesini sağlar.

Bu çalışmanın amacı, doktorun karar verme sürecini bilgisayara aktarmaktır. Bu nedenle doktorun kullandığı verilerin aynısını kullanmak gerekmektedir. Bununla birlikte, bilgisayarla öğrenme süreci çeşitli yapay öğrenme yöntemleri kullanılarak yapılır. Yapay öğrenme yöntemlerinde, algoritmalara verilen özniteliklerin sonuç üzerindeki etkileri önemli bir faktördür ve bu nedenle öznitelik seçme yöntemlerinin kullanılması esastır. Öznitelikleri seçerken, temel işlem, her bir öznitelik için ağırlıkların belirlenmesini ve özniteliklerin ağırlıklarına göre seçilmesini veya ortadan kaldırılmasını içerir. Bir özniteliğin ağırlığı genellikle (0,1) veya (-1, +1) aralığında hesaplanır. Hesaplanan ağırlık değeri 1 veya -1'e ne kadar yakınsa, ilgili özniteliğin sonuç üzerindeki etkisi o kadar önemli olacaktır. Bu değer 0'a ne kadar yakınsa sonuç üzerinde o kadar az etkisi olur ve ilgili öznitelik ortadan kaldırılabilir.

Bu çalışmada özniteliklerin seçimi için dört farklı yöntem kullanılmıştır. Bu yöntemler, bilgi kazancı, bilgi kazanç oranı, temel bileşen analizi (Principal component analysis-PCA) ve korelasyona dayalı öznitelik alt kümesi seçimini (CFS) içerir. Bu görev için RapidMiner ve WEKA veri madenciliği araçları kullanılmıştır [7, 8]. Bilgi kazancı, bilgi kazanç oranı ve PCA tabanlı öznitelik seçme işlemleri RapidMiner kullanılarak gerçekleştirilmiştir. WEKA kullanılarak çeşitli yöntemler denendi, ancak CFS dışındaki yöntemler 25 özelliğin tamamını seçti ve bu nedenle özellikler azaltılamadı. Bu dört yöntemi uygulayarak, orijinal veri setine ek olarak sekiz farklı veri seti daha elde edildi. Her bir veri setine dahil edilen öznitelikler Tablo 4.1.'de görülmektedir. Burada orijinal veri seti dışında, kullanılan veri setleri için seçilen öznitelikler görülmektedir.

Tablo 4.1. Öznitelik seçimi sonrası oluşturulan veri setleri.

Öznitelikler	CFS	PCA	Bilgi Kazancı			Bilgi Kazanç Oranı		
			>0,1	>0,2	>0,5	>0,3	>0,4	>0,5
Yaş	X	X	X					
Cinsiyet								
Kronik Hastalık								
Belirtiler								
CRP		X						
D.BİLİRUBİN						X	X	
DEMİR		X	X			X	X	
FERRİTİN	X	X	X	X	X	X	X	X
FOLAT	X	X	X	X	X	X	X	X
HCT		X	X			X	X	
HGB	X	X	X	X		X		
İ.BİLİRUBİN			X			X	X	X
CREATİNİN								
MCH		X	X			X	X	
MCHC						X		
MCV	X	X	X					
NEUT#						X		
PLT		X						
RBC		X	X			X	X	X
RDW		X				X		
T. BİLİRUBİN	X					X	X	X
TIBC		X	X			X	X	X
UIBC		X	X	X		X		
VİTAMİN B12	X	X	X	X	X	X	X	X
WBC		X						

Bütün veri setlerinde yer alan özniteliklerin Ferritin, Folat ve Vitamin B12 olduğu görülmektedir. Bu öznitelikler en önemli ve etkili özniteliklerdir. Sonuç olarak, sınıflandırma algoritmaları toplam dokuz veri seti için çalıştırılmış ve sonuçlar karşılaştırılmıştır. Bu bölümde, çalışmada kullanılan dört farklı öznitelik seçme yöntemi hakkında genel bilgiler verilmiştir.

4.1.1. Bilgi kazancı (Information gain)

Bir veri kümesindeki bir özniteliğin bilgi kazancı, ait olduğu sınıfı belirleme yeteneğidir. Örneğin, bir veri kümesindeki bir özniteliğin değeri, ait olduğu sınıfı bilmemizi sağlıyorsa, özniteliğin bilgi kazancı 1 olacaktır. Aksi durumda ise, yani bir özniteliğin değeri bize ait olduğu sınıf hakkında hiçbir bilgi vermiyorsa, o zaman bu özelliğin bilgi kazancı 0 olacaktır. Esasen, bilgi kazancını anlamak için entropi

teorisinin anlaşılması gerekir. Entropi, basitçe verilerin içerdiği bilgiler olarak tanımlanabilir. Shannon'un entropi formülü Denklem (4.1)'de ve Denklem (4.2) 'de verilmiştir [132].

$$E(sınıf) = - \sum_{i=1}^c p_i \log_2(p_i) \quad (4.1)$$

$$E(sınıf, öznitelik) = - \sum_{j=1}^v \frac{c_j}{c} E(C_j) \quad (4.2)$$

Denklem (4.1)'de c , sınıfların sayısıdır (hedef değişkenin alabileceği değerlerin sayısı da denebilir). p_i , rastgele bir verinin sınıf i 'den gelme olasılığıdır. Denklem (4.2)'deki v özniteliklerin sayısıdır (tahmin değişkeninin alabileceği değerlerin sayısı da denebilir). C ise sınıf değerlerini temsil etmektedir.

Veri kümesindeki bir öznitelik her sınıf için farklı bir değere sahipse entropisi 0 olur. Diğer bir deyişle, o özelliğin değerine göre sınıf belirlenebilir ve diğer özelliklere bakmaya gerek yoktur. Bu durumda, bilgi kazancı 1 olur. Bir özellik sınıf değeri ile ne kadar az ilişkili olursa, bilgi kazancı o kadar düşük olacaktır. Bilgi kazancı hesaplaması için kullanılan denklem Denklem (4.3) 'te verilmiştir.

$$Bilgi\ Kazancı(sınıf, öznitelik) = E(sınıf) - E(sınıf, öznitelik) \quad (4.3)$$

Sınıf ve öznitelik bu denklemde görüldüğü şekilde temsil edilmektedir. Veri kümesini sınıflara böldükten sonra, tüm sınıfların entropi değerinden belirlenen özniteliğin entropi değerinin çıkarılmasıyla bilgi kazancı elde edilir. Entropide, değeri 1'e yaklaştıkça özelliğin önemi azalırken, bilgi kazanımında değeri 1'e yaklaştıkça önemi artar.

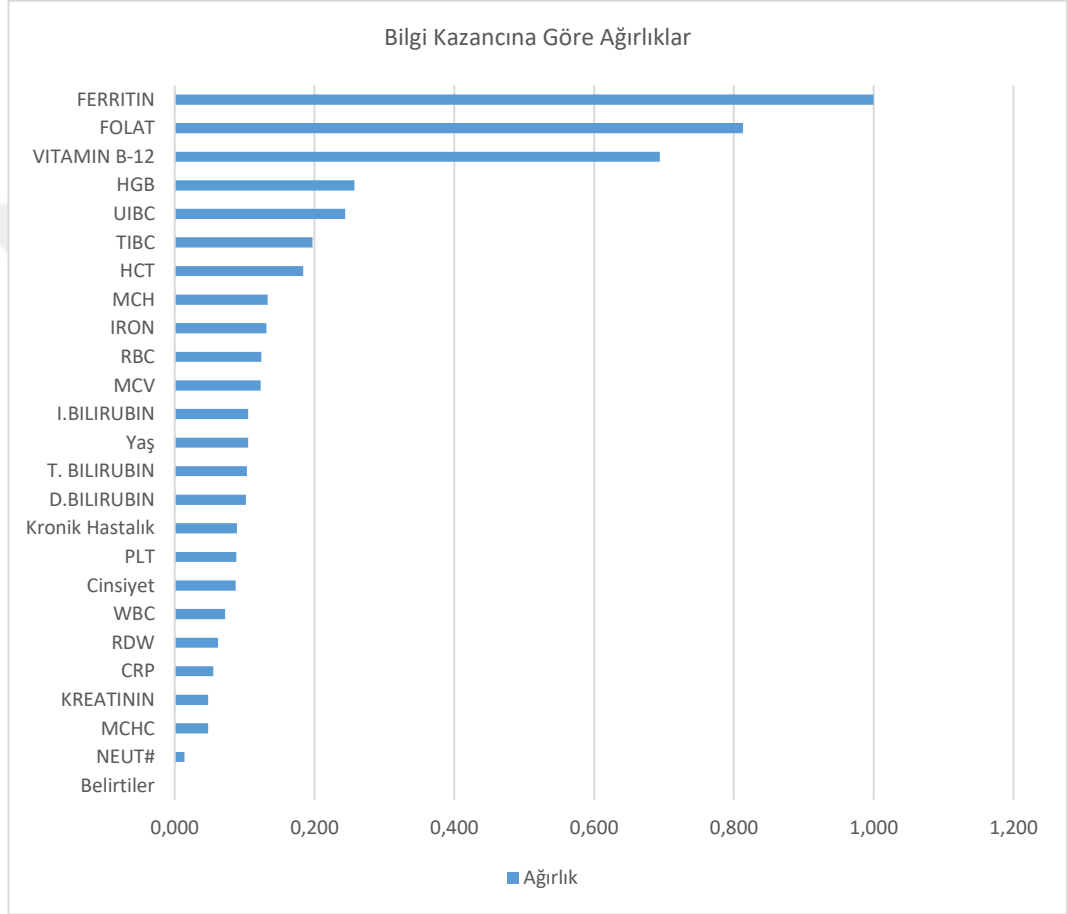
Bu çalışma için elde edilen bilgi kazanım değerleri Tablo 4.2. ve Şekil 4.2.'de gösterilmektedir. Bilgi kazancı yöntemi ile ağırlık değeri 0,5'ten büyük olan 3

özniteliğin seçildiği bir veri seti, 0,2'den büyük ağırlık değerine sahip 5 özniteliğin seçildiği bir veri seti ve ağırlık değeri 0,1'den büyük olan 13 özniteliğin seçildiği bir veri seti olmak üzere üç farklı veri seti üretilmiştir.

Tablo 4.2. Bilgi kazancı yöntemi ile bulunan öznitelik ağırlıkları.

Öznitelik	Ağırlık
Belirtiler	0,000
NEUT#	0,014
MCHC	0,048
KREATININ	0,048
CRP	0,055
RDW	0,062
WBC	0,072
Cinsiyet	0,087
PLT	0,088
Kronik Hastalık	0,089
D.BİLİRUBIN	0,102
T. BİLİRUBIN	0,103
Yaş	0,105
I.BİLİRUBIN	0,105
MCV	0,123
RBC	0,124
DEMİR	0,131
MCH	0,133
HCT	0,184
TIBC	0,197
UIBC	0,244
HGB	0,257
VİTAMİN B-12	0,694
FOLAT	0,813
FERRİTİN	1,000

Tablo 4.2. ve Şekil 4.2.'de görüldüğü gibi bilgi kazancı en fazla olan öz nitelikler sırasıyla Ferritin, Folat ve Vitamin B12 olarak bulunmuştur. Bu yöntem ile oluşturulan veri setlerinde yer alan öz nitelikler, ağırlıkları 0,1 değerinin üzerinde olan 13 adet, 0,2 değerinin üzerinde olan 5 adet ve 0,5 değerinin üzerinde olan 3 adet öz nitelik olarak seçilmiştir.



Şekil 4.2. Bilgi kazancına göre öz nitelik ağırlıkları.

4.1.2. Bilgi kazanç oranı (Information gain ratio)

Bilgi kazanç oranı, bilgi kazancının içsel değere oranıdır. Temel olarak, içsel değer, bir veri ögesinin ait olduğu sınıfı belirlemek için gereken bilgi miktarıdır. Bilgi kazanımının yeterli olmadığı durumlarda kazanç oranının değeri uygulanabilir. Denklem (4.4) bilgi kazanç oranının hesaplanmasını göstermektedir.

$$\text{Bilgi Kazanç Oranı}(\text{sınıf}, \text{öznitelik}) = \frac{E(\text{sınıf}) - E(\text{sınıf}, \text{öznitelik})}{E(\text{öznitelik})} \quad (4.4)$$

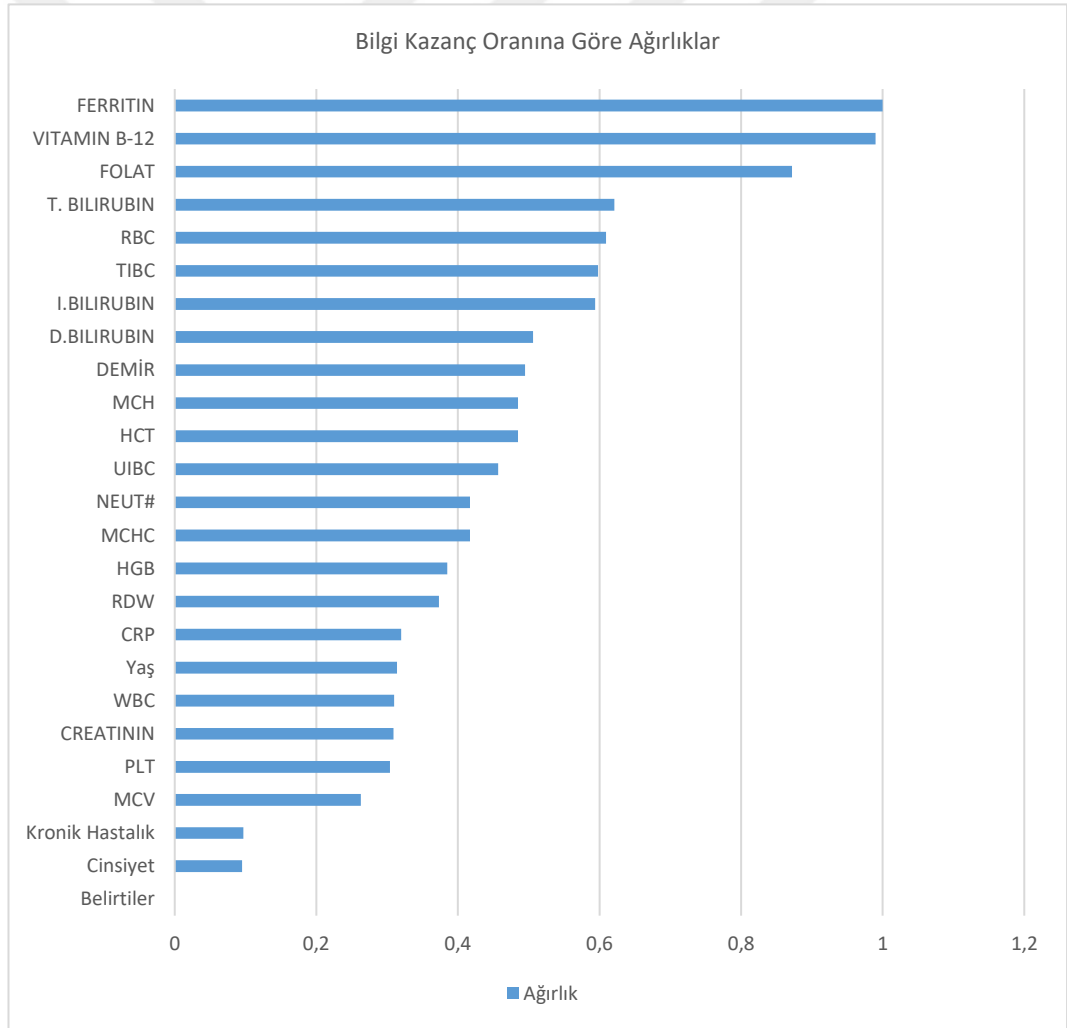
Bilgi kazanç oranı değerinin hesaplamasında, özniteliğin değeri 1'e yaklaştıkça önemi de artar [133]. Bu çalışma için elde edilen bilgi kazanım oranı değerleri Tablo 4.3. ve Şekil 4.3.'te görülebilir. Bilgi kazanç oranı yöntemi ile ağırlık değeri 0,5'ten büyük olan 7 özniteliğin seçildiği bir veri seti, ağırlık değeri 0,4'ten büyük olan 11 özniteliğin seçildiği bir veri seti ve ağırlık değeri 0,3'ten büyük olan 16 özniteliğin seçildiği bir veri seti olmak üzere üç farklı veri seti üretilmiştir.

Tablo 4.3. Bilgi kazanç oranı yöntemi ile bulunan öznitelik ağırlıkları.

Öznitelik	Ağırlık
Belirtiler	0,000
Cinsiyet	0,095
Kronik Hastalık	0,097
MCV	0,263
PLT	0,304
CREATININ	0,309
WBC	0,310
Yaş	0,314
CRP	0,320
RDW	0,373
HGB	0,385
MCHC	0,417
NEUT#	0,417
UIBC	0,457
HCT	0,485
MCH	0,485
DEMİR	0,495
D.BİLİRUBIN	0,506

Tablo 4.3. (Devamı)

Öznitelik	Ağırlık
I.BİLİRUBİN	0,594
TIBC	0,598
RBC	0,609
T. BİLİRUBİN	0,621
FOLAT	0,872
VİTAMİN B-12	0,990
FERRİTİN	1,000



Şekil 4.3. Bilgi kazanç oranına göre öznitelik ağırlıkları.

Tablo 4.3. ve Şekil 4.3.'te görüldüğü gibi bilgi kazancı en fazla olan öznitelikler sırasıyla Ferritin, Vitamin B12 ve Folat olarak bulunmuştur. Bu yöntem ile oluşturulan veri setlerinde yer alan öznitelikler, ağırlıkları 0,35 değerinin üzerinde olan 16 adet, 0,45 değerinin üzerinde olan 11 adet ve 0,6 değerinin üzerinde olan 7 adet öznitelik olarak seçilmiştir.

4.1.3. Temel bileşen analizi (Principal component analysis - PCA)

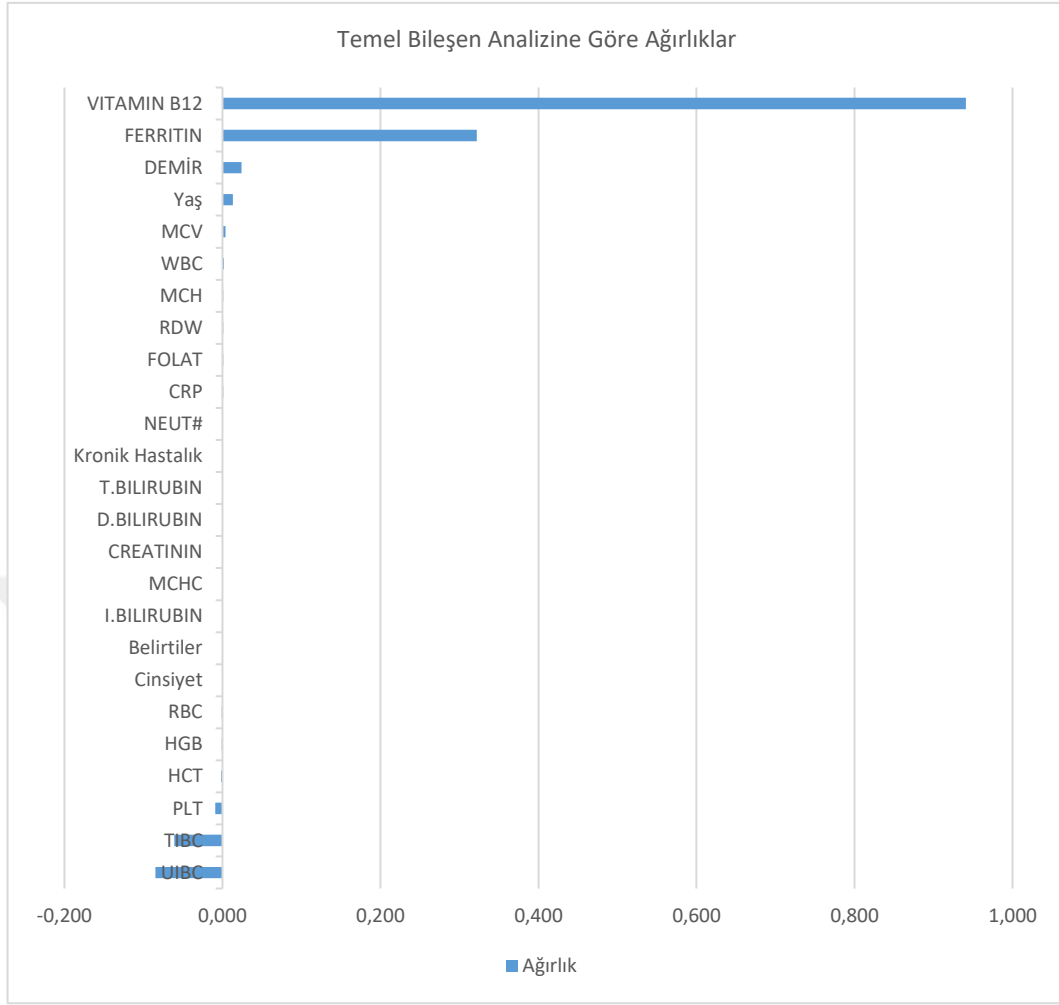
Temel bileşen analizi, yaygın olarak kullanılan bir öznitelik seçim yöntemidir. Temel bileşen analizini gerçekleştirmek için öncelikle öznitelikler arasındaki ilişkinin belirlenmesi gerekir. Sınıflandırma algoritmalarının ve makine öğrenmesi yöntemlerinin kullanımında birbiriyle ilişkili çok sayıda özniteliğe sahip olmak istenmeyen bir durumdur. Veri kümesinde bağımsız özniteliklere sahip olmak, sınıflandırma sürecinde daha güçlü sonuçlar elde edilmesini sağlar. Temel bileşen analizi kullanılarak, birbiriyle ilişkilendirilen çok sayıda öznitelik, korelasyon olmaksızın daha az özellik ile temsil edilebilir.

Bu yöntemde temel amaç birbiri ile ilişkili olan öznitelikleri, matematiksel prosedürler kullanarak birbiri ile ilişkili olmayan temel bileşen de denilen özniteliklere dönüştürmektir. Burada temel bileşen sayısı başlangıçtaki öznitelik sayısına eşit ya da daha az sayıda olmaktadır [7].

PCA ile yapılan ağırlık hesaplamasında, ağırlık değerleri 1'e yaklaştıkça özniteliğin önemi de artmaktadır. Bu çalışma için elde edilen ağırlık değerleri Tablo 4.4. ve Şekil 4.4.'te gösterilmiştir.

Tablo 4.4. Temel bileşen analizi (PCA) yöntemi ile bulunan öznelik ağırlıkları.

Öznelik	Ağırlık
UIBC	-0.085
TIBC	-0.061
PLT	-0.009
HCT	-0.002
HGB	-0.001
RBC	-0.001
Cinsiyet	-0.000
Belirtiler	0.000
I.BİLİRUBIN	0.000
MCHC	0.000
CREATININ	0.000
D. BİLİRUBIN	0.000
T. BİLİRUBIN	0.000
Kronik Hastalık	0.000
NEUT#	0.000
CRP	0.001
FOLAT	0.001
RDW	0.001
MCH	0.001
WBC	0.002
MCV	0.004
Yaş	0.013
DEMİR	0.024
FERRITIN	0.322
VİTAMİN B12	0.941



Şekil 4.4. Temel bileşen analizine göre öznitelik ağırlıkları.

Tablo 4.4. ve Şekil 4.4.'ten de anlaşılacağı gibi temel bileşen analizi yöntemi ile 0'dan farklı ağırlık değerine sahip tüm öznitelikler seçilmiş ve 16 öznitelik içeren yeni bir veri seti oluşturulmuştur.

4.1.4. Korelasyon tabanlı öznitelik altküme seçimi (Correlation-based feature subset selection -CFS)

Bu yöntem, sınıflandırma yöntemlerinin başarısını artırma konusunda yararlı olabilecek bir dizi öznitelik seçmeyi amaçlamaktadır. Bir özniteliğin etkili olabilmesi için, bu özniteliğin sınıfla yüksek bir korelasyonu olmalı, diğer özniteliklerle ise daha az ilişkili olmalıdır. Bu yöntemde her öznitelik seti ayrı ayrı ele alınır ve korelasyon ağırlık değerleri hesaplanır. Ele alınan alt kümelerden en yüksek ağırlığa sahip olan alt

küme, sınıflandırma algoritmasına sunulur [134]. Bu çalışma için elde edilen alt kümedeki öz nitelikler Şekil 4.5.'te görülebilir.

```

===Attribute Selection on all input data ===
Search Method:
    Best first.
    Start set: no attributes
    Search direction: forward
    Stale search after 5 node expansions
    Total number of subsets evaluated: 178
    Merit of best subset found: 0.549
Attribute Subset Evaluator (supervised, Class (nominal): Diagnosis):
    CFS Subset Evaluator
    Including locally predictive attributes
Selected attributes: 1,9,10,12,17,22,25 : 7
    Age
    FERRITIN
    FOLATE
    HGB
    MCV
    T. BİLİRUBIN
    VİTAMİN B-12

```

Şekil 4.5. CFS yöntemine göre seçilen öz niteliklerin ekran görüntüsü [131].

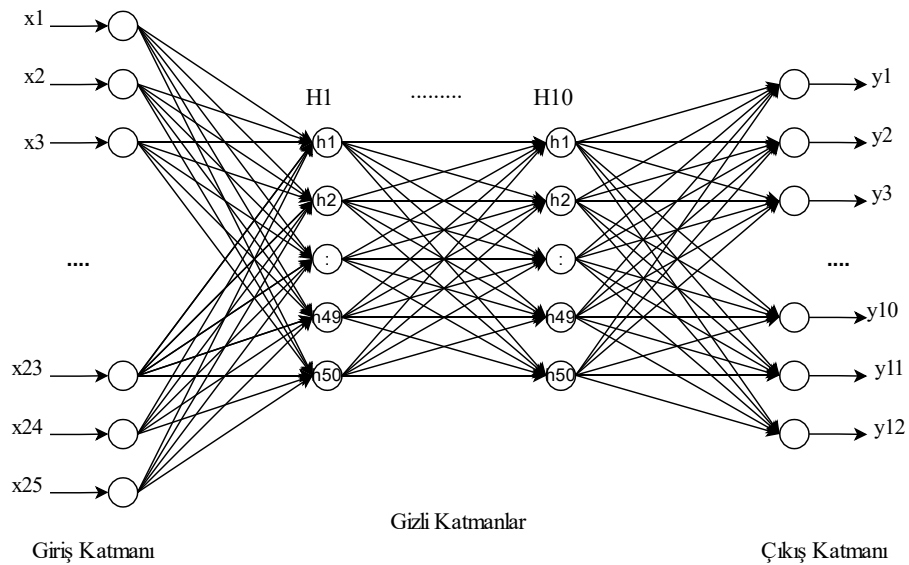
Toplam 178 farklı alt küme değerlendirilmiş ve en iyi alt kümenin ağırlık değeri 0,579 olarak bulunmuştur. Bu alt kümede, yedi öz nitelik (yaş, ferritin, folat, HGB, MCV, T. bilirubin ve vitamin B12) seçilmiş ve yeni bir veri seti oluşturulmuştur.

4.2. Sınıflandırıcılar

Bu çalışmada, iyi bilinen sınıflandırma yöntemlerinin performansı, tamamen orijinal bir veri seti oluşturularak değerlendirilmiştir. Literatürde yaygın olarak kullanılan yöntemlerden yapay sinir ağları, destek vektör makineleri, karar ağaçları ve Naive bayes yöntemleri sınıflandırıcılar olarak seçilmiştir. Bunlar, umut verici sonuçlar veren son teknoloji sınıflandırma yöntemleridir. Ayrıca yapılan kaynak araştırmasında bu yöntemlerin tıbbi verilerle birlikte kullanıldığında başarılı sonuçlar verdiği görülmüştür. Bu nedenle bu yöntemler uygulanmış ve sonuçlar karşılaştırılmıştır. Sınıflandırma işlemleri MATLAB® R2020a versiyonu kullanılarak gerçekleştirilmiştir [135].

4.2.1. Yapay sinir ağıları

Yapay sinir ağıları (YSA), kabaca insan beyninin öğrenme sürecini modelleyen güncel ve başarılı bir yöntemdir ve bu çalışma kapsamında uygun bir yöntem olduğu düşünülmüştür. İnsan beyninin örnekleri analiz ederek öğrendiği gibi, yapay sinir ağıları da örneklerden öğrenmektedir. YSA, nöron denilen insan beynindeki sinir hücreleri gibi birbirine bağlı hücrelerden oluşur. YSA modelinde, her nöronun girdileri, ağırlıkları, toplama ve aktivasyon işlevleri ve çıktıları olmalıdır [136]. Her giriş, nöronun aktivasyon seviyesini etkileyen bir ağırlığa sahiptir. Çıkış değeri, ağırlıklarla çarpılan giriş sinyallerinin toplamı olarak transfer fonksiyonuna yansıtılır. Yapay bir nöronun öğrenme kapasitesi, seçilen öğrenme algoritmasının ağırlıkları düzenlenerek belirlenir [15]. YSA yönteminde başlangıçta bir eğitim seti oluşturulur ve ağa hem girdi hem de çıktı verilir. Ağ tarafından üretilen çıktılar daha sonra gerçek çıktılarla karşılaştırılır. Hata hesabı yapıldıktan sonra ağırlıklar güncellenir ve bu süreç en düşük hata oranına ulaşıncaya kadar tekrarlanır ve eğitim süreci tamamlanır. Bir sonraki adımda eğitim sürecinde oluşturulan model tercihen farklı verilerden oluşan bir test seti ile tekrar çalıştırılır ve ağın öğrenmesi test edilir. Bu tez çalışması için çok katmanlı ileri beslemeli bir sinir ağı modeli kullanılmıştır. Önerilen sinir ağının temel yapısı Şekil 4.6.'da sunulmuştur.

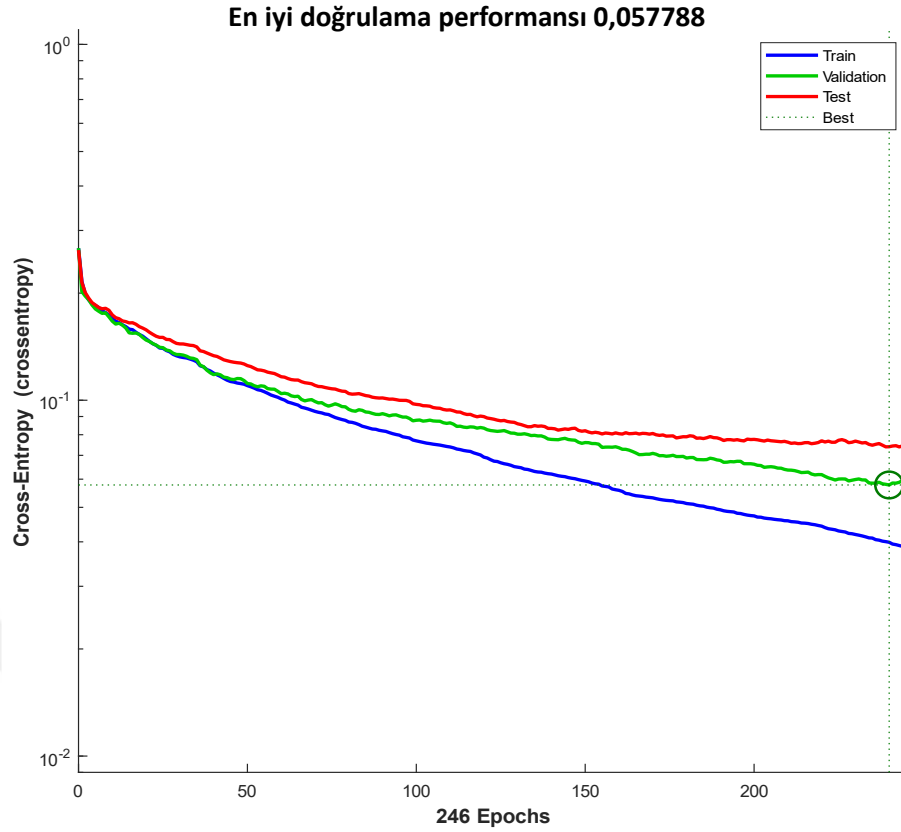


Şekil 4.6. Önerilen yapay sinir ağının yapısı.

Önceki bölümde anlatıldığı gibi veri setinde 1663 örnek ve 25 öznitelik vardır. Bu nedenle YSA'nın giriş katmanı da 25 nörondan oluşmaktadır. Çıktı olarak, her örnek 12 farklı sınıftan birine aittir. Dahası, sinir ağı modelinde her katmanı 50 nörondan oluşan 10 gizli katman vardır. Sigmoid transfer fonksiyonu aktivasyon fonksiyonu olarak seçilmiştir. Denklem (4.5) modellenen sinir ağının sigmoid transfer fonksiyonunu göstermektedir. Burada x girişleri ve $f(x)$ çıktıyı göstermektedir.

$$f(x) = \frac{1}{1+e^x} \quad (4.5)$$

Bu çalışma için tasarlanan sinir ağı modeli, gizli katmanlarda sigmoid fonksiyonları ve çıktı katmanında softmax fonksiyonları kullanılan çok katmanlı bir ileri beslemeli sinir ağıdır. Eğitim süreci, ölçekli bir eşlenik gradyan geri yayılım algoritması kullanılarak gerçekleştirilmiştir. Veri setinin yaklaşık %60'ı (997 örnek) eğitim işlemi için, %20'si (333 örnek) test işlemi için ve %20'si (333 örnek) doğrulama işlemi için kullanılmıştır. Doğrulama işlemi ağıın ezberleme yapmasını engellemek için kullanılmaktadır. Şekil 4.7 yapay sinir ağı modelinin eğitim sürecinde elde edilen eğitim durumu performansı grafiğini göstermektedir.



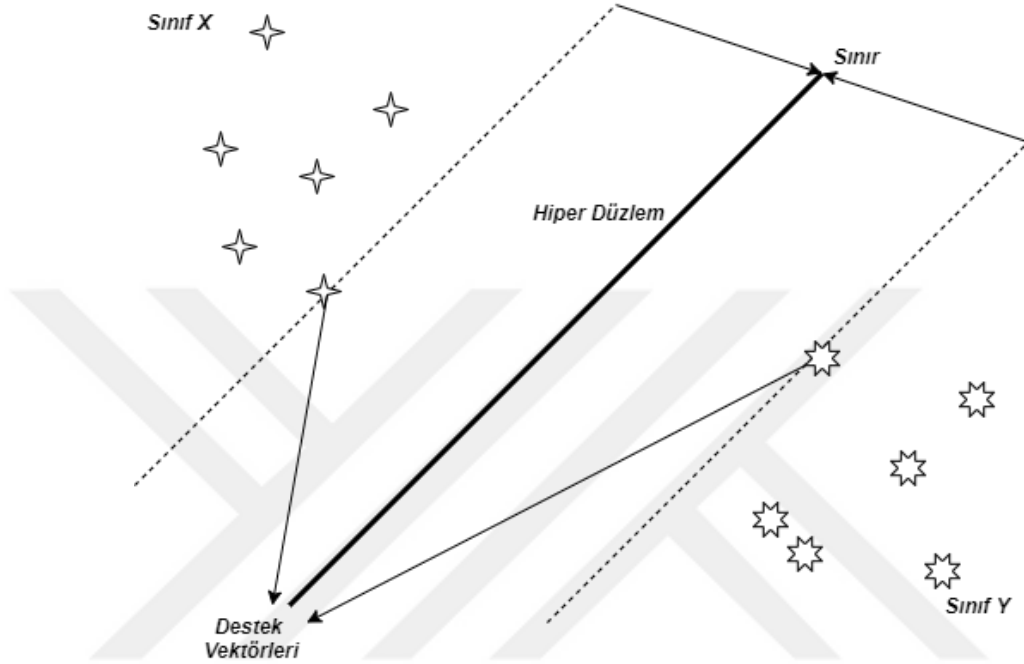
Şekil 4.7. Yapay sinir ağının eğitim durumu performans grafiği.

Kullanılan eğitim, doğrulama ve test veri setleri random olarak seçilmiştir. Ağın eğitimi için 246 iterasyon yapılmıştır. Eğitim işleminin performans ölçümü çapraz-entropi ile değerlendirilmiştir. Doğrulama kontrolü 6 kez yapılmıştır. En başarılı doğrulama performans değeri 240.iterasyonda elde edilmiştir.

4.2.2. Destek vektör makineleri

Destek vektör makineleri (DVM), sınıflandırma veya regresyon problemlerine uygulanabilecek denetimli öğrenme yöntemleri arasındadır. Sınıflandırma, veri kümesinin giriş uzayını doğrusal veya doğrusal olmayan şekilde bölerek gerçekleştirilir. Doğrusal karar çizgisi, izlenen örnekler arasında minimum bir mesafeye, ancak maksimum çizgi aralığına sahip olacak şekilde çizilir. Gerçek dünya

uygulamalarında iyi sonuçlar veren bir yöntemdir [15]. DVM yönteminin yapısı Şekil 4.8.'de görülmektedir.

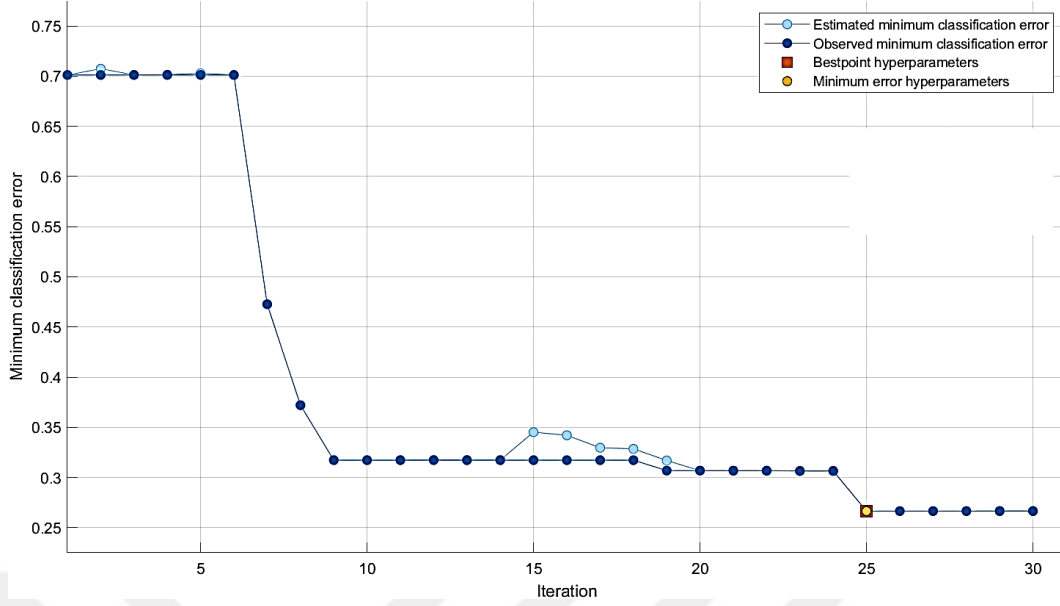


Şekil 4.8. Destek vektör makineleri yönteminin yapısı.

Hiper düzlem (H) hesaplanmasının formülü Denklem (4.6) 'da verilmiştir, burada w bir ağırlık setini, x giriş örneği özelliklerini ve b bias değerini belirtir.

$$H: w * x_i + b = 0 \quad (4.6)$$

DVM yönteminde, kernel fonksiyonu sınıflandırıcının başarısı için önemli parametrelerden biridir. Bu çalışma için üç farklı kernel fonksiyonu kullanılmıştır: doğrusal, kübik ve ikinci dereceden. Kernel ölçeği MATLAB tarafından otomatik olarak seçilmiştir. Veri düzenleme (regülizasyon) ve standartlaştırılmış parametreler de gerçek özellikleri temsil edecek şekilde ayarlanmıştır [6]. Şekil 4.9. DVM yönteminin eğitim sürecinde elde edilen minimum sınıflandırma hatası grafiğini göstermektedir.

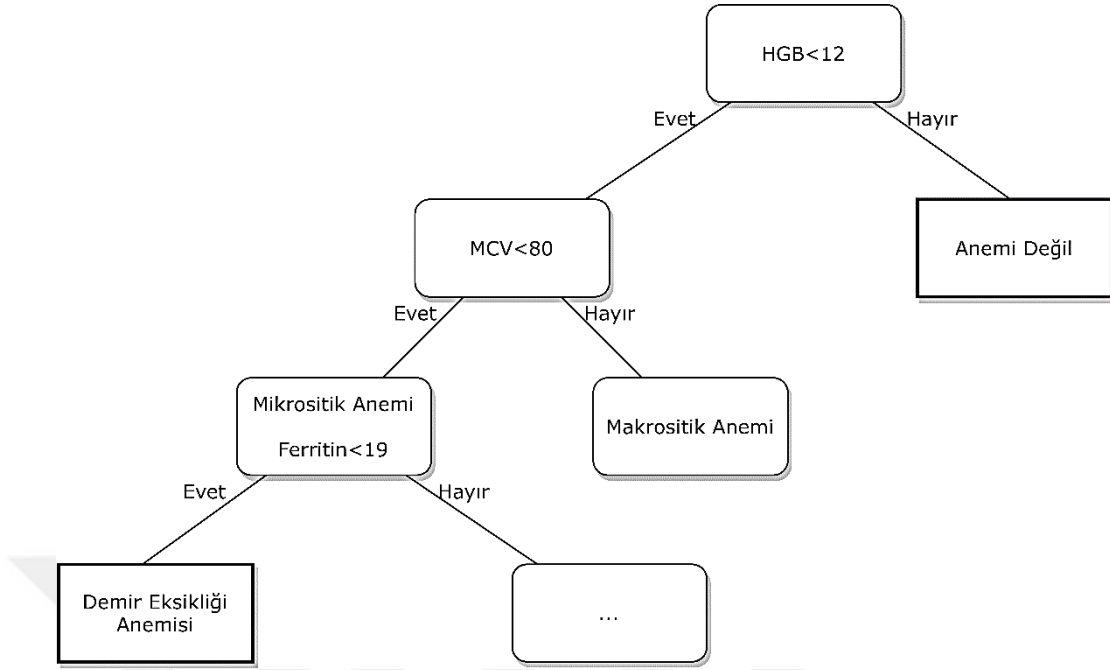


Şekil 4.9. DVM eğitim süreci minimum sınıflandırma hatası.

Minimum sınıflandırma hatası grafiğinde görüldüğü gibi eğitim aşamasında 25. iterasyonda hata sabitlenmiştir.

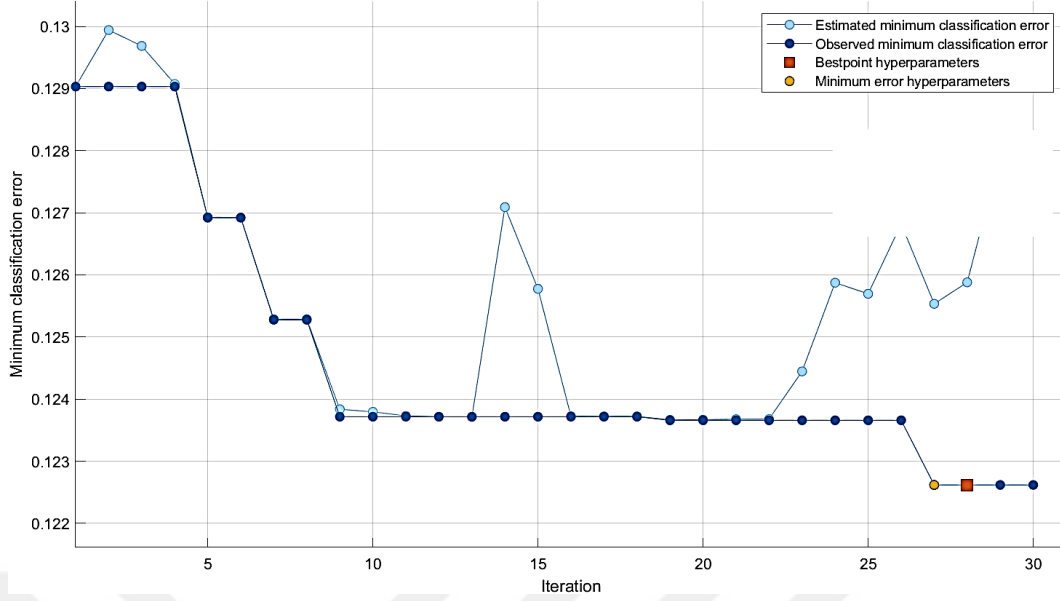
4.2.3. Karar ağaçları

Karar ağaçlarının sınıflandırma süreci, bir özelliğin bir veri kümesindeki sınıflar ile ayırt edilemeyeceğini test etme yöntemidir. Bulunan her özellik, ağacın dallanma durumunu oluşturur. Bu yöntem ile veri setindeki tüm verilerin sınıflardan birine yerleştirilmesi amaçlanır ve bu şekilde aynı anda bir sınıf tanımı yapılır. Sonuçların anlaşılması ve yorumlanması kolaydır [15]. Bu yöntem aynı zamanda iyi sonuçlar verir ve genellikle tıbbi verilerle birlikte kullanılır. Karar ağacının yapısının bir örneği Şekil 4.10.'da görülmektedir.



Şekil 4.10. Örnek bir karar ağacı yapısı.

Bu çalışma için kullanılan karar ağacı yöntemi iki farklı topluluk yöntemi kullanılarak gerçekleştirilmiştir: güçlendirilmiş (boosted) ve torbalanmış (bagged). Çok sınıflı problemlerin çözümünde topluluk teknikleri kullanılmakta ve başarıyı arttırmaktadır. Amaç, ikili sınıfları çoklu sınıflar oluşturmak üzere gruplayarak performansı artırmaktır. Kullandığımız AdaBoost yönteminde, her bir iterasyonda, karar ağacının yanlış sınıflandırılmış örneklerinin ağırlıkları azaltılır ve doğru sınıflandırılmış örneklerin ağırlıkları artırılır. Sonraki iterasyonlarda, güncellenmiş ağırlıklar kullanılır. Bu nedenle, algoritma doğru sınıflandırılmış ögelere odaklanır. Torbalama yönteminde, veriler alt gruplara bölünür ve her bir alt grup için bir öğrenme modeli uygulanır. Torbalanmış ağaçlar, çok sayıda karar ağacının birleştirilmesiyle oluşturulur ve sınıfların kategorik ve doğrusal olmadığı durumlarda başarılı sonuçlar verdiği bilinmektedir. Bu çalışmada, veri setinin her bir alt kümesi için farklı bir karar ağacı oluşturulmuştur. Doğruluk, her bir ağaç yapısının ortalama doğruluk performansı dikkate alınarak hesaplanmıştır [137]. Şekil 4.11. karar ağaçları yönteminin eğitim sürecinde elde edilen minimum sınıflandırma hatası grafiğini göstermektedir.



Şekil 4.11. Karar ağaçları eğitim süreci minimum sınıflandırma hatası.

Minimum sınıflandırma hatası grafiğinde görüldüğü gibi karar ağaçlarının eğitim aşamasında gözlenen hata oranı tahmin edilen hata oranının altındadır. Bu sınıflandırma işleminin başarılı olduğu anlamına gelmektedir.

4.2.4. Naive Bayes

Naive Bayes, Bayes'in olasılık teoremine dayalı bir makine öğrenmesi algoritmasıdır. Naive Bayes yönteminde, her bir veri örneğinin ait olduğu sınıf net olduğunda, amaç gelecek olan bir sonraki veri örneğinin sınıf etiketini belirleyecek bir kural oluşturmaktır [138]. Bu durum koşullu olasılık olarak da adlandırılmaktadır. Veri sınıfı etiketlendiğinde ilgili öznelik tarafından alınan değer ilkesine dayanmaktadır. Bu yöntem bir veri setine uygulandığında, yapılan işlemler Denklem (4.7) ve Denklem (4.8) 'de ifade edilmektedir.

$$P(PC | S) = \frac{P(S | PC) * P(C)}{P(S)} \quad (4.7)$$

$$P(PC|S) = P(S|PC) * P(PC) \quad (4.8)$$

Burada;

PC: Ana Kategori (Parent Category),

S: Başarılı (Successful),

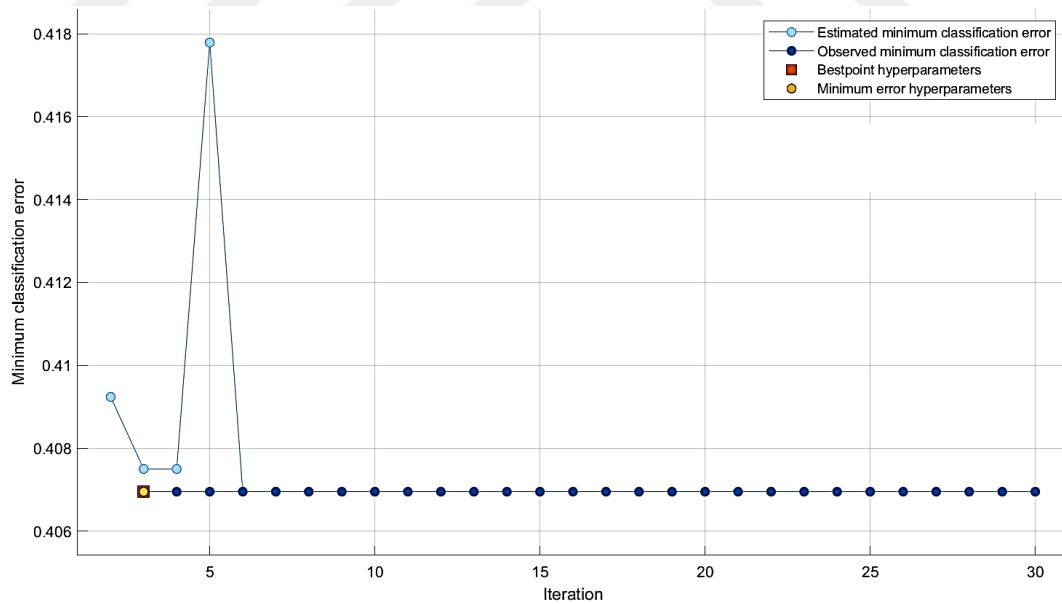
P (PC): Ana kategorinin başarılı olma olasılığı,

P (S): Sınıf etiketinin başarılı olma olasılığı,

P (PC | S): Sınıf etiketi başarılı olduğu durumda üst kategorinin olma olasılığı,

P (S | PC): Üst kategori durumunda sınıf etiketinin başarılı olma olasılığı'dır.

Şekil 4.11. karar ağaçları yönteminin eğitim sürecinde elde edilen minimum sınıflandırma hatası grafiğini göstermektedir.



Şekil 4.12. Naive Bayes yöntemi eğitim süreci minimum sınıflandırma hatası.

Minimum sınıflandırma hatası grafiğinde görüldüğü gibi Naive Bayes yönteminin eğitim aşamasında gözlenen hata oranı 6. İterasyondan itibaren sabitlenmiştir.

4.3. Değerlendirme

Bu çalışmada tüm yöntemler için değerlendirme işlemi olarak k-kat çapraz doğrulama yöntemi kullanılmıştır. K-kat çapraz doğrulama yönteminde, veri kümesi k adet farklı alt gruba bölünmektedir. Bir alt grup test seti olarak seçildiğinde, geriye kalan k-1 adet alt grup eğitim seti olarak kullanılır. Bu işlem bütün alt gruplar bir kere test seti seçilerek devam eder. Bu şekilde k adet veri setinin her biri bir kez test seti olacak şekilde tüm kombinasyonlar test edilir ve her sonucun ortalaması alınarak bir performans değeri bulunur. Bu çalışma için k değeri 10 olarak seçilmiştir.

Bu çalışmada performans ölçümü için Alıcı İşletim Karakteristiği (Receiver operating characteristics- ROC) analizi kullanılmıştır. ROC analizi, makine öğrenimi ve veri madenciliği tekniklerinin performansını ölçmek için etkili bir yöntemdir [130, 139]. ROC analizi için karmaşıklık matrisi Tablo 4.5.'te gösterilmektedir.

Tablo 4.5. ROC analizi için karmaşıklık matrisi.

Tahmin edilen		Gerçek	
		Pozitif	Negatif
		TP (Doğru Pozitif)	FP (Yanlış Pozitif)
Pozitif			
Negatif		FN (Yanlış Negatif)	TN (Doğru Negatif)

Doğruluk, geri çağırma (duyarlılık), özgüllük, kesinlik (güven), F1-skoru ve AUC (eğri altındaki alan) değerleri temelde Denklemler (4.9) - (4.14) 'te gösterildiği gibi hesaplanır.

$$\text{Doğruluk (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.9)$$

$$\text{Geri çağırma (Recall - R)} = \frac{TP}{TP+FN} \quad (4.10)$$

$$\text{Özgüllük (Specificity)} = \frac{TN}{TN+FP} \quad (4.11)$$

$$Kesinlik(Precision - P) = \frac{TP}{TP+FP} \quad (4.12)$$

$$F1 - skoru = \frac{2*P*R}{P+R} \quad (4.13)$$

$$Eğri altı alan (AUC) = \frac{TPR-TNR}{2} \quad (4.14)$$

Bu denklemlerde yer alan ifadelerin açıklamaları şu şekildedir;

TP (Gerçek Pozitif): Öngörülen değerin ve gerçek değerin pozitif olduğu örnek sayısıdır.

TN (Gerçek Negatif): Öngörülen değerin ve gerçek değerin negatif olduğu örnek sayısıdır.

FP (Yanlış Pozitif): Öngörülen değerin pozitif ve gerçek değerin negatif olduğu örnek sayısıdır.

FN (Yanlış Negatif): Öngörülen değerin negatif ve gerçek değerin pozitif olduğu örnek sayısıdır.

P: Kesinlik / Güven

R: Geri Çağırma / Hassasiyet

TPR (Gerçek Pozitif Oranı): Hassasiyet.

TNR (Gerçek Negatif Oranı): Özgüllük.

ROC eğrisi, doğruluk ve duyarlılık arasındaki dengeyi değerlendirmek için kullanılır. Eğri altındaki alan (AUC) olarak bilinen ROC eğrisinin altında kalan alan, ROC puanı olarak tanımlanır. ROC eğrisi, yanlış pozitiflerin bir fonksiyonu olarak gerçek pozitiflerin değişen sınıflandırma eşik değerlerine bağlı olarak çizilir. ROC puanı "1" olursa, pozitiflerin negatiflerden mükemmel bir şekilde ayrıldığını gösterir. ROC puanı "0" olduğunda ise, bu, hiçbir pozitifin bulunmadığı anlamına gelir [139]. ROC

analizi, ikili sınıflı problemler için yaygın olarak kullanılır, ancak aynı zamanda çok sınıflı problemler için de uygundur. Çok sınıflı problemler için iki sınıflı bir yaklaşım kullanılır. Bu yaklaşımlardan biri "bire karşı bir" yaklaşımı ve diğeri "bire karşı hepsi" yaklaşımıdır. "Bire karşı bir" yaklaşımında, her sınıf, diğeri sınıfların her biri ile çiftler halinde karşılaştırılır. "Bire karşı hepsi" yaklaşımında ise, "t" sınıfı için, diğeri tüm sınıflar "t değil" olarak işaretlenir ve "t" ile karşılaştırılır [140]. Bu çalışmada, "bire karşı bir" yaklaşımı kullanılarak çok sınıflı bir ROC analizi uygulanmıştır. Kesinlik, tahmin edilen pozitif sınıfın gerçekte pozitif olduğu orandır. Geri çağırma, doğru tahmin edilen gerçek pozitiflerin oranıdır. Bu iki ölçüm aynı zamanda önemli performans ölçümleri olduğundan, F1-skoru da bu iki ölçüm kullanılarak hesaplanır [141]. Sonraki bölümde, bu çalışmada kullanılan tüm veri setleri detaylıca açıklanmış ve tüm yöntemler için ortaya çıkan karmaşıklık matrisleri ve ROC eğrileri gösterilmiştir.

BÖLÜM 5. ARAŞTIRMA BULGULARI VE TARTIŞMA

Bu çalışmada, 12 farklı anemi türünü, deneyimli bir uzman doktorun yapacağı gibi teşhis etmek için dört farklı makine öğrenme yöntemi tamamen orijinal bir veri seti ile test edilmiştir. Bu anemi türleri, verilerin toplandığı Düzce ilinde en sık görülen anemi türleridir. Bu çalışma için orijinal veri setinin yanında öznitelik seçme yöntemleri yardımıyla oluşturulan 8 farklı veri seti daha kullanılmıştır. Öznitelik seçme yöntemleri olarak bilgi kazancı, bilgi kazanç oranı, korelasyon ve temel bileşen analizi kullanılmıştır. Sınıflandırıcı olarak ise YSA modeli, doğrusal, karesel ve kübik DVM modelleri, Naive Bayes modeli ve artırılmış ve torbalanmış karar ağacı modelleri kullanılmıştır. Modellerin başarılarını değerlendirmek için ROC analizinin yanısıra AUC, doğruluk, geri çağırma, kesinlik, F1-skoru ve sınıflandırma hatası ölçütleri kullanılmıştır.

Bu bölümde kullanılan veri setinin nasıl elde edildiği, hangi veri ve özniteliklerden oluştuğu gibi bilgiler detaylıca açıklanmıştır. Ayrıca kullanılan yapay öğrenme yöntemleri sonucunda elde edilen bulgular ve sonuçlar paylaşılmış ve tartışılmıştır.

5.1. Veri seti

Bu çalışmada kullanılan veriler Düzce Üniversitesi Tıp Fakültesi İnvaziv Olmayan Klinik Araştırmalar Etik Kurulu kararı ile alınmıştır. Etik kurul izni alabilmek için Kişisel Verilerin Korunması Kanunu kapsamında tüm hastaların tüm verileri alınamamakta, hangi hastalara ait hangi verileri istediğimiz bilgisi önem arz etmektedir. Bu nedenle Hematoloji Uzmanı Doç. Dr. Birgül ÖNEÇ ile yapılan görüşmeler neticesinde hastalık olarak anemi çeşitleri seçilmiş, veriler olarak ise anemi tanısında önemli olan 30 öznitelik belirlenmiştir. Anemi çeşitleri seçimi için

Hematoloji Polikliniğine başvuran hastalardan tanı olarak, uluslararası hastalık sınıflandırma (ICD-10) kodları D50 ile D64.9 arası girilmiş olan hasta verileri kullanılmıştır.

Hastalıkların sınıflandırılması için belirlenen uluslararası hastalık kodları DSÖ kılavuzunda D50-D89 arasında yer alan ICD-10 kodları kan ve bağışıklık sistemi hastalıklarını temsil etmektedir [142]. Bizim çalışmamız etik kurul tarafından anemi ve anemi ilişkili hastalıklar olarak sınırlandırıldığından D50 ve D64.9 arasındaki kodlar bizi ilgilendirmektedir. Bu kodlara baktığımızda; D50-D53 nutrisyonel anemiler, D55-D59 hemolitik anemiler ve D60-D64 aplastik ve diğer anemiler şeklinde özetlenebilir. Detaylandırmak gerekirse; D50 Demir eksikliği anemisi, D51 Vitamin B12 eksikliği anemisi, D52 folat eksikliği anemisi ve D53 diğer nutrisyonel anemilerdir. D55 enzim bozukluklarına bağlı anemiler, D56 talasemiler, D57 orak hücre bozuklukları, D58 diğer kalıtsal hemolitik anemiler ve D59 edinilmiş hemolitik anemilerdir. D60 edinilmiş saf kırmızı hücre aplazisi, D61 diğer aplastik anemiler, D62 akut posthemorajik anemiler, D63 Kronik hastalık anemileri ve D64 diğer anemiler şeklinde bir sınıflandırma yapılmıştır [143].

Bizim veri setimizde yer alan veriler; DSÖ hastalık sınıflandırma kılavuzunda demir eksikliği anemisi, vitamin B12 eksikliği anemisi ve folat eksikliği anemisi ile birlikte nutrisyonel anemiler grubuna (ICD kodları D50-53) dâhil edilmiştir. Talasemi, talasemi taşıyıcılığı ve enzim bozukluğuna bağlı kalıtsal ve edinsel hemolitik anemiler hemolitik anemiler grubundadır (ICD kodları D55-59). Aplastik ve diğer anemiler grubu (ICD kodları D60-64) aplastik anemi, kronik hastalık anemisi ve diğer anemileri içerir [142]. Ancak eritrosit morfolojisine göre sınıflandırıldığında, mikrositik anemiler içerisinde demir eksikliği anemisi, makrositik anemilerde ise vitamin B12 ve folat eksikliği anemileri yer almaktadır. Ayrıca klinikte bu hastalıkların bir arada görüldüğü durumlar yaygındır. Şekil 3.9.'da verilen anemilerin eritrosit morfolojisine göre sınıflandırılmasına göre mikrositik anemiler içerisinde demir eksikliği anemisi, talasemi ve talasemi taşıyıcılığı yer almaktadır. Bazı kronik hastalık anemileri mikrositik olsa da çoğu normositik anemiler arasında yer alır. Hemolitik anemiler de

normositik anemilere dâhildir. Vitamin B12 eksikliği ve folat eksikliği anemileri ise makrositik anemiler grubunda yer alır.

Demir eksikliği anemisi en yaygın görülen anemi türüdür. Kırmızı kan hücrelerinin akciğerlerden dokulara oksijen taşımalarını sağlayan, hemoglobinin merkezinde bulunan, demir içeren protein yapıdır. Vücut demiri kaybolduğunda bu işlev gerçekleşemez ve sonuç olarak halsizlik, yorgunluk, nefes darlığı gibi çeşitli belirtiler görülür. En çok kadınlarda ve küçük çocuklarda görülür. Tanısında MCV, demir, ferritin ve demir bağlama kapasite değerlerinin ölçülmesi önemlidir. Ayrıca iç kanama olasılığı da ortadan kaldırılmalıdır. Demir eksikliği anemisi, demir takviyeleri ve demir açısından zengin besinler içeren bir diyetle tedavi edilebilir [144, 145].

B12 vitamini, kırmızı kan hücresi üretiminde ve sinir sisteminin işleyişinde önemli bir rol oynar. Vücuttaki B12 vitamini yetersiz kaldığında sağlıklı üretim ve kırmızı kan hücrelerinin bölünmesi yapılamaz. Sonuç olarak, kırmızı kan hücrelerinin kemik iliğinden kana geçmesi ile ilgili sorunlar ortaya çıkar ve çeşitli bedensel ve zihinsel belirtilere neden olur. HGB, MCV ve vitamin B12 değerleri tanısında önemli ölçümlerdir. B12 vitamini eksikliği, yeterli beslenme ve B12 vitamini takviyesi ile tedavi edilebilir [144, 146].

Folik asit, meyvelerde, yeşil yapraklı sebzelerde ve ette bulunan bir maddedir. Yetersizlik, alımının yetersiz olması veya vücut tarafından yeterince emilmemesi durumunda ortaya çıkar. Hücre bölünmesinde önemli rolü olan folat vücutta eksik olduğunda çeşitli fonksiyon bozukluklarına neden olabilir. Özellikle hamilelikte ve çocukların büyüme dönemlerinde takibi önemlidir. Serum folat seviyesi tanısında önemli bir ölçümdür. Folik asit eksikliği, folik asit açısından zengin bir diyet ve takviyelerle tedavi edilebilir [146, 147].

Kronik hastalık anemisi, özellikle yaşlı bireylerde kanser türleri, diyabet, kalp, böbrek ve romatizmal hastalıklar, enfeksiyonlar ve iltihaplanma gibi kronik hastalıklara eşlik eden bir anemi türüdür. Düşük serum demir seviyeleri ve toplam demir bağlama

kapasitesi tanıda önemli ölçümlerdir. Bu tür aneminin tedavisi için öncelikle altta yatan hastalık tedavi edilmelidir [148].

Talasemi, hemoglobin zinciri çok az sayıda üretildiğinde veya hiç hemoglobin zinciri üretilmediğinde ortaya çıkan bir hastalıktır. Genetik geçişli bir hastalıktır. Bu nedenle heterozigotlar taşıyıcı olur ve homozigotlar hastalanır. HGB, HCT, eritrosit sayısı ve MCV, MCH ve MCHC indeks değerleri tanıda önemli ölçümlerdir. Transfüzyon, genellikle taşıyıcılar söz konusu olduğunda olmasa da hastaların tedavisinde uygulanır. Talasemi hastaları tüm yaşamları boyunca izlenmelidir [148].

Hemolitik anemi, kırmızı kan hücrelerinin üretildiklerinden daha hızlı bir şekilde yok edildikleri bir durum olarak tanımlanabilir. Nedeni kalıtsal veya edinilmiş olabilir. Tanıda hastanın öyküsü önemli olmakla birlikte tam kan sayımı ve periferik yayma, hemoglobin elektroforezi, kemik iliği testleri gibi laboratuvar yöntemleri de kullanılmaktadır. Tedavisinde ilaç tedavisi, cerrahi müdahale, kan transfüzyonu, kemik iliği ve kök hücre nakli uygulanabilir [144, 148].

Çalışmamızda anemi tanısı alan hastalara yukarıda anlatılanlar dışında anemi tanısı da konmuştur. DSÖ tanımına, ICD 10 kodu D64, “diğer anemiler” olarak dâhil edilmiştir. Tam kan sayımı ve özellikle HGB, HCT ve RBC parametreleri anemi tanısında kullanılır. Yaşam kalitesini önemli ölçüde etkileyen bir hastalık olduğu için anemi tipini tanımak ve tedaviyi buna göre organize etmek çok önemlidir [127].

Bu çalışma için uygun bir veri setinin oluşturulması amacıyla etik kurul izni ile Düzce Üniversitesi Sağlık Uygulama ve Araştırma Merkezi Bilgi İşlem Birimi’nden alınan verilerde yer alan tüm bilgiler şunlardır:

- Arşiv No: Her hasta için bir kez kullanılmaktadır. Hastanın her gelişinde sabittir. Maksimum 10 karakterlik rakam içerir.

- Kabul No: Hastanın her girişı için ayrı bir kabul numarası vardır. Maksimum 7 karakterlik rakam içerir.
- Kabul Tarihi: Hastanın her gelişı için geldiđi tarih yer alır. DateTime deđerler içerir.
- Birim Adı: Hastanın hangi birimden geldiđi bilgisi yer alır. String deđerler içerir.
- Başvuru No: Hastanın yaptıđı başvuru numarası yer alır. Harf ve rakam deđerleri içerir.
- İşlem Kodu: Hastaya yapılan işlemin kod deđerı yer alır. Maksimum 12 karakterlik rakam içerir.
- İşlem Adı: Hastaya yapılan işlemin adı yer alır. String ve rakamsal deđerler içerir.
- İşlem Tarihi: Hastaya yapılan her işlemin yapıldıđı tarih yer alır.
- Tanı Kodu: Her hastalık için farklı bir kod vardır. Bir hastanın birden fazla hastalıđı olabilir. Harf ve rakam deđerleri içerir. Anemi ve ilişkili tanılar için D50 ve D64.9 arası kodlar kullanılmaktadır.
- Tanı Adı: ICD kodlarına göre tanımlanmış hastalık isimleri yer alır. String deđerler içerir.
- Tahlil Adı: Hastaya uygulanan tahlillerin adı yer alır. String deđerler içerir.
- Sonuç: Uygulanan tahlil ve testlerin sonuçları yer alır. Rakam ve string deđerler içerir.
- Yaş: Hastanın yaşı yer alır. Rakamsal deđerler içerir.
- Cinsiyet: Hastanın cinsiyeti yer alır. Kadın ve Erkek olmak üzere 2 deđişken mevcuttur.
- Onay Tarihi: Hastaya yapılan tahlillerin tarihleri yer alır. DateTime deđerler içerir.

Bu bilgiler “başvuru”, “işlem”, “tanı” ve “tahlil” olmak üzere dört farklı tabloda yer almaktadır. Alınan veriler ve tutulan hasta dosyaları retrospektif olarak incelenmiş ve “epikriz” adında yeni bir veri tabanı tablosu daha oluşturulmuştur. Hematoloji Polikliniği’nde tutulan hasta dosyaları incelenerek oluşturduğumuz 5. Tablo olan “epikriz” tablosunda yer alan bilgiler ise şu şekildedir:

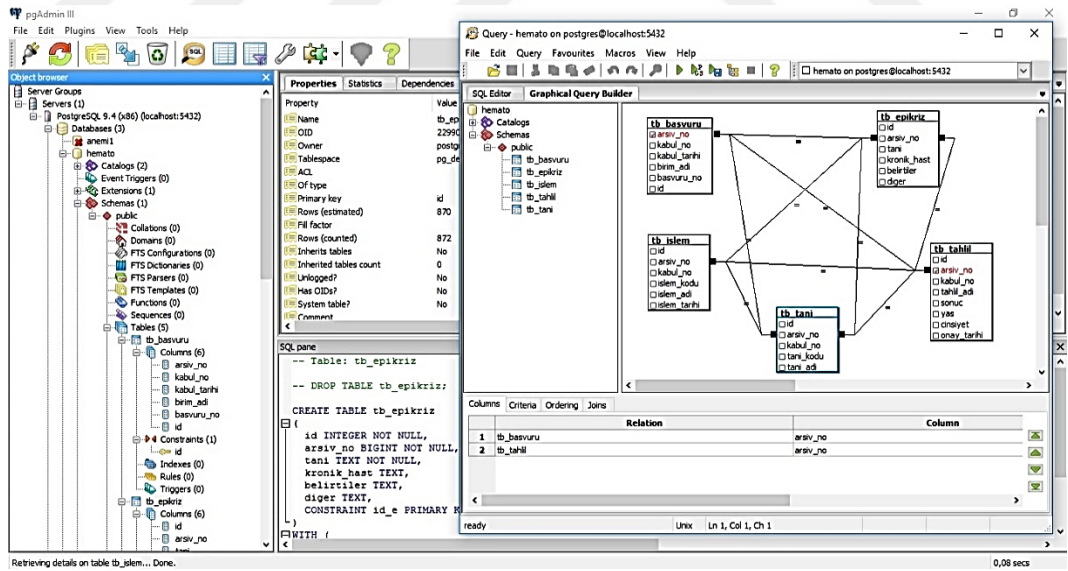
- Arşiv No: Bahsi geçen tüm tablolarda yer almaktadır.
- Tanı: Hastanın dosyasında yer alan uzmanın verdiği tanı kararıdır. String değerler içerir.
- Kronik Hastalıklar: Hastanın dosyasında yazılan kronik hastalıkları yer alır. String değerler içerir.
- Belirtiler: Hastanın dosyasında yazılan belirtiler yer alır. String değerler içerir.
- Diğer: Hastanın dosyasında yazılan diğer bilgiler (öyküsü, daha önce aldığı tedaviler vb) yer alır. String değerler içerir.

Alınan verilerin tutulduğu beş farklı tablo için veritabanı işlemleri yapılması gerekmiştir. Bu amaçla PostgreSQL veri tabanı yönetim sisteminin pgAdmin arayüz platformu kullanılarak veri tabloları oluşturulmuş, tablolar arasında ilişkiler kurulmuş ve çeşitli sorgular oluşturulmuştur. 27-29 Mayıs 2016 tarihleri arasında Sakarya Üniversitesi Sosyal Bilimler Enstitüsü ve Finansal Ekonometri Bölümünün Koordinatörlüğünde düzenlenen PostgreSQL eğitimine katılarak bu konudaki eksiklerimi tamamlamaya çalıştım. Şekil 5.1., Şekil 5.2. ve Şekil 5.3.’te gerçekleştirilen çeşitli veri tabanı işlemleri görülmektedir. Yeni verilerle oluşturulan “epikriz” tablosuna hastaneden temin edilen dosyalardaki verilerin girişi işlemi her bir dosya için detaylı inceleme gerektirdiğinden uzun sürmüştür.

id	arsv_no	tani	kronik_hast	belirtiler	diger_text
1	248740	Anemi yok demir eksikliği yok		septom yok	
2	289698	LAP için gelmiş		septom yok	
3	2011174795	kronik hast anemisi+dea	siroz	var	
4	2011174478	Polisitemia vera			
5	2011174508	tedavi alan hasta demir eksikliği			
6	149206	demir+folik asit eksikliği/transfüzyon listesinde yok	diyabet, böbrek, akciğer, k var	var	
7	2011140072	kronik hast. Anemisi			
8	324485	hasta kaydı bulunamadı			
9	192181	kronik hast. Anemisi	kronik böbrek hast.+malig		
10	2011175193	anemik değil		var	
11	118792	dağ merk.gelmış _kronik hast. Anemi var		var_Bağ dönmese, Bayılma	
12	140634	kronik hast. Anemisi	böbrek yetmezliği	var_ Halsizlik , sırt bacak ağr	
13	332425	demir eksikliği var anemisi yok		ağızda aft	
14	2011144753	kronik hepatit, talasemi			
15	189796	demir eks.anemisi+demir eksikliği		mens.düzensizliği, sürekli dem	
16	2011175394	koagülasyon bozukluğu		diş eti ve burun kanaması, gai iğinde tiner	
17	92279	trombositopeni		diyabet, kilo kaybı	
18	2011175794	Polisitemia vera		yok	dağ merk.den
19	142342	kalp kapak repl ilişkili kronik hem kalp kapak replasmanı yapıldı ilacı alıyor			
20	184627	demir+B12 eksikliği anemisi (lab.sonucu bulunmadı)		adet düzensizliği,demir eks..b	kontrolde çağır
21	248764	lökosit bozuklukları ??		grip+öksürük+wbv düşüklüğü	
22	163893	koagülasyon bozukluğu, pıhtılaşma t tahlilde FVII-S1 çıkmış		bacakta morluk	
23	136543	anemi	h.pylori+ ülser	demir eks.	3ay tedavi gö

Şekil 5.1. Veri tablolarının oluşturulması ve içeriğinin düzenlenmesi.

Şekil 5.1. veri tablolarının oluşturulmasından sonra veri girişi yapılan ekranı göstermektedir.



Şekil 5.2. Tabloların düzenlenmesi ve ilişkilerin kurulması.

Şekil 5.2. veri tablolarının düzenlenmesi ve ilişkilerin kurulduğu ekranı göstermektedir. Şekil 5.3. sorgu ekranını ve sorgu sonucu elde edilen veri tablosunu göstermektedir.

Query - hemato on postgres@localhost:5432 - [C:\Users\TuBa\Desktop\]

File Edit Query Favourites Macros View Help

SQL Editor Graphical Query Builder

```

SELECT DISTINCT
  tb_tani.arsiv_no,
  tb_tani.tani_kodu,
  tb_tani.tani_adi,
  tb_epikriz.tani,
  tb_epikriz.kronik_hast,
  tb_epikriz.belirtiller,
  tb_tahsil.diger,
  tb_tahsil.tahsil_adi,
  tb_tahsil.sonuc,
  tb_tahsil.yes,
  tb_tahsil.cinsiyet,
  tb_tahsil.onay_tarihi
FROM
  public.tb_tani,
  public.tb_tahsil,
  public.tb_epikriz
WHERE
  tb_epikriz.arsiv_no = tb_tahsil.arsiv_no AND
  tb_epikriz.arsiv_no = tb_tani.arsiv_no AND
  tb_tani.arsiv_no = tb_tahsil.arsiv_no AND
  tb_tani.tani_adi ILIKE 'Anemi';

```

Data Output

arsiv_no	tani_kodu	tani_adi	tani	kronik_hast	belirtiler	diger	tahsil_adi	sonuc	yes	cinsiyet	onay_f
integer	character	text	text	text	text	text	text	text	text	text	date
36009	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	BODYUM	142	82	KADIN	2014-
36010	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.BILIR	0.08	82	KADIN	2014-
36011	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.BILIR	0.15	82	KADIN	2014-
36012	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.BILIR	0.22	82	KADIN	2014-
36013	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.BILIR	0.23	82	KADIN	2014-
36014	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.BILIR	0.24	82	KADIN	2014-
36015	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.PROTE	6.34	82	KADIN	2014-
36016	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.PROTE	6.97	82	KADIN	2014-
36017	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.PROTE	7.51	82	KADIN	2014-
36018	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.PROTE	7.66	82	KADIN	2014-
36019	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	T.PROTE	8.01	82	KADIN	2014-
36020	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	TIBIC	274	82	KADIN	2014-
36021	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	TIBIC	314	82	KADIN	2014-
36022	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	TIBIC	321	82	KADIN	2014-
36023	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	TIBIC	330	82	KADIN	2014-
36024	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	TIBIC	351	82	KADIN	2014-
36025	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	TSH	1.02	82	KADIN	2014-
36026	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	TSH	1.17	82	KADIN	2014-
36027	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	UIBC	297	82	KADIN	2014-
36028	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	UIBC	243	82	KADIN	2014-
36029	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	UIBC	244	82	KADIN	2014-
36030	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	UIBC	262	82	KADIN	2014-
36031	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	UIBC	266	82	KADIN	2014-
36032	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	URE	51.8	82	KADIN	2014-
36033	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	URE	57	82	KADIN	2014-
36034	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	URE	68.4	82	KADIN	2014-
36035	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	URE	73.5	82	KADIN	2014-
36036	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	URE	97.6	82	KADIN	2014-
36037	244908 D63	Anemi, başka y	MDS	HT, KBY	bel ve bacak ağrıları	Dış merkezde Hb	URİK AS	3.5	82	KADIN	2014-

OK

Unix ln 1, Col 1, Ch 1

74958 rows 48013 ms

Şekil 5.3. Sorgu işlemleri ve sorgu sonucu elde edilen veri seti.

Programın yazılımında kullanılan programlama dili olarak C# dili kullanımına karar verilmiştir. C# programlama dilinin nesne yönelimli, taşınabilir, multi-threaded (birçok farklı işlemin aynı zamanda gerçekleştirilmesi olayı), dinamik ve güvenilir olması ayrıca Türkçe kaynaklarının mevcut olması gibi avantajları vardır. C# dili ile yazılım yapmak için gerekli geliştirme ortamı olarak Microsoft Visual Studio 2019 platformu kullanılmaktadır. Lisanslı sürümü Sakarya Üniversitesi Bilgi İşlem Daire Başkanlığı tarafından sağlanan Dreamspark platformundan temin edilmiştir. Elde edilen veriler ve Sayın Doç. Dr. Birgül ÖNEÇ ile yapılan görüşmeler sonucunda arayüz tasarımları Şekil 5.4., 5.5. ve 5.6.'da görüldüğü şekilde gerçekleştirilmiştir.

Şekil 5.4. Açılış formu.

Şekil 5.4. ve Şekil 5.5. tasarlanan form arayüzlerini göstermektedir.

Şekil 5.5. Veri giriş ve kayıt formu.

Yapılan ara yüz tasarımı hastanın kimlik bilgileri olmaksızın gerekli bilgilerin (arşiv no, cinsiyet, doğum tarihi gibi) girişini sağlamak üzere Hasta Bilgileri paneli oluşturulmuştur. Hastanın öyküsü ve anemiye yol açabilecek diğer hastalıkları varsa bunların belirlenmesi için bir “Özgeçmiş” paneli oluşturulmuştur. Fizik muayene bulguları ve anamnez sonucu belirlenecek çeşitli belirtilerin girişi için “Fizik

Muayene” ve “Semptomlar” olmak üzere iki panel eklenmiştir. Laboratuvar işlemlerini belirlemek üzere bir panel oluşturulmuştur. Formda gerçekleştirilebilecek çeşitli işlemler için butonlar eklenmiştir. Forma girilen verilerin veritabanına kaydedilmesi için “Bilgileri Kaydet”, ilk defa gelen bir hastanın kaydı için “Yeni Kayıt”, kontrol için gelen hastanın bilgilerini görmek için “Kontrol Formu”, ilgili formu kapatmak için “Kapat” ve uygulamadan çıkmak için “Uygulamayı Kapat” butonları oluşturulmuştur. Formda yer alan “Sistemi Eğit” butonu ise tasarım sürecinde çeşitli denemeleri yapmak için geçici olarak oluşturulmuştur.

Veri setinin oluşturulması amacıyla veritabanından verileri çekmek için hazırlanan arayüz tasarımı Şekil 5.6.’da görülmektedir.

The screenshot shows a window titled 'Form4' with a light gray background. It contains two data tables, a search bar, and several buttons.

Table 1 (Left):

id	arsiv_no	kabul_no	tani_kodu	tani_adi
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer
22	101254	837330	D64	Anemi, diğer

Table 2 (Right):

arsiv_no	tani_adi	tani
286976	Trombositopeni, ...	trombositopeni
140208	Anemi, diğer	Anemi _ Demir e...
2011199365	Anemi, diğer	Anemi
2011104560	Demir eksikliği a...	Anemi, DEA
2011170926	Kronik myeloprol...	Pansitopeni
2011199982	Anemi, diğer	daha önce gelmi...
2011175981	Anemi, tanımlan...	Anemi
391924	Bakteriyel pnöm...	lenfoma ?
117695	Trombositopeni, ...	b12 anemi

Form Elements:

- Search Bar:** "Tanı ismi giriniz:" with a dropdown arrow.
- Buttons:** "Verileri Getir", "Sonucu Excele Aktar", "EĞİTİM", "TEST", "Sonucu Excele Aktar", "KAPAT".
- Input Field:** A text box containing "anemi".

Şekil 5.6. Veri seti hazırlama formu.

Burada “Verileri Getir” butonu ile girilen tanı ismine göre veritabanında bir sorgu oluşturularak veriler beş farklı tablodan çekilmektedir. “Sonucu Aktar” butonu ile oluşturulan veri seti sınıflandırıcılar ile kullanılmaya hazır hale getirilmektedir.

Elimizdeki tablolar veritabanına aktararak çeşitli sorgular ile incelendiğinde tekrarlayan veriler olmaksızın anemi ve ilişkili tanıları olan kayıtlar elde edilmiştir. Bu

kayıtların bir kısmı için yapılan ikinci basamak test ve muayeneler sonucunda başka hastalıklara ait olduğu anlaşılmış ve geri kalan 1663 adet hasta verisi işleme alınmıştır. Tablo 5.1.'e bakıldığında bu verilerin yer aldığı veri setinde karşımıza çıkan 24 farklı anemi tipi görülmektedir. Basit olarak D.50-D.53 ICD koduna sahip tanılar demir, Vitamin B12 ve folat eksikliği anemileri; D.55-D.59 ICD koduna sahip tanılar kalıtsal, kazanılmış ve otoimmün hemolitik anemiler ve talasemiler; D60-D64 ICD koduna sahip tanılar diğer anemiler şeklinde gruplandırılabilir [9].

Tablo 5.1. Veri setinin ham halinde yer alan anemi ilişkili tanılar ve ICD kodları.

ICD Kodu	Tanı Adı
D50	Demir eksikliği anemisi
D50.8	Demir eksikliği anemileri, diğer
D50.9	Demir eksikliği anemisi, tanımlanmamış
D51	Vitamin B ₁₂ eksikliği anemisi
D51.3	Vitamin B ₁₂ eksikliği anemisi, diğer, diyetle bağlı
D51.8	Vitamin B ₁₂ eksikliği anemisi, diğer
D52	Folat eksikliği anemisi
D52.8	Folat eksikliği anemileri, diğer
D56	Talasemiler
D56.0	Alfa talasemi
D56.1	Beta talasemi
D56.8	Talasemiler, diğer
D56.9	Talasemi, tanımlanmamış
D57	Talasemi taşıyıcılığı
D59.1	Otoimmün hemolitik anemiler, diğer
D59.4	Hemolitik anemiler, diğer, otoimmün olmayan
D59.5	Paroksizmal nokturnal hemoglobinüri
D59.8	Hemolitik anemiler, diğer kazanılmış
D59.9	Hemolitik anemi kazanılmış, tanımlanmamış
D63	Anemi, kronik hastalıklarda sınıflanmış
D63.8	Kronik hastalık anemisi
D64	Anemi, diğer
D64.8	Anemiler diğer, tanımlanmamış
D64.9	Anemi, tanımlanmamış

Tablo 5.2. bu çalışmada kullanılan veri setinde bulunan öznitelikleri ve özelliklerini göstermektedir.

Tablo 5.2. Veri setinde bulunan öznitelikler ve özellikleri.

Öznitelik Adı	Veri Tipi	Min.	Maks.	Ort.
Yaş	Sayısal	20	109	55.4
Cinsiyet	Sayısal	0	1	0.6
Kronik Hastalık	Sayısal	0	1	0.6
Belirtiler	Sayısal	0	1	0.5
CRP (C Reaktif Protein)	Sayısal	0	27.6	1.2
D. Bilirubin (Direkt Bilirubin)	Sayısal	0	11.3	0.2
Demir	Sayısal	4	377	77.8
Ferritin	Sayısal	0	2338.4	166
Folat	Sayısal	1	99.6	11.7
HCT (Hematokrit)	Sayısal	11	64.5	35.4
HGB (Hemoglobin)	Sayısal	1	22.9	11.6
I. Bilirubin (İndirekt Bilirubin)	Sayısal	-0.1	5.01	0.4
Kreatinin	Sayısal	0.2	8	0.9
MCH (Ortalama Eritrosit Hemoglobini)	Sayısal	13.9	45.2	27.4
MCHC (Ortalama Eritrosit Hemoglobin Konsantrasyonu)	Sayısal	25.6	38.2	33.1
MCV (Ortalama Eritrosit Hacmi)	Sayısal	49	126.6	82.7
NEUT (Neutrofil Sayısı)	Sayısal	0	47.8	3.9
PLT (Platelet Sayısı)	Sayısal	1	1239	260.7
RBC (Kırmızı Kan Hücresi Sayısı)	Sayısal	1.2	45.2	4.3
RDW (Kırmızı Kan Hücresi Dağılım Genişliği)	Sayısal	11.2	38.2	17.2
T. Bilirubin (Total Bilirubin)	Sayısal	-0.02	5.7	0.7
TIBC (Total Demir Bağlama Kapasitesi)	Sayısal	104	697	353.8
UIBC (Bağlanmamış Demir Bağlama Kapasitesi)	Sayısal	-9	676	273.1
Vitamin B-12	Sayısal	13.1	2000	512.4
WBC (Beyaz Kan Hücresi Sayısı)	Sayısal	0.7	431.3	7.6
Tanı	Metinsel	-	-	-

Tablo 5.2.'de görülen özniteliklere ek olarak, ham veri setinde verilerimizi düzenlememizi sağlayan başka öznitelikler de vardır. Arşiv Numarası, yukarıda da belirtildiği gib hastane kayıtlarında bir hastayı tanımlamak için kullanılan benzersiz bir özelliktir. Her hasta için sadece bir kere tanımlanır. Hastanın Başvuru Numarası ve Onay Tarihi özniteliklerine bakarak aynı hastanın kliniğe ne sıklıkla başvurduğunu ve ilk gelişindeki tahlil sonucu verilerini temin edebildik. Diğer Bilgiler, bu çalışma için öznitelikler olarak dâhil edilmeyen hasta geçmişi verilerini içermektedir. Bu bilgiyi tez çalışmasında hastanın anemi tipleri tanısını aldığını onaylamak için kullandık. Almamıza izin verilen veriler yalnızca 30 öznitelikten oluşan verilerdir. Yukarıda açıklandığı gibi bu dört öznitelik yapay öğrenme yöntemleri ile kullanılmamaktadır.

Kullanılan özniteliklerin detaylarını açıklamak gerekirse; Hastanın yaşı, cinsiyeti, belirtilerinin ve kronik hastalıklarının varlığı gibi bilgiler, anemi tipinin belirlenmesinde önemli rol oynayan özelliklerdir. Kan analizindeki Bilirubin değerleri (Direkt, İndirekt ve Total Bilirubin), karaciğer ve safra kesesi fonksiyonlarını değerlendirmek için kullanılır. C-reaktif protein (CRP), vücuttaki iltihaplanma varlığı hakkında bilgi sahibi olmamızı sağlar. Kandaki demir değerleri her türlü kansızlık, demir eksikliği ve demir zehirlenmesi durumlarının değerlendirilmesinde kullanılır. Ferritin değeri, demir eksikliği anemisi, kronik hastalık anemisi ve talasemi tanısında önem arz etmesinin yanısıra, demir yükleme tedavisinin izlenmesinde de önemli bir değerdir. Folat, kandaki folik asit değerini ifade eder ve megaloblastik ve makroskopik anemilerin değerlendirilmesinde ve ayrıca folat eksikliği anemisinin tedavisinin izlenmesinde kullanılır. Hematokrit (HCT) kanda bulunan hemoglobin ve eritrosit miktarını gösterir. Hemoglobin (HGB) kanda bulunan toplam hemoglobin miktarını gösterir ve tam kan sayımı parametrelerinin araştırılmasında anemiye gösteren ilk değerdir. Kandaki kreatinin değeri böbrek fonksiyonlarının değerlendirilmesinde kullanılır. Ortalama eritrosit hemoglobini (MCH), eritrositlerdeki toplam hemoglobin miktarını gösterir. Ortalama eritrosit hemoglobin konsantrasyonu (MCHC), eritrositlerdeki hemoglobin konsantrasyonunun yüzdesidir. Ortalama eritrosit hacmi (MCV), oksijen taşıyan kırmızı kan hücrelerinin ortalama boyutudur. NEUT, kandaki nötrofillerin sayısıdır. PLT, işlevi kanın pıhtılaşmasını sağlamak olan trombositlerin sayısıdır. Kırmızı kan hücresi sayısı (RBC), kanda bulunan eritrositlerin sayısıdır ve

kırmızı hücre dağılım genişliği (RDW), kandaki eritrositlerin dağılım genişliğini gösterir. Toplam demir bağlama kapasitesi (TIBC) ve bağlanmamış demir bağlama kapasitesi (UIBC) de anemi tiplerini teşhis etmek için kullanılan önemli parametrelerdir. B12 vitamini, hematopoez ve normal nöronal fonksiyonlar için gerekli bir vitamindir. Düşük B12 vitamini durumunda, B12 vitamini eksikliği anemisi düşünülebilir. Beyaz kan hücresi sayısı (WBC), kandaki lökositlerin sayısıdır. Bunlar vücudun savunması görevi görür ve bağışıklık sisteminden sorumludur [55, 73, 127,129].

Hazırlanan veri setinin son halinde 1663 adet hasta verisi vardır. Bu verilerin tanıya göre dağılımı Tablo 5.3.'te verilmiştir. Anemiye bağlı hastalıkların dağılımı düzensiz ve dengesizdir. Veri setimizde 1109 kadın ve 554 erkek hasta vardır. Kadınlarda anemi prevalansının yüksek olduğu bilinmektedir ve bu veriler bu durumu doğrulamaktadır. Veri setinin %21'ini oluşturan demir eksikliği anemisi, bölgede en sık görülen anemi türüdür, en az görülen anemi türünün ise talasemi taşıyıcılığı olduğu görülmüştür.

Tablo 5.3. Veri setinde bulunan verilerin tanıya göre dağılımı.

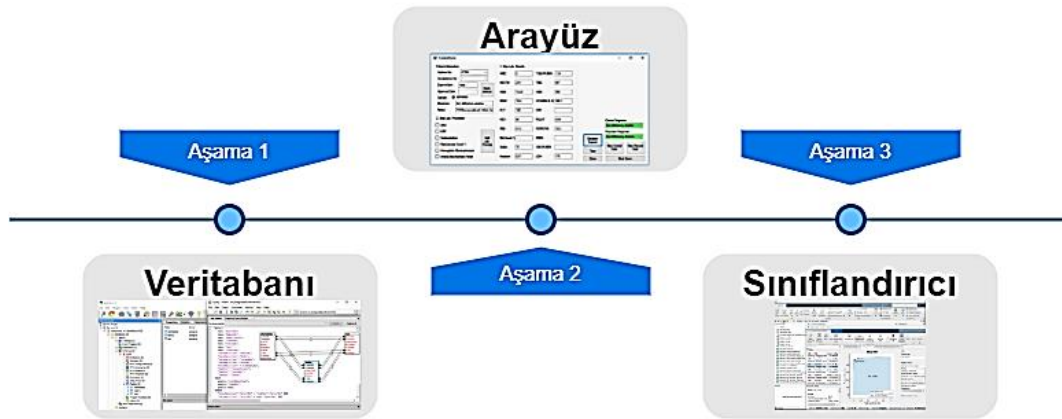
ICD-10 Kodları	Tanı	Sayı	%
D50	Demir Eksikliği Anemisi	351	21.10
D51	Vitamin B12 Eksikliği Anemisi	50	3.006
D52	Folat Eksikliği Anemisi	234	14.07
D50-D51	Demir ve Vit.B12 Eksikliği Anemisi	164	9.86
D50-D52	Demir ve Folat Eksikliği Anemisi	187	11.24
D51-D52	Folat ve Vit. B12 Eksikliği Anemisi	55	3.30
D56	Talasemi	80	4.81
D57	Talasemi Taşıyıcılığı	23	1.38
D59	Hemolitik Anemi	42	2.52
D63	Kronik Hastalık Anemisi	170	10.22
D64	Anemi, Diğer	123	7.39
-	Anemi Değil	184	11.06

Bu çalışma Türkiye'nin Batı Karadeniz Bölgesi'nde yer alan Düzce ilinde yapılmıştır. Burada listelenen anemi türleri, ilde en sık görülen 12 farklı anemi türüdür. Veriler bir hastaneden alındığı için, veri setindeki tüm hastalarda en az bir hematolojik bozukluk vardır. “Anemi Değil” şeklinde belirtilen grup sağlıklı bireylerden oluşmamaktadır ve bu nedenle sağlıklı olarak değerlendirilmemelidir. Anemi grubu dışında hematolojik rahatsızlıklar yaşadıklarından “Anemi Değil” ifadesinin kullanılması uygun kabul edilmiştir.

Özetle bu çalışmada, Düzce Üniversitesi Araştırma ve Uygulama Hastanesi Hematoloji Polikliniğine başvuran, tanı veya ön tanı olarak anemi ve ilişkili hastalıkların (ICD kodları D50.0-D64.9) girildiği hastaların verileri kullanılmıştır. Kullanılan hasta verileri şu öznetelikleri içermektedir: yaş, cinsiyet, kronik hastalık, belirtiler, CRP (C reaktif protein), D. bilirubin (direkt bilirubin), demir, ferritin, folat, HCT (hematokrit), HGB (hemoglobini), I. bilirubin (indirekt bilirubin), kreatinin, MCH (ortalama eritrosit hemoglobini), MCHC (ortalama eritrosit hemoglobin konsantrasyonu), MCV (ortalama eritrosit hacmi), NEUT (nötrofil sayısı), PLT (trombosit sayısı), RBC (kırmızı kan hücresi sayısı), RDW (kırmızı hücre dağılım genişliği), T. bilirubin (total bilirubin), TIBC (toplam demir bağlama kapasitesi), UIBC (bağlanmamış demir bağlama kapasitesi), B-12 vitamini ve WBC (beyaz kan hücresi sayısı). Anemiye bağlı hastalıkların tanısı konurken öncelikle hemogramın değerlendirilmesi gerekmektedir. Hemogramın yorumlanması, DSÖ anemi tanımlarına [9] ve Türk Hematoloji Derneği (THD) tarafından Ekim 2014'te yayınlanan Hematoloji Laboratuvar Rehberi [73] tavsiyelerine dayanmaktadır. Bu önerilere göre, veriler incelenirken bir hastada ilk olarak hemoglobin değerleri dikkate alınır, erkek hastalarda HGB <13 g / dL ve kadın hastalarda HGB <12 g / dL ise hastanın durumu anemi olarak tanımlanır. Bunu takiben hastalar, MCV değerine göre mikrositik (MCV<80), normositik (80<MCV<100) veya makrositik (MCV>100) olarak sınıflandırılmalıdır. Mikrositik anemisi olanların ferritin değeri sorgulanır ve buna göre demir eksikliği veya talasemi teşhisleri ihmitalleri değerlendirilir. Demir eksikliğinden şüphelenilmeyen hastalarda (THD anemi kılavuzlarına göre Ferritin>15 ise) demir ve demir bağlama kapasitesi değerlendirilir. Bu değerlendirme ile demir eksikliği veya kronik hastalık anemisinin ayırıcı tanısı amaçlanmaktadır. Anemili her

hastada B12 vitamini ve folik asit değerlerinin de değerlendirilerek diğer anemi türlerine (demir eksikliği, kronik hastalık anemisi, talasemi vb.) eşlik eden B12 vitamini eksikliği veya folat eksikliği anemileri veya özellikle makrositik tanımlı anemiler olarak belirlenmesi gerekmektedir. Ek olarak, anemili hastaların diğer serileri (beyaz kan hücreleri ve trombositler) değerlendirilmeli ve bunlar normal değilse (yüksek veya düşük değerler) periferik yayma işlemi yapılmalıdır. DSÖ kriterlerine göre, şiddetli anemisi olanlarda acil transfüzyon gerektirebilecek herhangi bir belirti ve bulguya çok dikkat edilmelidir. Ek olarak, aneminin etiyojisine ilişkin önemli kanıtlar varsa, her anemi türü için tetikte olmak hayati önem taşır.

Bilgisayarın anemiye uzman bir tıbbi danışman gibi teşhis edebilmesi için gerçek hasta verilerine ve deneyimli bir tıp uzmanının tavsiyesine ihtiyaç vardır. Etik kurul onayı alındıktan sonra, veriler Şekil 5.7.'de gösterildiği gibi, ilk aşamada hastane veri tabanından program arayüzüne aktarılmıştır. İkinci aşamada oluşturulan arayüz, deneyimli bir tıbbi uzmanın görüşüne dayanmaktadır.



Şekil 5.7. Çalışmanın aşamaları.

Böylece, Şekil 5.7.'de görüldüğü gibi, ilk aşamada veriler veritabanından arayüz yardımıyla çekilmiştir. İkinci aşamada arayüz yardımıyla gelen veriler bir veri seti haline getirilmiştir. Üçüncü ve son aşamada ise veriler çeşitli sınıflandırıcılar tarafından işlenmiş ve sonuçlar yorumlanmıştır. Amaç, deneyimli tıp uzmanının karar vereceği şekilde karar vermek olduğu için verilerde niteliksel bir değişiklik

yapılmamasına özen gösterilmiştir. Ön işlemde sadece verilerin sayısallaştırılması ve normalleştirilmesi işlemleri gerçekleştirilmiştir.

5.2. Bulgular ve Tartışma

Bu bölümde verilerin elde edilip hazırlanmasından sonra gerçekleştirilen öznitelik seçimi, k-kat çapraz doğrulama yöntemi ve sınıflandırma işlemlerinin sonuçları yer almaktadır. Sonuçlar değerlendirilirken ROC eğrileri ve karmaşıklık matrisleri oluşturulmuştur. Ayrıca AUC, doğruluk, hassasiyet, kesinlik, f1-skoru ve sınıflandırma hatası ölçütleri de kullanılmıştır. Bu bilgileri içeren tablolar Tablo 5.4. ile Tablo 5.14. arasında verilmiştir. Tüm veri setleri ile elde edilen en başarılı yöntemlerin ROC eğrileri ve karmaşıklık matrisleri Şekiller 5.8. – 5.21. boyunca gösterilmektedir.

Tablo 5.4.'te görülen performans ölçütleri, 25 öznitelik içeren orijinal veri seti üzerinde YSA, DVM türleri, Naive Bayes ve karar ağacı modelleri kullanılarak sınıflandırıldığında elde edilen değerlerdir.

Tablo 5.4. Modellerin 25 öznitelik içeren orijinal veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
Orijinal Veri seti (25 öznitelik)	Doğrusal DVM	68.1	0.97	70.1	77.7	0.737	31.93
	Karesel DVM	69.3	0.97	73.9	75.1	0.748	30.72
	Kübik DVM	64.3	0.96	72.8	70.5	0.716	35.71
	Arttırılmış KA	83.2	0.98	96.7	78.1	0.864	16.77
	Torbalanmış KA	80.6	0.98	92.9	79.2	0.855	19.42
	Naive Bayes	59.8	0.88	76.6	52.8	0.625	40.22
	YSA	79.6	-	78.2	79.9	0.790	20.4

Tablo 5.4.'te görüldüğü gibi, en yüksek doğruluk oranı artırılmış karar ağaçları ile elde edilmiştir. Bu çalışmada kullanılan AdaBoost yönteminde, karar ağacının doğru sınıflandırılmış örneklerinin ağırlıkları artırılmış ve her yinelemede yanlış sınıflandırılmış örneklerinin ağırlıkları azaltılmıştır. Sonraki yinelemelerde, güncellenmiş ağırlıklar kullanılmıştır. Böylece, her yinelemede, algoritma doğru sınıflandırılmış durumlara odaklanır.

Tablo 5.5. - 5.14.'te yer alan performans değerleri, dört farklı yöntem kullanılarak gerçekleştirilen özellik seçme işlemi ile elde edilen 8 farklı veri seti üzerinde yapay öğrenme yöntemleri çalıştırıldığında elde edilen sonuçları gösterir.

Tablo 5.5. Modellerin 7 öznitelik içeren CFS veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
CFS Veri seti (7 öznitelik)	Doğrusal DVM	76.1	0.98	83.2	82.7	0.829	23.93
	Karesel DVM	75.8	0.96	82.1	79.9	0.809	24.23
	Kübik DVM	71.8	0.95	78.3	74.6	0.764	28.20
	Arttırılmış KA	81.9	0.98	96.7	77.7	0.861	18.09
	Torbalanmış KA	85.6	0.98	90.8	82.3	0.863	14.37
	Naive Bayes	74	0.93	81.5	66.7	0.733	26.03
	YSA	76.4	-	74.5	85.9	0.798	23.6

Veri setinde özellik seçimi tamamlandığında, neredeyse tüm yöntemler için sınıflandırma başarısının arttığı açıktır. Tablo 5.5.'te görüldüğü üzere en başarılı yöntem torbalanmış ağaçlardır. Fakat 7 özellikli CFS veri setinin performans analizi, güçlendirilmiş ağaçların ve sinir ağlarının başarısının orijinal verilere göre azaldığını göstermesi bakımından ilginç bir sonuç olmuştur.

Tablo 5.6. – 5.8.'de yer alan performans değerleri, öznitelik seçme yöntemi olarak bilgi kazancının seçilmesi sonucu elde edilen 3 farklı veri seti üzerinde yapay öğrenme yöntemleri çalıştırıldığında elde edilen sonuçları gösterir.

Tablo 5.6. Modellerin 3-öznitelik içeren bilgi kazancı veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
Bilgi Kazancı VS (3 öznitelik)	Doğrusal DVM	71.1	0.9	88.6	37.1	0.523	28.86
	Karesel DVM	71.1	0.89	88.6	38.6	0.537	28.92
	Kübik DVM	66.9	0.86	22.3	34.7	0.271	33.07
	Arttırılmış KA	71.6	0.9	83.7	36.4	0.507	28.44
	Torbalanmış KA	71.6	0.88	64.7	37.8	0.477	28.38
	Naive Bayes	66.5	0.75	49.5	39.2	0.437	33.49
	YSA	63.8	-	36.5	67.9	0.475	36.2

Tablo 5.6.'da görülen, 3 öznitelikli bilgi kazancı veri setinin performans analizi Tablo 5.5.'te görülen 7 özellikli CFS veri setine kıyasla başarının azaldığını göstermiştir. Bununla birlikte, elde edilen orijinal veri setinden daha başarılı sonuçlar elde

edilmiştir. YSA ve karar ağacı temelli yöntemlerin başarılarında düşüş olmasına rağmen, Naive Bayes ve DVM yöntemlerinin başarısının arttığı görülmüştür.

Tablo 5.7. Modellerin 5-öznitelik içeren bilgi kazancı veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
Bilgi Kazancı VS (5 öznitelik)	Doğrusal DVM	75.5	0.98	86.4	78.7	0.823	24.53
	Karesel DVM	74.6	0.97	83.7	80.2	0.819	25.37
	Küçük DVM	72	0.97	81.5	78.5	0.799	27.96
	Arttırılmış KA	79.9	0.98	96.7	78.4	0.865	20.14
	Torbalanmış KA	81.2	0.97	90.2	83.4	0.866	18.82
	Naive Bayes	70.7	0.94	81.5	68.5	0.744	29.34
	YSA	72.9	-	78.6	82.1	0.803	27.1

Tablo 5.7.'de görülen 5 öznitelikli bilgi kazancı veri setinin performans analizi, başarının Tablo 5.6.'da görülen 3 öznitelikli bilgi kazancı veri setine kıyasla arttığını göstermiştir. Bu artış tüm yöntemler için belirgin bir şekilde gözlenmektedir. Özniteliklerin azalmasının başarıyı arttıracak düşünülmesine ve 3 öznitelikli veri kümesindeki öznitelikler en yüksek bilgi kazancına sahip öznitelikler olmasına rağmen, sadece bu özniteliklerin kullanılması yöntemlerin başarısını arttırmamıştır.

Tablo 5.8.'de 13 öznitelikten oluşan bilgi kazancı veri setinin performans analizi görülmektedir.

Tablo 5.8. Modellerin 13-öznitelik içeren bilgi kazancı veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
Bilgi Kazancı VS (13 öznitelik)	Doğrusal DVM	72.5	0.98	76.6	80.6	0.785	27.48
	Karesel DVM	73.6	0.97	82.1	80.3	0.812	26.39
	Küçük DVM	68.7	0.96	78.8	71.8	0.751	31.26
	Arttırılmış KA	82.2	0.98	97.3	78.2	0.867	17.79
	Torbalanmış KA	83.8	0.98	91.8	80.5	0.857	16.23
	Naive Bayes	65.1	0.91	79.3	58.9	0.675	34.87
	YSA	71.1	-	78.5	69.6	0.738	28.9

Görüldüğü gibi, Tablo 5.8.'deki 13 öznitelikli bilgi kazancı veri seti için performans tablosunun analizi yapıldığında, başarının önceki veri setlerine göre biraz daha düşük olduğu görülmektedir. Torbalanmış ve güçlendirilmiş karar ağacı yöntemlerinde

başarı artışı ön plana çıkmaktadır. Öznitelik sayısının artması ile DVM ve Naive Bayes yöntemlerinin başarısında hafif bir düşüş gözlenmektedir.

Tablo 5.9. – 5.11.’de yer alan performans değerleri, öznitelik seçme yöntemi olarak bilgi kazanç oranı yönteminin kullanılması sonucu elde edilen 3 farklı veri seti üzerinde yapay öğrenme yöntemleri çalıştırıldığında elde edilen sonuçları göstermektedir.

Tablo 5.9. Modellerin 7 öznitelik içeren bilgi kazanç oranı veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
Bilgi Kazanç Oranı VS (7 öznitelik)	Doğrusal DVM	72.1	0.94	72.3	60.7	0.659	27.90
	Karesel DVM	71.3	0.91	66.3	55.7	0.605	28.74
	Kübik DVM	67.6	0.88	59.2	53.4	0.561	32.35
	Arttırılmış KA	77.6	0.94	78.3	64.6	0.707	22.36
	Torbalanmış KA	78.7	0.95	71.2	61.5	0.659	21.34
	Naive Bayes	65.8	0.81	54.3	46.3	0.499	34.21
	YSA	65.1	-	48.1	28.3	0.356	34.9

Bilgi kazanç oranı yöntemi ile elde edilen 7 öznitelikli veri setinin performans analizi, başarının önceki veri setlerine yakın değerlerde olduğunu göstermiştir.

Tablo 5.10. yapay öğrenme modellerinin 11 öznitelikli bilgi kazanç oranı veri seti üzerindeki performans analizini göstermektedir.

Tablo 5.10. Modellerin 11 öznitelik içeren bilgi kazanç oranı veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
Bilgi Kazanç Oranı VS (11 attributes)	Doğrusal DVM	74	0.98	75.5	79	0.772	26.03
	Karesel DVM	72.5	0.96	76.1	76.5	0.762	27.54
	Kübik DVM	67.6	0.95	73.4	69.2	0.712	32.41
	Arttırılmış KA	80	0.97	91.3	68.9	0.785	20.02
	Torbalanmış KA	80.8	0.97	83.7	74.4	0.787	19.24
	Naive Bayes	65.1	0.87	66.3	58.4	0.620	34.87
	YSA	76.2	-	71.4	85.3	0.777	23.8

Bilgi kazanç oranı yöntemi ile elde edilen 11 öznitelikli veri setinin performans analizi, önceki veri setlerinin başarılarına kıyasla yöntemlerin başarısında kayda değer bir fark olmadığını ortaya koymuştur.

Tablo 5.11. yapay öğrenme modellerinin 16 öznitelikli bilgi kazanç oranı veri seti üzerindeki performans analizini göstermektedir.

Tablo 5.11. Modellerin 16 öznitelik içeren bilgi kazanç oranı veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
Bilgi Kazanç Oranı VS (16 attributes)	Doğrusal DVM	71	0.98	75	80.2	0.775	29.04
	Karesel DVM	70	0.97	79.3	78.5	0.788	30.01
	Kübik DVM	65.1	0.96	73.4	70.3	0.718	34.87
	Arttırılmış KA	82	0.98	96.7	78.1	0.864	18.03
	Torbalanmış KA	84.2	0.98	94	79.4	0.860	15.75
	Naive Bayes	61.4	0.88	72.3	53.8	0.616	38.60
	YSA	74.7	-	74	77.2	0.756	25.3

Bilgi kazanç oranı yöntemi ile elde edilen 16 öznitelikli veri setinin performansının analizi yine yöntemlerin başarı oranında kayda değer bir artış olmadığını göstermektedir. Ancak genel olarak özellik sayısının azalması YSA ve KA temelli yöntemlerin başarı oranının düşmesine neden olmuştur. Bu sonuç bu iki yöntemin başarılarının öznitelik sayısından etkilenmediğini düşündürmektedir.

Tablo 5.12. yapay öğrenme modellerinin, temel bileşen analizi yöntemi ile elde edilmiş 16 öznitelik içeren veri seti üzerindeki performans analizini göstermektedir.

Tablo 5.12. Modellerin 16 öznitelik içeren temel bileşen analizi veri seti için elde edilen performans değerleri.

	Yöntem	Doğruluk	AUC	Kesinlik	Geri çağırma	F1-skoru	Hata
PCA Dataset (16 attributes)	Doğrusal DVM	71.1	0.97	76.6	80.6	0.785	28.86
	Karesel DVM	70.8	0.97	76.1	76.9	0.764	29.22
	Kübik DVM	66.3	0.96	76.6	73.8	0.751	33.67
	Arttırılmış KA	83	0.98	96.7	79.5	0.872	17.01
	Torbalanmış KA	85.4	0.98	91.3	81.6	0.861	14.61
	Naive Bayes	59	0.9	80.4	54.8	0.651	41.01
	YSA	76.7	-	78.3	84.2	0.811	23.3

Tablo 5.12.'de görülen PCA-16 veri seti için performans tablosunun analizi, önceki veri setlerine benzer bir başarı oranı göstermektedir. Yapay öğrenme yöntemlerinde bu 16 öznitelikli veri setiyle elde edilen sonuçlar, 25 öznitelikli orijinal veri setiyle elde edilen sonuçlara yakın ya da daha başarılı sonuçlar vermiştir. Bu durumun tek istisnası ise YSA yöntemidir.

Tablo 5.13. ve Tablo 5.14., en başarılı performansı veren veri setlerini ve yapay öğrenme modellerini göstermektedir.

Tablo 5.13. Modellerin veri setlerine göre doğruluk değerleri.

Yöntem	Veri seti	Doğruluk
Doğrusal DVM	CFS -7	76.1 %
Karesel DVM	CFS -7	75.8 %
Kübik DVM	Info. Gain -5	72.0 %
Arttırılmış KA	PCA -16	83.0 %
Torbalanmış KA	CFS -7	85.6 %
Naive Bayes	CFS -7	74.0 %
YSA	Original -25	79.6 %

Doğrusal ve Karesel DVM, Torbalanmış KA ve Naive Bayes yöntemleri, en yüksek başarıyı korelasyon tabanlı bir öznitelik seçimi olan CFS yöntemi ile elde edilen 7 öznitelikli veri seti (CFS-7) ile göstermişlerdir. Bu dört yöntemde de başarıyı arttıran sonuçlar vermesi CFS-7 veri setini en başarılı veri seti yapmaktadır. Kübik SVM

yöntemi 5 öznitelikli bilgi kazancı (InfoGain-5) veri setiyle en yüksek başarıyı göstermiştir. Arttırılmış KA yöntemi en yüksek başarıyı 16 öznitelikli temel bileşen analizi (PCA-16) veri setiyle gösterdi. YSA yöntemi, orijinal veri seti ile en başarılı sonucu veren tek yöntemdir. Bu önemli ve dikkat çekici bir bulgu olarak karşımıza çıkmaktadır.

Tablo 5.14. Veri setlerinin modellere göre doğruluk değerleri.

Veri setleri	DVM	Doğruluk	Topluluk KA	Doğruluk	Naive Bayes Doğruluk	YSA Doğruluk
Original -25	Karesel	69.3 %	Arttırılmış	83.2 %	59.8 %	79.6 %
CFS -7	Doğrusal	76.1 %	Torbalanmış	85.6 %	74.0 %	76.4 %
Info. Gain -3	Doğrusal/ Karesel	71.1 %	Arttırılmış / Torbalanmış	71.6 %	66.5 %	63.8 %
Info. Gain -5	Doğrusal	75.5 %	Torbalanmış	81.2 %	70.7 %	72.9 %
Info. Gain -13	Karesel	73.6 %	Torbalanmış	83.8 %	65.1 %	71.1 %
Info. Gain Ratio -7	Doğrusal	72.1 %	Torbalanmış	78.7 %	65.8 %	65.1 %
Info. Gain Ratio -11	Doğrusal	74.0 %	Torbalanmış	80.8 %	65.1 %	76.2 %
Info. Gain Ratio -16	Doğrusal	71.0 %	Torbalanmış	84.2 %	61.4 %	74.7 %
PCA -16	Doğrusal	71.1 %	Torbalanmış	85.4 %	59.0 %	76.7 %

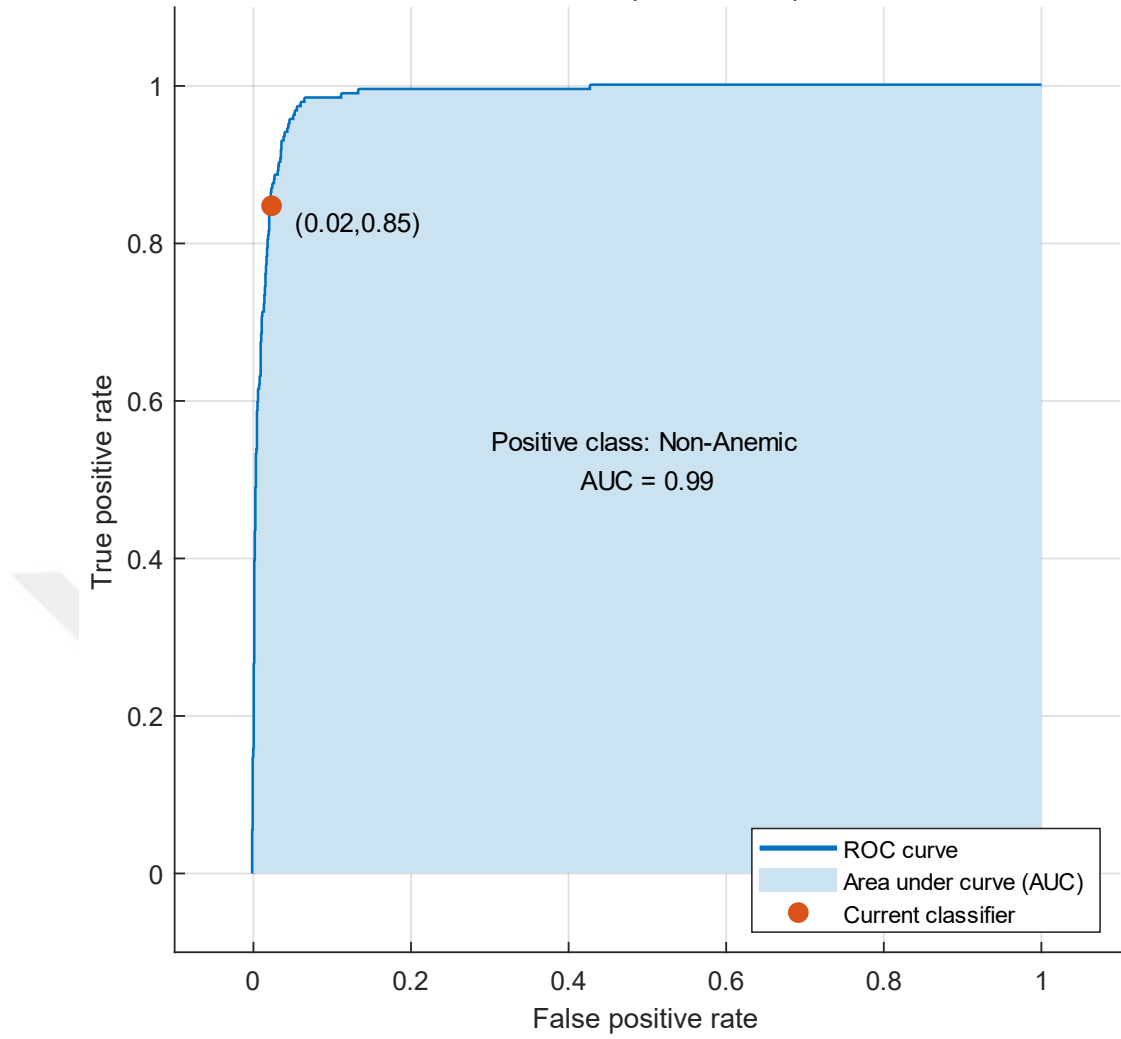
Tablo 5.14. incelendiğinde DVM yöntemleri içinde en başarılı olan Doğrusal DVM olarak karşımıza çıkmaktadır. Topluluk KA yöntemlerinden Torbalanmış KA yönteminin Arttırılmış KA yöntemine göre daha başarılı olduğu söylenebilir.

Bu çalışmada kullanılan en başarılı yöntemlerin karmaşıklık matrisleri ve ROC eğrileri Şekil 5.8. ile Şekil 5.21. boyuncagösterilmektedir.

Şekil 5.8. ve Şekil 5.9. Doğrusal DVM yöntemi ile en başarılı sonuçların alındığı CFS-7 veri seti ile elde edilen karmaşıklık matrisini ve ROC eğrisini göstermektedir.

True Class	Anemia of Chronic Disease	110	23		8		1		1	11	13	1	2
	Anemic	66	37		4		1			7	1	7	
	Carriage of Thalassemia	1	3	7	2	1		2	2	4		1	
	Folate Deficiency anemia	1	1	1	197			19		6	6	2	1
	Folate and Vit. B12 Deficiency anemia				5	39			10				1
	Hemolytic anemia	11	1		6	1	9	2	1	1	7	1	2
	Iron and Folate deficiency anemia				5			172	4	6			
	Iron and Vit. B12 deficiency anemia							17	134	13			
	Iron deficiency anemia		1					7	3	338	2		
	Non-Anemic	10		1	1		1			15	156		
	Thalassemia		13		7	3		2	4	5	1	44	1
	Vit. B12 deficiency anemia	3	3		5	2			9		4	2	22
Predicted Class													

Şekil 5.8. CFS-7 veri seti ile doğrusal DVM yönteminin karmaşıklık matrisi.



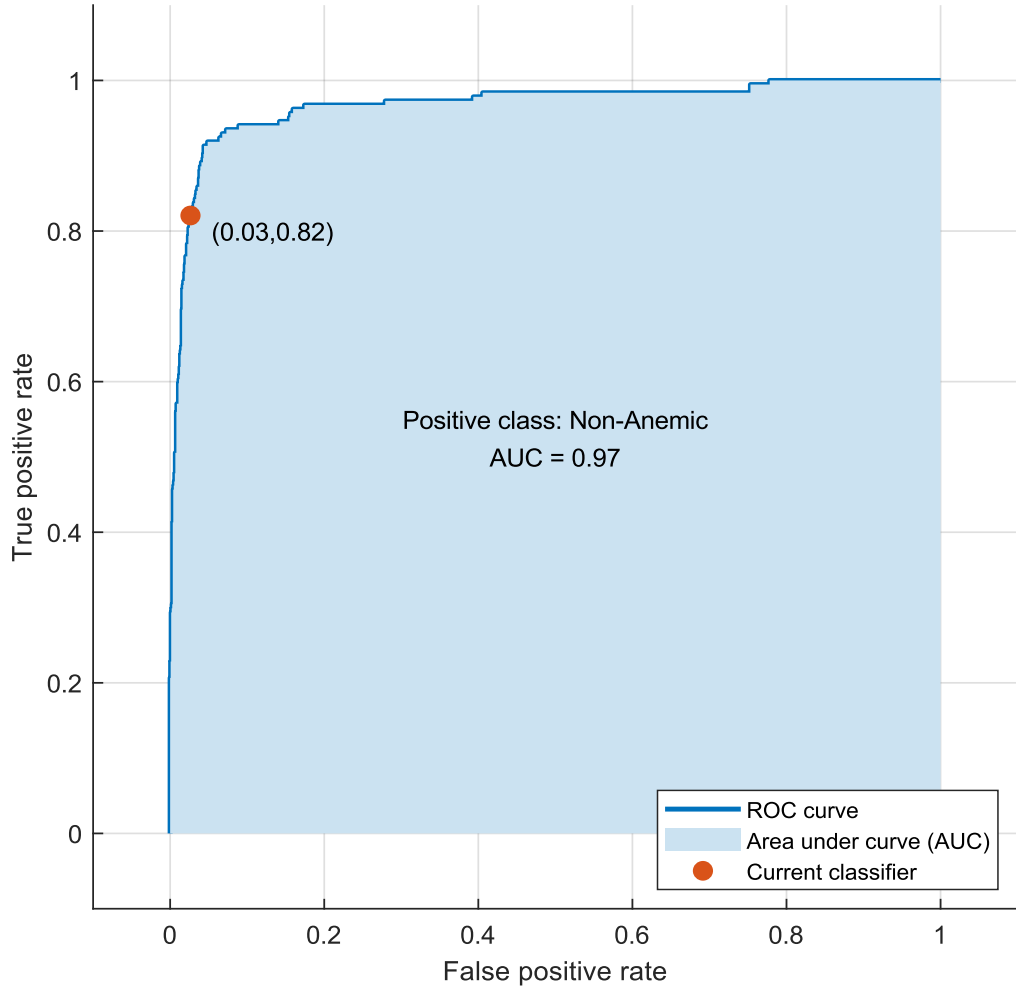
Şekil 5.9. CFS-7 veri seti ile doğrusal DVM yönteminin ROC eğrisi grafiği.

Şekil 5.8 ve Şekil 5.9'da görüldüğü gibi, doğrusal DVM anemi tiplerini %76,1 doğruluk oranında tahmin etmiştir. En yüksek başarı oranı demir eksikliği anemisi tanısında (%96,1), en düşük başarı oranı ise hemolitik anemi tanısında (%21,4) elde edilmektedir.

Şekil 5.10 ve 5.11, karesel DVM yönteminin en yüksek sonucu verdiği CFS-7 veri seti ile elde edilen karmaşıklık matrisini ve ROC eğrisini göstermektedir.

True Class	Anemia of Chronic Disease	101	30	1	8		2			7	18		3
	Anemic	34	68	1	4		4			6	3	3	
	Carriage of Thalassemia	1	3	10	2			1	3	3			
	Folate Deficiency anemia	2	2	3	195	2	1	16	1	5	3	3	1
	Folate and Vit. B12 Deficiency anemia			1	5	35	1	1	10			1	1
	Hemolytic anemia	5	6		2	1	16	1		2	6	1	2
	Iron and Folate deficiency anemia				13			150	8	14		2	
	Iron and Vit. B12 deficiency anemia				1	4		11	131	13		2	2
	Iron deficiency anemia	3	1	3	1			6	11	320	5	1	
	Non-Anemic	15	3		4		1			8	151		2
	Thalassemia		7	2	3	2	5	3	5	2	1	50	
	Vit. B12 deficiency anemia	3	2	2	7	1	1		4	1	3	2	24
Predicted Class													

Şekil 5.10. CFS-7 veri seti ile karesel DVM yönteminin karmaşıklık matrisi.



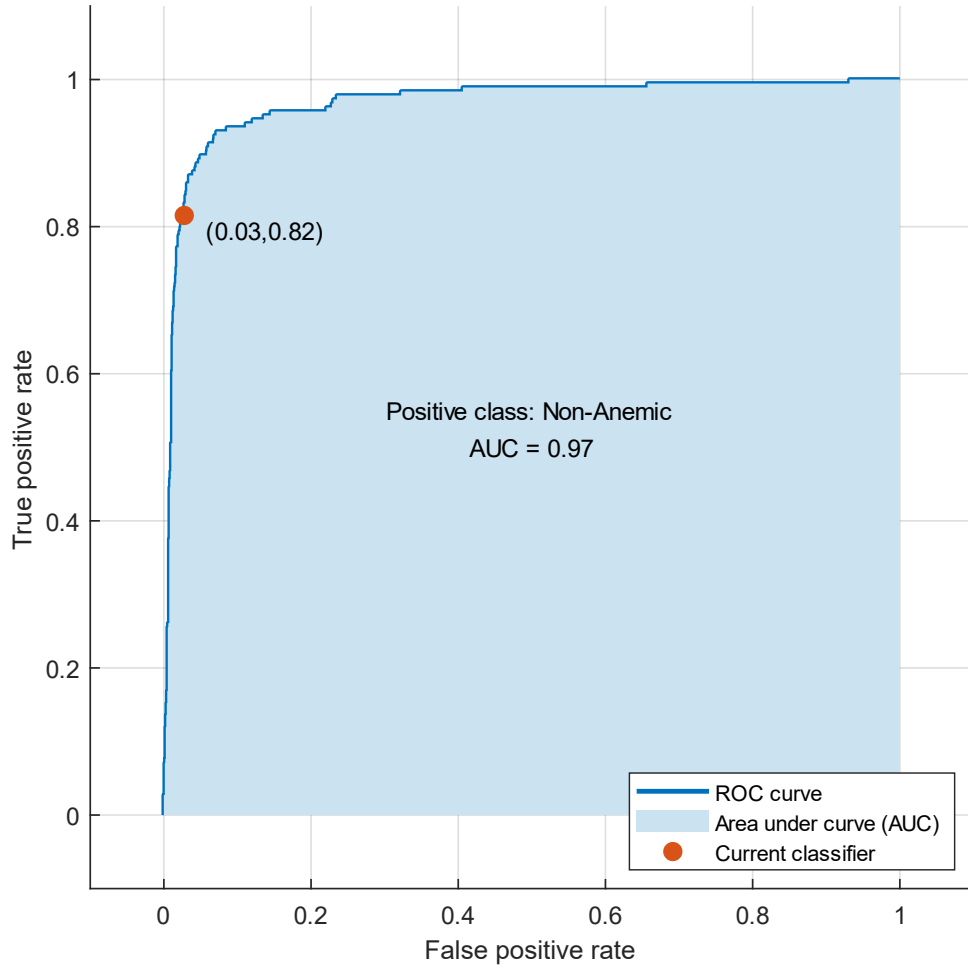
Şekil 5.11. CFS-7 veri seti ile karesel DVM yönteminin ROC eğrisi grafiği.

Şekil 5.10. ve Şekil 5.11.'de görüldüğü gibi, karesel DVM, anemi tiplerini %75,8'lik bir doğruluk oranında tahmin etmiştir. En yüksek başarı oranı demir eksikliği anemisi tanısında (%94,1), en düşük başarı ise hemolitik anemi tanısında (%42,8) görülmektedir.

Şekil 5.12. ve 5.13., kübik DVM yönteminin en yüksek sonucu verdiği InfoGain-5 veri seti ile elde edilen karmaşıklık matrisini ve ROC eğrisini göstermektedir.

True Class	Anemia of Chronic Disease	87	33	1	8		11			4	20	2	4
	Anemic	51	58	2	2		5			1	1	3	
	Carriage of Thalassemia	6		1	6	1			3	3	1		2
	Folate Deficiency anemia	5	3	3	200	4	2	9	1	2	1	2	2
	Folate and Vit. B12 Deficiency anemia			2	7	31			9			2	4
	Hemolytic anemia	12	7		6	1	3	1		3	7	1	1
	Iron and Folate deficiency anemia				7	1		157	11	11			
	Iron and Vit. B12 deficiency anemia		1	1		5	1	10	131	11		1	3
	Iron deficiency anemia	5			2		3	10	7	312	8	2	2
	Non-Anemic	18	4		2		2			7	150	1	
	Thalassemia	6	8	1	10	4		2	4	5	2	36	2
	Vit. B12 deficiency anemia	5	4		3				3	1	1	1	32
Predicted Class													

Şekil 5.12. InfoGain-5 veri seti ile kübik DVM yönteminin karmaşıklık matrisi.



Şekil 5.13. InfoGain-5 veri seti ile kübik DVM yönteminin ROC eğrisi grafiği.

Şekil 5.12. ve Şekil 5.13. kübik DVM yönteminin anemi tiplerini %72'lik bir doğruluk oranında tahmin ettiğini göstermektedir. En yüksek başarı oranı demir eksikliği anemisi tanısında (%88,9), en düşük başarı oranı ise talasemi taşıyıcılığı tanısında (%4,3) karşımıza çıkmaktadır.

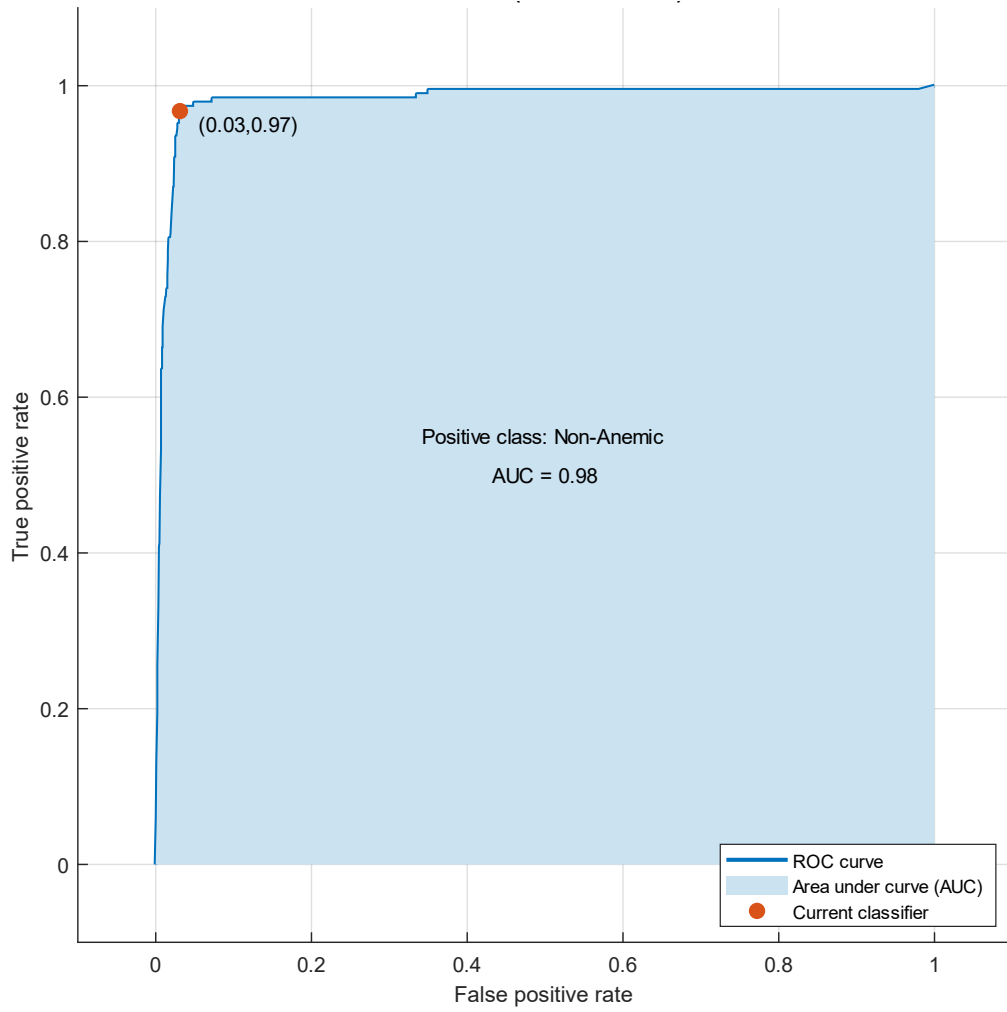
Sonuçların analizi, DVM yöntemleri kullanılarak en yüksek sınıflandırıcı başarı oranının çekirdek fonksiyonu doğrusal olarak seçildiğinde (%76,1) elde edildiğini göstermiştir. Çekirdek fonksiyonu karesel olarak seçildiğinde, başarı oranı (%75,8) ile doğrusal orana yakın çıkmaktadır. En düşük sınıflandırıcı başarı oranı, çekirdek fonksiyonu kübik olarak seçildiğinde (%72) elde edilmiştir. DVM modelleri kullanılarak en yüksek oranda sınıflandırılan hastalık demir eksikliği anemisi iken, en

düşük oranda sınıflandırılanlar hemolitik anemi ve talasemi taşıyıcılığı tanıları olmuştur.

Şekil 5.14. ve Şekil 5.15., arttırılmış karar ağacı yönteminin en başarılı sonucu verdiği PCA-16 veri seti ile elde edilen karmaşıklık matrisini ve ROC eğrisini göstermektedir.

True Class	Anemia of Chronic Disease	123	12		4					1	27	1	2
	Anemic	74	43		1					1	3	1	
	Carriage of Thalassemia	2	4	4	5			1	2	4			1
	Folate Deficiency anemia	4		1	224	1		3			1		
	Folate and Vit. B12 Deficiency anemia					48			6				1
	Hemolytic anemia	19	1		3	1		1		3	8		6
	Iron and Folate deficiency anemia				3			176	7	1			
	Iron and Vit. B12 deficiency anemia							1	157	1		1	4
	Iron deficiency anemia	2			1			1	1	342	4		
	Non-Anemic	2			3					1	178		
	Thalassemia	2	8		8	2		2	5	9	3	38	3
	Vit. B12 deficiency anemia					1						2	47
		Predicted Class											
		Anemia of Chronic Disease											
		Anemic											
		Carriage of Thalassemia											
		Folate Deficiency anemia											
		Folate and Vit. B12 Deficiency anemia											
		Hemolytic anemia											
		Iron and Folate deficiency anemia											
		Iron and Vit. B12 deficiency anemia											
		Iron deficiency anemia											
		Non-Anemic											
		Thalassemia											
		Vit. B12 deficiency anemia											

Şekil 5.14. PCA-16 veri seti ile arttırılmış KA yönteminin karmaşıklık matrisi.



Şekil 5.15. PCA-16 veri seti ile arttırılmış KA yönteminin ROC eğrisi grafiği.

Şekil 5.14. ve Şekil 5.15.'te görüldüğü gibi, arttırılmış KA yöntemi, anemi tiplerini %83 doğruluk oranında tahmin etmiştir. En yüksek başarı oranı demir eksikliği anemisi hastalarının tanısında (%97,4), en düşük başarı ise hemolitik anemi hastalarının tanısında (%0) görülmektedir.

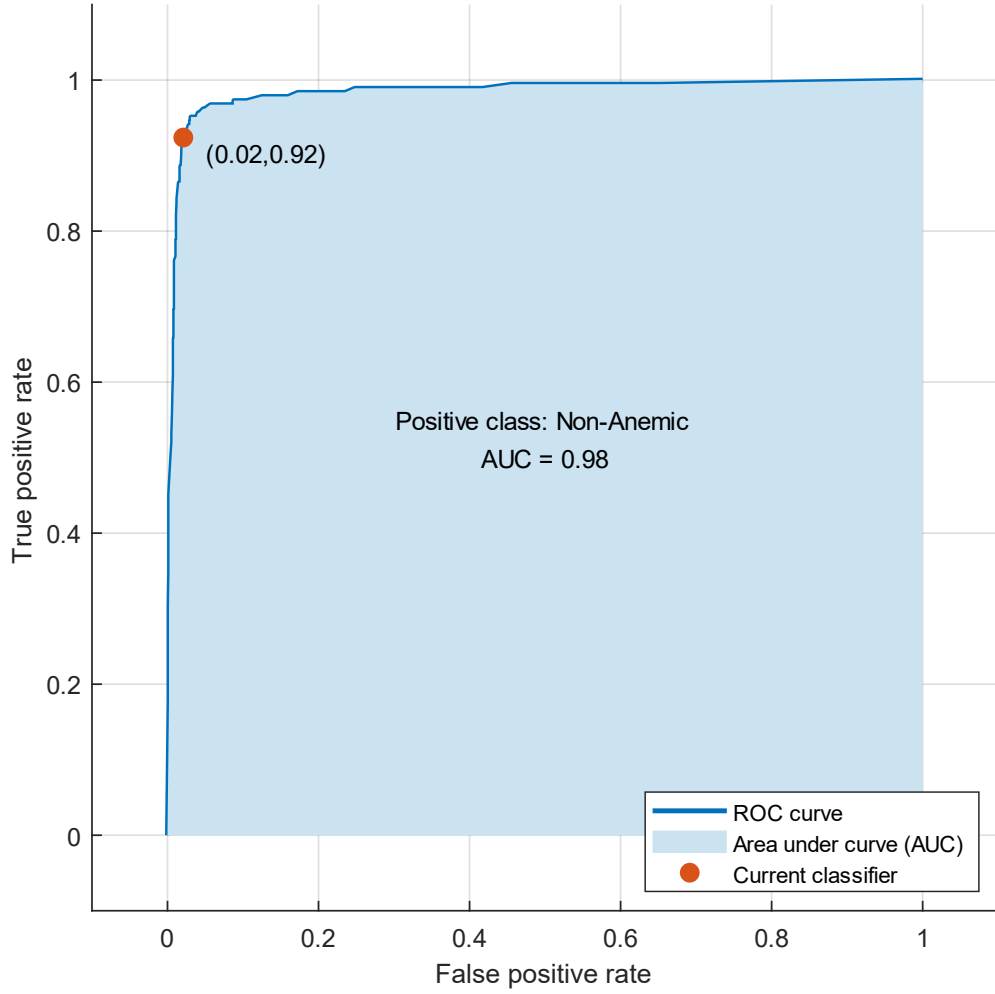
Şimdiye kadar tüm yöntemlerin en başarılı tahmini demir eksikliği anemisi tanısında yapmalarının nedeni örnek sayısı ile ilgili olabilir. Veri setinde yer alan verilerin tanılara göre dağılımına bakıldığında en yüksek veri sayısı demir eksikliği anemisi tanısında görülmektedir. Talasemi taşıyıcılığı ve hemolitik anemi ise veri setinde en

az veri sayısı bulunan tanılardır. Bu nedenle sınıflandırıcılar, demir eksikliği anemisi tanılarını en yüksek başarı oranı ile tahmin ederken diğer hastalık tanılarını en düşük başarı oranı ile tahmin etmektedirler.

Şekil 5.16. ve Şekil 5.17., torbalanmış karar ağacı yönteminin en başarılı sonucu verdiği CFS-7 veri seti ile elde edilen karmaşıklık matrisini ve ROC eğrisini göstermektedir.

True Class	Anemia of Chronic Disease	120	26	1	2					1	19		1
	Anemic	39	77	2			1			1	2	1	
	Carriage of Thalassemia	2	3	8	2	1		1	2	2		1	1
	Folate Deficiency anemia	2	3		225			3	1				
	Folate and Vit. B12 Deficiency anemia			1		45		1	8				
	Hemolytic anemia	8	5		4	1	12	1		2	4	1	4
	Iron and Folate deficiency anemia				3	1		176	5	2			
	Iron and Vit. B12 deficiency anemia			2		5		3	150	1			3
	Iron deficiency anemia	2						1	1	343	4		
	Non-Anemic	9	1		3					1	170		
	Thalassemia	1	5	1	6	2		2	4	3	1	53	2
	Vit. B12 deficiency anemia		1		2	1	1		1		1	1	42
		Predicted Class											
		Anemia of Chronic Disease Anemic Carriage of Thalassemia Folate Deficiency anemia Folate and Vit. B12 Deficiency anemia Hemolytic anemia Iron and Folate deficiency anemia Iron and Vit. B12 deficiency anemia Iron deficiency anemia Non-Anemic Thalassemia Vit. B12 deficiency anemia											

Şekil 5.16. CFS-7 veri seti ile torbalanmış KA yönteminin karmaşıklık matrisi.



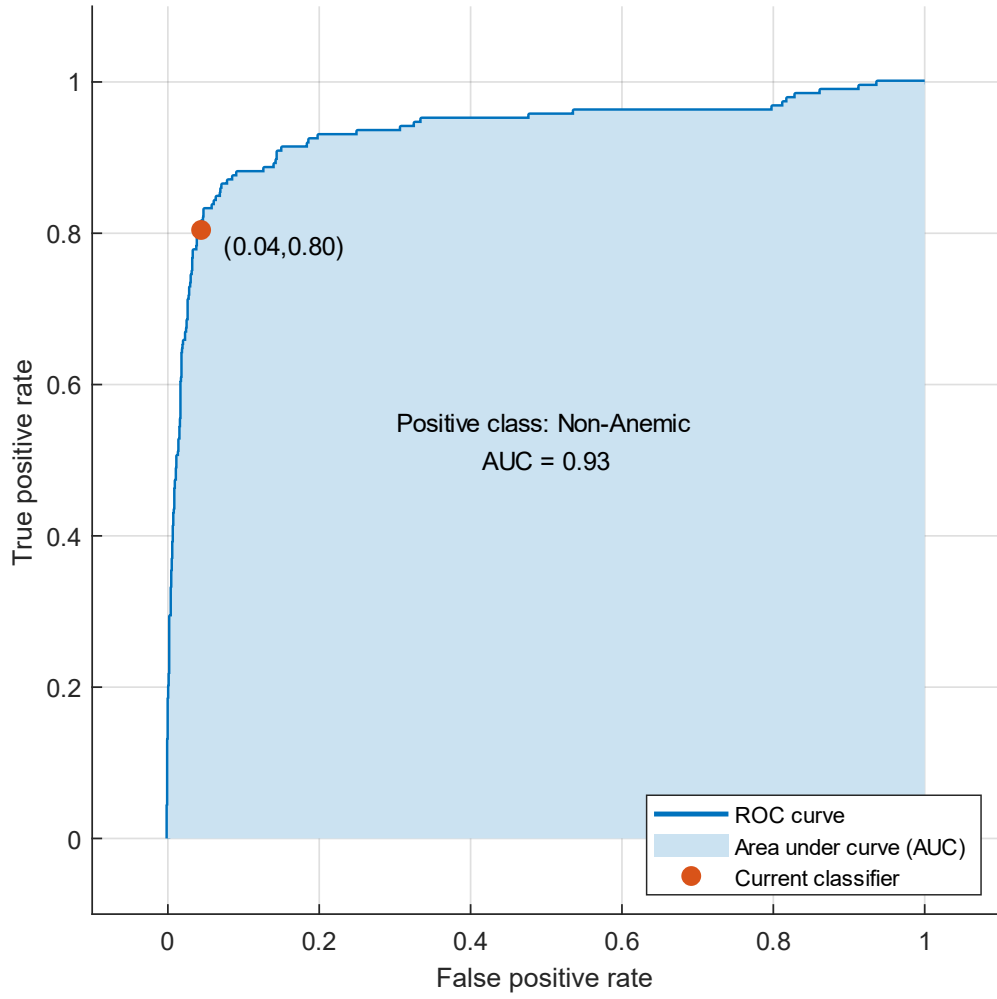
Şekil 5.17. CFS-7 veri seti ile torbalanmış KA yönteminin ROC eğrisi grafiği.

Şekil 5.16. ve Şekil 5.17.'de görüldüğü gibi, torbalanmış karar ağacı yöntemi, anemi tiplerini %85,6 doğruluk oranında tahmin etmiştir. En yüksek başarı oranı demir eksikliği anemisi hastalarının tanısında (%97,7), en düşük başarı ise hemolitik anemi hastalarının tanısında (%35,7) görülmektedir. Bunun nedeni, yukarıda da bahsedildiği gibi örnek sayısı olabilir.

Şekil 5.18. ve Şekil 5.19., Naive Bayes yönteminin en başarılı sonucu verdiği CFS-7 veri seti ile elde edilen karmaşıklık matrisini ve ROC eğrisini göstermektedir.

True Class	Anemia of Chronic Disease	103	28		7				4	22	2	4
	Anemic	48	48	2	3				4	7	10	1
	Carriage of Thalassemia	1	3	8	1	2		1	3	4		
	Folate Deficiency anemia	3	3	1	209	3	1	6		5	1	1
	Folate and Vit. B12 Deficiency anemia		1		6	42			5			1
	Hemolytic anemia	3	8		2	1	13	1		2	6	1
	Iron and Folate deficiency anemia				13	1		159	9	5		
	Iron and Vit. B12 deficiency anemia					3		15	129	14		3
	Iron deficiency anemia	7	15					12	8	282	26	1
	Non-Anemic	7	15		3		1			8	148	2
	Thalassemia	4	8	5	3	1		3	5	4	1	43
	Vit. B12 deficiency anemia	3	2		4	1	1			1	2	3
												33
Predicted Class												

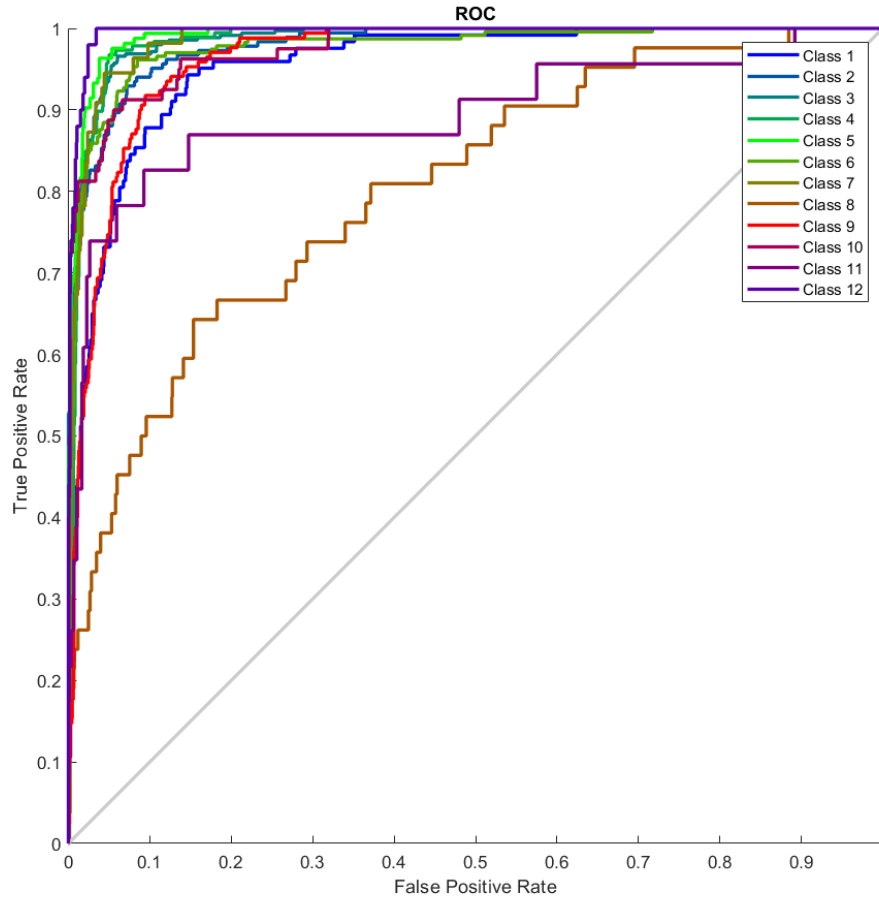
Şekil 5.18. CFS-7 veri seti ile Naive Bayes yönteminin karmaşıklık matrisi.



Şekil 5.19. CFS-7 veri seti ile Naive Bayes yönteminin ROC eğrisi grafiği.

Şekil 5.18. ve Şekil 5.19.'da görüldüğü gibi, Naive Bayes yöntemi, anemi tiplerini %74 doğruluk oranında tahmin etmiştir. En yüksek başarı oranı folat eksikliği anemisi hastalarının tanısında (%87,6), en düşük başarı ise hemolitik anemi hastalarının tanısında (%26,1) görülmektedir. Burada farklı bir şekilde en başarılı sınıflandırılan tanı olarak karşımıza folat eksikliği anemisi tanısı çıkmaktadır.

Şekil 5.20. ve Şekil 5.21., YSA yönteminin en başarılı sonucu verdiği orijinal-25 veri seti ile elde edilen karmaşıklık matrisini ve ROC eğrisini göstermektedir.



Şekil 5.21. Orijinal veri seti ile YSA yönteminin ROC eğrisi grafiği.

Şekil 5.20. ve Şekil 5.21.'de görüldüğü gibi, YSA yöntemi, anemi tiplerini %79,6 doğruluk oranında tahmin etmiştir. En yüksek başarı oranı vitamin B12 eksikliği anemisi hastalarının tanısında (%86,4), en düşük başarı oranı ise anemi hastalarının tanısında (%60,8) görülmektedir. Demir eksikliği anemisi tanısında %85,4 başarı ile ikinci en yüksek başarı oranı elde edilmiştir.

Yapay sinir ağlarının yapısından dolayı en yüksek başarı, veri setinde en fazla sayıda bulunan demir eksikliği anemisi tanısında ve en düşük başarı da veri setinde daha az sayıda bulunan anemi tanısında elde edilmiştir. Burada farklı bir şekilde en yüksek başarı ile sınıflandırılan tanı ile en düşük başarı ile sınıflandırılan tanı arasında çok yüksek puanlarda başarı farkı olmadığı gözlenmektedir.

Bu çalışma için elde edilen bu sonuçlara göre YSA ve topluluk karar ağacı yöntemleri Naive Bayes ve SVM yöntemlerine göre daha anlamlı sonuçlar verdikleri ve her sınıfın tahmin oranlarının dağılımlarının daha dengeli olduğu söylenebilir. Sonuçlar ayrıca en yüksek başarı oranında sınıflandırılan hastalıkların demir eksikliği anemisi ve folat eksikliği anemisi olduğunu ve en düşük başarı oranında sınıflandırılan hastalıkların hemolitik anemi ve talasemi taşıyıcılığı olduğunu göstermektedir. Bunun nedenlerinden biri örneklerin sayısı olabilir. Veri setinde en fazla örnek sayısına sahip sınıf, 351 örnekle demir eksikliği anemisidir ve bunu 234 örnek ile folat eksikliği anemisi sınıfı izlemektedir. Veri setinde en az sayıda örneğe sahip sınıf, yalnızca 23 örnekle talasemi taşıyıcılığı olup, bunu 42 örnek ile hemolitik anemi izlemektedir.

Tüm sonuçlar incelendiğinde, bu verilerle, karar ağacı temelli topluluk yöntemlerinin YSA, Naive Bayes veya DVM yöntemlerine göre daha başarılı sonuçlar ürettiğini ve tahmin oranlarının dağılımının daha dengeli olduğu belirtilmiştir. Veri setinde öznitelik seçimi yapmak, sınıflandırıcılarda başarıyı artıran önemli bir faktördür. Fakat bu bulgulara dikkat çeken bir nokta YSA yönteminde en başarılı sonucun orijinal veri seti kullanılarak elde edilmiş olmasıdır. Bu bulguya dayanarak ileride bu konuda yeni çalışmalar yapmayı planlamaktayız.

BÖLÜM 6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında anemi tanısının genel klinik şartlarında koyulabilmesi için bir sistem geliştirilmesi amaçlanmıştır. Diğer bir deyişle bir uzman doktorun hastalığa karar verme sürecinin bilgisayar programına öğretilmesi amaçlanmıştır. Bu amaçla en yaygın görülen kan hastalıklarından olan anemi tanısında kullanılan veriler kullanılmıştır. Hematoloji kliniğine başvuran anemi tanılı hastaların verileri tasarlanan bir program yardımıyla toplanmıştır. Elde edilen verilerin öznitelik seçimi korelasyon, bilgi kazancı, bilgi kazanç oranı ve temel bileşen analizi yöntemleri kullanılarak yapılmıştır. Bu yöntemler sonucunda dokuz farklı veri seti elde edilmiştir. Bu veri setleri üzerinde literatürde sıkça kullanılan destek vektör makineleri, karar ağaçları, naive bayes ve yapay sinir ağı olmak üzere dört temel yapay öğrenme yöntemi kullanılmıştır. Eğitim ve test aşamaları k-kat çapraz doğrulama yöntemi ile gerçekleştirilmiştir. Yöntemlerin başarımları analizleri ROC analizi yöntemi kullanılarak gerçekleştirilmiştir. Her model doğruluk değerlerine ek olarak sınıflandırma hatası, ROC eğrisi altında kalan alan (AUC), kesinlik, geri çağırma ve F1-skoru metrikleri kullanılarak değerlendirilmiştir. Yapılan bu çalışmalar makale olarak da yayınlanmıştır [149].

Bu çalışmada, DSÖ Uluslararası Hastalık Sınıflandırması Kodlarında (ICD-10) tanımlanan 12 farklı anemi türüne tanı konması modellenmiştir. Çalışmanın önemli bir noktası şudur; çalışmada yer alan veriler ve öznitelikler deneyimli bir uzman doktor tarafından anemileri teşhis etmek için kullanılan 25 farklı özniçeği içermektedir. Ayrıca kullanılan veriler tamamen orijinaldir ve sayısal kan parametrelerinin yanı sıra yaş, cinsiyet, kronik hastalıklar ve belirtiler gibi hastanın diğer bilgilerini de içermektedir. Bilgi kazancı, bilgi kazanç oranı, temel bileşen analizi ve korelasyon tabanlı öznitelik alt küme seçimi gibi bilinen öznitelik seçme yöntemleri kullanılarak

yeni veri kümeleri de oluşturulmuştur. Bu veri setleri kullanılarak yapılan deneylerde değerli sonuçlar elde edilmiştir.

Bir hastayı tedavi etmek için öncelikle hastalığın belirlenmesi gerekir. Uzmanların bu teşhisi koymalarına yardımcı olmak için çeşitli yöntemler geliştirilmiştir. Kaynak araştırmasında incelenen çalışmaların yöntemleri ve sonuçları incelenerek yeni veri seti için özgün bir çalışma yürütülmüştür. İncelenen çalışmalar örtüşen profillere sahip olabilse de veriler aynı yaklaşımla kullanılmadığı için amaç performans karşılaştırması yapmak değildir. Bu çalışmada kullanılan tüm veriler hasta bireylerin dosyalarından alınmıştır. Bu nedenle sağlıklı bireylerden veri alınmadığından “anemi değil” terimi, anemi türlerinden birine ait olmayan diğer kan hastalıkları olan hastalar için uygun kabul edilmektedir. Çalışma genişletildiğinde, anemi dışındaki diğer hematolojik bozukluklar da hastalıkların tahminine dâhil edilebilir. Bu çalışmada, en bilinen yapay öğrenme yöntemlerinden dördü (yapay sinir ağları, destek vektör makineleri, Naive bayes ve karar ağaçları) tamamen orijinal bir veri setine uygulanmış ve sonuçlar incelenmiştir. İleride yeni verilere ek olarak yeni yöntemler de eklenerek çalışma genişletilebilir.

Bu çalışmada gerçek hasta verilerinin kullanılması için Etik Kurul izni gerekmiştir. Etik Kurul onayının alınabilmesi için çalışma hakkındaki her detayın belirlenmesi gerekmektedir. Bu nedenle, verilerin tüm niteliklerinin Etik Kurul Komitesi'ne bildirilmesi gerekmiştir. Alınan veriler ile elde edilen veri setinde bazı verilerin sayısallaştırılması dışında herhangi bir ön işlem veya verilere herhangi bir müdahaleye yapılmamıştır. Verilerin içeriğini ve kalitesini değiştirmeye yönelik bir işlem bulunmadığından, bu çalışmada kullanılan veri seti sayısal müdahalenin yapılmadığı tamamen orijinal bir veri seti olmuştur. Dahası, gerçek hasta verilerinin kullanılması bu çalışmanın alaka düzeyini önemli hale getirmiştir.

Çalışma 12 farklı anemi türünün teşhisinde doktorlara ve uzman olmayan personele yardımcı olarak karar verme desteği sağlayabilir. Bu çalışmanın sınırlarından biri, tüm veri setinin hematolojik hastalıkları olan bireylerden alınan verilerden oluşmasıdır. Bir

diğeri aneminin aynı zamanda diğer hastalıklara da eşlik edebilen bir belirti olmasıdır. Bu sınırlamalara rağmen elde ettiğimiz sınıflandırıcı performanslarının kabul edilebilir düzeyde olduğu açıktır. Bu çalışmanın literatüre olumlu katkılar yaptığına inanıyoruz.

Bu çalışma için elde edilen sonuçlara göre YSA ve topluluk karar ağacı yöntemleri Naive Bayes ve SVM yöntemlerine göre daha anlamlı sonuçlar verdikleri ve her sınıfın tahmin oranlarının dağılımlarının daha dengeli olduğu söylenebilir. Sonuçlar ayrıca en yüksek başarı oranında sınıflandırılan hastalıkların demir eksikliği anemisi ve folat eksikliği anemisi olduğunu ve en düşük başarı oranında sınıflandırılan hastalıkların hemolitik anemi ve talasemi taşıyıcılığı olduğunu göstermektedir. Bunun nedenlerinden biri örneklerin sayısı olabilir. Veri setinde en fazla örnek sayısına sahip sınıf, 351 örnekle demir eksikliği anemisidir ve bunu 234 örnek ile folat eksikliği anemisi sınıfı izlemektedir. Veri setinde en az sayıda örneğe sahip sınıf, yalnızca 23 örnekle talasemi taşıyıcılığı olup, bunu 42 örnek ile hemolitik anemi izlemektedir.

Tüm sonuçlar incelendiğinde, bu verilerle, karar ağacı temelli topluluk yöntemlerinin YSA, Naive Bayes veya DVM yöntemlerine göre daha başarılı sonuçlar ürettiğini ve tahmin oranlarının dağılımının daha dengeli olduğu belirtilmiştir. Veri setinde öznitelik seçimi yapmak, sınıflandırıcılarda başarıyı artıran önemli bir faktördür. Fakat bu bulgularda dikkat çeken önemli bir nokta YSA yönteminde en başarılı sonucun orijinal veri seti kullanılarak elde edilmiş olmasıdır. Bu bulguya dayanarak ileride bu konuda yeni çalışmalar yapmayı planlamaktayız.

Son olarak, başarı oranını artırmak için, gelecekteki çalışmalar farklı yöntemleri birleştirmeye ve yeni hibrit yöntemler geliştirmeye odaklanmalıdır. Ayrıca veri setinin genişletilmesi ile anemi dışındaki hematolojik hastalıklar da bu çalışmalara dahil edilebilir.

Sonuç olarak çalışmamızdan elde edilen önemli bulgular şu şekilde özetlenebilir:

- a. Çalışmada kullanılan anemi hastalarının verileri üzerinde hiçbir işlem yapılmaması veri setini tamamen orijinal kılmaktadır. Bu orijinal verilerin

sınıflandırılması için kullanılan 4 öznitelik seçme yönteminin ve 4 sınıflandırıcının performansları umut vericidir.

- b. Genel olarak en başarılı yöntem topluluk karar ağaçları yöntemi olmuştur.
- c. Özellikle YSA yönteminin en başarılı sonucu orijinal veri seti ile vermesi dikkat çekicidir.
- d. Veri setinde bulunan sınıflardan Demir eksikliği anemisi yöntemlerin hemen hepsinde en başarılı sınıflandırılan sınıf olmuştur. Bunun sebebi olarak örnek sayısının fazla olması düşünülmektedir.
- e. İleride yapılacak çalışmalarda kullanılan veri seti daha kapsamlı ve daha dengeli bir hale getirilebilir. Ayrıca kullanılan yöntemler çeşitlendirilebilir, birden fazla yöntemin birarada kullanılması gibi hibrit yöntemler geliştirilebilir.

KAYNAKLAR

- [1] www.ttb.org.tr/udek/images/stories/file/bilisim_calistay/osaka.ppt, Eriřim tarihi: 22.04.2014.
- [2] http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.536b22c5467940.91357806 Eriřim: 08/05/2014.
- [3] Shortliffe E. H., Cimino J. C., Biomedical Informatics: Computer Applications in Health Care And Biomedicine, Part 20: Clinical Decision Support Systems, Musen M. A., Shahar Y., Shortliffe E. H., 3rd Edition, ISBN 978-0-387-36278-6, Springer, 2006.
- [4] <https://www.postgresql.org/docs/>, Eriřim: 08.01.2021
- [5] <https://visualstudio.microsoft.com/tr/vs/>, Eriřim: 08.01.2021
- [6] https://www.mathworks.com/products/new_products/release2020a.html, Eriřim tarihi: 10.01.2021
- [7] <https://docs.rapidminer.com/>, Eriřim: 08.01.2021
- [8] <https://www.cs.waikato.ac.nz/ml/weka/index.html> Eriřim tarihi: 10.01.2021.
- [9] T. Vos et al., Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010, *Lancet*, vol. 380, no. 9859, pp. 2163-2196, 2012, DOI: 10.1016/S0140-6736(12)61729-2.
- [10] WHO, Haemoglobin concentrations for the diagnosis of anemia and assessment of severity, Vitamin and Mineral Nutrition Information System, World Health Organization, Geneva, 2011.
- [11] N.J. Kiassebaum et al., A systematic analysis of global anemia burden from 1990 to 2010, *Blood*, vol. 123, pp. 615-624, 2014, DOI: 10.1182/blood-2013-06-508325.

- [12] Shortliffe E.H., The science of biomedical computing, *Med Inform* 9:185-93,1984.
- [13] Van Bommel J.H. The structure of medical informatics. *Med Inform* 9:175-80,1984.
- [14] Hsu W., Markey M.K., Wang M.D. Biomedical Imaging Informatics in The Era of Precision Medicine: Progress, Challenges, and Opportunities, *J Am Med Inform Assoc.*, 20(6):1010-1013, November 2013.
- [15] Ghista D.N., *Biomedical Engineering: Yesterday, Today, And Tomorrow*, IEEE Engineering In Medicine And Biology, 2000.
- [16] Hassanien A.E., Al-Shammari E.T., Ghali N., *Computational Intelligence Techniques in Bioinformatics, Computational Biology and Chemistry*, <http://dx.doi.org/10.1016/j.compbiolchem.2013.04.007>,2013.
- [17] Ledley R. S. Lusted L. B., Reasoning Foundations of Medical Diagnosis, *Science*, 130(3366):9-21, 1959.
- [18] Warner H. R., Toronto A. F., Veasey L.G., and Stephenson R., A mathematical approach to medical diagnosis, Application to congenital heart disease, *JAMA: The Journal of the American Medical Association*, 177:177–183, 1961.
- [19] Özata M. Aslan Ş., Klinik karar destek sistemleri ve örnek uygulamalar, *Kocatepe Tıp Dergisi* 5:11-17, 2004.
- [20] Beck J.R., Bell J.R., Hirai F., Simmons J.J., Lyon H.C. Jr., Computer-Based Exercises in Cardiac Diagnosis (PlanAlyzer), *PROC. Annu. Symp. Computer Applications in Medical Care*, 1988 Nov.9: 403-408, PMID: PMC2245328, 1988.
- [21] Beck J.R., O'Donnell J.F., Hirai F., Simmons J.J., Healy J.C., Lyon H.C.Jr., Computer-based exercises in anemia diagnosis (PlanAlyzer), *Methods Inf. Med.*, 28(4):364-9 PMID:2695787, 1989.
- [22] Lyon H.C.Jr., Bell J.R., O'Donnell J.F., Hirai F., Healy J.C., Beck J.R., The PlanAlyzer Cases for Teaching Clinical Reasoning: A Demonstration of the Cases, Discussion of the Research & Development PROCess, Lessons Learned and Strategies for Introducing Computer-Based Programs into Medical School Courses as a Vehicle for Curriculum Reform, *PROC. Annu. Symp. Computer Applications in Medical Care*, 1993:952, PMID: PMC2850769, 1993.

- [23] Waghlikar K.B., Sundararajan V., Deshpande A.W., Modelling Paradigms For Medical Diagnostic Decision Support: A Survey And Future Directions, *J. Med. Syst.*, 36:3029-3049, Doi 10.1007/ S 10916-011-9780-4, 2012.
- [24] Evans R.S., The Help System: A Review Of Clinical Applications in Infectious Diseases and Antibiotic Use, *M. D. Computing: Computers in Medical Practice*, 8(5), 1991.
- [25] De Dombal F. T., Computers, diagnoses and patients with acute abdominal pain. *Archives of emergency medicine*, 9(3):267–270, 1992.
- [26] Barnett G. O., Cimino J. J., Hupp J. A., Hoffer E. P., Dxplain: An Evolving Diagnostic Decision-Support System, *JAMA: The Journal Of The American Medical Association*, 258(1):67–74, 1987.
- [27] Engle R. L., Flehinger B. J., Allen S., Friedman R., Lipkin M., Davis B. J., Leveridge L. L., HEME: A Computer Aid to Diagnosis of Hematologic Disease, *Bulletin of the New York Academy of Medicine*, 52(5):584–600, 1976.
- [28] Shortliffe E. H., Mycin: A Knowledge-Based Computer Program Applied To Infectious Diseases, In *Annu Symp Comput Appl Med Care*, pp. 66–69, 1977.
- [29] Goldberg R. N., Weiss S. M., An Experimental Transformation Of A Large Expert Knowledge Base, *Journal of Medical Systems*, 6(1):41–52, 1982.
- [30] Politakis P. G., Using Empirical Analysis To Refine Expert System Knowledge Bases (SEEK). PhD thesis, New Brunswick, NJ, USA, 1983.
- [31] Amarel S., Introduction To The Comtex Microfiche Edition Of The Rutgers University Artificial Intelligence Research Reports: The History Of Artificial Intelligence At Rutgers University, *AI Magazine*, 6(3):192–202, 1985.
- [32] Chokhani S., Correspondences Between Biomathematical And Causal Models For Clinical Decision Making, *Journal of Medical Systems*, 5(4):249–264, 1981.
- [33] Myers J. D., The computer as a diagnostic consultant, with emphasis on use of laboratory data, *Clin Chem*, 32(9):1714–1718, 1986.
- [34] Banks G., Artificial intelligence in medical diagnosis: the internist/caduceus approach, *Critical Reviews in Medical Informatics*, 1(1):23–54, 1986.
- [35] Leitich H., Kiener H. P., Kolarz G., Schuh C., Graninger W., Adlassnig K. P., A Prospective Evaluation Of The Medical Consultation System Cadiag-

II/Rheuma in a Rheumatological Outpatient Clinic, *Methods Inf Med*, 40(3):213–220, 2001.

- [36] Faust O., Acharya U.R., Tamura T., Formal Design Methods for Reliable Computer-Aided diagnosis: A Review, *IEEE Reviews in Biomedical Engineering*, Vol.5, 2012.
- [37] Sepulveda J.L., Young D.S., The Ideal Laboratory Information System, *ARCH Pathol Lab.Mat.*, 137:1129-1140, 2013.
- [38] Wang X., Ling J., Multiple valued logic approach for matching patient records in multiple databases, *Journal of Biomedical Informatics*, 45: 224–230, 2012.
- [39] Stantica B., Terenzianib P., Governatoric G., Bottrighib A., Sattara A., An implicit approach to deal with periodically repeated medical data, *Artificial Intelligence in Medicine*, 55: 149–162, 2012.
- [40] Seera M., Lim C.P., A hybrid intelligent system for medical data classification, *Expert Systems with Applications*, 41:2239–2249, 2014.
- [41] Trindadea R., Ferreiraa N.C., Caramelo F., Development of a Medical Image Database for Data Mining and Group Analysis Applications, *PROcedia Technology (CENTERIS 2012)*, 5:912-921, 2012.
- [42] Polat K., Güneş S., Medical decision support system based on artificial immune recognition immune system (AIRS), fuzzy weighted pre-pROcessing and feature selection, *Expert Systems with Applications*, 33:484–490, 2007.
- [43] Polat K., Güneş S., A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-pROcessing andAIRS, *Computer Methods and Programs in Biomedicine*, 88(2):164-174, 2007.
- [44] Shouzhong X., Fanglu W., Building Knowledge-Based Systems with an Assembling Technique, *IEEE Engineering in Medicine and Biology*, 0739-5175 Pg.81-83, 1997.
- [45] Kawamoto K., Lobach D.F., Design Implementation, Use And Preliminary Evaluation Of SEBASTIAN, A Standards-Based Web Service For Clinical Decision Support, *Amia Annual Symposium PROceedings Archive 2005*:380-384, 2005.
- [46] Yan H., Jiang Y., Zheng J., Fu B., Xiao S., Peng C., The Internet-Based Knowledge Acquisition And Management Method To Construct Large-Scale Distributed Medical Expert Systems, *Computer Methods And Programs in Medicine*, 74:1-10, 2004.

- [47] Dragusin R., Petcu P., Lioma C., Larsen B., Jorgensen H.L., Cox I.J., Hansen L.K., Ingwersen P., Winther O., FindZebra: A search Engine for rare diseases, *International Journal of Medical Informatics*, 82(6): 528-538, 2013.
- [48] Tradigoa G., Veltria P., Grecob S., Geomedica: managing and querying clinical data distributions on geographical database systems, *PROCedia Computer Science (International Conference on Computational Science 2010)*, 1:979-986, 2012.
- [49] De Maio C., Loia V., Fenza G., Gallo M., Linciano R., Morrone A., Fuzzy Knowledge Approach to Automatic Disease Diagnosis, *IEEE International Conference on Fuzzy Systems*, June 27-30, 2011, Taipei, Taiwan].
- [50] Isern D., Moreno A., Computer-based execution of clinical guidelines: a review, *International Journal of Medical Informatics*, 77:787-808, 2008.
- [51] Gerhard Weiss, *Multi Agent Systems 2nd edition*, Editor: Gerhard Weiss, MIT Press, USA, ISBN:978-0-262-01889-0, 2013.
- [52] Mago V.K., Devi M.S., A Multi-Agent Medical System For Indian Rural Infant And Child Care, *Ijcai'07 PROCEEDINGS OF THE 20th International Joint Conference On Artificial Intelligence*, San Francisco, Pp:1396-1401, 2007.
- [53] Bennett C.C., Hauser K., Artificial Intelligence Framework For Simulating Clinical Decision-Making: A Markov Decision PROCess Approach, *Artificial Intelligence in Medicine*, 57:9-19, 2013.
- [54] Bennett C.C., Doub T.W., Selove R., EHRs Coonect Research And Practice: Where Predictive Modeling, Artificial Intelligence, and Clinical Decision Support Intersect, *Health Policy and Technology*, 1:105-114, 2012.
- [55] Bennett C.C., Clinical Productivity System-A Decision Support Model, *International Journal Of Productivity and Performance Management*, 60(3):311-319, 2011.
- [56] Hillman R. S., Ault K. A., Leporrier M., Rinder H. M., *Hematology in Clinical Practice/Klinik uygulamada hematoloji*, 5.Baskı, ISBN: 978-975-277-404-9, Çeviri: SAKACI S., Çeviri Editörleri: Özcan M., Ören H., Arat M., Demir M., Soysal T., Özkalemkaş F., Sönmez M., Güneş Tıp Kitabevleri, Ankara, 2012.
- [57] Bunn F., Aster J.C., *Kan Hastalıklarının Patofizyolojisi (LANGE)*, ISBN: 978-605-4499-54-0, Çeviri: SAKACI S., Çeviri Editörleri: Soysal T., Ören H., Demir M., Haznedaroğlu İ., Özkalemkaş F., Bolaman Z., Sönmez M., İstanbul Tıp Kitabevi, İstanbul, 2013.

- [58] Patel N., Automation in Hematology: Here's The State Of The Art in 2014, www.mlo-online.com, Sayfa:6-8-11, 2014, <http://www.mlo-online.com/articles/201401/automation-in-hematology-heres-the-state-of-the-art-in-2014.php>, Erişim:12.05.2014.
- [59] Linder J., Zahniser D.J., Digital imaging in hematology: new approaches to the CBC and WBC differential, www.mlo-online.com, Sayfa:14-16-18, <http://www.mlo-online.com/articles/201205/digital-imaging-in-hematology.php>, Erişim:12.05.2014.
- [60] Zini G., Artificial Intelligence in Hematology, Hematology, 10(5):393-400, ISSN:1024-5332, 2005.
- [61] Lipkin M., Correlation of Data with a Digital Computer in the Differential Diagnosis of Hematological Diseases, IRE Transactions on Medical electronics, P.243-246, 1960.
- [62] Slonka J., Alrifai M., Bein G., Sachs Y.J., A Highly Specialised Self-Made Computer Program Enhances Efficiency And Safety Of Immunohaematology Reports, Official Journal Of The British Blood Transfusion Society Transfusion Medicine, 23:207-214, 2013.
- [63] Venkatalakshmi B., Thilagavathi K., Automatic Red Blood Cell Counting Using Hough Transform, IEEE Conference on Information and Communication Technologies (ICT2013), JeJu Island, India, p.267-271, ISBN:978-1-4673-5759-3, 2013.
- [64] Haun D.E., Foley A.B., Jarreau P.C., Development and feasibility of an electronic white blood cell identification trainer, Clin Lab Sol., 26(1), PMID: 23479833, 2013.
- [65] Wicht A., Wetter T., Klein U., A web-based system for clinical decision support and knowledge maintenance for deterioration monitoring of hematological patients, computer methods and programs in biomedicine, 111:26-32, 2013.
- [66] Maity M., Sakar P., Chakraborty C., Computer Assisted Approach to Anemic Erythrocyte Classification Using Blood Pathological Information, Emerging Applications of Information Technology (EAIT), India, pp.116-121, ISBN:978-1-4673-1827-3, 2012.
- [67] Rupp S., Schlarb T., Haßmeyer E., Zerfaß T., Fully automated detection of the counting area in blood smears for computer aided hematology, 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts USA, pp.7759-7762, 2011.

- [68] Mustafee N., Williams M.D., Hughes F., Katsaliaki K., Simulation Based Study of Hematology Outpatient Clinics With Focus on Model Reusability, PROceedings of the 2011 Winter Simulation Conference, pp.1178-1189, 2011.
- [69] Diamond L.W., Mishka V.G., Seal A.H., Nguyen D.T., Multiparameter Interpretative Reporting in Diagnostic Laboratory Hematology, Elsevier B.V., 1994.
- [70] Lee H., Chen Y.P., Cell morphology based classification for red cells in blood smear images, Pattern Recognition Letters 49, p 155-161, 2014.
- [71] www.thd.org.tr/thdData/userfiles/file/CEKIRDEK_Programi_Tanitim.pdf, Erişim tarihi:15/12/2014
- [72] www.thd.org.tr/thdData/userfiles/file/KUR_Programi_Tanitim.pdf, Erişim tarihi:15/12/2014
- [73] www.thd.org.tr/thdData/userfiles/file/RAPOR_Programi_Tanitim.pdf, Erişim tarihi:15/12/2014
- [74] Türk Hematoloji Derneği Eritrosit Hastalıkları ve Hemoglobin Bozuklukları Tanı ve Tedavi Kılavuzu, Version 1, Ankara, Temmuz 2011.
- [75] Quaglini S., Stefenalli M., Barosi G., Berzuini A., anaemia: an expert consultation system, Computers and Biomedical Research vol.19, Issue.1, p 13-27, 1986.
- [76] Quaglini S., Stefenalli M., Barosi G., Berzuini A., A performance evaluation of the expert system ANEMIA, Computers and Biomedical Research vol.21, Issue 4, p 307-323, 1988.
- [77] Birndorf N.I., Pentecost J.O., Coakley J.R., Spackman J.R., An Expert System to Diagnose Anemia and Report Results Directly on Hematology Forms, Computers And Biomedical Research, 29(2): 6–26, 1996.
- [78] Ortiz M.T., Jaslow R., Rothman M., Rudelli R., A Computerized Anemia Study Of Large Institutionalized Populations, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 12, No. 3, P.1264-1266, 1990.
- [79] Yeh J.S., Cheng C.H., Using Hierarchical soft computing method to discriminate micROCyte anemia, Expert Sysytems with applications 29,P515-524, 2005.

- [80] Doğan Ş., Türkoğlu İ., Iron-Deficiency Anemia Detection From Hematology Parameters By Using Decision Trees, *International Journal of Science & Technology*, 3(1):85-92, 2008.
- [81] Setsirichok D., Piroonratana T., Wongseree W., Usavanarong T., Paulkhaolarn N., Kanjanakorn C., Sirikong M., Limwongse C., Chaiyaratana N., Classification of complete blood count and haemoglobin typing data by a C4.5 decision tree, a naive Bayes classifier and a multilayer perceptron for thalassaemia screening, *Biomedical signal PROCessing and Control*, 7:202-212, 2012.
- [82] I. Yılmaz, Demir eksikliği anemisi ile beta talasemi minörün ayırıcı tanısında eritrosit indekslerin rolü, *Expertise Thesis, Dept. of Internal Medicine, Faculty of Medicine, Pamukkale University, Denizli, Turkey*, 2010.
- [83] E. Urrechaga, U. Aguirre and S. Izquierdo, Differential Diagnosis of MicROCyctic Anemia, *Anemia*, 2013. DOI: 10.1155/2013/457834.
- [84] Jamei M.K., Talarposhti K.M., Discrimination Between Iron Deficiency Anaemia (IDA) and β -Thalassemia Trait (β -TT) Based on Pattern-Based Input Selection Artificial Neural Network (PBIS-ANN), *Journal of Advances in Computer Research*, pISSN: 2345-606x, eISSN: 2345-6078, (Research Gate), 2016.
- [85] Barnhart-Magen G., Gotlib V., Marilus R., Einav Y., Differential Diagnostics of Thalassemia Minor by Artificial Neural Networks Model, *Journal of Clinical Laboratory Analysis*, 27:481-486, 2013. Doi:10.1002/jela.21631.
- [86] Panneerselvam, P., Application of embedded system for a genetic disease, sickle cell anemia, *Advances in Electrical Engineering (ICAEE)*, 2014 International Conference on, DOI: 10.1109/ICAEE.2014.6838446, Page(s): 1 - 4, 2014.
- [87] Yılmaz A., Dağlı M., Allahverdi N., A Fuzzy Expert System Design for Iron Deficiency Anemia, *IEEE 7th International Conference on Application of Information and Communication Technologies (AICT)*, Baku, Azerbaijan, Page(s): 1- 4, DOI: 10.1109/ICAICT.2013.6722707, 2013.
- [88] Wongseree W., Chaiyaratana N., Vichittumaros K., Winichagoon P., Fucharoen S., Thalassaemia classification by neural networks and genetic programming, *Information Sciences* 177, P. 771–786, 2007.

- [89] Amendolia S.R., Cossu G., Ganadu M.L., Golosio B., Masala G.L., Mura G.M., A comparative study of K-Nearest Neighbour, support Vector Machine and Multi-Layer Perceptron for Thalassemia screening, *Chemometrics and Intelligent Laboratory Systems* 69, P. 13-20, 2003.
- [90] Hirimutugoda Y.N., Wiyayarathna G., Artificial Intelligence-Based Approach for Determination of Haematologic Diseases, 2009.
- [91] Sanap S.A., Nagori M., Kshirsagar V., Classification of Anemia Using Data Mining Techniques, *SEMCCO 2011, Part II*, pp:113-121, Berlin Heidelberg 2011.
- [92] Allahverdi N., Tunali A., Işık H., Kahramanli H., A Takagi–Sugeno type neuro-fuzzy network for determining child anemia, *Expert Systems with Applications* 38 P. 7415–7418, 2011.
- [93] Ahmed F., Sitara A., Exploration Of Co-Relation Between Depression And Anemia In Pregnant Women Using Knowledge Discovery And Data Mining Algorithms And Tools, *Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Health Informatics*, September edition, 2012.
- [94] Solanki A.V., Data Mining Techniques Using WEKA Classification for Sick Cell Disease, *International Journal of Computer Science and Information Technologies*, Vol.5(4), pp:5857-5860, 2014.
- [95] Kishore R., Rao K.P., Murthy G.R.S., Performance Evaluation of Entorpy and Gini using Threaded and Non Threaded ID3 on Anaemia Dataset, 2015 Fifth International Conference on Communication Systems and Network Technologies, IEEE, 2015.
- [96] Amin N., Habib A., Comparison of Different Classification Techniques Using WEKA for Hematological Data, *American Journal of Engineering Research*, Vol.4, Issue.3, pp:55-61, e-ISSN: 2320-0847, 2015.
- [97] Kimura K., Tabe Y., Ai T., Takehara I., Fukuda H., Takahashi H., Naito T., Komatsu N., Uchihashi K., Ohsaka A., A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA, *Scientific Reports*, 9, 1-9, DOI: <https://doi.org/10.1038/s41598-019-49942-z> (2019).
- [98] Alsheref F. K., Gomaa W. H., Blood Diseases Detection using Classical Machine Learning Algorithms, *International Journal of Advanced Computer Science and Applications*, 10:7, 77-81, (2019).

- [99] K. Meena, D. K. Tayal, V. Gupta, A. Fatima, Using Classification Techniques for Statistical Analysis of Anemia, *Artificial Intelligence in Medicine*, Vol. 94, pp.138-152, <https://doi.org/10.1016/j.artmed.2019.02.005>, 2019.
- [100] Çil B., Ayyıldız H., Tuncer T., Discrimination of β -thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine-based decision support system, *Medical Hypotheses*, 138, 1-6, DOI: <https://doi.org/10.1016/j.mehy.2020.109611>, 2020.
- [101] Varghese N., Machine Learning Techniques for the Classification of Blood Cells and Prediction of Diseases, *International Journal of Computer Science Engineering*, 9:1, 66-75, 2020.
- [102] Alzubaidi L., Fadhel M. A., Al-Shamma O., Zhang J., Duan Y., Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sick Cell Anemia Diagnosis, *Electronics*, 9:427, 1-20, DOI:10.3390/electronics9030427 2020.
- [103] Shaik M. F., Subashini M., Anemia Diagnosis by Fuzzy Logic Using LabVIEW, Presented at IEEE International Conference on Intelligent Computing and Control (I2C2), DOI: 10.1109/I2C2.2017.8321790, 2017.
- [104] Dalvi P. T., Vernekar N., Anemia Detection using Ensemble Learning Techniques and Statistical Models, Presented at IEEE International Conference on Recent Trends in Electronics Information Communication Technology, May 20-21, 2016.
- [105] Belginova S., Uvaliyeva I., Ismukhamedova A., Decision Support System for Diagnosing Anemia, Presented at 4th International Conference on Computer and Technology Applications, DOI: 10.1109/CATA.2018.8398684, 2018.
- [106] Dimauro G., Caivano D., Girardi F., A New Method and a Non-Invasive Device to Estimate Anemia Based on Digital Images of the Conjunctiva, *IEEE Access-Special Section on Human-Centered Smart Systems and Technologies*, DOI: 10.1109/ACCESS.2018.2867110, 2018.
- [107] Ludwig H., Belle S.V., Barrett-Lee P., ve diğerleri, The european Cancer Anaemia Survey (ECAS): A large, multinational, prospective survey defining the prevalence, incidence and treatment of anaemia in cancer patients, *European Journal of Cancer* 40, p. 2293-2306, 2004.
- [108] Repetto L., Moeremans K., Annemans L., European guidelines for the management of chemotherapy-induced anaemia and health economic aspects of treatment, *Cancer Treatment Reviews* 32, p.5-9, 2006.

- [109] Aapro M., Van Erps J., MacDonald K., Soubeyran P., Turner M., Muenzberg M., Dunlop R., Warrinnier H., Abrahamc I., Promoting evidence-based management of anemia in cancer patients: Background, development, and scientific validation of RESPOND, a web-based clinical guidance system based on the EORTC guidelines, *Critical Reviews in Oncology/Hematology* 65 P.32–42, 2008.
- [110] Bokemeyer C., Aapro M.S, Courdi A., Foubert J., Link H., Österborg A., Repetto L., Soubeyran P., EORTC guidelines for the use of erythropoietic proteins in anaemic patients with cancer: 2006 update, *European Journal of Cancer* 43, P258-27, 2007.
- [111] Ostrovsky E., Zelig U., Gusakova I, Ariad S., Mordechai S., Nisky I., Kapilushnik J., Detection of Cancer Using Advanced Computerized Analysis of Infrared Spectra of Peripheral Blood, *IEEE Transactions On Biomedical Engineering*, Vol. 60, No. 2, February 2013.
- [112] Gaweda A.E., Muezzinoglu M.K., Aronoff G.R., Jacobs A.A., Zurada J.M., Brier M.E., Using Clinical Information in Goal-Oriented Learning, *IEEE Engineering in medicine and biology magazine*, 0739-5175/07/2007 pp. 27-36, 2007.
- [113] Chait Y., Horowitz J., Nichols B., Shrestha R.P., Hollof C.V., Germain M.J., Control-Relevant Erythropoiesis Modeling in End-Stage Renal Disease, *IEEE Transactions on Biomedical Engineering*, ISSN:0018-9294, 61(3):658-664, 2014.
- [114] Guerreroa J.D.M., Olivasa E. S., Vallsa G. C., L'opeza A. J. S., Ruixob J. J. P., TorresN. V. J., Use of neural networks for dosage individualisation of erythropoietin in patients with secondary anemia to chronic renal failure, *Computers in Biology and Medicine* 33, P. 361–373, 2003.
- [115] Montero P.E., Chermisi M., Martínez J. M. M., Gómez-Sanchis J., Barbieri C., Olivas E. S., Mari F., Vila-Francés J., Stopper A., Gatti E., Martín-Guerrero J. D., Optimization of anemia treatment in hemodialysis patients via reinforcement learning, *Artificial Intelligence in Medicine* 62, P. 47–60, 2014.
- [116] Martínez J. M. M., Montero P.E., Barbieri C., Olivas E. S., Mari F., Martínez-Sobera M., Amato C., López A. J. S., Bassi M., Magdalena-Benedito R., Stopper A., Martín-Guerrero J. D., Gatti E., Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques, *Computer Methods and Programs in Biomedicine* 117, P.208–217, 2014.

- [117] Akabua E., Inanc T., Gaweda A., Brier M. E., Kim S, Zurada J. M., Individualized model discovery: The case of anemia patients, Computer Methods and Programs in Biomedicine 118, P. 23–33, 2015.
- [118] Çarklı Yavuz B., Karagül Yıldız T., Yurtay N., Yılmaz Z., Comparison of K Nearest Neighbours and Regression Tree Classifiers Used With Clonal Selection Algorithm to Diagnose Haematological Diseases,” AJIT-e: Online Academic Journal of Information Technology, vol. 5, no.16, DOI: 10.5824/1309-1581.2014.3.001.x, 2014.
- [119] Uzoka F.M.E., Osuji J., Obot O., Clinical decision support system (DSS) in the diagnosis of malaria: A case comparison of two soft computing methodologies, Expert Systems with Applications, Vol.38, pp:1537–1553, 2011.
- [120] Arslan A.K., Colak C., Sarihan M.E., Different Medical Data Mining Approaches Based Prediction Of İschemic Stroke, Computer Methods And Programs in Medicine, Vol: 130, pp: 87-92, 2016.
- [121] Karagül Yıldız T., Demirci H., Yurtay N., Effect of the clonal selection algorithm on classifiers, Artificial Intelligence and Applied Mathematics in Engineering Problems, DOI:10.1007/978-3-030-36178-5_84, 2020.
- [122] <http://www.hematoloji.org.tr/content.php?gid=XX> Erişim tarihi:15.12.2014.
- [123] Ünüvar A., Hematolojik Hastalıkların Tanısında Yeni Laboratuvar Yöntemleri, XXX. Ulusal Hematoloji Kongresi, III. Hematoloji İlk Basamak Kursu, İstanbul, 10-14 Ekim 2003.
- [124] Ayaz S., Türkiye Yüksek İhtisas Hastanesi, Hematoloji Laboratuvarı, Laboratuvar Test Rehberi. www.tyih.gov.tr/Eklenti/319,testrehberihematolojipdf.pdf?0, Erişim tarihi:15.12.2014.
- [125] Şencan M., Düğenci N., Hematoloji Laboratuvarı Test Rehberi, 2012. hastaneler.cumhuriyet.edu.tr/hematolojirehber.pdf, Erişim tarihi:15.12.2014.
- [126] Hematoloji Laboratuvarı Kılavuzu 1, II. Bölüm: Tam Kan Sayımı, Türk Hematoloji Derneği, 2014.
- [127] Klinik Uygulamada Hematoloji, Kısım 1: Eritrosit Bozuklukları, Bölüm 2. Anemiye Klinik Yaklaşım, 5.baskı, ISBN: 978-975-277-404-9, Güneş Tıp Kitabevi, 2012.

- [128] Hoffman R., Jr. Benz E.J., Silberstein L.E., Heslop H.E, Weitz J.I., Anastasi J., Hematology: Basic Principles and Practice, 6th edition, ISBN: 978-1-4377-2928-3, Elsevier, 2013.
- [129] Klinik Hematoloji, Ankara Ünv. Tıp Fakültesi Hematoloji Bilim Dalı, ISBN:975-7226-35-1, ANTIP A.Ş. Yayınları, 1997.
- [130] Anemi Ders Notları, Dr. Birgül ÖNEÇ, 2014.
- [131] Fawcett T., An Introduction to ROC Analysis, Pattern Recognition Letters, vol 27 pp.861-874, 2006.
- [132] C. E. SHANNON, A Mathematical Theory of Communication, Reprinted with corrections from The Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, July, October 1948.
- [133] MATLAB User's Guide, MATLAB R2017a, Natick, MA, USA, The MathWorks Inc., 2020.
- [134] M. A. Hall, Correlation-based feature subset selection for machine learning, Doctor of Philosophy thesis in the University of Waikato, Hamilton, New Zealand, 1999.
- [135] <https://www.mathworks.com/help/matlab/>, Erişim: 08.01.2021
- [136] Öztemel E., Artificial Neural Networks, Papatya Press, Istanbul, Turkey, ISBN: 975-67-97-39-8, 2006.
- [137] Quinlan J. R., Induction of Decision Trees, Machine Learning, Vol.1, pp: 81-106, 1986.
- [138] Wu, X., Kumar, V., Quinlan, JR, Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z.H., Top 10 algorithms in data mining, Knowledge and information systems, 14 (1), 1-37, 2008.
- [139] Hand D. J., Till R. J., A simple generalization of the area under the ROC curve for multiple class classification problems, Machine Learning, Vol.45 pp:171-186, 2001.
- [140] Wandishin M. S., Mullen S.J., Multiclass ROC analysis, Weather and Forecasting Vol.24, DOI: 10.1175 / 2008WAF2222119.1, 2008.

- [141] Ramzan F., Khan M. U. G., Rehmat A., Iqbal S., Saba T., Rehman A., Mahmood Z., A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks, Journal of Medical Systems vol.44(37), 2020.
- [142] International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)-WHO Version for 2019, <https://icd.who.int/browse10/2019/en> Erişim tarihi: 11.07.2020.
- [143] <https://www.icd10data.com/ICD10CM/Codes/D50-D89>, Erişim tarihi: 10.01.2021
- [144] http://www.kanhastaliklari.org.tr/icerik.php?id=127&alt_id=196&tab=126, Erişim tarihi: 11.07.2020.
- [145] <https://www.mayoclinic.org/diseases-conditions/iron-deficiency-anemia/symptoms-causes/syc-20355034>, Erişim tarihi: 11.07.2020.
- [146] <https://www.mayoclinic.org/diseases-conditions/vitamin-deficiency-anemia/symptoms-causes/syc-20355025>, Erişim tarihi: 11.07.2020.
- [147] http://www.thd.org.tr/thdData/userfiles/file/9_MSEK_09.pdf, Erişim tarihi: 11.07.2020.
- [148] <http://thd.org.tr/thdData/userfiles/file/Eritrosit-Tani-ve-tedavi-Kilavuzu-2019.pdf>, Erişim tarihi: 11.07.2020.
- [149] Karagül Yıldız T., Yurtay N., Öneç B., Classifyin anemia types using artificial learning methods, Engineering Science and Technology, an International Journal, Vol.24 (1), 2021.

**EK 1: Düzce Üniversitesi Tıp Fakültesi İnvaziv Olmayan Klinik Araştırmalar
Etik Kurulu Karar Formu**

*; *Engelhardtia* *Bursera*

ASL GIBBIE

Etik Kurul Başkanının
Unvanı/Adı/Soyadı: Prof. Dr. Ahmet ATAÖĞÜ
İmza:

Note: Enk' has no diphthongs; however, we have written /e/ as /ai/ or /ei/ where appropriate.

İNVAZİF OLMAYAN KLİNİK ARAŞTIRMALAR ETİK KURULU KARAR FORMU

ARAŞTIRMANIN AÇIK ADI	Hematolojik Hastalıkların Tanısında Sergisel Yaklaşımın B- Yöntem Tasarımı, KlonDoktor.
VARSA ARAŞTIRMANIN PROTOKOL KODU	

ETİK KURUL BİLGİLERİ	ETİK KURULUN ADI	Düzce Üniversitesi Tıp Fakültesi İnvaziv Olmayan Klinik Araştırmalar Etik Kurulu
	AÇIK ADRESİ:	Düzce Üniversitesi Tıp Fak. Morfoloji Binası 4. Kat Konuralp-Düzce
	TELEFON	0380 542 14 16
	FAKS	0380 542 13 02
	E-POSTA	duzceetik@duzce.edu.tr

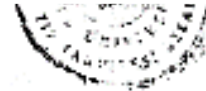
BAŞVURU BİLGİLERİ	KOORDİNATÖR SORUMLU ARAŞTIRMACI UNVANI ADI SOYADI	Yrd. Doç. Dr. Nilüfer YURTAY		
	KOORDİNATÖR SORUMLU ARAŞTIRMACININ UZMANLIK ALANI	Bilgisayar ve Bilişim Bilimleri		
	KOORDİNATÖR SORUMLU ARAŞTIRMACININ BULUNDUĞU MERKEZ	Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi		
	VARSA İDARI SORUMLU UNVANI ADI SOYADI			
	DESTEKLEYİCİ			
	PROJE YÜRÜTÜCÜSÜ UNVANI ADI SOYADI (TUBİTAK vb. gibi kaynaklardan destek alanlar için)			
	DESTEKLEYİCİNİN YASAL TEMSİLCİSİ			
	ARAŞTIRMANIN FAZİ VE TÜRÜ	FAZ 1	☐	
		FAZ 2	☐	
		FAZ 3	☐	
		FAZ 4	☐	
		Gözlemsel ilaç çalışması	☐	
		Tıbbi cihaz klinik araştırması	☐	
		İn vitro tıbbi tanı cihazları ile yapılan performans değerlendirme çalışmaları	☐	
		İlaç dışı klinik araştırma	☐	
	Diğer ise belirtiniz Bilgi-işlem Verilerinin Analizi			
		TEK MERKEZ ☒	ÇOK MERKEZLİ ☐	ULUSAL ☐
		ULUSLARARASI ☐		

DEĞERLEN DİRİLEN BELGİLER	Belge Adı	Taribi	Versiyon Numarası	Dil		
	ARAŞTIRMA PROTOKOLÜ			Türkçe ☐	İngilizce ☐	Diğer ☐
	III GİLENDİRİLMİŞ GÖNCÜLLÜ OLUR FORMU			Türkçe ☐	İngilizce ☐	Diğer ☐

Etik Kurul Başkanının
Unvanı/Adı/Soyadı: Prof. Dr. Ahmet ATAÖĞLU
İmza:

ASLI GİBİ

Not: Etik Kurul Başkanı, imzasının her alındığı her sayfaya imza atmalıdır.



ÖZGEÇMİŞ

Tuba KARAGÜL YILDIZ Düzce Üniversitesi'nde Araştırma Görevlisi olarak görev yapmaktadır. Lisans eğitimlerini Sakarya Üniversitesi Bilgisayar Sistemleri Eğitimi Bölümü'nde ve Düzce Üniversitesi Bilgisayar Mühendisliği Bölümü'nde tamamladı. Düzce Üniversitesi Elektrik-Elektronik ve Bilgisayar Mühendisliği alanında yüksek lisans eğitimini tamamladı. Millî Eğitim Bakanlığı'nda Bilişim Teknolojileri Öğretmeni olarak görev yaptı. Çalışma alanları yapay öğrenme, veri madenciliği, tıp bilişimi ve optimizasyon yöntemleridir.