T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

YAPAY ZEKA YÖNTEMLERİ KULLANILARAK ST SEGMENT YÜKSELMELİ MİYOKART İNFARKTÜSLÜ HASTALARIN ÖLÜM RİSK ANALİZİ VE TAHMİNİ

Bahar ÖZYILMAZ

Ağustos-2024
BATMAN

T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

YAPAY ZEKA YÖNTEMLERİ KULLANILARAK ST SEGMENT YÜKSELMELİ MİYOKART İNFARKTÜSLÜ HASTALARIN ÖLÜM RİSK ANALİZİ VE TAHMİNİ

Bahar ÖZYILMAZ

Danışman
Dr. Öğr. Üyesi Abdulkerim ÖZTEKİN

Ağustos-2024
BATMAN

TEZ KABUL VE ONAYI

Bahar ÖZYILMAZ tarafından hazırlanan "Yapay zeka yöntemleri kullanılarak ST segment yükselmeli miyokart infarktüslü hastaların ölüm risk analizi ve tahmini" adlı tez çalışması 28/08/2024 tarihinde aşağıdaki jüri üyeleri tarafından oy birliği ile Batman Üniversitesi Lisansüstü Eğitim Enstitüsü Elektrik Elektronik Mühendisliği Ana Bilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Doç. Dr. Melih KUNCAN	……..
Danışman Dr. Öğr. Üyesi Abdulkerim ÖZTEKİN	……..
Üye Doç. Dr. Emrullah ACAR	……..

Yukarıdaki sonucu onaylarım.

Lisansüstü Eğitim Enstitüsü Müdürü
Dr. Öğr. Üyesi Ömer Murat ÖTER

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Bahar ÖZYILMAZ
28/08/2024

ÖZET

YÜKSEK LİSANS TEZİ

YAPAY ZEKA YÖNTEMLERİ KULLANILARAK ST SEGMENT YÜKSELMELİ MİYOKART İNFARKTÜSLÜ HASTALARIN ÖLÜM RİSK ANALİZİ VE TAHMİNİ

Bahar ÖZYILMAZ

Batman Üniversitesi Lisansüstü Eğitim Enstitüsü
Elektrik Elektronik Mühendisliği Ana Bilim Dalı

Danışman: Dr. Öğr. Üyesi Abdulkerim ÖZTEKİN

2024, 57 Sayfa

Jüri
Dr. Öğr. Üyesi Abdulkerim ÖZTEKİN
Doç. Dr. Melih KUNCAN
Doç. Dr. Emrullah ACAR

Kalp krizi, yani akut miyokard infarktüsü (Acute Myocardial Infarction – AMI), kalbi besleyen damarların zamanla daralıp tıkanmasıyla kalp kasının bir kısmının yeteri kadar kan alamaması durumudur. Bu durumun geç fark edilmesi ve hemen müdahale edilememesi birey için ölüm ve bazı kalıcı hasarlara neden olabilmektedir. İstatistiklere bakıldığında kardiyovasküler hastalıkların neden olduğu ölümler her geçen gün daha da artmaktadır. Akut miyokart infarktüsünün en ciddi ve ölümcül türlerinden biri olan ST yükselmeli akut miyokart infarktüsü (ST-segment Elevation Acute Myocardial Infarction - STEMI) acil tanı ve müdahale gerektirir. Sağlıkta kullanılan yapay zekâ tabanlı uygulamalar yaygınlaşarak erken teşhis ve tedavinin önünü açmıştır. Modern tıpta, STEMI hastalarının doğru ve hızlı bir şekilde tanımlanarak tedavi edilmesi hayati bir öneme sahiptir. Bu süreç içerinde hastaların ölüm riskini önceden belirlemek, klinik kararların verilmesinde çok büyük rol oynamaktadır. Geleneksel risk değerlendirme yöntemleri genellikle zaman alıcı ve subjektif süreçler olup, klinik verilerin manuel olarak analiz edilmesine dayanmaktadır.

Bu çalışmada, yapay zekâ yöntemlerinden makine öğrenmesi algoritmaları kullanılarak hastalarının ölüm risk analizi ve tahminlemesi yapılmıştır. Kullanılan veri seti üzerinden kalp krizi geçirenlerin parametrelerinin neler olduğu, bu parametrelerin veri seti üzerindeki dağılımı ve birbirleriyle ilişkili olduğu durumlar farklı analiz yöntemleriyle incelenip bireylerde ölüme sebebiyet açısından ne derecede etkili olduğu belirlenmiştir. Bu tezin amacı, kalp krizi geçirmiş bireylerde mortalite riskinin yapay zekâ tabanlı algoritmalar yardımıyla önceden tahmin edilmesini kolaylaştırarak hastaların erken tedavi edilmelerini sağlayacak faydalı bir hekim karar destek modeli geliştirmektir. Çalışmanın, bu yönüyle STEMI hastalarının yönetiminde klinik karar desteği sağlaması ve sağlık hizmetlerinin kalitesini artırmaya yönelik katkılar sunması beklenmektedir.

Anahtar Kelimeler: Kalp hastalığı, Miyokard infarktüsü, Mortalite/morbidite teşhisi, Yapay zeka

ABSTRACT

MS THESIS

DEATH RISK ANALYSIS AND PREDICTION OF PATIENTS WITH ST-SEGMENT ELEVATED MYOCARDIAL INFARCTION USING ARTIFICIAL INTELLIGENCE METHODS

Bahar ÖZYILMAZ

Batman University Graduate Education Institute

Electrical and Electronics Engineering Department of Science

Advisor: Asst. Prof. Dr. Abdulkerim ÖZTEKİN

2024, 57 Pages

Jury
Asst. Prof. Dr. Abdulkerim ÖZTEKİN
Doç. Dr. Melih KUNCAN
Doç. Dr. Emrullah ACAR

Heart attack, acute myocardial infarction (AMI), is a condition in which a part of the heart muscle cannot receive enough blood due to the narrowing and blockage of the vessels feeding the heart over time. Noticing this situation late and failing to intervene immediately may cause death and some permanent damage for the individual. Looking at the statistics, deaths caused by cardiovascular diseases are increasing day by day. ST-segment elevation acute myocardial infarction (ST-segment Elevation Acute Myocardial Infarction - STEMI), one of the most serious and fatal types of acute myocardial infarction, requires urgent diagnosis and intervention. Artificial intelligence-based applications used in health have become widespread, paving the way for early diagnosis and treatment. In modern medicine, it is vital that STEMI patients are identified and treated accurately and quickly. In this process, determining the risk of death of patients in advance plays a major role in making clinical decisions. Traditional risk assessment methods are often time-consuming and subjective processes and rely on manual analysis of clinical data.

In this study, death risk analysis and prediction of patients will be made using machine learning algorithms, which is one of the artificial intelligence methods. Based on the data set used, the parameters of those who had a heart attack, the distribution of these parameters on the data set and the situations in which they are related to each other will be examined with different analysis methods and it will be determined to what extent they are effective in causing death in individuals. The aim of this thesis is to develop a useful physician decision support model that will facilitate early treatment of patients by making it easier to predict the risk of mortality in individuals who have had a heart attack with the help of artificial intelligence-based algorithms. In this respect, the study is expected to provide clinical decision support in the management of STEMI patients and contribute to improving the quality of healthcare services.

Keywords: Mortality/morbidity diagnosis, Heart disease, Myocardial infarction, Artificial intelligence

ÖNSÖZ

Çalışmalarım boyunca bilgisini benden esirgemeyen, her konuda tüm bilgi birikimini paylaşan, kıymetli vaktini ayırıp yardım ve katkılarıyla beni yönlendiren, her alanda destek olan ve mesleki ilgimi sürdürmeme yardımcı olup, zirvede tutmamı sağlayan değerli hocam Sayın Dr. Öğr. Üyesi Abdulkerim ÖZTEKİN 'e canı gönülden teşekkürlerimi sunarım. Ayrıca yüksek lisans eğitimim süresince beni her zaman destekleyen ve yanımda olan aileme teşekkür ederim.

Bahar ÖZYILMAZ
BATMAN-2024

İÇİNDEKİLER

ÖZET ..... **iv**

ABSTRACT ..... **v**

ÖNSÖZ ..... **vi**

İÇİNDEKİLER ..... **vii**

KISALTMALAR ..... **ix**

ŞEKİLLER DİZİNİ ..... **xi**

TABLOLAR DİZİNİ ..... **xii**

1. GİRİŞ **1**

2. KAYNAK ARAŞTIRMASI **4**

3. GENEL BİLGİLER **8**

3.1. Koroner Arter Hastalığı 8
3.2. Miyokard İnfarktüsü 9
3.3. ST-Segment Yükselmeli Akut Miyokard İnfarktüsü 10

4. MATERYAL VE YÖNTEM **14**

4.1. Veri Setinin Oluşturulması 14
4.2. Makine Öğrenmesi Yöntemleri 18
4.2.1. Yapay sinir ağları 18
4.2.2. k-en yakın komşuluk 21
4.2.3. Destek vektör makineleri 23
4.2.4. Karar ağacı 24
4.2.5. Rassal orman 25
4.2.6. Lojistik model ağacı 26
4.3. Algoritmaların Performans Ölçütleri 27

5. ARAŞTIRMA BULGULARI VE TARTIŞMA 31

5.1. Makine Öğrenmesi Yöntemlerinin Orjinal Veri Setine Uygulanması..... 32

5.2. Makine Öğrenmesi Yöntemlerinin Arttırılmış Veri Setine Uygulanması 36

5.3. Makine Öğrenmesi Yöntemlerinin Artırılmış Veri Setinde Önemli Görülen Parametrelere Uygulanması 41

6. SONUÇ 47

KAYNAKLAR 51

ÖZGEÇMİŞ 57

KISALTMALAR

AHA	: Amerikan Kalp Cemiyeti
AKS	: Akut Koroner Sendrom
AUC	: Alan Altı Eğri
AV	: Atriyoventriküler
CABGO	: Koroner Bypass Cerrahisi
CX	: Sirkümfleks Arter
DM	: Diabetes Mellitus (Şeker Hastalığı)
EF	: Ejeksiyon Fraksiyonu
EKG	: Elektrokardiyogram
EL	: Ensemble Learning (Topluluk Öğrenmesi)
ESC	: Avrupa Kardiyoloji Derneği
GLM	: Genel Doğrusal Modeller
HB	: Hemoglobin
HDL	: Yüksek Dansiteli Lipoprotein
HPL	: Hiperlipidemi
İKH	: İskemik Kalp Hastalığı
KAH	: Koroner Arter Hastalığı
KY	: Kalp yetmezliği
KBY	: Kronik Böbrek Yetersizliği
k-NN	: k- Nearest Neighbourhood (k-En Yakın Komşuluk)
LAD	: Sol Koroner Arter
LDL	: Düsük Dansiteli Lipoprotein
LMCA	: Sol Ana Koroner Arter
LMT	: Logistic Model Tree (Lojistik Model Ağacı)
LVEF	: Left Ventricular Ejection Fraction
MI	: Miyokard İnfarktüsü
ML	: Machine Learning (Makine Öğrenmesi)
MLP	: Multilayer Perceptron
NSTEMI	: ST Segment Yükselmesiz MI
PAH	: Periferik Arter Hastalığı
PCI	: Koroner Anjıoplasti
PKG	: Perkütan Koroner Girişim
PPKG	: Primer Peruktan Koroner Girişim
RCA	: Sağ Koroner Arter
RF	: Random Forest (Rastgele Orman)
ROC	: Alıcı Çalışma Karakteristik Eğrisi
RT	: Random Tree (Karar Ağacı)
SMOTE	: Sentetik Azınlık Aşırı Örnekleme Tekniği
STEMI	: ST Segment Yükselmeli MI
SVM	: Destek Vektör Makinaları
SVO	: Selebro Vasküler Hastalık
TG	: Trigliserid
TÜİK	: Türkiye İstatistik Kurumu
VF	: Ventriküler Fibrilasyon
VT	: Ventriküler Taşikardi
WBC	: Beyaz Kan Hücreleri
WHO	: World Health Organization (Dünya Sağlık Örğütü)

WEKA : Waikato Environment for Knowledge Analysis
YSA : Yapay Sinir Ağları

ŞEKİLLER DİZİNİ

Şekil 3.1. Plak birikimi 11
Şekil 3.2. Amerikan Kalp Cemiyetinin (AHA) vasküler lezyon sınıflamasına göre plak progresyonu şekilde gösterilmiştir (Stary, H. C., vd., 1995). 12
Şekil 3.3. Dünya çapında önde gelen ölüm nedenleri 2019 verileri (WHO, 2019)..... 13
Şekil 4.1.YSA tahmin yöntemi 19
Şekil 4.2. Yapay sinir hücresi (Öztürk, K., & Şahin, M. E., 2018).**Hata! Yer işareti tanımlanmamış.**
Şekil 4.3. K-En yakın komşu algoritma görüntüsü 22
Şekil 4.4. SVM ile lineer veri sınıflandırması (Kavzoğlu, T., ve Çölkesen, İ., 2010) ... 24
Şekil 4.5. Öğrenilmiş ağaç örneği (Aydemir, H., 2020)..... 25
Şekil 4.6. Random forest (TseKiChun, 2021) 26

TABLOLAR DİZİNİ

Tablo 4.1. Veri setindeki bazı parametrelerin öznitelik bilgileri 14
Tablo 4.2. Hastaların klinik ve fonksiyonel özellikleri 16
Tablo 4.3. Karışıklık Matrisi Örneği 27
Tablo 5.1. Orijinal Veri Seti 31
Tablo 5.2. MLP algoritması sınıflandırma sonucu 32
Tablo 5.3. k-NN algoritması sınıflandırma sonucu 33
Tablo 5.4. SVM algoritması sınıflandırma sonucu 33
Tablo 5.5. Random tree algoritması sınıflandırma sonucu 34
Tablo 5.6. Random forest algoritması sınıflandırma sonucu 34
Tablo 5.7. LMT algoritması sınıflandırma sonucu 35
Tablo 5.8. Algoritmaların tahminleme sonuçlarının özeti 36
Tablo 5.9. MLP algoritması sınıflandırma sonucu 37
Tablo 5.10. k-NN algoritması sınıflandırma sonucu 38
Tablo 5.11. SVM algoritması sınıflandırma sonucu 38
Tablo 5.12. Random tree algoritması sınıflandırma sonucu 39
Tablo 5.13. Random forest algoritması sınıflandırma sonucu 39
Tablo 5.14. LMT algoritması sınıflandırma sonucu 40
Tablo 5.15. Artırılımış veride algoritmaların tahminleme sonuçlarının özeti 41
Tablo 5.17. k-NN algoritması sınıflandırma sonucu 42
Tablo 5.18. SVM algoritması sınıflandırma sonucu 43
Tablo 5.19. Random tree algoritması sınıflandırma sonucu 43
Tablo 5.20. Random forest algoritması sınıflandırma sonucu 44
Tablo 5.21. LMT algoritması sınıflandırma sonucu 44
Tablo 5.22. Algoritmaların veri setindeki 5 önemli paramatreye uygulanmasından elde edilen tahmin sonuçlarının özeti 46
Tablo 6.1. Elde edilen sonuçların literatüdeki sonuçlarla karşılaştırılması 48

1. GİRİŞ

Dünya genelinde en yaygın ölüm nedenlerinden biri olan kardiyovasküler hastalıklar ciddi halk sağlığı sorunlarına yol açan açmaktadır. Bu hastalıkların en kritik formlarından biri olan Akut Miyokard İnfarktüsü (Acute Myocardial Infarction – AMI) kalp krizinin tıbbi adıdır. Kalp kasına kan akışı birden durduğunda yaşamı tehdit eden komplikasyonlar ortaya çıkabilir ve genellikle yüksek mortalite (ölüm) ve morbidite (hastalık) oranları ile ilişkilidir. Bu durum genellikle kalpteki koroner arter adı verilen bir veya daha fazla arterdeki tıkanmadan kaynaklanır. Genellikle yağ, kolesterol ve hücresel atık ürünlerinden oluşan maddelerin plak biriktirmesi nedeniyle bir tıkanıklık meydana gelebilir.

Miyokard infarktüsü, ST yükselmeli akut miyokart infarktüsü (ST-segment elevation acute myocardial infarction - STEMI) ve ST yükselmesiz akut miyokart infarktüsü (Non-ST segment elevation acute myocardial infarction - NSTEMI) olarak iki başlık altında incelenir. Akut miyokart infarktüsünün en ciddi türlerinden biri olan ST yükselmeli akut miyokart infarktüsü acil müdahale gerektirir. Majör koroner arterlerden birinin tıkanması ciddi sağlık sorunlarına STEMI'ye yol açar. Koroner tıkanma ve koroner kan akışındaki azalma, sıklıkla aterosklerotik bir plağın parçalanması sonucu meydana gelen trombosit agregasyonu nedeniyle gelişen tıkayıcı trombüs oluşmasına bağlıdır. STEMI, tanı ve tedavisinin hemen başlanması gereken bir akut koroner sendromdur. Akut koroner sendromlar arasında en tehlikeli olanı ST-segment yükselmeli miyokart infarktüsüdür (Paralta De Figueiredo, M., vd., 2023).

İskemik kalp hastalığı (dokuya giden kan akışının azalması ve oksijen ihtiyacının karşılanamaması), dünya genelindeki ölümlerin en sık nedenidir ve aynı zamanda sıklığı giderek artmaktadır. Dünya Sağlık Örgütü'nün (DSÖ) 2017 yılı "Dünya' da İlk 10 Ölüm Nedenleri Raporu"na göre kardiyovasküler hastalıklar ilk sırada yer almaktadır ve 2015 yılında 8.756.000 birey iskemik kalp hastalıkları sebebiyle hayatını kaybetmiştir (WHO, 2018). Avrupa Kardiyoloji Derneği (ESC) 2017 verilerine göre de kardiyovasküler hastalıklar ESC üyesi ülkelerinde birinci ölüm sebebi olup, bu hastalıklardan dolayı her yıl ortalama 3.800.000 birey hayatını kaybetmiştir (Timmis, A., vd., 2017). Türkiye İstatistik Kurumu'nun (TÜİK) ölüm nedeni istatistiklerine göre ise Türkiye' de ölüme sebep olan ilk altı hastalık grubundan, dolaşım sistemi hastalıkları birinci sırada yer alıp, 2016 yılında 162.876 kişi bu hastalıklardan dolayı hayatını kaybetmiştir (TÜİK, 2015).

DSÖ 2030 yılı tahminlerine göre Dünya genelinde ölüm nedenleri sıralamasında iskemik kalp hastalıklarının ilk sırada yer alması öngörülmektedir (WHO, 2015-2030).

Modern tıpta, STEMI hastalarının doğru ve hızlı bir şekilde tanımlanarak tedavi edilmesi hayati bir öneme sahiptir. Bu süreç içerinde hastaların ölüm riskini önceden belirlemek, klinik kararların verilmesinde çok büyük rol oynamaktadır. Geleneksel risk değerlendirme yöntemleri genellikle zaman alıcı ve subjektif süreçler olup, klinik verilerin manuel olarak analiz edilmesine dayanmaktadır. Son zamanlarda, yapay zekâ (Artifical Intelligence – AI) ve makine öğrenmesi (Machine Learning - ML) tekniklerinin tıbbi alanda kullanımı ve yaygınlaşması, klinik verilerin analiz edilmesinde devrim niteliğinde gelişmelere yol açmıştır. Bu tekniklerin büyük veri setlerini işleyebilme ve karmaşık örüntüleri tespit edebilme yetenekleri, hastalıkların tanı ve prognozunda önemli bir araç haline gelmiştir. Özellikle, STEMI hastalarının mortalite riskinin tahmin edilmesinde AI tabanllı yöntemlerinin kullanılması, klinik sonuçların iyileştirilmesi ve sağlık hizmetlerinin etkinliğinin artırılması açısından büyük bir potansiyele sahiptir.

Günümüzde teknolojinin geçmişe kıyasla büyük bir hızla gelişmesiyle, yapay zekâ sistemlerinin kullanım alanları genişlemiş ve insanların işlerini kolaylaştıran bir özellik haline gelmiştir. Yapay zekâ, makinelerin insan zekasını veya davranışını taklit etmesine, oluşturmasına ve göstermesine olanak tanıyan bir algoritma, kod veya teknolojidir. AI, tıpkı insanlar gibi seçilmiş görevleri yerine getirebilen ve problemleri çözebilen bir sistemdir. AI sistemlerinin görevleri öğrenmeyi, planlamayı, akıl yürütmeyi, karar vermeyi ve problem çözmeyi kapsar. Bu sebeple birçok alanda yaygın olarak kullanılan yapay zekânın sıklıkla uygulandığı alanlarından biri de sağlık alanı olmuştur.

Bilgisayarların hayatımıza girmesiyle birlikte günlük hayatta insan beyniyle yapılan işlemler makinalar tarafından oldukça kısa zamanlarda yapılarak daha hızlı sonuçlar alınmasını sağlamaktadır. Makinaların işlemleri hızlı yapabilmesi ve doğru sonuçlar üretmesine yönelik çalışmalar her geçen gün artmaktadır (Cruz, J. A., & Wishart, D. S., 2006). Bu gelişmeler yapa zekâyla birlikte insan hayatına inanılmaz kolaylıklar sunmaktadır. Kolaylık sunmanın yanı sıra, insan beyninin günlük çalışma gücünün sonucunda ulaşmış olduğu sonuçlarda zaman maliyeti sağlayarak oldukça büyük avantajlar getirmiştir. Makine öğrenimi, çeşitli karmaşık görevleri gerçekleştirebilecek uyarlanabilir modeller üretmek için veriler üzerinde eğitilmiş algoritmaları kullanan bir yapay zekâ alt alanıdır. Geliştirilen makine öğrenmesi algoritmaları bu maliyeti milisaniyelere kadar indirmeyi başarmaktadır (Kolay, N., & Erdoğmuş, P., 2016).

Günümüzde sağlık alanında geniş bir uygulama alanı bulan makine öğrenimi yöntemleri, objektif ve optimum çözümler sunarak hekimlerin doğru ve güncel bilgiye ulaşmasında karar destek aracı olarak kullanılabilmektedir. Bu karar destek araçları sayesinde ön görülemeyen bilgi ve bulgular ortaya çıkarılabilmekte, yeni bir örnek sınıf bilgisi tahmin edilebilmekte veya alınan hekim kararları desteklenebilmektedir. Böylece, hastalıklarda erken tanı ve tedaviye başlanarak hastaların hastalıklarından kurtulma oranının yükseltilip ölüm riskinin azaltılması sağlanmaktadır.

Bu tez çalışmasında, miyokart infarktüslü hastalarda hastaneye başvuru değerlerinin hastane içi ölümü öngörülmesi açısından makine öğrenmesi algoritmaları kullanılarak, ölüm riski analizi ve tahminlerinin yapılması amaçlanmaktadır. Yapılacak tahminleme sonucunda, mortalite riski olan hastaların sınıflandırılmasıyla birlikte hekime kararında destek olacak bir sistemin geliştirilmesi ve böylece erken teşhisle birlikte gerekli önlemlerin alınması sağlanmış olacaktır. Hastaların ölü/sağ tahminleri gerçekleştirilirken eğiticili sınıflandırma yöntemlerinden Yapay Sinir Ağları (Artificial Neural Network – ANN), k-En Yakın Komşuluk, (k-Nearest Neighborhood – k-NN), Destek Vektör Makinesi (Support Vector Machine - SVM), Karar Ağacı (Random Tree – RT), Rassal Orman (Random forest - RF) ve Lojistik Model Ağacı (Logistic Model Tree – LMT) algoritmaları kullanılmıştır. Kullanılan bu yöntemlerin performansları farklı metrikler ile kıyaslanarak en başarılı yöntem belirlenmeye çalışılmıştır.

Bu çalışmayla birlikte dünyada ve ülkemizde ciddi can kaybına neden olan ve hayat kalitesini etkileyen kalp krizi hastalığının teşhisinin yapılması ve tedavi yöntemlerini artırabilecek çalışmalar ortaya koyulması hedeflenmektedir. Çalışmamızın bu bağlamda STEMI hastalarının yönetiminde klinik karar desteği sağlaması ve sağlık hizmetlerinin kalitesini artırmaya yönelik katkılar sunması beklenmektedir.

2. KAYNAK ARAŞTIRMASI

Literatürde, miyokard infarktüsü hastalarının ölüm risk analizi üzerine önceden yapılmış farklı çalışmalar yer almaktadır. Yapılan bu çalışmalarla ilgili araştırmalar aşağıda kısaca özetlenmiştir.

Cheng vd., yaptıkları çalışmada**,** ST-segment yükselmeli miyokart enfarktüsü hastalarının hastane içi mortalite riskini tahmin etmeye yönelik faktörleri belirlemeyi ve makine öğrenimi modelleri geliştirmeyi amaçlamıştır (Cheng, C. Y., vd., 2022). Bu çalışma, Tayvan'daki Kaohsiung Chang Gung Memorial Hastanesi acil servisine başvuran non-travma yetişkinleri (≥20 yaş) içeren tek merkezli, retrospektif bir çalışma olup çalışmaya akut STEMI tanısı konmuş ve kriterlere uyan toplam 1567 hasta dahil edilmiştir. Lojistik Regresyon (Logistic regression - LR) için Alıcı Çalışma Karakteristik Eğrisi (Receiver Operating Characteristic Curve – ROC) Eğri Altındaki Alan (Area Under the Curve – AUC) 0.839, Rassal Orman (Random forest - RF) için ise 0.825 olarak bulunmuş, doğruluk oranları LR ve RF için sırasıyla 0.821 ve 0.812 olarak tespit edilmiştir, LR'de duyarlılık ve özgüllük sırasıyla 0.883 ve 0.815; RF'de ise 0.875 ve 0.806 olarak belirlenmiş olup sonuç olarak, LR ve RF algoritmaları kullanılarak geliştirilen tahmin modellerinin STEMI hastalarının hastane içi ölüm riskini tahmin etmede kullanılabilecek etkili araçlar olarak öne çıktığı belirtilmiştir.

Kumar vd., yaptıkları çalışmada, k-NN algoritmasını kullanarak kalp inmesi (kalp krizi) tespiti yapmıştır. Veri madenciliği teknikleri aracılığıyla önemli parametreleri analiz ederek, kalp hastalıklarının teşhisini iyileştirmek için makine öğrenimi teknikleri kullanılmıştır (Kumar, K., vd., 2022). k-NN algoritması, 11 belirlenen parametre ile eğitildiğinde, mevcut algoritmalar olan Rastgele Orman ve Karar Ağacı gibi algoritmalarla karşılaştırıldığında erken kalp inmesi tespitini daha etkili bir şekilde öngörülmüştür. Önerilen k-NN algoritması, kalp inmelerini tespit etmede ortalama %91,4'lük bir doğruluk oranı elde edilmiştir. Makine öğrenimi teknikleri, özellikle k-NN algoritması, kalp hastalıklarının teşhisini geliştirmek için kullanılmış olup 11 parametre ile eğitilen k-NN algoritması, diğer algoritmalarla karşılaştırıldığında erken kalp inmesi tespitinde üstün doğruluk göstermiştir.

Zhao vd., yaptıkları çalışmada, makine öğrenmesi algoritmalarını kullanarak, hastaneye STEMI ile başvuran 8158 hasta verileri kaydedilmiş ve hastane sonuçlarına göre iki gruba ayırmışlardır (Zhao, P., vd., 2023). Hasta verileri rastgele olarak eğitim seti (%75) ve test seti (%25) olarak bölünmüş ve eğitim seti, adaptif sentetik (ADASYN)

örnekleme ile ön işlemden geçirilmiştir. Model geliştirmek için dört yaygın kullanılan makine öğrenmesi algoritması seçilmiştir. Kadınlarda STEMI'ye bağlı hastane içi mortaliteyi tahmin etmede, genel popülasyondan türetilen modellerin ve kadınlara özgü modellerin performansı, doğruluk, seçicilik, duyarlılık, G-mean (geometrik ortalama) ve AUC gibi çeşitli metriklerle karşılaştırılmıştır. ADASYN ile modellerin performansı önemli ölçüde iyileştirilmiştir. Genel popülasyonda, ADASYN ile birleştirilen destek vektör makinesi (SVM) en iyi performansı göstermiştir. Ancak, kadınlarda STEMI durumunda kötü performans göstermiş olup buna karşılık, önerilen kadınlara özgü modeller, kadınlarda STEMI durumunda iyi performans göstermiştir ve en iyi performans gösteren model %72.25 doğruluk, %82.14 duyarlılık, %71.69 seçicilik, %76.74 G-mean ve %79.26 AUC değerlerine ulaşmıştır. Kadınlara özgü modelin doğruluk ve G-mean değerleri, genel popülasyondan türetilen modelden sırasıyla %34.64 ve %9.07 ‘dan daha yüksek olduğu görülmüştür.

Oliveira vd., yaptıkları çalışmada, AMI olan hastalarda hastaneye kabul anındaki mortaliteyi tahmin etmek için bir makine öğrenimi modeli geliştirmeye odaklanmıştır (Oliveira, M., vd., 2023). Farklı makine öğrenimi teknikleri ve değişkenler kullanılarak üç deney yapılmıştır. Deney 1; Stokastik Gradyan İnişi'nin en yüksek sınıflandırma doğruluğuna %80 ve %79 AUC'ye sahip olduğunu göstermiştir. Deney 2; ek değişkenler içeriyor ve SVM yöntemini kullanarak %81 AUC elde etmiştir. Deney 3; özellik seçimi ve Sentetik Azınlık Aşırı Örnekleme Tekniği (Synthetic Minority Over-sampling Technique - SMOTE) kullanmış ve Stokastik Gradyan İnişi ile %88 AUC ve %80 hatırlama elde etmiştir. Özellikle laboratuvar verilerinin eklenmesi, modellerin performansını artırmıştır. Bu çalışma mevcut bilgi ve bağlam temelinde uygun yaklaşımların seçilmesinin önemini vurgulamaktadır. Yapay zeka ve makine öğrenimini klinik karar alma süreçlerine entegre etmek, sağlık hizmetlerinin verimliliğini ve etkinliğini artırabileceğini ifade etmiştir.

Kasim vd., yaptıkları çalışmada, makine öğrenimi ve yığılmış ensembl öğrenme EL teknikleri (makine öğreniminde birden fazla modelin bir araya getirilerek tahmin performansını artırmayı amaçlayan yöntem) kullanarak STEMI olan Asyalı kadınlarda hastane içi mortaliteyi tahmin etmeyi amaçlamışlardır (Kasim, S., vd., 2023). ML modelleri, AUC değerleri 0.60 ile 0.93 arasında değişen ve seçilen özelliklerle en iyi performansı gösteren SVM Linear modeli (AUC: 0.934) ile geleneksel risk skoru ile karşılaştırıldığında daha iyi performans göstermiştir. Çalışma, STEMI olan Asyalı kadınlarda mortalitenin ortak belirleyicilerini saptamış, bunlar arasında sistolik kan

basıncı, killip sınıfı, açlık kan şekeri, beta-bloker, ACE inhibitörü ve oral hipoglisemik ajan yer almıştır. LIME analizini kullanarak (makine öğrenimi modellerinin tahminlerini anlamayı ve açıklamayı kolaylaştırmak için kullanılan teknik) yaş, PCI müdahalesi, kan basıncı ve farmakolojik müdahaleler gibi değişkenlerin hayatta kalma sonuçlarını etkilediğini göstererek bireysel hasta tahminlerine yönelik içgörüler sağlamışlardır. STEMI olan kadınlarda genel mortalite oranı %12.8 değerinde olup; sistolik kan basıncı, killip sınıfı ve ilaç kullanımı gibi çeşitli belirleyiciler arasında hayatta kalanlar ve hayatta kalamayanlar arasında önemli farklılıklar bulunmuştur.

Tao vd., yaptıkları çalışmada, STEMI hastalarında hastane içi mortaliteyi erken evrede tahmin etmek için elektronik tıbbi kayıt (HPEMR) ana sayfasına dayalı, verilerini kullanarak bir makine öğrenimi algoritması geliştirmiştir (Tao, D., vd., 2022). 2013-2017 yılları arasında Çin'in Shenzhen kentindeki 7 hastaneden 5865 STEMI hastasının verileri kullanılmıştır. Dört makine öğrenimi algoritması eğitilmiştir. Bunlar: Lojistik Regresyon, Destek Vektör Makinesi, Karar Ağacı ve Yapay Sinir Ağıdır. Model, demografik verileri, tanı, eşlik eden hastalıkları ve hastanede yatış bilgilerini dikkate almıştır. Atriyal fibrilasyon, akut böbrek yetmezliği, tip 2 diyabetik nefropati, akut kalp yetmezliği ve kalp fonksiyon sınıfı IV gibi eşlik eden hastalıklar, ölüm riskinin daha yüksek olması ile ilişkilendirilmiştir. Model, test veri setinde 0.879 AUC ile iyi bir ayırt etme yeteneği göstermiştir.

Yadav vd., yaptıkları çalışmada, Kuzey Hindistan'da ST elevasyonlu miyokard enfarktüsü kaydına dahil edilen hastalarda 30 günlük tüm nedenlere bağlı ölümü tahmin etmek için bir makine öğrenimi modeli geliştirmeyi ve değerlendirmeyi amaçlamıştır (Yadav, R., vd., 2022). ML modeli, 31 değişken kullanılarak geliştirilmiş ve Ekstra Ağaç ML modeli, doğrulama veri setinde en iyi tahmin yeteneğini göstermiştir. Modelin duyarlılığı %85, AUC %79.7 ve doğruluk oranı %75'tir. ML modeli ölümün önemli belirleyicileri olarak; gecikmiş revaskülarizasyon süresi, başlangıçtaki kardiyojenik şok, sol ventrikül ejeksiyon fraksiyonu (EF) <%30, yaş, serum kreatinin, sunumda kalp yetmezliği, kadın cinsiyet ve orta-şiddetli mitral regürjitasyonu tanımlamıştır. Çalışmanın sonuçları şunları göstermektedir: ML modelleri, STEMI sonrası ölümü tahmin etmede gelişmiş performans sağlamıştır. Ekstra Ağaç ML modeli, diğer ML modellerine kıyasla daha iyi tahmin yeteneğine sahip olup, STEMI sonrası 30 günlük tüm nedenlere bağlı ölümü tahmin etmede en etkili model olarak tanımlanmıştır. Geliştirilen ML modeli, yüksek riskli bireyleri tanımlamak ve yoğun izleme ve yakın takip gerektirebilecek bireyleri belirlemek için kullanılabilir.

Li vd., yaptıkları çalışmada, Çin'deki akut miyokard enfarktüsü kayıt verilerinden yararlanılarak, STEMI hastaları için hastane içi mortalite tahmin modelleri geliştirmiştir (Li, X., vd., 2017). Veri seti olarak, 2013 ve 2014 yıllarında hastaneye yatırılan 18,744 STEMI hastasının verileri kullanılmıştır. Genel doğrusal modeller, karar ağaçları ve Bayes modelleri gibi yorumlanabilir makine öğrenimi yöntemleri kullanılmıştır. Demografik bilgiler, semptomlar, tıbbi geçmiş, fiziksel muayene ve laboratuvar test sonuçları gibi 132 özellik kullanılmıştır. Rastgele orman ve Bayes ağı modelleri, tüm aday özellikler kullanıldığında en iyi performansı gösterdi. Özellik seçimi: Genel Doğrusal Modeller (GLM - Generalized Linear Models) yöntemlerinin performansını artırdı. Rastgele orman modeli için AUC değeri 0.917, Bayes ağı modeli için AUC değeri 0.846 olarak bulunmuştur. Ölüm oranı 2013 yılında hastaneye yatırılan hastalar için hastane içi mortalite oranı %6.70, 2014 yılı için %6.78 olarak belirlenmiştir.

Literatürde yapılan benzer çalışmalardan elde edilen bulgular ile tez çalışmamızdan elde edilen sonuçlar incelendiğinde, çalışılan tüm performans ölçütlerinde çok iyi bir sınıflandırma performansının sağlandığı görülmüştür. Yapılan testlerde kullanılan tüm ML algortimalarında %94,9 - %99,9 aralığında doğruluk, %90,5-%100 aralığında duyarlılık, %98,8-%100 aralığında kesinlik, %98,7-%100 aralığında seçicilik, %95,4-%99,9 aralığında F-skor ve %95,3-%100 aralığında AUC değerleri elde edilmiştir. Elde edilen bu bulgular, çalışmamızın aynı sınıflandırma algoritmaları kullanan çalışmalardan bile daha yüksek bir sınıflandırma performansına sahip olduğunu ortaya koymuştur.

3. GENEL BİLGİLER

3.1. Koroner Arter Hastalığı

Koroner arter hastalığı (KAH), kolesterol birikiminin veya plak adı verilen yağlı maddelerin kalp kasını besleyen koroner arterlerde daralma veya tıkanmaya neden olduğu bir kalp damar hastalığıdır. Koroner arterler daraldığında veya tıkandığında, kalbe yeterli miktarda kan, oksijen ve besin sağlayamadığından kan akımı azalır veya tamamen kesilir.

Türkiye İstatistik Kurumu'nun verilerine göre, ülkemizde 2017 yılı ölüm nedenleri sınıflandırmasında tüm ölüm sebepleri arasında %39,6 ile en sık ölüm nedeni dolaşım sistemi hastalıkları, dolaşım sistemi hastalıklarından ise KAH %39,7 oranıyla ölüm nedenleri arasında birincil neden olarak dikkat çekmektedir. KAH içerisinde %71 ile MI Türkiye'nin en sık ölüm sebebi olarak belirlenmiştir (TÜİK, 2017).

Kardiyovasküler hastalıklar ülkemizde olduğu gibi dünyada da mortalite ve morbiditenin en önemli nedenidir. Kardiyovasküler hastalıklar; koroner kalp hastalıkları, inme, kalp yetmezliği, periferik arter hastalığı, karotis arter hastalıkları ve aortailiak hastalıklarını kapsar. Koroner kalp hastalığı, koroner arter hastalığı, kan damarlarında daralma ve ateroskleroz sonucu kalbe giden oksijenin azalması sonucunda ortaya çıkar. KAH; unstabil anjina, miyokard infarktüsü ve kalp yetmezliğine yol açabilir (Shrivastava A. K., vd., 2015).

KAH; yaş, erkek cinsiyet, aile hikayesi, sigara kullanımı , hipertansiyon, diyabet obezite, yüksek total kolesterol, düşük yüksek yoğunluklu lipoprotein kolesterol (HDL), yüksek düşük yoğunluklu lipoprotein kolesterol (LDL), yüksek trigliserid (TG), hareketsiz yaşam tarzı ve artmış bel çevresi gibi risk faktörleri bilinmesine rağmen hala dünya çapında ölümün en önde gelen nedenidir (Torpy, J. M., vd., 2009). KAH için risk faktörleri düzeltilebilir ve düzeltilemeyen diye 2'ye ayrılır.

Düzeltilebilir risk faktörleri (tedavi ya da kontrol edebilecekler);

-Sigara içmek veya sigara dumanına maruz kalmak

-Hipertansiyonunun olması

-Doymuş yağ, kolesterol ve tuz miktarı yüksek beslenme alışkanlıklarının olması

-Diyabet (şeker hastalığının) olması

-Kolesterol düzeylerinin fazla olması

-Fiziksel aktivite düzeyinin az olması

-Alkol tüketimi

-Stres durumu

-Obezite (kilolu olmak)

Düzeltilemeyen risk faktörleri;

-İleri yaş

-Erkek olmak

-Ailede KAH öyküsünün olması

Bahsedilen risk faktörleri ne kadar çok ise hastalık riski de o kadar fazladır. Risk faktörleri hakkında bilgi sahibi olmak, yaşam tarzı alışkanlıklarınızı değiştirerek kardiyovasküler hastalıklara yakalanma riskinizi azaltmanıza yardımcı olabilir (Ebrahim, S., vd., 2011).

3.2. Miyokard İnfarktüsü

Kalp krizi (myocardial infarction–MI), kalbi besleyen koroner damarlarının oksijen veya besin eksikliği nedeniyle tıkanması veya daralması durumudur (Cleveland Clinic, 2024). Bu daralma ve tıkanma sonucunda kalp kasına kan akışı yavaşlamakta ve kalbin yeteri kadar beslenmemesi durumunda kalp krizine ve kalp kriziyle birlikte kalıcı hasarlara neden olabilmektedir. DSÖ'nün güncel verileri dikkate alındığında Türkiye'de ve dünyada ölüm oranlarının %31'den fazlasının kalp krizi geçirerek gerçekleştiği görülmektedir (Katarya, R., & Meena, S.K., 2021).

Bu oran her geçen gün artmakla beraber kurtarılan hastalarda dahi farklı hasarlara sebep olma potansiyeli oldukça yüksektir. Kalp krizi; baş dönmesi, nefes darlığı, hazımsızlık, göğüs-kol-boyunda ağrı ve yorgunluk gibi durumlarda şüphe uyandıracak bir hastalık olsa da tıpta birçok neden ve durumların değerlerine bakılarak teşhis konulan hastalıktır (Bashir, S., vd., 2019). Teşhis konulması süreci, bakılan parametre ve yoğunluğa bağlı olarak uzun süreç gerektirmektedir.

İskemik kalp hastalıkları dünya genelinde olduğu gibi ülkemizde de erişkinlerde mortalite (ölüm) ve morbiditenin (hastalıkların) en önemli nedenlerinden biridir. Aterosklerotik damar hastalığı ve burada oluşan komplikasyonlar kardiyovasküler hastalıkların en önemli kısmını oluşturur. Koroner ateroskleroz genellikle kronik KAH sebep olsa da, koroner damarlar içindeki tıkayıcı lezyonlar ve aterosklerotik plaklar miyokartta ani iskemiye de sebep olabilmektedir. Kısa sürede oluşan ve miyokart dokusunu tehdit eden iskemik durumları tümü günümüzde akut koroner sendromlar (AKS) olarak adlandırılmaktadır (Oto, A., vd., 2009).

Miyokard infarktüsü, ST segment elevasyonlu ve Non-ST segment elevasyonlu olarak 2 başlık altında incelenir. STEMI, tanı ve tedavisinin hemen başlanması gereken bir akut koroner sendromdur. Daralan ya da tıkanan damarın, nekroz oluşmadan koroner girişim ile revaskulerizasyonun sağlanması için perkütan koroner girişim (PKG) gerekmektedir.

Akut koroner sendromlar arasında en tehlikeli ve ölümcül olanı: ST-segment yükselmeli miyokart infarktüsüdür.

3.3. ST-Segment Yükselmeli Akut Miyokard İnfarktüsü

ST segment yükselmeli miyokard infarktüsü, esas olarak koroner arterde trombüs oluşumundan kaynaklanan en ciddi ve ölümcül kardiyovasküler hastalıklardan biridir ve kısmi veya tam tıkanmaya neden olabilir. STEMI hastaları için erken ve etkili reperfüzyon çok önemlidir ve mevcut kılavuzlarda primer perkütan koroner girişim (PPKG) önerilmektedir (İbanez, B., vd., 2018).

Akut miyokart infarktüsü Kuzey Amerikada ve Avrupada ölümün önde gelen nedenlerinden biridir. ABD’de her sene 1 milyondan fazla bireyde akut miyokart infarktüsü görülmektedir. Bu duruma ek olarak 300.000’den fazla bireyin hastaneye başvurmadan önce akut MI nedeniyle hayatını kaybettiği tahmin edilmektedir. Son 30 yıl içerisinde koroner yoğun bakımların artması, fibrinolitik ile kateter bazlı reperfüzyon tedavilerindeki gelişmelere rağmen, mortalitede bir azalma gözlenmiş olsa da, akut miyokart infarktüsü (MI) nedeniyle hastaneye başvurmadan önce ölenler de dahil olmak üzere toplam mortalite oranı %30'dan fazladır (Griffin, P., ve Topol, E. J., 2009).

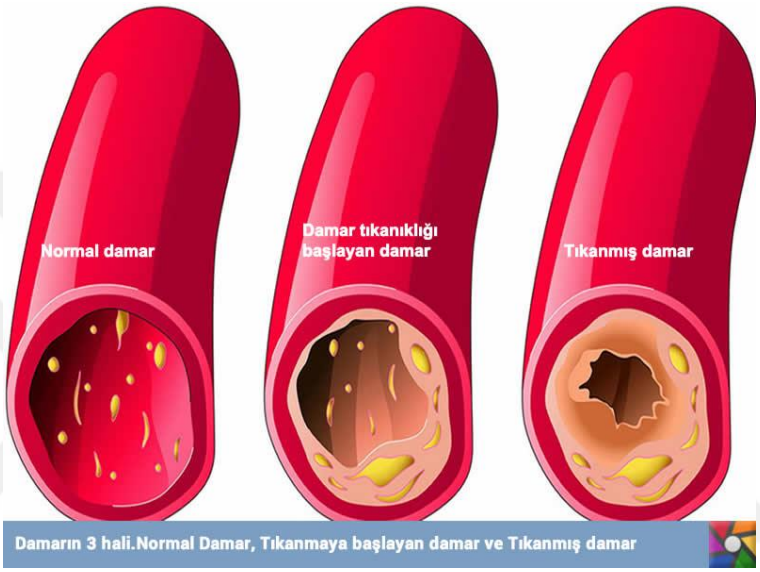
Fizyopatolojisi, hassas plakların rüptür veya erozyonu sonucu tam epikardiyal koroner arter tıkanması ST segment yükselmeli MI’ın en yaygın nedenidir (Falk, E., vd., 1995; Libby, P., 2001).

Ateroskleroz; kronik bir arter hastalığıdır ve vasküler ölümlerin önemli bir nedenidir. Arter duvarındaki yağlı çizgilenmeler zamanla ateroma ve karakteristik plaklara dönüşür (Bentzon, J. F., vd., 2014). Bu ateromatöz plakların akut rüptürü sorumlu arterde kısmi veya total tıkanıklıklara yol açan bölgesel trombüslere neden olur. Bu plakların klinik sonuçları plağın yerleşim yerine damarın tıkanma hızına ve tıkanma derecesine bağlıdır. Bu hastalık uzun yıllar boyunca varlığını sürdürmektedir ve sıklıkla birden fazla damar yatağında bir arada bulunmaktadır. Başlıca klinik belirtileri arasında

iskemik kalp hastalığı (İKH), iskemik inme ve periferik arter hastalığı (PAH) bulunur (Herrington, W., vd., 2016).

KAH, damar sertliği yani ateroskleroz nedeniyle oluşmaktadır. Ateroskleroz arterlerde, zamanla plak birikimiyle karakterize bir durumdur. Bu plaklar koroner arterlerde kan akışını etkileyerek koroner arter hastalığına yol açar.

Şekil 3.1'de soldan sağa doğru; normal damar, damar tıkanıklığı başlayan damar, tıkanmış damar olarak damarların zamanla oluşan, plak birikimi seyri verilmiştir.



Şekil 3.1. Plak birikimi

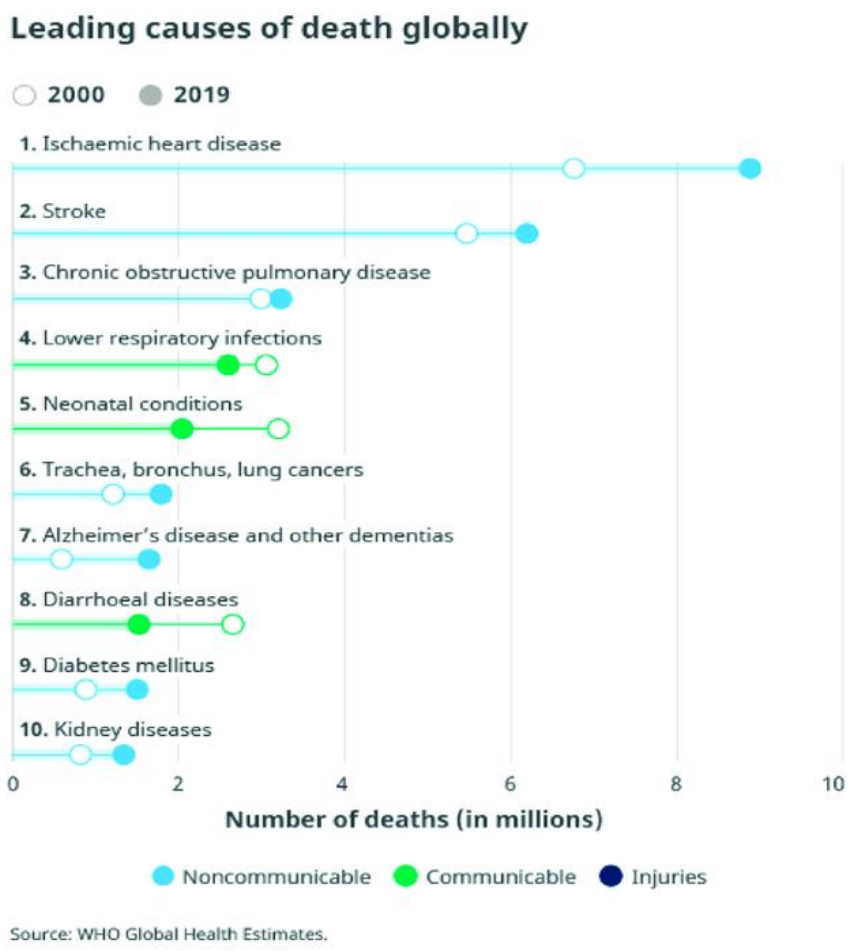
Aterosklerozun patofizyolojisi; yağlı çizgilenmeler, düz kas hücreleri, bağ dokusu hücreleri ve lipidlerin yağ bileşenlerini içeren başlıca üç ana komponentten oluşan aterosklerotik plaklara dönüşür (Ridker, P. M., vd., 2009). Aterosklerozun zamana göre seyri Şekil 3.2'de gösterilmiştir. Endotel disfonksiyonu: 1.dekadan itibaren, ilk 10 yıllık süre; 3. dekadan itibaren, 10-30 yıl arası; 4. dekadan itibaren, 40 yıl ve sonrası olarak 3'e ayrılır.

Şekil 3.2. Amerikan Kalp Cemiyetinin (AHA) vasküler lezyon sınıflamasına göre plak progresyonu şekilde gösterilmiştir (Stary, H. C., vd., 1995).

Plaktaki aktive makrofajlar tarafından salgılanan enzimler plağın rüptüre olmasına neden olur. Plak rüptüre olmasıyla plak içeriğinin dolaşıma karışması tromboz oluşumuna yol açar (Jonason, T., & Bergstrom, R., 1987). Oluşan tromboz plağın şeklini değiştirebilir, damar lümenini tıkayabilir, emboli yapabilir veya plak içeriği embolize olabilir. Stabil plaklar anstabil plaklara göre daha çok fibröz yapı ve daha düşük lipid içeriğine sahiptir. Stabil plaklar %100 blokaja neden olmazlar. Anstabil plaklar kalın lipid çekirdek ve ince fibröz kapsül yapısından oluşur ve öngörülemeyen şekilde rüptüre olma eğilimindedir (Rodriguez-Granillo G. A., vd., 2007).

Damar lümeninde plakların sebep olduğu darlık; AKS, MI, ölümcül aritmiler ve ani kardiyak ölümlerin meydana gelmesinde etkilidir. Erkek bireyler, kadın bireylerden daha erken yaşta ve daha sık bu hastalığa yakalanırlar. Kadınlar menapoza kadar, ilerlemiş aterosklerozdan korunurlar. Ciddi risk faktörlerinin bulunmadığı premenapozal dönemde hastalığın görülmesi kadınlarda nadirdir ve erkeklerden ortalama 10 yıl sonra bu hastalığa yakalanırlar (Awtry, E. H., & Loscalso, J., 2001).

Dünya sağlık örgütünün 2019 yılında, dünya çapında önde gelen ölüm nedenleri verileri Şekil 3.3'te verilmiştir. İskemik kalp hastalığı 9 milyon civarında 1. sırada yer almaktadır, daha sonra felç 2. sırayı alarak 6 milyon civarındadır, 3. sırayı kronik akciğer hastalığı 3 milyon civarı olarak almıştır ve diğer hastalıklarda Şekil 3.3'te görüldüğü gibi birbirini takip etmektedir.



Şekil 3.3. Dünya çapında önde gelen ölüm nedenleri 2019 verileri (WHO, 2019).

4. MATERYAL VE YÖNTEM

Bu tez çalışmasında Dicle Üniversite Hastanesinden temin edilen veri seti kullanılarak (Ozbek, M., vd., 2023), hastane içi ölümü ön görülmesi açısından mortalite riski tahmini ve analizi makine öğrenmesi yöntemleri kullanılarak yapılmıştır. Kullanılan veri seti ve yöntemlere ilişkin bilgiler aşağıda verilmiştir.

4.1. Veri Setinin Oluşturulması

Bu çalışmada Dicle Üniversite Hastanesinde 735 hastadan alınan veri seti kullanılmıştır (Ozbek, M., vd., 2023). Hastalardan alınan anamnez ve laboratuvar parametreleri arasında; yaş, cinsiyet, sigara öyküsü, kalp hızı - hipertansiyon (vital bulgular) kan şekeri miktarı diyabet (Diabetes Mellitus - DM), Ejeksiyon Fraksiyonu (EF), MI öyküsü, MI saati, MI tipi, killip sınıflandırılması, koroner anjıografi (PCI) öyküsü, perkütan koroner girişim sonuçları (Sorumlu damar, Stent-sayı, Lezyon-Sayı), klinik sonuç (hastane mortalitesi) verileri, laboratuvar kan sonuçları (albumin, trigiliserid, kolesterol, HDL, LDL, WBC, HB, lenfosit, monosit, nötrofil), CABGO öyküsü, SVO öyküsü, KBY öyküsü, aile öyüsü, Kalp Yetersizliği (KY), akciğer ödemi, kardiyojenik şok öyküsü, AV tam blok, Ventriküler Taşikardi (VT) öyküsü, Ventriküler Fibrilasyon (VF) öyküsü, arrestonadmission (arrestte başvuru), saat latency (gecikme süresi) HPL öyküsü ve eşlik eden hastalıklar, olmak üzere toplam 38 farklı parametre sonuçları bulunmaktadır. Bu veri setinde yer alan hastalardan, 644 kişi sağ iken geri kalan 91 kişinin hastane içi ölümü gerçekleşmiştir. Bu parametrelerden bazıları ve ifade ettikleri durumlar Tablo 4.1'de verilmiştir.

Tablo 4.1. Veri setindeki bazı parametrelerin öznitelik bilgileri

Değişken	Anlamı	Açıklama
Age	Yaş	Hastanın Yaşı
Sex	Cinsiyet	0:Kadın 1:Erkek
Heart attack	MI öyküsü	0: Yok 1: Var

Killip classification	Killip sınıflandırması	1: killip 1 (kalp yetersizliği yok) 2: killip 2 (hafif ve orta derecede kalp yetersizliği) 3: killip 3 (pulmoner ödem) 4: killip 4 (kardiyojenik şok)
Chol	Kolesterol düzeyi HDL öyküsü	0: Yok 1: Var Mg/dl cinsinden değeri
Diabetes	DM öyküsü	0: Yok 1: Var
Death	Hastane içi ölümü	0: Yok 1: Var
Thalach	Max. kalp atış hızı	Değer
Cardiac arrest on admission	Bilinçkaybı (baygınlık)	0: Yok 1: Gelişte arrest
Smoke	Sigara öyküsü	0: Yok 1: Var
Heart	Kalp yetersizliği	0: Gelişmedi 1: Gelişti
Heart attack	MI tipi	1: Anterior MI 2: İnferior MI 3: İnferolateral MI 4: Anteroseptal MI 5: Diğer (Belirtiniz)
Vein	Sorumlu damar	1: LAD 2: RCA 3: CX 4: Diğer (İMA, Diagonal)
Target	Hedef sonuç	0:Ölüm yok 1:Ölüm var

Bu hastaların hastaneye başvuru sırasında elde edilen bilgileri ile hastane içi ölümü ön görülmesi açısından mortalite riski tahmini ve analizi makine öğrenmesi yöntemleriyle yapılmıştır. Çalışmanın ilk aşamasında veri seti incelenerek problem üzerinde tartışılmıştır. Çalışma için toplanan veriler SPSS® (IBM Statistical Package for the Social Sciences, sürüm 24.0) istatistik programı kullanılarak değerlendirilmiştir. Veri sayısı, verilerin parametreleri ve bu parametlerin sonucu etkileme durumu üzerinde durulmuştur. Orijinal ham veri setinin temel tanımlayıcı istatistikleri, sürekli özelliklerin ortalama ± standart sapma olarak verildiği ve kategorik parametrelerin sayısal ve yüzde

olarak ifade edildiği durum Tablo 4.2'de verilmiştir. Hastaların yaşı 61,74 ± 14,10 yıl olup popülasyonun %29,25'ini kadın hastalar oluşturmaktadır. Sürekli değişkenlerin ortalamalarının karşılaştırılması Student'ın t-testi kullanılarak hesaplanmış ve kategorik değişkenler için χ2 testi kullanılmıştır. İki taraflı p-değerleri tabloda verilmiş olup, burada p-değerleri < 0,05 olan parametreler istatistiksel olarak anlamlı kabul edilmiştir.

Tablo 4.2. Hastaların klinik ve fonksiyonel özellikleri

Parametreler (37)	**Tümü (735)**	**Sağ (644)**	**Ölü (91)**	**p-value**
Cinsiyet (Kadın)	215 (29.25%)	180 (27.95%)	35 (38.46%)	**0.039**
Yaş	61.74±14.10	60.59±13.88	69.86±13.07	**0.004**
Kalp hızı (bpm)	82.90±18.11	82.24±16.94	87.58±24.51	**<0.001**
Hastaneye başvuru saati ve semptom başlangıcı arasındaki süre (0-24 saat)	6.53±4.24	6.44±4.33	7.10±3.57	0.462
MI hastalarının hastaneye geliş saati (0-24 saat)	13.24±5.21	13.23±5.17	13.29±5.53	**0.005**
Bilinç kaybı	63 (8.57%)	27 (4.19%)	36 (39.56%)	**<0.001**
HT öyküsü	260 (35.4%)	228 (35.5%)	32 (35.2%)	0.964
DM öyküsü	216 (29.38%)	181 (28.10%)	35 (38.46%)	**0.042**
HPL öyküsü	38 (5.17%)	33 (5.12%)	5 (5.49%)	0.881
Sigara öyküsü	366 (49.79%)	335 (52.01%)	31 (34.06%)	**0.001**
MI öyküsü	35 (4.76%)	31 (4.81%)	4 (4.39%)	0.861
PCI öyküsü	49 (6.66%)	40 (6.21%)	9 (9.89%)	0.188
CABGO öyküsü	17 (2.31%)	14 (2.17%)	3 (3.29%)	0.505
SVO öyküsü	28 (3.80%)	21 (3.26%)	7 (7.69%)	**0.039**
Aile öyküsü	97 (13.19%)	91 (14.13%)	6 (6.59%)	**0.047**
KBY öyküsü	7 (0.95%)	6 (0.93%)	1 (1.09%)	0.878
Kalp yetersizliği	83 (11.29%)	68 (10.55%)	15 (16.48%)	0.095
Akciğer ödemi	34 (4.62%)	20 (3.10%)	14 (15.38%)	**<0.001**
Kardiyojenik şok	63 (8.57%)	35 (5.43%)	28 (30.76%)	**<0.001**
Av tam blok	30 (4.08%)	23 (3.57%)	7 (7.69%)	0.063
VT	68 (9.25%)	52 (8.07%)	16 (17.58%)	**0.003**
VF	56 (7.61%)	38 (5.90%)	18 (19.78%)	**<0.001**
Kan şekeri	174.05±95.36 (4 kayıp)	169.67±92.74 (4 kayıp)	204.67±107.78	**<0.001**
Albumin	3.53±0.48 (215 kayıp)	3.55±0.47 (176 kayıp)	3.27±0.57 (39 kayıp)	**<0.001**

Trigliserid (mmol/L)	152.20±110.20 (32 kayıp)	152.39±111.80 (27 kayıp)	150.87±98.54 (5 kayıp)	0.928
Kolesterol (mmol/L)	184.35±47.41 (32 kayıp)	185.23±47.47 (26 kayıp)	178.00±46.84 (6 kayıp)	0.469
HDL (mmol/L)	34.91±10.34 (31 kayıp)	34.92±9.37 (26 kayıp)	34.84±15.72 (5 kayıp)	**<0.001**
LDL (mmol/L)	120.82±38.25 (42 kayıp)	121.46±38.80 (35 kayıp)	116.22±33.82 (7 kayıp)	0.843
WBC	13.44±5.52 (201 kayıp)	12.94±4.60 (163 kayıp)	18.04±9.63 (38 kayıp)	0.419
HB	13.68±1.95 (11 kayıp)	13.70±1.90 (7 kayıp)	13.53±2.31 (4 kayıp)	0.062
Lenfosit	2.03±1.60 (11 kayıp)	2.02±1.59 (7 kayıp)	2.14±1.68 (4 kayıp)	0.138
Monosit	0.68±0.38 (11 kayıp)	0.68±0.39 (7 kayıp)	0.69±0.31 (4 kayıp)	0.661
Nötrofil	10.86±5.35 (11 kayıp)	10.51±4.87 (7 kayıp)	13.46±7.60 (4 kayıp)	0.465
EF	44.47±10.40 (56 kayıp)	45.39±9.70 (46 kayıp)	38.17±11.97 (10 kayıp)	**<0.001**
Sorumlu damar				
LAD	304 (41.36%)	268 (41.61%)	36 (39.56%)	0.351
RCA	273 (37.14%)	243 (37.73%)	30 (32.96%)	
CX	116 (15.78%)	96 (14.90%)	20 (21.91%)	
Diğer (İMA, DİAGONAL)	22 (2.99%) (30 kayıp)	11 (1.70%) (26 kayıp)	1 (1.09%) (4 kayıp)	
Killip sınıflandırılması				
Killip I	555 (75.51%)	522 (81.05)	32 (35.16)	**<0.001**
Killip II	96 (13.06%)	82 (12.73%)	14 (15.38%)	
Killip III	28 (3.80%)	15 (2.32%)	13 (14.28%)	
Killip IV	57 (7.75%)	25 (3.88%)	32 (35.16%)	
MI tipi				
Anterior MI	237 (32.24%)	201 (31.21%)	36 (39.56%)	**0.042**
Inferior MI	281 (38.23%)	250 (38.81%)	31 (34.06%)	
Inferolateral MI	70 (9.52%)	58 (9.00%)	12 (13.18%)	
Anteroseptal MI	59 (8.02%)	58 (9.00%)	1 (1.09%)	
Diğer	59 (8.02%) (29 kayıp)	51 (7.91%) (26 kayıp)	8 (8.79%) (3 kayıp)	

Akut miyokart infarktüsü hastasında ölüm riskini tahmin etmek hem mümkündür hem de oldukça faydalıdır (Lee, K. L., vd., 1995; Morrow, D. A., vd., 2000). Bu tahmin

tedavinin kararında verilmesinde önerilerin yapılmasında, hasta ile ailelerine danışmanlık sağlanmasında yardımcı olur. Klinik çalışmalarda ve literatürde erken mortaliteyi öngörmenin en önemli bağımsız parametreleri; ileri yaş, yüksek killip derecesi, yüksek kalp hızı, düşük sistolik kan basıncı, anterior infarkt lokalizasyonu olduğu belirlenmiştir (Lee, K. L.,vd., 1995; Morrow, D. A., vd., 2000). Bu beş gösterge 1 aylık mortalite için %90'ın üzerinde bilgi verir (Kırbaş, Ö., 2011).

4.2. Makine Öğrenmesi Yöntemleri

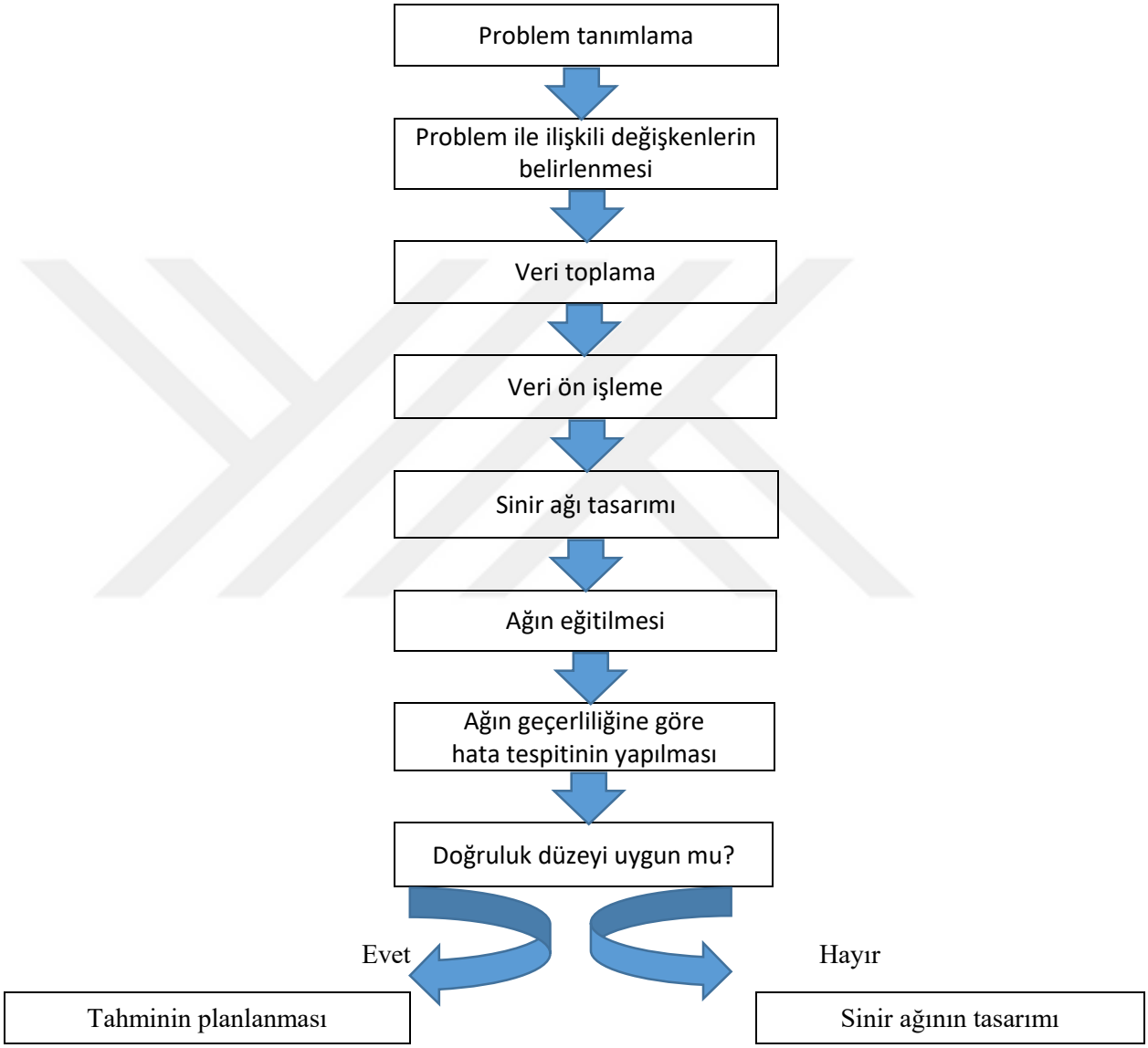
Bu çalışmada kullanılan makine öğrenmesi yöntemleri bir açık kaynak kodlu (opensource) yazılım olan WEKA (**W**aikato **E**nvironment for **K**nowledge **A**nalysis) platformu kullanılarak uygulanmıştır. Bu araç, Yeni Zelanda'daki Waikato Üniversitesi'nde 1993 yılında C diliyle yazılmaya başlanan bir veri madenciliği yazılımıdır. İlk resmi sürümü 1996'da yayınlanmış, 1997'de ise gelişim göstererek Java diliyle güncellenmiştir. Sonraki sürümleri de Java diliyle geliştirilmeye devam etmektedir. Weka açıldığında bir ara yüz ile karşılaşımıza çıkar, bu ara yüzde bulunan ve WEKA'nın logosu olan kuş, programa adını verdiği ve yalnızca Yeni Zelanda adalarında bulunan, nesli tükenmiş olan bir kuş türüdür (Şeker, Ş. E., 2016).

4.2.1. Yapay sinir ağları

Yapay sinir ağları (YSA) insan beynini taklit eden bir öğrenme yapısına sahiptir ve giriş ile sonuç arasındaki en iyi tahmini yapabilir. Sınıflama, çıkarım, tahmin, büyük veri setlerinden mantıksal sonuçlar çıkarma gibi alanlarda kullanılmaktadır. Birçok tanımı olan YSA temel olarak, beynin sinirlerinin işleyişinden esinlenilerek, sistemlere; öğrenme, hatırlama ve bilgiler arasında ilişkiler kurma gibi yetenekleri kazandırmayı amaçlayan bir bilgi işleme algoritmasıdır.

Yapay sinir ağında, insanlarda olan biyolojik sinir hücresindeki nöronlara benzer yapılar bulunur. Bu yapılar bir araya geldiğinde, yapay sinir hücresini oluşturur ve yapay sinir ağını meydana getirir (Graupe, D., 2013). Beyindeki sinir hücrelerinin çalışma şeklini taklit ederek geliştirilen YSA, birbirine sinaptik bağlantılara benzer yapılarla bağlı yapay sinir hücrelerinden oluşur. Her bir nöron arasında bulunan bağlantılar sayesinde, bir nörona gelen bilgi diğer nörona iletilir. (Alan, M., 2020).

İlk YSA modeli 1943'te bir sinir hekimi Warren McCulloch ve bir matematikçi olan Walter Pitts tarafından "Sinir Aktivitesinde Düşüncelere Ait Bir Mantıksal Hesap (A Logical Calculus of Ideas Immanent in Nervous Activity)" başlıklı makale ile ortaya koyulmuştur (Öztürk, K., & Şahin, M. E., 2018). YSA biyolojik sinir ağlarını taklit eden sentetik yapılardır (Eğrioğlu, E., vd., 2009).

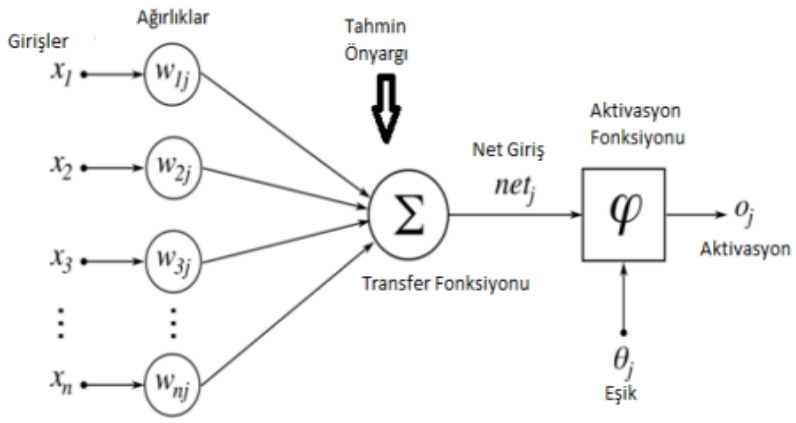


Şekil 4.1. YSA tahmin yöntemi

YSA, insan beyninin yapısını ve nöronları arasındaki bağlantılardan yararlanılarak geliştirilmişse de nöronların bağlandığı gerçek beyinden farklı olarak katmanlara, bağlantılara ve veri iletme yeteneklerine sahip olup, biyolojik sinir sisteminin matematiksel bir modellemesi olarak da tanımlanabilir Gershenson, C., 2021). Bu

sistemler, kendi kendilerine öğrenebilme yeteneğine sahiptirler. Genellikle problemin çözümünün ya da tespitinin karmaşık olduğu uygulamalarda en iyi çalışma performansını sağlayabilmektedir (Copeland, M., 2016).

Bir sinir ağı tecrübelerden faydalanarak bilgi biriktirmek ve verileri kullanılabilir düzeye getirmek için doğal bir eğilime sahip basit işlem öğelerinden oluşan paralel bir işlemcidir. Bu durumda sinir ağları belki de insan düşünce ve öğrenme süreçlerine dair en anlamlı MI tekniklerinden biri olabilir. Ayrıca, insan odaklı problem çözümlemeleri ve yapay zekâ çözümlerinin felsefi tartışmaları açısından bilimsel literatürde çok yönlü etkiler yaratmaktadır. (Boz, H., & Kose, U., 2018).



Şekil 4.2. Yapay sinir hücresi (Öztürk, K., & Şahin, M. E., 2018).

Şekil 4.2'de görüldüğü gibi hücreye n tane veri girişi yapılmaktadır (X_n veri girişi). Bu girilen veriler ağırlıklarla çarpılırak tüm veriler toplanır ve sonrasında önyargı eklenir bunun sonucunda ise net yargı bulunur. Net girdi aktivasyon fonksiyonundan geçirilir ve bir veri çıktısı elde edilmiş olur.

Çok katmanlı ileri beslemeli yapay sinir ağları Rumelhart ve arkadaşları (Rummelhart vd., 1986) tarafından geliştirilmiş ve bu modele literatürde çok katmanlı algılayıcı (Multilayer Perceptron - MLP), ileri beslemeli yapay sinir ağı (Feedforward Neural Network - FFNN), geriye yayılım modeli (Back Propogation Network - BPN) ve bazı çalışmalarda ise çok katmanlı ileri beslemeli yapay sinir ağı modeli (Multilayered Feed Forward Neural Networks - MFNN) adı verilmiştir. Bu çalışmada, literatürdeki yaygın kullanıma uygun olarak kısaca MLP olarak anılacaktır.

MLP; girdi katmanı, çıktı katmanı ve bu iki katman arasındaki gizli katman(lar)dan oluşan ileri beslemeli bir yapıya sahip YSA modelidir. Girdi katmanından

alınan girişler, girdi katmanı ve gizli katman arasında bulunan bağlantı ağırlıkları ile çarpılıp gizli katmana iletilmektedir. Gizli katmandaki nöronlara gelen girişler toplanarak aynı şekilde gizli katman ile çıktı katmanı arasındaki bağlantı ağırlıkları ile çarpılarak çıktı katmanına iletilir. Çıktı katmanındaki nöronlar da, kendisine gelen bu girişleri toplayarak buna uygun bir çıkış üretirler (Efe, M. Ö., ve Kaynak, O., 2000). Bir MLP'nin temel bileşenleri şunlardır:

Giriş katmanı: Girdi verilerini alır ve ağa aktarır. Her bir girdi özelliği, bu katmandaki bir düğüme karşılık gelir.

Gizli katman(lar): Girdileri işleyerek karmaşık ilişkileri öğrenir. Her gizli katman, önceki katmanın çıkışını alır, ağırlıklarla çarparak ve bir aktivasyon fonksiyonuna sokarak çıktı üretir.

Çıkış katmanı: Son çıktıları sağlar. Bu genellikle bir sınıflandırma problemi ise sınıf etiketlerini içerir veya bir regresyon problemi ise sürekli bir değeri temsil eder.

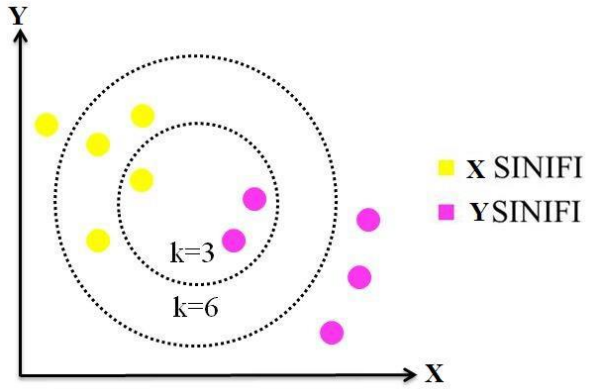
MLP'lerin eğitimi, giriş verileriyle beklenen çıktılar arasındaki farkı minimize ederek gerçekleştirilir. Bu genellikle bir optimizasyon algoritması ve bir geri yayılım (backpropagation) algoritması kullanılarak yapılır. Geri yayılım, ağın çıkışındaki hata miktarını geriye doğru hesaplayarak ağırlıkların güncellenmesini sağlar, böylece hatayı azaltır ve ağın performansını artırır.

4.2.2. k-en yakın komşuluk

k-en yakın komşuluk yönteminin tarihçesi onlarca yıl öncesine dayanmaktadır (Chakrabarti, S., vd., 2008). k-NN algoritması 1950'li yılların başlarında ilk kez tanımlanmış olmasına rağmen 1960'lı yıllara kadar popüler olamamıştır. Çünkü o zamanlardaki hesaplama gücü, k-en yakın komşu algoritmasının ihtiyaç duyduğu büyük çapta bir eğitim veri seti için henüz yeterli değildi. Ancak artan hesaplama gücü ile birlikte, algoritma özellikle örüntü tanıma alanında yaygın bir şekilde kullanılmaktadır (Han, J., & Kamber, M., 2001).

Örüntü tanımada, özel uzayda en yakın eğitim örneklerine dayalı nesneleri sınıflandırmak için k-NN algoritması kullanılan bir yöntem olup bu sınıflandırma algoritmasında; hangi sınıfa ait olduğu bilinmeyen verilerin (test örnekleri), örneği önceden sınıflandırılmış öğrenme örneklerine olan uzaklıkları hesaplanır ve en yakın olduğu düşünülen k tane veride bulunan en fazla sınıfa göre sınıflandırma yapılır (Lee, S., vd., 2014).

Kısaca; k-NN olarak ifade edilen k-en yakın komşu sınıflandırma algoritmasının temelinde “birbirine yakın olan nesneler muhtemelen aynı kategoriye aittir” mantığı yatmaktadır (Eren, Ö., 2008). Şekil 4.3’te k-en yakın komşuluk algoritmasına ait bir sınıflandırma modeli gösterilmiştir.



Şekil 4.3. K-En yakın komşu algoritma görüntüsü

k-NN algoritması örnek tabanlı öğrenmenin örneğidir. Örnek tabanlı öğrenmede eğitim örnekleri birebir saklanmakta ve bilinmeyen bir test örneğinin eğitim setinin hangi üyesine en yakın olduğuna karar vermek için uzaklık fonksiyonu kullanılmaktadır. En yakın eğitim örneği bulunduktan sonra, test örneği için sınıf tahmin edilmektedir (Chakrabarti, S., vd., 2008). Bunun için de test örneği eldeki mevcut eğitim kümesindeki en benzer olduğu kayıtlarla karşılaştırılmaktadır (Larose, D. T., 2005).

k-NN algoritmasının en büyük avantajlarından birisi uygulama açısından kolay olmasıdır. Üstelik kayıtların özellikleri dikkatli seçilmiş ve uzaklık hesaplamalarında dikkatli ağırlıklandırılmış ise, oldukça iyi performans vermektedir (Nisbet, R., vd., 2009).

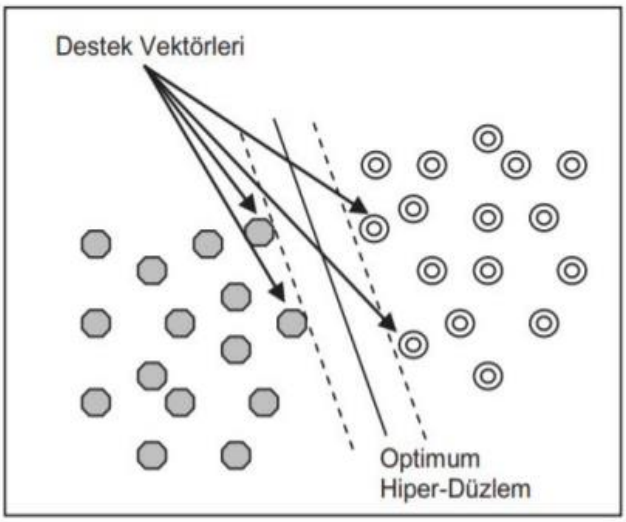
Yeni veri örneği k en yakın komşunun çoğunluğunda olduğu belli bir sınıfa atanır. Örneğin; k=1 seçilmişse sınıflandırılmaya çalışılan yeni veri örneği en yakında olan 1 komşusunun ait olduğu sınıfa atanır, k=3 seçilmişse, sınıflandırılmaya çalışılan yeni veri için en yakında olan 3 komşusu sınıf bilgileri dikkate alınarak sınıflandırılır. Şekil 4.3’te k=3 olarak belirlendiğinde yeni gelen örneğin sınıflandırılması durumu görülmektedir. Burada d1, d2, d3 karşılaşılan yeni verinin sınıflardaki verilere olan en yakın 3 komşusuna olan mesafesidir. Bu işlem sonrasında en yakın komşulukta ikisinin Yıldız Sınıfı’na daha yakın olduğundan veri Yıldız Sınıfı’na atanır (Polat, H., 2016).

4.2.3. Destek vektör makineleri

Temelde istatistiksel öğrenme algoritmalarından biri olan ve Vapnik-Chervonenkis tarafından tanımlanan, Destek Vektör Makinaları (Support Vector Machine – SVM) birçok gerçek sorunda başarılı sonuçlar vermiştir. SVM'nin dayandığı teori, Vladimir Vapnik ve Alexey Chervonenkis tarafından 1960'lı yıllarda başlatılıp 1970'li yıllarda gelişen başarılı bir çalışmanın eseridir. Fakat ilk başarılı uygulamaları 1990'lı yıllarda gerçekleştirilmiş olup, bu uygulamalardan sonra matematikçilerin ve yapay zeka bilim adamlarının ilgi odağı haline gelmiştir. SVM'nin asıl mantığı doğrusal olarak ayrıştırılabilen veri yapıları için en iyi ayırıcı düzlemin belirlenmesidir. Doğrusal olarak ayrıştırılamayan veri yapıları dönüşüm tekniği ile farklı bir boyuta taşınarak çözülür. Destek vektörü öğrenme, basit fikirler üzerine kurulma ve pratik uygulamalarda yüksek performans göstermesi bakımından oldukça kullanışlıdır. SVM'lerde kullanılacak örnek sayısı önemli değildir. SVM eğitim esnasında görülmemiş verileri de sorunsuz olarak sınıflandırır. Bu SVM'nin genelleştirebilme yeteneğini gösterir. Genelleştirebilme özelliği SVM'yi diğer tekniklere göre (YSA, karar ağacı vs.) iyi bir alternatif sunmaktadır. Son zamanlarda ise örüntü tanıma, yüz bulma-tanıma, veri madenciliği, dil yapısını inceleyen mantıksal programlamalarda, uçak alt basınç profillerinin modellenmesi, biyoloji ve diğer bioinformatik uygulamalarda, gen analizlerinde ve proteinlerin sınıflandırılmasında SVM kullanılmaya başlanmıştır. (Kecman, V., 2001).

SVM iki boyutlu uzayda doğrusal, üç boyutlu uzayda düzlemsel ve çok boyutlu uzayda hiperdüzlemsel olarak ayırma mekanizmaları ile veriyi iki veya daha çok sınıfa ayırma yeteneğine sahip denetimli bir sınıflandırma algoritmasıdır (Güran, A., vd., 2014). SVM algoritması, biyoenformatik, görüntü, konuşma, tanıma, çevre bilimleri ve duygu analizi gibi çeşitli uygulama alanlarında kullanılmaktadır. SVM kullanılarak sınıflandırılan bir veri kümesinin tek bir doğru ile ayrılma durumu lineer ayrılma olarak tanımlanmaktadır (Cortes, C., & Vapnik, V., 2001).

Şekil 4.4'te hiperdüzlem üzerinde destek vektör makinesi tarafından lineer olarak sınıflandırılan veri örneği gösterilmektedir.



Şekil 4.4. SVM ile lineer veri sınıflandırması (Kavzoğlu, T., ve Çölkesen, İ., 2010)

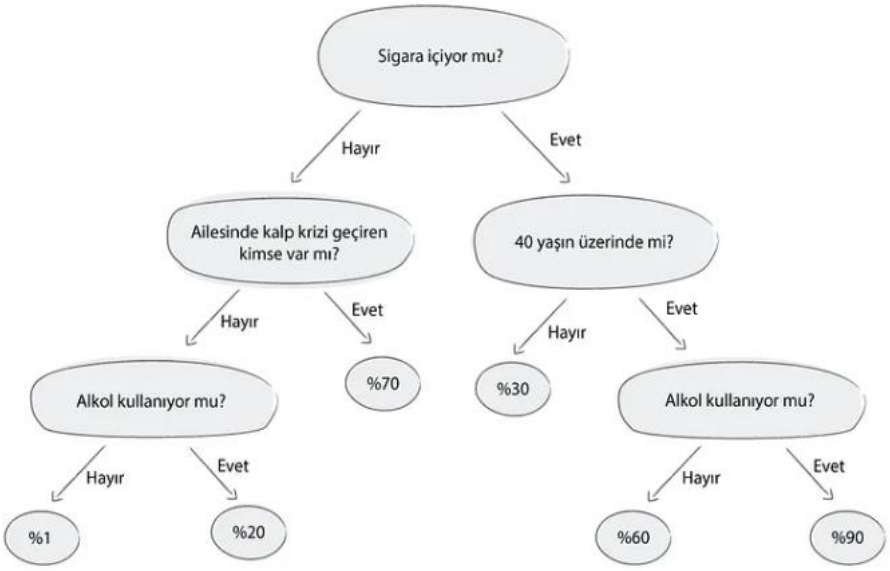
4.2.4. Karar ağacı

Karar ağacı (Random Tree – RT) algoritması, olası ağaç kümelerinden rastgele bir ağaç seçer. Ağaç kümesi içindeki her bir ağaç eşit bir şekilde denenme şansına sahiptir ve ağaçların dağılımı eşitlik gösterir. Rastgele ağaçlar, etkin bir şekilde oluşturulabilir ve birçok rastgele ağacın biraraya geldiği modeller genellikle yüksek doğruluk oranına sahip olur (Onan, A., 2015). RT algoritması, her düğümde belirli miktarda rastgele seçilmiş özellikleri kullanarak ağacı oluşturur (Breiman, L., 2001).

Karar ağaçları, giriş verilerinin bir sınıflandırma ya da kümeleme algoritması ile, tüm elemanların aynı sınıf etiketine ait olana kadar ağaç dallanmasına benzer biçimde alt gruplara ayırma işlemidir (Orhan, U., 2012). Bu ayrıştırma işlemi için entropi, bilgi kazancı veya doğruluk oranı vs. gibi ölçütler kullanılır. Karar ağaçları sınıflandırma, özellik seçimi, özellik çıkarımı ve karar kurallarının oluşturulmasında kullanılabilir. Ayrıca farklı algoritmalarla melez (hibrit) teknikler kullanılarak daha iyi sonuçlar elde edilir. RT başarılı bir performans ortaya koyan, düşük maliyetli, güvenilir, uygulaması kolay, anlaşılabilir ve yorumlanabilir olması nedeniyle çok tercih edilen veri madenciliği metodudur. Bu nedenle veri madenciliği yöntemlerinden karar ağaçları en sık kullanılan algoritmalardandır.

Karar ağacı öğrenmesinde bir ağaç yapısı oluşturulup ağacın yaprakları seviyesinde sınıf etiketleri, bu yapraklara giden ve başlangıçtan ayrılan kollar ile de özellikler üzerindeki işlemeler tanımlanmaktadır. Karar ağacı öğrenmesi esnasında

öğrenilen bilgi bir ağaç üzerinde modellenir. Bu ağacın bütün iç düğümleri (interior nodes) birer girdiyi ifade etmektedir. Şekilde örnek bir öğrenilmiş ağaç verilmiş olup MI geçiren bireylerin çeşitli özellikleri ele alınarak veriler üzerinden öğrenilmesi durumunda hayali olarak elde edilecek sonuçlar verilmiştir.



Şekil 4.5. Öğrenilmiş ağaç örneği (Aydemir, H., 2020)

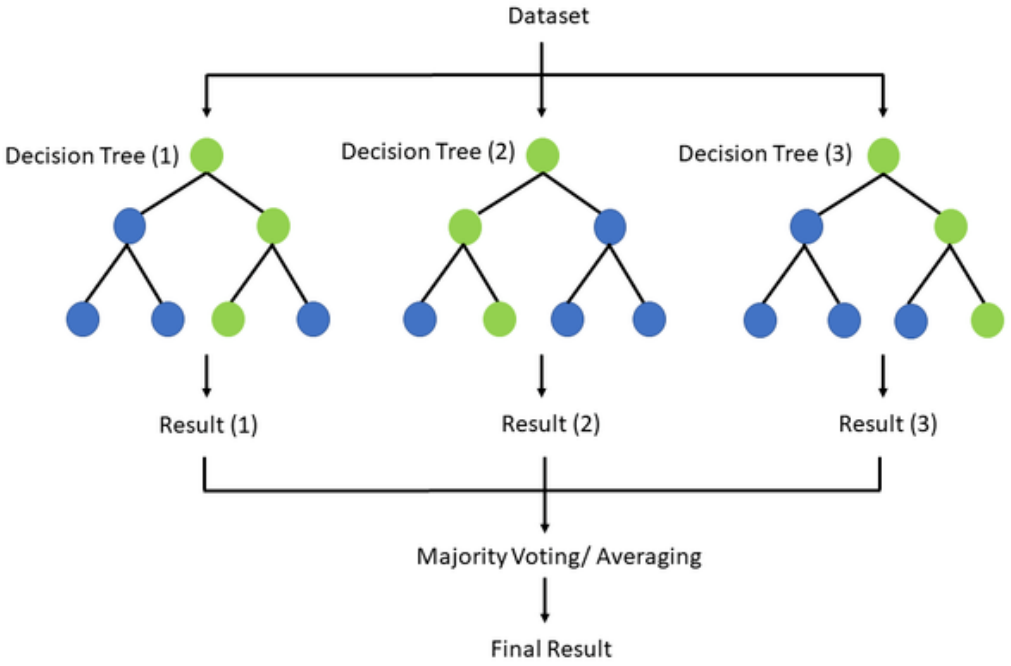
4.2.5. Rassal orman

Karar ağaçları algoritmasının anlaşılması, rassal orman (Random Forest – RO) algoritmasını daha iyi anlamamıza yardımcı olur. Birden fazla karar ağacı kullanarak daha uyumlu modeller üretilmesine fırsat verir, bu şekilde elimizdeki veri setinin daha uyumlu bir sınıflandırma yapmasını bu teknik sayesinde sağlayabiliriz (Hegde, C., vd., 2015).

Karar ağacı algoritmasını n defa çalıştırılarak daha iyi tahminler yapılabilir. Karar ağaçları veri setinden rastgele seçilmiş alt kümelerdir. Bu karar ağaçları bir araya gelerek karar ormanlarını oluşturur. Kayıp verili veya dengesiz veri setlerinde iyi sonuçlar üretir (Ayyadevara, V. K., 2018).

RO algoritması 2001 yılında Breiman tarafından geliştirilmiştir. Bu algoritma, tek bir karar ağacı oluşturmak yerine, birçok karar ağacının kararlarını birleştirir. Bu karar ağaçlarının her biri ayrı ayrı eğitim kümelerinde eğitilmiştir. Ayrı eğitim kümeleri için aynı dağılımlı rasgele özellik seçimi kullanılır. Karar ağaçlarını oluştururken, her seviyedeki özniteliği belirlemede önce tüm ağaçlarda çeşitli bir takım hesaplamalar

yapılır ve nitelikler belirlenir. Daha sonra diğer ağaçlardaki nitelikler birleştirilir ve en çok kullanılan öznitelik seçilir. Bu öznitelik karar ağacına dâhil edildikten sonra diğer aşamalarda da aynı işlemler tekrarlanır. Random forest, karar ağacı oluşturmak için CART (Classification and Regression Tree) algoritmasını kullanır. Karar ağacındaki düğüm ve dallar bu algoritmanın biçimine göre oluşturulur (Daş, B., ve Türkoğlu, İ., 2014).



Şekil 4.6. Random forest (TseKiChun, 2021)

4.2.6. Lojistik model ağacı

Lojistik model ağacı (Logistic Model Tree – LMT) algoritması, lojistik regresyon ve karar ağacı öğrenmesi algoritmalarının birleşiminden meydana gelen, gözetimli bir sınıflandırma biçimidir (Daş, B., ve Türkoğlu, İ., 2014).

Sıradan karar ağaçları yaprakları parçalı bir sabit modeli oluştururken, lojistik modeli ağacı yaprakları parçalı bir doğrusal regresyon sağlayan, doğrusal regresyon modeline sahip bir karar ağacıdır (Landwehr, N., vd., 2003).

Lojistik varyantta, LogitBoost algoritması, ağacın her düğümünde bir lojistik regresyon modeli üretmek için kullanılır, düğüm daha sonra random tree algoritmassını

kullanılarak ayrılır. Her LogitBoost yürütmesinde kendi sonuçlarından üst düğüm üzerinden yeniden başlatılır. En son olarak, ağaç budanır (Sumner, M., vd., 2005).

4.3. Algoritmaların Performans Ölçütleri

Makine öğrenmesi algoritmalarıyla elde edilen sonuçlar nesnel değerlendirme ölçütleri kullanılarak performansları karşılaştırılmıştır. Kullanılan modellerin başarımlarını ölçmede, literatürde yaygın bir şekilde kullanılan metriklerden faydalanılmıştır. Bu metrikler karmaşıklık/hata matrisinden (confusion matrix) elde edilen doğruluk (accuracy), duyarlılık/hatırlama (sensitivity/recall), seçicilik (specifity), kesinlik (precision), F-skor (F-measure), Alıcı Çalışma Karakteristik Eğrisi (Receiver Operating Characteristic Curve – ROC) ve Eğrinin Altındaki Alan (Area Under the Curve – AUC) gibi metriklerden oluşmaktadır.

Karmaşıklık matrisi çıktının iki veya daha fazla sınıf olabileceği, makine öğrenimi sınıflandırma problemi için bir performans ölçümüdür (Narkhede, S., 2018). Tahmin edilen ve gerçek değerlerin 4 farklı kombinasyonunu içeren bir tablodur. Bu çalışmada iki sınıf için (ölü/sağ) oluşturulan karmaşıklık matrisi örneği Tablo 4.3'de gösterilmişitr. Tablodan görüleceği üzere, matrisin sütunları örneklere ait gerçek sınıfları, satırları ise öngörülen/tahmin edilen sınıfları ifade etmektedir.

Tablo 4.3. Karışıklık Matrisi Örneği

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	**TP**	**FP**
	Ölü Negatif (1)	**FN**	**TN**

Karmaşıklık matrisi doğruluk, duyarlılık, seçicilik, kesinlik ve en önemlisi AUC-ROC eğrilerini ölçmek için son derece kullanışlıdır. Öncelikle TP, FP, FN ve TN kavramlarını ölü/sağ analojisi açısından anlamaya çalışalım:

TP (True Positive): Doğru/gerçek pozitif. Olumlu tahminde bulunulduğunu ve bunun doğru olduğunu ifade eder. Yani hastanın sağ olduğu tahmin edildi ve bu doğru.

TN (True Negative): Doğru/gerçek negatif. Olumsuz tahminde bulunulduğunu ve bunun doğru olduğunu ifade eder. Yani hastanın ölü olduğu tahmin edildi ve bu doğru.

FP (False Positive): Yanlış pozitif. Olumlu tahminde bulunulduğunu ve bunun yanlış olduğunu ifade eder. Yani hastanın sağ olduğu tahmin edildi ve bu yanlış. Bu hata literatürde Tip I hata olarak adlandırılmaktadır.

FN (False Negative): Yanlış negatif. Olumsuz tahminde bulunulduğunu ve bunun yanlış olduğunu ifade eder. Yani hastanın ölü olduğu tahmin edildi ve bu yanlış. Bu hata literatürde Tip II hata olarak adlandırılmaktadır.

Tahmin edilen değerlerin pozitif-negatif, gerçek değerlerin ise doğru-yanlış olarak tanımlandığı görülmektedir. Şimdi doğruluk, duyarlılık, seçicilik, kesinlik, F-skor ve en önemlisi AUC-ROC değerlerinin nasıl hesaplandığı aşağıda özetlenmiştir.

Doğruluk (Accuracy): Doğruluk Denklem (4.1)'de verildiği üzere, tüm sınıflardan (olumlu ve olumsuz), kaç tanesini doğru tahmin edildiğini ifade eder. Yani burada tüm hastalardan (sağ ve ölü) kaçının gerçekten doğru tahmin edildiğini belirtir. Doğruluk mümkün olduğunca yüksek olmalıdır.

$$Doğruluk = \frac{TP+TN}{TP+FP+FN+TN} \tag{4.1}$$

Duyarlılık/Hatırlama (Sensitivity/Recall): Duyarlılık Denklem (4.2)'de verildiği üzere, tüm pozitif sınıflardan kaç tanesinin doğru tahmin edildiğini ifade eder. Yani burada sağ olan hastalardan kaçının gerçekten sağ olduğunu belirtir. Duyarlılık mümkün olduğunca yüksek olmalıdır.

$$Duyarlılık = \frac{TP}{TP+FN} \tag{4.2}$$

Kesinlik (Precision): Kesinlik Denklem (4.3)'de verildiği üzere, pozitif olarak tahmin ettiğimiz tüm sınıflardan kaçının gerçekten pozitif olduğunu ifade eder. Yani burada sağ olarak tahmin edilen tüm hastalardan kaçının gerçekten sağ olduğunu belirtir. Kesinlik mümkün olduğunca yüksek olmalıdır.

$$Kesinlik = \frac{TP}{TP+FP} \tag{4.3}$$

Seçicilik (Specifity): Seçicilik Denklem (4.4)'de verildiği üzere, tüm negatif sınıflardan kaç tanesinin doğru tahmin edildiğini ifade eder. Yani burada ölü olan hastalardan kaçının gerçekten ölü olduğunu belirtir. Seçicilik mümkün olduğunca yüksek olmalıdır.

$$Seçicilik = \frac{TN}{TN+FP} \tag{4.4}$$

F-skor (F-measure): F-skor Denklem (4.5)'de verilmiştir. Düşük kesinlik ve yüksek duyarlılık ile iki modeli karşılaştırmak zordur veya bunun tersi de geçerlidir. Bu yüzden onları karşılaştırılabilir hale getirmek için F-skor kullanılmaktadır. F-skoru, duyarlılık ve kesinliği aynı anda ölçmeye yardımcı olur ve uç değerleri daha fazla cezalandırarak aritmetik ortalama yerine harmonik ortalama kullanır.

$$F - skor = \frac{2*duyarlılık*kesinlik}{duyarlılık+kesinlik} \tag{4.5}$$

ROC ve AUC: Alıcı Çalışma Karakteristik Eğrisi analizi ve Eğrinin Altındaki Alan farklı parametrelendirmeler altında bir modelin kalitesini değerlendirmek, iki veya daha fazla modelin performansını karşılaştırmak için sinyal işlemeden ödünç alınan ve veri biliminde yaygın olarak kullanılan araçlardır (Bento, C., 2022). ROC, kesinlik ve duyarlılık arasındaki dengeyi görselleştirmek için kullanılan bir özet aracıdır (Gareth, J., vd., 2013). ROC eğrisi, x eksenindeki yanlış pozitif oranın y eksenindeki gerçek pozitif orana karşı çizilmesiyle elde edilir. Gerçek pozitif oran bir sinyali algılama olasılığı ve yanlış pozitif oran yanlış alarm olasılığı olduğundan, ROC analizi, hastalıkları veya diğer davranışları güvenle tespit eden eşikleri belirlemek için tıbbi çalışmalarda da yaygın olarak kullanılmaktadır (Cook, N. R., 2007).

Kesinlik ve duyarlılık gibi geleneksel performans ölçümleri, büyük ölçüde olumlu gözlemlere dayanır. Bunun yerine, ROC ve AUC, kaliteyi değerlendirmek için hem olumlu hem de olumsuz gözlemleri hesaba katan gerçek pozitif ve yanlış pozitif oranları kullanır.

AUC tıbbi ve istatistiksel analizlerde sıklıkla kullanılan bir ölçümdür. Özellikle sınıflandırma modellerinin performansını değerlendirmek için kullanılır. ROC eğrisi altında kalan alanı ifade etmektedir. AUC'nin değeri 0-1 arasında olur, daha yüksek bir AUC değeri, modelin daha iyi performans gösterdiği anlamına gelir. AUC'nin değeri 0.5'e

yaklaşıyorsa, modelin rastgele tahmin yapma yeteneği ile benzer olduğunu, 1'e yaklaşıyorsa mükemmel bir sınıflandırma yeteneği olduğunu gösterir. Bu nedenle, AUC değeri, bir sınıflandırma modelinin doğruluğunu ve ayrıca sınıflar arasındaki ayrım yeteneğini ölçmek için yaygın olarak kullanılan bir ölçümdür.

5. ARAŞTIRMA BULGULARI VE TARTIŞMA

Çalışma için toplanan veriler SPSS® (IBM Statistical Package for the Social Sciences) istatistik programı kullanılarak değerlendirilmiştir. Toplam 735 hastayı içeren veri setine ait parametreler ve bu parametrelere ait bazı istatistiki değerler (min, max, ortalama ve standart sapma) Tablo 5.1'de verilmiştir.

Tablo 5.1. Orijinal Veri Seti

Veri türleri	Hasta Sayısı	Minimum	Maxsimum	Ortalama Değer	Standart Sapma
Hastaneiçiölümü	735	0	1	0,12	0,33
MIsaati	735	0	23	13,24	5,21
Yaş	735	6	93	61,74	14,10
KalpHızı	735	28	174	82,90	18,11
Killip	735	1	4	1,44	0,89
Saatlatency	735	1	24	6,53	4,24
Arrestonadmission	735	0	1	0,09	0,28
Cinsiyet	735	0	1	0,71	0,46
HTöyküsü	735	0	1	0,35	0,48
DMöyküsü	735	0	1	0,29	0,46
HPLöyküsü	735	0	1	0,05	0,22
Sigaraöyküsü	735	0	1	0,50	0,50
MIöyküsü	735	0	1	0,05	0,21
PCIöyküsü	735	0	1	0,07	0,25
CABGöyküsü	735	0	1	0,02	0,15
SVOöyküsü	735	0	1	0,04	0,19
AileÖyküsü	735	0	1	0,13	0,34
KBYöyküsü	735	0	1	0,01	0,10
Kalpyetersizliği	735	0	1	0,11	0,32
Akciğerödemi	735	0	1	0,05	0,21
KardiyojenikŞok	735	0	1	0,09	0,28
AVTamBlok	735	0	1	0,04	0,20
VT	735	0	1	0,09	0,29
VF	735	0	1	0,08	0,27
Kanşekeri	731	29	729	174,05	95,36
Albumin	520	0,6	4,6	3,53	0,48
Trigliserid	703	14	1297	152,20	110,20
Kolesterol	703	40	499	184,35	47,41
HDL	704	5	128	34,91	10,34
LDL	693	36	400	120,82	38,25
WBC	534	0,06	54,3	13,44	5,52

HB	724	2,87	19,93	13,68	1,95
Lenfosit	724	0,07	22,3	2,03	1,60
Monosit	724	0,03	5,08	0,68	0,38
Nötrofil	724	0,05	44,4	10,86	5,35
EF	679	0	65	44,47	10,40
MItipi	706	1	5	2,18	1,22
Sorumludamar	705	1	4	1,77	0,78

5.1. Makine Öğrenmesi Yöntemlerinin Orjinal Veri Setine Uygulanması

Tablo 5.1'de görüleceği üzere, özellikle bazı parametrelerde (albumin 215, WBC 201, EF 56, LDL 42 vb.) ciddi oranda eksik/kayıp değerlerin olduğu görülmektedir. Bazı hastalara ait bu kayıp değerler, SPSS programı yardımıyla, verinin özelliğine bağlı olarak program tarafından otomatik olarak seçilen çoklu veri atama (multiple imputation) işlemiyle tamamlanmıştır. Böylece tüm veri setindeki hastalara ait parametreler eksiksiz bir şekilde oluşturulmuştur. Tamamlanan bu orjinal veri seti incelendiğinde, 735 hastaya (644 sağ - 91 ölü) ait veride, mortalitesi gerçekleşmiş hasta sayısının çok düşük olduğu görülmektedir.

Bu orjinal veri seti kullanılarak yapılan testlerde, kullanılan makine öğrenmesi algoritmalarının hemen hemen hepsinde sınıflandırma performansının çok düşük olduğu gözlemlenmiş olup, elde edilen testlerden çıkan sonuçlar Tablo 5.2-Tablo 5.7'de verilmiş ve sonrasında karşılaştırma amaçlı ayrı bir tablo oluşturulup sonuçlar Tablo 5.8' de gösterilmiştir.

Tablo 5.2. MLP algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	590	41
	Ölü Negatif (1)	54	50

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**

MLP Algoritması	%87,1	%91,6	%93,5	%54,9	%92,5	%81

Tablo 5.2'de MLP algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; MLP algoritmasının ortalamada doğru sınıflandırma oranı % 87,1 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 91,6 %93,5 ve %54,9 olarak bulunmuştur. F-skor %92,5 ve AUC değeri de %81 olarak elde edilmiştir.

Tablo 5.3. k-NN algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	614	66
	Ölü Negatif (1)	30	25

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
k-NN Algoritması	%86,9	%95,3	%90,3	%27,5	%92,7	%61,5

Tablo 5.3'te k-NN algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; k-NN algoritmasının ortalamada doğru sınıflandırma oranı % 86,9 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 95,3 %90,3 ve %27,5 olarak bulunmuştur. F-skor %92,7 ve AUC değeri de %61,5 olarak elde edilmiştir.

Tablo 5.4. SVM algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	644	91
	Ölü Negatif (1)	0	0

	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F Skor	AUC
SVM Algoritması	%87,6	%100	%87,6	%0	%93,4	%50

Tablo 5.4'te SVM algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; SVM algoritmasının ortalamada doğru sınıflandırma oranı % 87,6 olarak elde edilmiştir. Duyarlılık %100, kesinlik %87,6, seçicilik %0, F-skor %93,4 ve AUC değeri de %50 olarak elde edilmiştir.

Tablo 5.5. Random tree algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	615	73
	Ölü Negatif (1)	29	18

	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
RT Algoritması	%86,1	%95,5	%89,4	%19,8	%92,3	%57,7

Tablo 5.5'te RT algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; RT algoritmasının ortalamada doğru sınıflandırma oranı % 86,1 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 95,5, %89,4 ve %19,8 olarak bulunmuştur. F-skor %92,3 ve AUC değeri de %57,7 olarak elde edilmiştir.

Tablo 5.6. Random forest algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	644	89
	Ölü Negatif (1)	0	2

	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
RO Algoritması	%87,9	%100	%87,9	%2,2	%93,5	%88,3

Tablo 5.6'da RO algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; RO algoritmasının ortalamada doğru sınıflandırma oranı % 87,9 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 100, %87,9 ve %2,2 olarak bulunmuştur. F-skor %93,5 ve AUC değeri de %88,3 olarak elde edilmiştir.

Tablo 5.7. LMT algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	625	63
	Ölü Negatif (1)	19	28

	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
LMT Algoritması	%88,9	%97	%90,8	%30,8	%93,8	%88,6

Tablo 5.7'de LMT algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; LMT algoritmasının ortalamada doğru sınıflandırma oranı % 88,9 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 97, %90,8 ve %30,8 olarak bulunmuştur. F-skor %93,8 ve AUC değeri de %88,6 olarak elde edilmiştir.

Tablo 5.8'de 735 hasta üzerinden (sağ hasta sayısı 644 ve ölü hasta sayısı 91) yapılan tahminleme sonuçlarının başarı yüzdeleri verilmiştir. Tabloda verilen sonuçlardan da görüleceği üzere algortimaların doğruluk yüzdeleri birbirine yakın olup, en yüksek doğruluk LMT algoritması ve en düşük doğruluk ise RT algoritmasında elde edilmiştir. Algoritmaların performansı genel olarak değerlendirildiğinde, neredeyse tüm algoritmaların duyarlılık-kesinlik göstergelerinin ve dolayısıyla F-skor değerinin yüksek olduğu, yani sağ hasta sayısının tahmin edilmesinde yüksek bir başarım gösterdiği görülmektedir. Ancak, ölümü gerçekleşmiş bireylerdeki tahmin performansının tüm algoritmalarda genel olarak çok düşük olduğu (seçicilik göstergesi) görülmektedir. Tablo

5.8'den de görüleceği üzere, seçicilik yönünden tüm algoritmaların kötü bir performans sergilediği, en yüksek performansa MLP algoritmasının ve en düşük performansa da SVM algoritmasının sahip olduğu görülmektedir.

AUC göstergesi bir sınıflandırma modelinin doğruluğunu ve ayrıca sınıflar arasındaki ayrım yeteneğini ölçmek için yaygın olarak kullanılan bir ölçümdür. Büyük ölçüde olumlu gözlemlere dayanan kesinlik ve duyarlılık gibi geleneksel performans ölçümler yerine, kaliteyi değerlendirmek için hem olumlu hem de olumsuz gözlemleri hesaba katan gerçek pozitif ve yanlış pozitif oranları kullanan AUC ölçümü (ROC eğrisi altındaki alan) ile performans değerlendirmesi yapmak bu analizin daha rasyonel bir şekilde yapılmasını sağlayacaktır. Bu yönüyle bakıldığında en yüksek AUC değeri LMT ve en düşük AUC değeri ise SVM algoritması ile elde edilmiştir.

Tüm algoritmalarda, ölümlü bireylerdeki başarım oranlarının bu denli düşük olmasının temel olarak dengesiz veri setinden kaynaklandığı düşünülmektedir. Orjinal veri setinde yer alan ölü/sağ hasta sayısının dengesiz bir biçimde dağılması algoritmaların performansını ciddi oranda etkilemiştir. Veri setinde yer alan ölümlü hasta sayısının az olması sebebiyle öğrenme işlemi yeterli bir şekilde gerçekleşememiştir. Bu sebeple bir sonraki bölümde veri artırma yoluna gidilmiş ve artırılmış veri ile algoritmaların performansı ele alınmıştır.

Tablo 5.8. Algoritmaların tahminleme sonuçlarının özeti

Algoritma	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
MLP	%87,1	%91,6	%93,5	%54,9	%92,5	%81
k-NN	%86,9	%95,3	%90,3	%27,5	%92,7	%61,5
SVM	%87,6	%100	%87,6	%0	%93,4	%50
RT	%86,1	%95,5	%89,4	%19,8	%92,3	%57,7
RO	%87,9	%100	%87,9	%2,2	%93,5	%88,3
LMT	%88,9	%97	%90,8	%30,8	%93,8	%88,6

5.2. Makine Öğrenmesi Yöntemlerinin Arttırılmış Veri Setine Uygulanması

Orijinal veri seti üzerinden yapılan testlerin sonuçları Bölüm 5.1 'de verilmişti. Elde edilen bu sonuçlar incelendiğinde, orijinal veri setinin verimli bir sınıflandırma için

yeterli olmadığı görülmüştür. Bunun üzerine veri setindeki hasta sayısının artırımı yoluna gidilmiştir. Orijinal veri seti üzerinde rezervuar örnekleme yöntemi kullanılarak (Vitter, J. S., 1985), veri sayısı 644'ü sağ ve 546'sı ölü olmak üzere, toplam 1190'a çıkarılmıştır. Rezervuar örneklemesinin makine öğrenimi, veri analizi ve bilgisayar grafikleri gibi alanlarda birçok uygulaması vardır. Örneğin, bir makine öğrenimi modelini eğitmek için büyük bir veri kümesinin temsili bir alt kümesini seçmek veya rastgele bir veri oluşturmak için kullanılabilir. Burada rezervuar ile rassal örnekleme yöntemi kullanılarak veri artırımı yapılmıştır. Böylece arttırılmış veriyle veri seti dengeli hale getirilmiştir. Arttırılmış veri seti üzerinde yapılan sınıflandırma işleminin sonuçları Tablo 5.9-Tablo 5.14'te verilmiştir. Burada ki bütün tahminleme sonuçları tek bir tablo haline getirilip başarı yüzdesi verilmiştir (Tablo 5.15).

Tablo 5.9. MLP algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	594	7
	Ölü Negatif (1)	50	539

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
MLP Algoritması	%95,2	%92,2	%98,8	%98,7	%95,4	%97,3

Tablo 5.9'da MLP algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre MLP algoritmasının ortalamada doğru sınıflandırma oranı % 95,2 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 92,2, %98,8 ve %98,7 olarak bulunmuştur. F-skor %95,4 ve AUC değeri de %97,3 olarak elde edilmiştir.

Tablo 5.10. k-NN algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	611	1
	Ölü Negatif (1)	33	545

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
k-NN Algoritması	%97,1	%94,9	%99,8	%99,8	%97,3	%97,4

Tablo 5.10'da k-NN algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; k-NN algoritmasının ortalamada doğru sınıflandırma oranı % 97,1 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 94,9, %99,8 ve %99,8 olarak bulunmuştur. F-skor %97,3 ve AUC değeri de %97,4 olarak elde edilmiştir.

Tablo 5.11. SVM algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	644	1
	Ölü Negatif (1)	0	545

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
SVM Algoritması	%99,9	%100	%99,8	%99,8	%99,9	%99,9

Tablo 5.11'de SVM algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; SVM algoritmasının ortalamada doğru sınıflandırma oranı %99,9 olarak elde edilmiştir.

Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 100, %99,8 ve %99,8 olarak bulunmuştur. F-skor %99,9 ve AUC değeri de %99,9 olarak elde edilmiştir.

Tablo 5.12. Random tree algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	583	0
	Ölü Negatif (1)	61	546

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
RT Algoritması	%94,9	%90,5	%100	%100	%95	%95,3

Tablo 5.12'de RT algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; RT algoritmasının ortalamada doğru sınıflandırma oranı %94,9 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla %90,5, %100 ve %100 olarak bulunmuştur. F-skor %95 ve AUC değeri de %95,3 olarak elde edilmiştir.

Tablo 5.13. Random forest algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	621	1
	Ölü Negatif (1)	23	545

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F Skor**	**AUC**
RO Algoritması	%98	%96,4	%99,8	%99,8	%98,1	%100

Tablo 5.13’te RO algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; RO algoritmasının ortalamada doğru sınıflandırma oranı %98 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla %96,4, %99,8 ve %99,8 olarak bulunmuştur. F-skor %98,1 ve AUC değeri de %100 olarak elde edilmiştir.

Tablo 5.14. LMT algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	598	1
	Ölü Negatif (1)	46	545

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
LMT Algoritması	%96,1	%92,9	%99,8	%99,8	%96,2	%97,7

Tablo 5.14’te LMT algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; LMT algoritmasının ortalamada doğru sınıflandırma oranı %96,1 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 92,9, %99,8 ve %99,8 olarak bulunmuştur. F-skor %96,2 ve AUC değeri de %97,7 olarak elde edilmiştir.

Tablo 5.15’te kullanılan tüm sınıflandırma algoritmalarının performans göstergeleri özet bir tablo olarak verilmştir. Tablo 5.15’te görüldüğü üzere, algoritmaların sınıflandırmadaki ortalama doğruluk yüzdeleri birbirine yakın olup, en yüksekten en düşüğe göre sırasıyla SVM, RO, k-NN, LMT, MLP ve RT algoritmaları ile elde edilmiştir. Algoritmaların performansı genel olarak değerlendirildiğinde, neredeyse tüm algoritmaların duyarlılık, kesinlik ve F-skor göstergelerinin yüksek olduğu ve dolayısıyla sağ hasta sayısının tahmin edilmesinde yüksek bir başarım gösterdikleri görülmektedir.

Aynı şekilde ölümü gerçekleşmiş bireylerdeki tahmin performansının, yani seçicilik parametresinin, tüm algoritmalarda birbirine yakın olmakla birlikte sağ hasta sayısının tahmin edilmesine kıyasla daha da yüksek olduğu göze çarpmaktadır. Bu da algoritmaların genel olarak ölümlü hastalarda da çok iyi bir sınıflandırma performansına sahip olduğunu göstermektedir. Tabloda verilen algoritmalara ait AUC göstergeleri de

bunu teyit etmektedir. Algoritmaların hem F-skor hem de AUC değerleri dikkate alınarak yapılan değerlendirmede, en iyi performansın birbirine çok yakın değerlere sahip SVM (%99,9 F-ksor ve AUC) ve RO (%98,1 F-skor ve %100 AUC) algoritmaları ile elde edildiği; ve en düşük performansın da %95 F-skor ve %95,3 AUC değerleri ile RT algoritması ile elde edildiği görülmektedir.

Tablo 5.15. Artırılımış veride algoritmaların tahminleme sonuçlarının özeti

Algoritma	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
MLP	%95,2	%92,2	%98,8	%98,7	%95,4	%97,3
k-NN	%97,1	%94,9	%99,8	%99,8	%97,3	%97,4
SVM	%99,9	%100	%99,8	%99,8	%99,9	%99,9
RT	%94,9	%90,5	%100	%100	%95	%95,3
RO	%98	%96,4	%99,8	%99,8	%98,1	%100
LMT	%96,1	%92,9	%99,8	%99,8	%96,2	%97,7

5.3. Makine Öğrenmesi Yöntemlerinin Artırılmış Veri Setinde Önemli Görülen Parametrelere Uygulanması

Klinik çalışmalarda ve literatürde erken mortaliteyi öngörmenin en önemli bağımsız parametreleri; ileri yaş, yüksek killip derecesi, yüksek kalp hızı, düşük sistolik kan basıncı, anterior infarkt lokalizasyonu olduğu belirlenmiştir (Lee, K. L., vd., 1995; Morrow, D. A., vd., 2000). Literatürde yer alan bu bilgiler ışığında, ölüme sebep olduğu düşünülen en önemli bağımsız parametrelerden ileri yaş, yüksek killip derecesi, yüksek kalp hızı, düşük sistolik kan basıncı ve anterior infarkt lokalizasyonu dikkate alınarak algoritmaların sınıflandırmadaki performansı gözlemlenmiştir. Bu sebeple artırılmış veri setimizde yer alan sadece bu önemli 5 paramatreyi kullanılarak, algoritmaların sınıflandırma performansları test edilmiştir. Elde edilen test sonuçları sırasıyla Tablo 5.16-Tablo 5.21'de verilmiştir.

Tablo 5.16. MLP algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	546	139
	Ölü Negatif (1)	98	407

	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
MLP Algoritması	%80,1	%84,8	%79,7	%74,6	%82,2	%86,4

Tablo 5.16‘da MLP algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; MLP algoritmasının ortalamada doğru sınıflandırma oranı %80,1 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik sırasıyla %84,8, %79,7 ve %74,6 olarak bulunmuştur. F-skor %82,2 ve AUC değeri de %86,4 olarak elde edilmiştir.

Tablo 5.16. k-NN algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	594	0
	Ölü Negatif (1)	50	546

	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
k-NN Algoritması	%95,8	%92,2	%100	%100	%96	%96,3

Tablo 5.17’de k-NN algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; k-NN algoritmasının ortalamada doğru sınıflandırma oranı % 95,8 olarak elde

edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 92,2, %100 ve %100 olarak bulunmuştur. F-skor %96 ve AUC değeri de %96,3 olarak elde edilmiştir.

Tablo 5.17. SVM algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	616	3
	Ölü Negatif (1)	28	543

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
SVM Algoritması	%97,4	%95,7	%99,5	%99,5	%97,5	%97,6

Tablo 5.18'de SVM algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; SVM algoritmasının ortalamada doğru sınıflandırma oranı %97,4 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 95,7, %99,5 ve %99,5 olarak bulunmuştur. F-skor %97,5 ve AUC değeri de %97,6 olarak elde edilmiştir.

Tablo 5.18. Random tree algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	581	0
	Ölü Negatif (1)	63	546

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
RT Algoritması	%94,7	%90,2	%100	%100	%94,9	%95,2

Tablo 5.19'da RT algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; RT algoritmasının ortalamada doğru sınıflandırma oranı %94,7 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla %90,2, %100 ve %100 olarak bulunmuştur. F-skor %94,9 ve AUC değeri de %95,2 olarak elde edilmiştir.

Tablo 5.19. Random forest algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	580	0
	Ölü Negatif (1)	64	546

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**
RO Algoritması	%94,6	%90,1	%100	%100	%94,8	%99,6

Tablo 5.20'de RO algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; RO algoritmasının ortalamada doğru sınıflandırma oranı %94,6 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla %90,1, %100 ve %100 olarak bulunmuştur. F-skor %94,8 ve AUC değeri de %99,6 olarak elde edilmiştir.

Tablo 5.20. LMT algoritması sınıflandırma sonucu

		Gerçek Değerler	
		Sağ Pozitif (0)	Ölü Negatif (1)
Tahmin Edilen Değerler	Sağ Pozitif (0)	548	11
	Ölü Negatif (1)	96	535

	Doğruluk	**Duyarlılık**	**Kesinlik**	**Seçicilik**	**F-Skor**	**AUC**

LMT Algoritması	%91	%85,1	%98	%97,9	%91,1	%92,9

Tablo 5.21'de LMT algoritmasının sınıflandırma sonucu verilmiştir. Sonuçlara göre; LMT algoritmasının ortalamada doğru sınıflandırma oranı %91 olarak elde edilmiştir. Duyarlılık, kesinlik ve seçicilik ise sırasıyla % 85,1, %98 ve %97,9 olarak bulunmuştur. F-skor %91,1 ve AUC değeri de %92,9 olarak elde edilmiştir.

Sınıflandırma algoritmalarının veri setinde yer alan ve erken mortaliteye sebep olduğu düşünülen en önemli 5 parametreye uygulanması ile elde edilen performans göstergeleri özet olarak Tablo 5.22'de sunulmuştur. Tablo 5.22'de görüldüğü gibi algoritmaların sınıflandırmadaki ortalama doğruluk yüzdeleri değerlendirildiğinde, MLP algoritması hariç diğer tüm algoritmaların %90'ın üzerinde bir tahminleme yeteneğine sahip olduğu ve %97,4 doğruluk oranına sahip SVM algoritmasının en yüksek performansı sergilediği görülmektedir. Tablodaki sonuçlardan MLP dışında diğer tüm algoritmaların duyarlılık, kesinlik ve F-skor göstergelerinin yüksek olduğu ve dolayısıyla sağ hasta sayısının tahmin edilmesinde yüksek bir başarım gösterdikleri görülmektedir.

Aynı şekilde algoritmaların ölümü gerçekleşmiş bireylerdeki tahmin performansı incelendiğinde, MLP dışındaki algoritmaların çoğunda %100 başarım sağlandığı görülmektedir. Elde edilen bu sonuçlar, literatürde erken mortaliteye sebep olduğu belirtilen ileri yaş, yüksek killip derecesi, yüksek kalp hızı, düşük sistolik kan basıncı ve anterior infarkt lokalizasyonu parametrelerinin en önemli değişkenler olduğu gerçeğini ortaya koymaktadır ve tabloda verilen algoritmalara ait seçicilik göstergesi değerleri bunu teyit etmektedir.

Algoritmaların hem F-skor hem de AUC değerleri dikkate alınarak yapılan değerlendirmede, en iyi performansın yine önceki testlerde olduğu gibi SVM (%97,5 F-ksor ve %97,6 AUC) ve RO (%94,8 F-skor ve %99,6 AUC) algoritmaları ile elde edildiğini ve bu performansı sırasıyla k-NN, RT ve LMT algoritmalarının takip ettiğini söylemek mümkündür. Elde edilen bu bulgular (doğruluk, F-skor ve AUC değerleri) bir önceki bölümde algoritmaların veri setinde yer alan tüm parametrelere (38 parametre) uygulanmasıyla elde edilen sonuçlarla kıyaslandığında, algortimaların sınıflandırma performanslarının MLP dışında benzerlik gösterdiğini ortaya koymaktadır. Bu da erken mortaliteye sebep olan en yüksek etkiye sahip olduğu ileri sürülen bu 5 parametrenin sınıflandırma işlemi için yeterli olduğu ve çok daha az parametre ile çok yüksek doğrulukta sınıflandırma yapabilmenin mümkün olduğunu göstermektedir.

Tablo 5.21. Algoritmaların veri setindeki 5 önemli paramatreye uygulanmasından elde edilen tahmin sonuçlarının özeti

Algoritma	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
MLP	%80,1	%84,8	%79,7	%74,6	%82,2	%86,4
k-NN	%95,8	%92,2	%100	%100	%96	%96,3
SVM	%97,4	%95,7	%99,5	%99,5	%97,5	%97,6
RT	%94,7	%90,2	%100	%100	%94,9	%95,2
RO	%94,6	%90,1	%100	%100	%94,8	%99,6
LMT	%91	%85,1	%98	%97,9	%91,1	%92,9

6. SONUÇ

Bu çalışmada, Dicle Üniversite Hastanesinden temin edilen veri seti kullanılarak, hastane içi ölümü ön görülmesi açısından mortalite riski tahmini ve analizi makine öğrenmesi yöntemleri kullanılarak yapılmıştır. Kullanılan veri seti üzerinden, kalp krizi geçiren hastaların parametrelerinin neler olduğu bilgileri verilerek bu parametrelerin veri seti üzerindeki dağılımı ve birbirleriyle ilişkili olduğu durumlar farklı analiz yöntemleriyle incelenerek bireylerde mortalite oranına ne derece etki ettiği gözlemlenmiştir.

Literatürde birçok alanda yaygın olarak kullanılan makine öğrenme yöntemlerinden k-NN, yapay sinir ağları (MLP), destek vektör makineleri (SVM), rassal orman (RF), karar ağacı (RT) ve lojistik model ağacı (LMT) gibi karar destek araçları kullanılmıştır. Bu araçlar sayesinde ön görülemeyen bilgi ve bulgular ortaya çıkarılabilmekte, yeni bir örnek sınıf bilgisi tahmin edilebilmekte veya alınan hekim kararları desteklenebilmektedir. Böylece, hastalıklarda erken tanı ve tedaviye başlanarak hastaların hastalıklarından kurtulma oranının yükseltilip ölüm riskinin azaltılması sağlanmaktadır.

Veri seti üzerinde koşturulan bu algoritmaların sınıflandırma performansları, yani teşhis güçleri, literatürde yaygın olarak kullanılan değerlendirme ölçütlerinden faydalanılarak birer birer hesaplanmış olup sonuçlar birbirleriyle karşılaştırılmıştır. Tezde kullanılan makine öğrenme algoritmaları orjinal veri setine uygulanmadan önce, veri setinde bir takım veri ön işleme adımları gerçekleştirilmiştir. İlk olarak veri setindeki eksik değerler ortalama yöntemiyle giderilerek eksiksiz hale getirilmiştir. Sonrasında dengesiz veri seti rezervuar ile rassal örnekleme yöntemi kullanılarak hasta sayısı arttırılarak dengeli hale getirilmiştir.

Bu çalışma kapsamında orjinal ve artırılmış veri setine uygulanan bu sınıflandırma yöntemlerinden elde edilen karışıklık matrisi oluşturulmuş olup doğruluk, duyarlılık, kesinlik, seçicilik, F-skor ve AUC gibi performans ölçütleri hesaplanmıştır. Hem orijinal ver setine hem de artırılmış veri setine uygulanan testlerden elde edilen sonuçlar, kullanılan algortimaların genel olarak yüksek sınıflandırma perfomansına sahip olduğunu göstermiştir. Yapılan testlerden elde edilen bulgular incelendiğinde, orijinal veri setine uygulanan algortimaların sağ hasta sayısının tahmin edilmesinde yüksek bir başarım gösterdiğini ancak ölümü gerçekleşmiş bireylerdeki tahmin performansının yeterli olmadığını göstermiştir. Bunun üzerine orijinal veri setindeki hasta sayısının

artırımı yoluna gidilmiştir. Artırılmış veri seti üzerinde yapılan testlerden elde edilen bulgular incelendiğinde, tüm algoritmaların yüksek doğrulukta sınıflandırma yeteneğine sahip olduklarını, hem sağ hem de ölü hasta sayısının tahmin edilmesinde yüksek bir başarım gösterdiklerini ortaya koymuştur. Algoritmaların doğruluk, F-skor ve AUC değerleri dikkate alınarak yapılan değerlendirmede; en iyi performansın birbirine çok yakın değerlere sahip SVM (%99,9 doğruluk, %99,9 F-ksor ve %99,9 AUC) ve RO (%98 doğruluk %98,1 F-skor ve %100 AUC) algoritmalarıyla elde edildiğini; ve en düşük performansın da %94,9 doğruluk, %95 F-skor ve %95,3 AUC değerleri ile RT algoritması ile elde edildiğini göstermiştir.

Literatürde yapılan benzer çalışmalardan elde edilen bulgular ile bu tez çalışmasından elde edilen sonuçlar Tablo 6.1'de verilmiştir. Tablo 6.1'deki sonuçlar incelendiğinde, çalışılan tüm performans ölçütlerinde en iyi başarımın bu tez çalışmasında elde edildiğini göstermektedir. Elde edilen bulgular, çalışmamızın aynı sınıflandırma algoritmaları kullanan çalışmalardan bile daha yüksek bir sınıflandırma performansına sahip olduğunu ortaya koymaktadır. Bunun nedeninin kullanılan veri setlerinin ve içeriklerinin farklı olmasından kaynaklandığı düşünülmektedir.

Tablo 6.1. Elde edilen sonuçların literatüdeki sonuçlarla karşılaştırılması

Literatür Çalışmaları	Algoritma	Doğruluk	Duyarlılık	Kesinlik	Seçicilik	F-Skor	AUC
Cheng vd. 2022	LR	%82,1	%88,3	-	%81,5	-	%83,9
	RO	%81,2	%87,5	-	%80,6	-	%82,5
Kumar vd. 2022	k-NN	%91,4	-	-	-	-	-
Zhao vd. 2023	SVM	%72,25	%82,14	-	%71,69	-	%79,26
Oliveira vd. 2023	SVM	%80	%80	-	-	-	%81
Kasim vd. 2023	SVM	-	-	-	-	-	%93,4
Tao vd. 2022	LR/SVM/RT/ k-NN	-	-	-	-	-	%87,9
Yadav vd. 2022	RT	%75	%85	-	-	-	%79,7
Li vd. 2017	RO	-	-	-	-	-	%91,7
Tez çalışmamız	SVM	%99,9	%100	%99,8	%99,8	%99,9	%99,9

	RO	%98	%96,4	%99,8	%99,8	%98,1	%100
	k-NN	%97,1	%94,9	%99,8	%99,8	%97,3	%97,4
	LMT	%96,1	%92,9	%99,8	%99,8	%96,2	%97,7
	MLP	%95,2	%92,2	%98,8	%98,7	%95,4	%97,3
	RT	%94,9	%90,5	%100	%100	%95	%95,3

Klinik çalışmalarda ve literatürde erken mortaliteyi öngörmenin en önemli bağımsız parametreleri; ileri yaş, yüksek killip derecesi, yüksek kalp hızı, düşük sistolik kan basıncı, anterior infarkt lokalizasyonu olduğu belirlenmiştir (Lee, K. L., vd., 1995; Morrow, D. A., vd., 2000). Literatürde yer alan bu bilgiler doğrultusunda, ölüme eşlik ettiği ifade edilen bu en önemli 5 bağımsız parametre dikkate alınarak artırılmış veri setine sınıflandırma testleri yapılmıştır. Algoritmaların doğruluk, F-skor ve AUC değerleri dikkate alınarak yapılan karşılaştırmada; en iyi performansın yine önceki testlerde olduğu gibi SVM (%97,4 doğruluk, %97,5 F-skor ve %97,6 AUC) ve RO (%94,6 doğruluk, %94,8 F-skor ve %99,6 AUC) algoritmaları ile elde edildiği ve bu performansı sırasıyla k-NN, RT ve LMT algoritmalarının takip ettiği görülmüştür.

Elde edilen bu sonuçlar kıyaslandığında, veri setinde yer alan hastalara ait toplam 38 parametre yerine literatürde erken mortaliteye neden olduğu belirtilen sadece en önemli 5 parametre kullanılarak, benzer sınıflandırma performansının daha az hesaplama yüküyle ve daha kısa sürede elde edilebileceği gözlemlenmiştir. Elde edilen bu sonucun, literatürde yer alan, klinik çalışmalarda erken mortaliteyi öngörmenin en önemli bağımsız parametreleri; ileri yaş, yüksek killip derecesi, yüksek kalp hızı, düşük sistolik kan basıncı, anterior infarkt lokalizasyonu olduğu gerçeğiyle örtüşmesi bakımından çok önemlidir.

Yapılan bu çalışma bizlere göstermiştir ki, sağlık alanındaki birçok uygulamada sağlık personelinin teşhis ve yorumlama süreçleriyle beraber yapay zekâ destekli bir karar destek sürecini de kullanması artık vazgeçilmez olacaktır. Bu yöntemlerin, kalp krizi, kanser vb. birçok farklı sağlık alanında kullanımının günden güne artarak hastalıkların tanısında ve erken teşhisinde önemli bir role sahip olacağı öngörülmektedir. Günümüzde biyomedikal alanda yapay zekâ algoritmaları kullanılarak birçok proje yapılmaktadır. Sağlıkta yapay zekâ uygulamalarının bizlere karar verme ve uygulama noktasında elimizi güçlendirecek argümanlar sunmaya devam edeceği kaçınılmazdır. Bu tez çalışmasında

elde ettiğimiz sonuçlar, literatürde mevcut benzer çalışmalara paralel olarak, makine öğrenme algoritmalarının yüksek doğrulukta sınıflandırma performansına sahip olduğunu ve ST-segment elevasyonlu miyokart enfarktüsü (STEMI) hastalarının hastane içi mortalite riskini tahmin etmede çok etkili ve güvenilir bir karar destek mekanızması olduğunu göstermektedir.

Bu çalışma kapsamında STEMI hastalarında hastane-içi mortalite risk tahmini yapılmış olup, 30-günlük ve 1-yıllık mortalite risk analizinin ileri bir çalışma olarak yapılmasının faydalı olacağı düşünülmektedir. Bununla birlikte, bu çalışmada kullanılan ML algoritmalarının literatürde kullanılan erişilebilir veri setlerine de uygulanarak, algoritmaların tahminleme performanslarının kyaslanmasında daha gerçekçi bir değerlendirme imkanı sunacağı öngörülmektedir.

KAYNAKLAR

Alan, M., 2020, Makine Öğrenmesi Temelli Fiziksel Aktivite Kestirim Sistemi Tasarım Ve Uygulaması, *Fırat Üniversitesi Uzay ve Savunma Teknolojileri Dergisi* 1(1), 115-120, 2022.

Awtry, EH., & Loscalso, J., 2001, Koroner Kalp Hastalığı. Andreoli, TE., Carpenter, CCJ., Griggs, RC., Loscalzo, J., editors. Çeviri Editörü: Çavuşoğlu, H., Cecil Essentials of Medicine Türkçesi. Beşinci Edisyon. Philadelphia: WB Saundres, 79; 2001.

Aydemir, H., 2020, Karar ağaçları, https://aydemirhamza.medium.com/karar a%C4%9Fa%C3%A7lar%C4%B1-52d0f1d46ab2 [Erişim tarihi: 26 Nisan 2024]

Ayyadevara, V. K., 2018, Random Forest, In: Pro Machine Learning Algorithms, Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-3564-5_5_[Erişim tarihi: 27 Nisan 2024]

Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K., (2019, January). Improving Heart Disease Prediction Using Feature Selection Approaches, *In 2019 16th international bhurban conference on applied sciences and technology (IBCAST)*, (pp. 619-623). IEEE.

Bento, C., 2022, ROC Analysis and the AUC — Area Under the Curve, Towards Data Science, https://towardsdatascience.com/roc-analysis-and-the-auc-area-under-the-curve-404803b694b9

Bentzon, J.F., Otsuka, F., Virmani, R., & Falk, E., 2014, Mechanisms of plaque formation and rupture. *Circulation Research.* 2014;114(12):1852–66.

Boz, H., & Kose, U., 2018, Emotion extraction from facial expressions by using artificial intelligence techniques. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *9*(1), 5-16.

Breiman, L., 2001, Random Forests, *Machine Learning*, 45(1), 5-32.

Chakrabarti, S., Cox, E., Frank, E., Güting, R. H., Han, J., Jiang, X., Kamber, M., Lightstone, S. S., Nadeau, T. P., Neapolitan, R. E., Pyle, D., Refaat, M., Schneider, M., Teorey, T. J., Witten, I. H., 2008, "Data Mining Know It All", Elsevier, *United States*, 2008.

Cheng, C. Y., Chiu, I. M., Lin, C. H. R., Lin, X. H., Chen, F. C., & Hsu, T. Y., (2022, December). Stepwise Regression Machine Learning Models for In-Hospital Mortality Prediction in Patients After ST-Segment Slevation Myocardial Infarction (STEMI). In *2022 IEEE/ACIS 23rd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)* (pp. 14-18). IEEE.

Cleveland Clinic, 2024, What is a Heart Attack?, https://my.clevelandclinic.org/health/diseases/16818-heart-attack-myocardial-infarction [Erişim tarihi: 18 Mayıs 2024]

Cook, N. R., 2007, Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation,* 115(7), 928-935.

Copeland, M., 2016, Nvidia, What's the Difference Between Artificial Intelligence, *Machine Learning and Deep Learning,* 29.

Cortes, C., & Vapnik, V., 2001, Dispensing system boosts throughput 50 percent. *Assembly,* 44(13), 97.

Cruz, J. A., & Wishart, D. S., 2006, Applications of machine learning in cancer prediction and prognosis, *Cancer informatics,* 2006; 2: 59–77.

Daş, B., ve Türkoğlu, İ., 2014, DNA Dizilimlerinin Sınıflandırılmasında Karar Ağacı Algoritmalarının Karşılaştırılması. *Eleco 2014 Elektrik – Elektronik – Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu,* 27 – 29 Kasım 2014, Bursa.

Ebrahim, S., Taylor, F., Ward, K., Beswick, A., Burke, M., Smith G. D., 2011, Multiple risk factor interventions for primary prevention of coronary heart disease. *Cochrane Database of Systematic Reviews,* Issue 1. Art. 2011 Jan 19:(1):CD001561. http://dx.doi.org/10.1002/14651858.CD001561.pub 3

Efe, M. Ö., ve Kaynak, O., 2000, Yapay Sinir Ağları ve Uygulamaları, *Boğaziçi Üniversitesi Yayınevi,* İstanbul.

Egrioglu, E., Aladag, C. H., Yolcu, U., Uslu, V. R., & Basaran, M. A., 2009, A new approach based on artificial neural networks for high order multivariate fuzzy time series. *Expert Systems with Applications,* 36(7), 10589-10594.

Eren, Ö., 2008, Alerjen proteinlerin otomatik olarak sınıflandırılması, Yüksek Lisans Tezi, *Başkent üniversitesi Fen Bilimleri Enstitüsü*, Ankara, 18-20.

Falk, E., Shah, P. K., ve Fuster, V., 1995, Coronary plaque disruption, *Circulation,* 1995;92:657–671.

Gareth, J., Daniela, W., Trevor, H., & Robert, T., 2013, An Introduction to Statistical Learning: With Applications in R. Springer, Berlin. *Springer Science & Business Media,* 24 Haz 2013 - 426 sayfa

Gershenson, C., 2021, Artificial Neural Networks for Beginners [PDF belgesi], Lecture Notes Online Web site: https://www.uv.mx/mia/files/2012/10/ [Erişim tarihi: 18.01.2021]

Graupe, D., 2013, Principles of Artificial Neural Networks (3rd edition), World Scientific, Advanced Series in Circuits and Systems: Volume 7, *Chicago, University of Illinois, USA.* https://doi.org/10.1142/8868

Griffin, P., ve Topol, E. J., 2009, Manual of cardiovascular medicine, *Philadelphia Pa:* Lippincott Williams&Wilkins 2009.

Güran, A., Uysal, M., & Doğrusöz, Ö., 2014, Destek vektör makineleri parametre optimizasyonunun duygu analizi üzerindeki etkisi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen Ve Mühendislik Dergisi,* 16(48), 86-93.

Han, J., & Kamber, M., 2001, "Data Mining Concepts and Techniques", Academic Press, *New York,* 2001.

Hegde, C., Wallace, S., and Gray, K., 2015, Using trees, bagging, and random forests to predict rate of penetration during drilling, *In SPE Middle East Intelligent Oil and Gas Conference and Exhibition.* OnePetro.

Herrington, W., Lacey, B., Sherliker, P., Armitage, J., and Lewington, S., 2016, Epidemiology of Atherosclerosis and the Potential to Reduce the Global Burden of Atherothrombotic Disease, *Circulation Research,* 2016;118(4):535–46.

İbanez, B., James, S., Agewall, S., Antunes, M. J., Bucciarelli-Ducci, C., and Bueno, H., et al, 2017, ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation, *European Heart Journal*, 2018;39(2):119–77.

Jonason, T., & Bergstrom, R., 1987, Cessation of smoking in patients with intermittent claudication, Effects on the risk of peripheral vascular complications, myocardial infarction and mortality, *Acta Medica Scandinavica,* 1987;221:253–260

Kasim, S., vd., 2023, In-Hospital Mortality Prediction using Machine Learning and Stacked Ensemble Learning of Asian Women with ST-Elevation Myocardial Infarction (STEMI), doi: 10.21203/rs.3.rs-2611510/v1

Katarya, R., & Meena, S.K., 2021, Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. *Health Technol.* 11, 87–97, 2021.

Kavzoğlu, T., ve Çölkesen, İ., 2010, Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi (Investigation of the Effects of Kernel Functions in Satellite Image Classification Using Support Vector Machines), April 2015, 73–82.

Kecman, V., 2001, "Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models (Complex Adaptive Systems)," The MIT Press, *Cambridge,* 2001.

Kırbaş, Ö., 2011, ST segment yükselmeli akut miyokard infarktüslü hastalarda başvuru ortalama trombosit hacmi ile fibrinolitik tedavi sonrası ST segment rezolüsyonu arasındaki ilişki, Uzmanlık Tezi, *Sağlık Bakanlığı Ankara Atatürk Eğitim ve Araştırma Hastanesi,* Ankara, 2011.

Kolay, N., & Erdoğmuş, P., 2016, The classification of breast cancer with Machine Learning Techniques, In Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 1-4, 2016.

Kumar, K., & Ananth, S., 2022, Heart Stroke Detection Using KNN Algorithm, ECS transactions, doi: 10.1149/10701.18385ecst.

Landwehr, N., Hall, M., & Frank, E., 2003, Logistic Model Trees, In N. Lavrač, D. Gamberger, H. Blockeel & L. Todorovski (Eds.), Machine Learning: ECML 2003 (Vol. 2837, pp. 241-252): Springer Berlin Heidelberg.

Larose, D. T., 2005, "Discovering knowledge in data : an introduction to data mining", John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

Lee, K. L., Woodlief, L. H., & Topol, E. J., et al 1995, Predictors of 30-day mortality in the era of reperfusion for acute myocardial infarction, Results from an international trial of 41,021patients, GUSTO-I Investigators, *Circulation*, 1995;91:1659–68.

Lee, S., Kang, P., & Cho S., 2014, Probabilistic Local Reconstruction for k-NN Regression and Its Application to Virtual Metrology in Semi Conductor Manufacturing, Neurocomputing, 131,427–439, 2014.

Li, X., Liu, H., Yang, J., Xie, G., & Xu, M., 2017, Using Machine Learning Models to Predict In-Hospital Mortality for ST-Elevation Myocardial Infarction Patients, MEDINFO 2017: Precision Healthcare through Informatics, 476-480. doi: 10.3233/978-1-61499-830-3-476

Libby, P., 2001, Current concepts of the pathogenesis of the acute coronary syndromes, *Circulation*, 2001;104(3):365-72.

Paralta De Figueiredo, M., Carias, M., Picarra, B., Rocha De Almeida, A., Claudio, F., Rocha, R., & Trinca, M., 2023, ST-segment elevation is associated with a higher risk of cardiac arrest in patients with acute coronary syndromes. *European Journal of Preventive Cardiology*, 30(Supplement_1) doi: 10.1093/eurjpc/zwad125.057

Morrow, D. A., Antman, E. M., & Charlesworth, A., et al 2000, Braunwald E. TIMI risk score for STelevationIHmyocardial infarction: a convenient, bedside, clinical score for risk assessment at presentation: an intravenous tPA for treatment of infarcting myocardium early II trial substudy, *Circulation* 2000;102:2031–7.

Narkhede, S., 2018. Understanding Confusion Matrix, Towards Data Science, [https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62]

Nisbet, R., Elder, J., & Miner, G., 2009, "Handbook of Statistical Analysis and Data Mining Applications", *Elsevier*, 2009.

Oliveira, M., Seringa, J., Pinto, F. J., Henriques, R., & Magalhães, T., 2023, Machine learning prediction of mortality in acute myocardial infarction, *BMC Medical Informatics and Decision Making*, *23*(1), 70.

Onan, A., 2015, Şirket iflaslarının tahmin edilmesinde karar ağacı algoritmalarının karşılaştırmalı başarım analizi, *Bilişim Teknolojileri Dergisi,* 8(1), 9-19.

Orhan, U., 2012, Makine Öğrenmesi Dersi, Entropi, Karar Ağaçları (ID3 ve C4.5 algoritmaları), Sınıflandırma ve Regresyon Ağaçları, Ders Notları, Müh. Mim. Fakültesi Bilgisayar Mühendisliği, Çukurova Universitesi, Adana, Retrieved from http://bmb.cu.edu.tr/uorhan/DersNotu/Ders03.pdf

Oto, A., Tokgözoğlu, L., İlkay, E., & Aytemir, K., 2009, Akut Koroner Sendrom, Ankara, *Erkem Yayınevi* 2009; 2-16.

Ozbek, M., Ildirimli, K., Arik, B., Aktan, A., Coskun, M. S., Evsen, A., Guzel, T., Acet, H., Demira, M., 2023, Dependence of clinical outcomes on time of hospital admission in patients with ST-segment elevation myocardial infarction. Ann Saudi Med. 2023 Jan-Feb;43(1):25-34. doi: 10.5144/0256-4947.2023.25. Epub 2023 Feb 2. PMID: 36739499; PMCID: PMC9899343.

Polat, H., 2016, Görsel – İşitsel Uyaranlar Kaynaklı Oluşan Duyguların EEG İşaretleri ile Sınıflandırılması, Yüksek Lisans Tezi, *Dicle Üniversitesi Fen Bilimleri Enstitüsü,* Diyarbakır, c. 7, sy. 1, ss. 33–40.

Ridker, P. M,, Danielson, E., Fonseca, F. A., Genest, J., Gotto, A. M., Kastelein, J. J., et al, 2009, Reduction in C-reactive protein and LDL cholesterol and cardiovascular event rates after initiation of rosuvastatin: a prospective study of the JUPITER trial. Lancet. 2009;373(9670):1175–82.

Rodriguez-Granillo GA, Agostoni P, Garcia-Garcia HM, Biondi-Zoccai GGL, Mc Fadden E, Amoroso G, et al, 2007, Meta-Analysis of the Studies Assessing Temporal Changes in Coronary Plaque Volume Using Intravascular Ultrasound. *American Journal of Cardiology,* 2007 J;99(1):5–10.

Shrivastava, A. K., Singh, H. V., Raizada, A., & Singh, S. K., 2015, C-reactive protein, inflammation and coronary heart disease, *The Egyptian Heart Journal,* 2015;67(2):89-97.

Stary, H. C., Chandler, A. B., & Dinsmore, R. E., et al 1995, A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis, A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, *American Heart Association Circulation,* 1995 Sep 1;92(5):1355-74.

Sumner, M., Frank, E., & Hall, M., 2005, Speeding Up Logistic Model Tree Induction. In A. Jorge, L. Torgo, P. Brazdil, R. Camacho & J. Gama (Eds.), Knowledge Discovery in Databases: PKDD 2005 (Vol. 3721, pp. 675-683): Springer Berlin Heidelberg.

Şeker, Ş.E., 2016, WEKA ile Veri Madenciliği, *Bilgisayar Kavramları Yayınları,* 158. İstanbul.

Tao, D., Liu, C., & Wan, S., 2022, Machine Learning Algorithms for Predicting in-Hospital Mortality in Patients with ST-Segment Elevation Myocardial Infar. arXiv preprint arXiv:2211.14480.

Timmis, A., Townsend, N., Gale, C., Grobbee, R., Maniadakis, N., Flather, M., et al 2017, European Society of Cardiology: CardiovasAkbulut ve Kahraman, Koroner Arter Hastalığı Eğitim Anketi-2 / doi: 10.5543/khd.2021.60362 33 cular disease statistics 2017. Eur Heart J 2018;39(7):508-79.

Torpy JM, Burke AE, Glass RM. JAMA patient page. Coronary heart disease risk factors. JAMA. 2009 Dec 2;302(21):2388. doi: 10.1001/jama.302.21.2388. PMID: 19952328.

TseKiChun, 2021, Random Forest Explain, https://en.m.wikipedia.org/wiki/File:Random_forest_explain.png [erişimtarihi:26.04.2024]

TÜİK, 2017, Türkiye İstatistik Kurumu, 2017 verileri, [Available from: http:// www.tuik.gov.tr

Türkiye İstatistik Kurumu (TÜİK) Ölüm Nedeni İstatistikleri, 2015, Available from: http://www.webcitation.org/query?url=http%3A%2F%2Fwww.tuik.gov.tr%2FPreHaberBultenleri.do%3Fid%3D21526&date=2018-01-09

Öztürk, K., & Şahin, M. E. (2018). Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. *Takvim-I Vekayi*, 6(2), 25-36.

Vitter, J. S. 1985, Random sampling with a reservoir. *ACM Transactions on Mathematical Software (TOMS)*, *11*(1), 37-57.

World Health Organization (WHO), 2019, The top 10 causes of death, Available at: https:// www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [Erişim tarihi: 23 Nisan 2024].

World Health Organisation, Global Health Estimates, projection of mortality and causes of death, (2015-2030), Available from: http://www.webcitation.org/query?url=http%3A%2F%2Fwww.who.int%2Fhealthinfo%2Fglobal_burden_disease%2Fen%2F & date=2018-01-09

Yadav, R., Bansal, A., Zachariah, G., Batra, V., Bhatt, D. L., Hendrickson, M., ... Gupta, M., 2022, Machine learning based model for risk prediction after ST-Elevation myocardial infarction: Insights from the North India ST elevation myocardial infarction (NORIN-STEMI) registry, *International Journal of Cardiology*, 10.1016/j.ijcard.2022.05.023.

Zhao, P., Liu, C., Zhang, C., Hou, Y., Zhang, X., Zhao, J., ... & Zhou, J., 2023, Using machine learning to predict the in-hospital mortality in women with ST-segment elevation myocardial infarction. *Reviews in Cardiovascular Medicine*, *24*(5), 126. doi: 10.31083/j.rcm2405126

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	**:**	BAHAR ÖZYILMAZ
Uyruğu	**:**	TC

EĞİTİM

Derece		**Adı, İlçe, İl**	**Bitirme Yılı**
Üniversite	:	Dicle Üniversitesi, Merkez, DİYARBAKIR	2016
Yüksek Lisans	:	Batman Üniversitesi, Merkez, BATMAN	Devam ediyor

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görevi**
2013-2024	Dicle Üniversitesi Hastanesi	Hemşire

UZMANLIK ALANI:

YABANCI DİLLER: İngilizce

YAYINLAR: