İZMİR BAKIRÇAY ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
AKILLI SİSTEMLER MÜHENDİSLİĞİ (DİSİPLİNLER ARASI) A.B.D.

SEPSİS'İN YAPAY ZEKA TEMELLİ ERKEN TANISI ÜZERİNE KARAR DESTEK SİSTEMİ

YÜKSEK LİSANS TEZİ

Pınar KAYA AKSOY

Tez Danışmanı: Doç. Dr. Orhan ER

EKİM 2022

SEPSİS'İN YAPAY ZEKA TEMELLİ ERKEN TANISI ÜZERİNE KARAR DESTEK SİSTEMİ

Pınar KAYA AKSOY

YÜKSEK LİSANS

Akıllı Sistemler Mühendisliği (Disiplinler arası) Anabilim Dalı

Danışman: Doç. Dr. Orhan ER

(İkinci Danışman: Doç. Dr. Deniz KILINÇ)

İzmir

İzmir Bakırçay Üniversitesi

Lisansüstü Eğitim Enstitüsü

Ekim 2022

JÜRİ VE ENSTİTÜ ONAYI

İzmir Bakırçay Üniversitesi Lisansüstü Eğitim Enstitüsü Akıllı Sistemler Mühendisliği (Disiplinler arası) Anabilim dalında öğrenim görmekte olan Pınar KAYA AKSOY'un "Sepsis'in Yapay Zeka Temelli Erken Tanısı Üzerine Karar Destek Sistemi" başlıklı tezi 04/10/2022 tarihinde aşağıdaki jüri tarafından değerlendirilerek "İzmir Bakırçay Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca, Akıllı Sistemler Mühendisliği (Disiplinler arası) Anabilim dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Jüri Üyeleri	**Unvanı Adı Soyadı**	**İmza**
Üye (Tez Danışmanı)	Doç. Dr. Orhan ER	
Üye	Prof. Dr. Ayşegül ALAYBEYOĞLU	
Üye	Dr. Öğr. Üyesi Okan BURSA	
Üye		
Üye		

Prof. Dr. Serkan ÇINARLI
Lisansüstü Eğitim Enstitüsü Müdürü

ÖZET

SEPSİS'İN YAPAY ZEKA TEMELLİ ERKEN TANISI ÜZERİNE KARAR DESTEK SİSTEMİ

Pınar KAYA AKSOY

Akıllı Sistemler Mühendisliği (Disiplinler arası) Anabilim Dalı

İzmir Bakırçay Üniversitesi, Lisansüstü Eğitim Enstitüsü, Ekim 2022

Danışman: Doç. Dr. Orhan ER (İkinci Danışman: Doç. Dr. Deniz KILINÇ)

Dünya Sepsis Derneği verilerine göre, her yıl 47-50 milyon kişinin sepsis olduğunu ve 11 milyon kişinin sepsisten öldüğünü tahmin ediyor, her 2.8 saniyede bir ölüm gerçekleşiyor. Sepsis, en yaygın hastalıklardan olan meme kanseri, prostat kanseri ve HIV-AIDS'in toplamından daha fazla ölüm oranına sahiptir. Enfeksiyona karşı düzensiz immün aracılı bir konak yanıtı olan sepsis, yaygın, ölümcül ve maliyetlidir. Bu tez çalışmasında, laboratuvar bulgularından makine öğrenimi algoritmalarını kullanarak Sepsis tahmini için karşılaştırmalı bir çalışma gerçekleştiriyoruz. Bu amaçla, sinir ağı sınıflandırıcısının yanı sıra farklı yapıları içeren otuz iki farklı makine öğrenmesi algoritması değerlendirilmiş ve karşılaştırılmıştır. Deneysel sonuçlar, diğer makine öğrenmesi algoritmaları arasında SVM (Cubic, Fine Gaussian), KNN (Fine, Weighted, Subspace), Trees (Weighted, Boosted, Baggged) ve sinir ağları tabanlı sınıflandırıcıların %100 ile daha iyi doğruluk elde ettiğini göstermektedir. Böylece sepsis hastasının hayatta kalıp kalamayacağını tahmin etmek için makine öğrenmesi algoritmalarının kullanılmasının uygun olduğu sonucuna varılmıştır. Bu tez çalışması, bir sepsis hastasının hayatta kalıp kalamayacağını tahmin ederken doktorların karar verme süreçleri için yeni bir destekleyici araç olarak kullanılma potansiyeline sahiptir.

Anahtar Kelimeler: Sepsis, Erken Tahmin, Yapay Zeka, Karar Destek Sistemleri

ABSTRACT

DECISION SUPPORT SYSTEM ON ARTIFICIAL INTELLIGENCE BASED EARLY DIAGNOSIS OF SEPSIS

Pınar KAYA AKSOY of student Department of Intelligent Systems Engineering (Interdisciplinary)Programme in

İzmir Bakırçay University, Graduate Education Institute, October 2022

Supervisor: Doç. Dr. Orhan ER (Co-Supervisor: Doç. Dr. Deniz KILINÇ)

According to World Sepsis Association data, it is estimated that 47-50 million people have sepsis and 11 million people die from sepsis each year, with one death occurring every 2.8 seconds. Sepsis has a higher death rate than the most common diseases of breast cancer, prostate cancer, and HIV-AIDS combined. Sepsis, a dysregulated immune-mediated host response to infection, is common, fatal, and costly. In this thesis, we perform a comparative study for Sepsis prediction using machine learning algorithms from laboratory findings. For this purpose, thirty-two different machine learning algorithms including different structures as well as neural network classifiers were evaluated and compared. Experimental results show that SVM (Cubic, Fine Gaussian), KNN (Fine, Weighted, Subspace), Trees (Weighted, Boosted, Bagged), and neural network-based classifiers, among other machine learning algorithms, achieve better accuracy with 100%. Thus, it was concluded that it is appropriate to use machine learning algorithms to predict whether the sepsis patient will survive. This thesis has the potential to be used as a new supportive tool for physicians when predicting whether a sepsis patient will survive.

Keywords: Sepsis, Early Forecasting, Artificial Intelligence, Decision Support Systems

01/11/2022

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın İzmir Bakırçay Üniversitesi tarafından kullanılan Turnitin bilimsel intihal tespit programıyla tarandığını ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Pınar KAYA AKSOY

İÇİNDEKİLER

Sayfa

JÜRİ VE ENSTİTÜ ONAYI ..... **ii**

ÖZET ..... **iii**

ABSTRACT ..... **iv**

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ ..... **v**

İÇİNDEKİLER ..... **vi**

TABLOLAR DİZİNİ ..... **viii**

ŞEKİLLER DİZİNİ ..... **ix**

SİMGELER VE KISALTMALAR DİZİNİ ..... **xi**

1. GİRİŞ **1**

2. MATERYAL METOT **4**

2.1. Literatür Çalışması **4**

2.2. Veritabanı Tanıtımı **7**

2.3. Kullanılan Yöntemler **9**

2.3.1. Makine öğrenimi algoritmaları **9**

2.3.1.1. Denetimli öğrenme (Supervised Learning) **12**

2.3.1.2. Denetimsiz öğrenme (Unsupervised Learning) **12**

2.3.1.3. Pekiştirmeli öğrenme (Reinforcement Learning) **12**

2.3.2. Yapay Sinir Ağları (YSA) **13**

2.3.3. Destek Vektör Makinesi (SVM) **13**

2.3.4. k-En Yakın Komşu Algoritması (KNN) **14**

2.3.5. Karar Ağacı Sınıflandırıcı Algoritması **14**

2.3.6. Naive Bayes **15**

2.3.7. Lojistik regresyon **15**

2.4. Başarım Metriklerinin Tanıtımı **16**

2.5. Özellik Mühendisliği **18**

3. SEPSİS TANIMI **21**

3.1. Ağır Sepsis **22**

3.2. Septik Şok **23**

3.3. Multiple Organ Disfonksiyonu Sendromu **23**

3.4. Sepsis Risk Faktörleri ve Klinik Bulguları **24**

3.4.1. Sepsis için Risk Faktörleri 24
3.4.2. Sepsis klinik bulguları 24
3.5. SOFA 27
3.6. Quick SOFA 28
4. DENEYSEL ÇALIŞMALAR 29
5. SONUÇLAR VE TARTIŞMA 52
ÖZGEÇMİŞ 60

TABLOLAR DİZİNİ

Sayfa

Tablo 2.1. İkili sınıflandırma için karışıklık matrisi 16
Tablo 2.2. Sınıflandırma değerlendirmeleri için eşik metrikleri 17
Tablo 3.1. 2016 kılavuzuna göre yenilenmiş Sepsis ve Septik Şok tanımı 22
Tablo 3.2. Sepsisle ilgili organ hipoperfüzyon bulguları..... 22
Tablo 3.3. MODS organ disfonksiyon parametreleri 23
Tablo 3.4. SOFA kriteri 27
Tablo 3.5. qSOFA kriterleri 28
Tablo 4.1. Karşılaştırma için makine öğrenmesi teknikleri (%) 29

ŞEKİLLER DİZİNİ

Sayfa

Şekil 1.1. Yaygın Sepsis kaynakları 2

Şekil 2.1. Yapay zekâ, makine öğrenmesi ve öğrenme teknikleri arasındaki ilişki..... 11

Şekil 2.2. Tek gizli katmanlı YSA 13

Şekil 3.1. Sepsis Tanı Algoritması 19

Şekil 4.1. En başarılı Fine KNN algoritması için Karışıklık Matrisi 31

Şekil 4.1. En başarılı Weighted KNN algoritması için Karışıklık Matrisi 31

Şekil 4.2. En başarılı Bagged Trees algoritması için Karışıklık Matrisi 32

Şekil 4.3. En başarılı Subspace KNN algoritması için Karışıklık Matrisi 32

Şekil 4.4. En başarılı Narrow Neural Network algoritması için Karışıklık Matrisi 33

Şekil 4.5. En başarılı Medium Neural Network algoritması için Karışıklık Matrisi 33

Şekil 4.6. En başarılı Wide Neural Network algoritması için Karışıklık Matrisi . 34

Şekil 4.7. En başarılı Bilayered Neural Network algoritması için Karışıklık Matrisi 34

Şekil 4.8. En başarılı Trilayered Neural Network algoritması için Karışıklık Matrisi 35

Şekil 4.9. Revize edilen başarılı Cubic SVM algoritması için Karışıklık Matrisi 36

Şekil 4.10. Revize edilen başarılı Fine Gaussian SVM algoritması için Karışıklık Matrisi 36

Şekil 4.11. Revize edilen başarılı Boosted Trees algoritması için Karışıklık Matrisi37

Şekil 4.12. En kötü başarılı Gaussian Naive Bayes algoritması için Karışıklık Matrisi 37

Şekil 4.13. En iyi performans gösteren Fine KNN algoritması için ROC eğrisi 38

Şekil 4.14. En iyi performans gösteren Weighted KNN algoritması için ROC eğrisi..... 39

Şekil 4.15. En iyi performans gösteren Bagged Trees algoritması için ROC eğrisi..... .39

Şekil 4.16. En iyi performans gösteren Subspace KNN algoritması için ROC eğrisi 40
Şekil 4.17. En iyi performans gösteren Narrow Neural Network algoritması için ROC eğrisi 40
Şekil 4.18. En iyi performans gösteren Medium Neural Network algoritması için ROC eğrisi 41
Şekil 4.19. En iyi performans gösteren Wide Neural Network algoritması için ROC eğrisi 41
Şekil 4.20. En iyi performans gösteren Bilayered Neural Network algoritması için ROC eğrisi 42
Şekil 4.21. En iyi performans gösteren Trilayered Neural Network algoritması için ROC eğrisi 42
Şekil 4.22. Revize edilen başarılı Cubic SVM algoritması için ROC eğrisi 43
Şekil 4.23. Revize edilen başarılı Fine Gaussian SVM algoritması için ROC eğrisi 43
Şekil 4.24. Revize edilen başarılı Boosted Trees algoritması için ROC eğrisi 44
Şekil 4.25. En kötü performans gösteren Gauss Naive Bayes algoritması için ROC eğrisi 44
Şekil 4.26. 3-CV en başarılı Cubic SVM algoritma için Karışıklık Matrisi..... 45
Şekil 4.27. 3-CV en başarılı Cubic SVM algoritması için ROC eğrisi 45
Şekil 4.28. 3-CV en başarısız Gaussian Naive Bayes algoritma için Karışıklık Matrisi 46
Şekil 4.29. 3-CV en başarısız Gaussian Naive Bayes algoritma için ROC eğrisi .. 46
Şekil 4.30. 5-CV en başarılı Fine KNN algoritma için Karışıklık Matrisi 47
Şekil 4.31. 5-CV en başarılı Fine KNN algoritması için ROC eğrisi 47
Şekil 4.32. 5-CV en başarısız Gaussian Naive Bayes algoritma için Karışıklık Matrisi 48
Şekil 4.33. 5-CV en başarısız Gaussian Naive Bayes algoritma için ROC eğrisi .. 48
Şekil 4.34. 10-CV en başarılı Bagged Trees algoritma için Karışıklık Matrisi 49
Şekil 4.35. 10-CV en başarılı Bagged Trees algoritma için ROC eğrisi 49
Şekil 4.36. 10-CV en başarısız Gaussian Naive Bayes algoritma için Karışıklık Matrisi 50
Şekil 4.37. 10-CV en başarısız Gaussian Naive Bayes algoritma için ROC eğrisi 50

SİMGELER VE KISALTMALAR DİZİNİ

ABD	Amerika Birleşik Devletleri
AI	Yapay Zekâ
AUROC	Alıcı çalışma özelliği/Area Under the Curve of The Receiver
CRP	C-reaktif protein
DN	Doğru Negatif
EHR	Elektronik Sağlık Kayıtlar
EMR	Elektronik Tıbbi Kayıtlar
GRADE	Önerilerin Değerlendirilmesi, Geliştirilmesi ve Değerlendirilmesi/ Grading of Recommendations Assessment, Development and Evaluation
HBYS	Hastane Bilgi Yönetim Sistemi
KNN	k-En Yakın Komşu Algoritması
LDH	Laktik Dehidrojenaz
MEWS	Modifiye Erken Uyarı Skoru
MIMIC	Çok Parametreli Akıllı Yoğun Bakımda İzleme/ Multiparameter Intelligent Monitoring in Intensive Care
	Operating Characteristic
QUADAS-2	Tanısal Doğruluk Çalışmalarının Kalite Değerlendirmesi/ Quality Assessment of Diagnostic Accuracy Studies
SAPS	Basitleştirilmiş Akut Fizyoloji Skoru
SIRS	Sistemik İnflamatuar Yanıt Sendromu
SOFA	Sıralı Organ Yetmezliği Değerlendirmesi
SVM	Destek Vektör Makinesi
WBC	Beyaz Kan Hücresi Sayımı
YBÜ	Yoğun Bakım Ünitesi

YN	Yanlış Negatif
YP	Yanlış Pozitif
YSA	Yapay Sinir Ağları

1. GİRİŞ

Dünyada sağlık alanında yapılan bilimsel araştırmaların ve yapay zekâ uygulamalarının sayısının hızla arttığı görülmektedir. Bu sebeple küresel büyük veri ve yapay zekâ şirketlerinin neredeyse tamamı sağlık alanına odaklanmış olup hasta hizmetlerinden tanı ve tedavi sistemlerine kadar geniş bir yelpazede çalışmalarını yürütmektedirler. Özellikle dijitalleşme teknolojilerin gelişmesi ile giderek artan biçimde elektronik hale gelen hasta verileri, bu verilerin analizinde yapay zekânın kullanım olanaklarını artırmıştır. Bu sayede, hastalık tanı ve teşhis sistemleri, yapay zekâ destekli robotlarla gerçekleştirilen ameliyatlar, kişiye özel tedaviler, ilaç geliştirme, gen düzenleme becerisi, karar destek sistemleri ve tekrar yatış tahmini gibi önemli uygulamaların geliştirilebilmesine olanak sağlanmıştır. Özellikle pandeminin dijital dönüşümü hızlandırdığı bu zorlu süreçten en çok etkilenen sektörlerin başında gelen sağlık sektöründeki gelişmeler yetersiz insan kaynağına bir çözüm alternatifi sunmaktadır. Sağlık çalışanlarının süreci yürütmeye yetişemediği bölümlerde daha hızlı sonuçların alınması gereken rutin işlerin yapay zekâ ile hızlandırılması ve yoğun çalışma temposunda hatanın azaltılmasına katkı sağlamaktadır.

Enfeksiyona karşı düzensiz immün aracılı bir konak yanıtı olan sepsis, yaygın, öldürücü ve maliyetlidir [1]. Genellikle kan zehirlenmesi olarak adlandırılan sepsis, vücudun bir enfeksiyona verdiği tepkinin gerçekten olumsuz bir sonucudur ve bu tepki organ hasarına neden olur. Sepsis riski altındakileri belirlemek ve herhangi bir klinik belirtiden önce uygun tedaviye başlamak, sepsisin mortalitesi ve sağlık bakım hizmetleri maliyet yükü üzerinde önemli bir etkiye sahip olacaktır. Dünyada her 5 ölümden 1'i sepsis ile ilişkilendiriliyor [2]. Ayrıca sepsis vakaların %40'ı 5 yaş altı çocuklardan oluşuyor. Dünyadaki tüm ölümlerin %20'si sepsis ile ilişkilidir, ülkeye bağlı olarak, ölüm oranı %15 ile %50 arasında değişmektedir. Dünya Sağlık Örgütü'nün 2018'de rapor ettiği, sepsis kaynaklı ölümler, kanser kaynaklı ölümlerden daha yüksek orana sahiptir. Sepsis de en yüksek ölüm oranı ise yoksul ve gelişmekte olan ülkelerdedir. Sepsisi teşhis etmek için spesifik bir test yoktur ve bir dizi farklı tarama araçları ve laboratuvar sonuçları kullanılmıştır. Bakteriler, mantarlar, virüsler ve parazitler dahil çoğu mikroorganizma türü sepsise neden olabilir. Bununla birlikte, mevsimsel grip virüsleri, dang virüsleri ve halk sağlığını ilgilendiren yüksek oranda bulaşıcı patojenler ile enfeksiyonlardan da kaynaklanabilir. Sepsis genellikle Şekil 1'de

gösterildiği gibi solunum, mide-bağırsak ve idrar yolları veya yaralar ve cilt gibi yaygın ve önlenebilir enfeksiyonların klinik olarak kötüleşmesi olarak ortaya çıkar.



Şekil 1.1. Yaygın sepsis kaynakları (https://www.worldsepsisday.org/toolkits)

Sepsis ilişkili mortalite (ölüm oranı) risk faktörleri, sağlık merkezinin büyüklüğü ve gelişmişliğine bağlı olarak değişmektedir. Amerika Birleşik Devletleri'nde (ABD) sepsis tanılı hastaların bakım maliyeti 50.000 doları geçmekte ve ülke ekonomisine yıllık 17 milyar dolarlık bir yük oluşturmaktadır. Etkilenen hastalarda %20 - 50 mortaliteye neden olmakta ve iyileşen hastalarda belirgin olarak yaşam kalitesini düşürmektedir. Ülkemizde yapılan geniş çok merkezli epidemiyolojik çalışmada sepsis ve septik şoka bağlı ölüm oranları kabul edilemez oranda yüksek bulunmuştur (%55 ve %70). Ölüm oranındaki ve yıllık sağlık bakımı harcamalarındaki kayda değer artış (artan kalış süresinden etkilenir), sepsis tedavisi ve araştırmasını tıbbi internet araştırmaları ve tıbbi bilişim alanında kritik bir alan haline getirmiş ve ilgili literatürde yakın zamanda bir artışa neden olmuştur.

Bu tez çalışmasının ana hedefi; hem hastaların sağ kalımını tahmin etmek hem de en önemli risk faktörlerine karşılık gelen özellikleri hesaplamaktır. Tez çalışmasının çıktısı olarak, sepsis şüphesi ile gelen hastaya tanı konulurken doktorlar için yeni bir destekleyici araç olarak kullanılma potansiyeli oluşturmaktır. Erken sepsis tahmini, tedavi ve hasta sonuçlarını iyileştirmek için doktorların müdahale etmesine ve önceden teşhis koymasına yardımcı olabilir. Sepsis teşhisi ve erken tahmin için

mevcut yöntemlerin çoğu, yalnızca elektronik tıbbi kayıtlar (EMR) sistemi depolanan yapılandırılmış verilerden yararlanır. Klinik karar destek araçları, sepsis için en yüksek risk altında olanları belirlemeye yardımcı olabilir. EMR ve laboratuvar verileri üzerine mevcut çalışmalar umut verici görünmektedir. Sepsisin daha erken bir aşamada saptanması hayat kurtarabilir ve hasta için mali harcamaları azaltabilir. Sepsis tanısı, diğer yüksek riskli durumlardan ayırt edilmesi zor olduğu için sıklıkla gecikir ve bu gecikme hastanın hızla kötüleşmesine neden olabilir. Bu soruna potansiyel olarak dönüştürücü bir yaklaşım, Klinik Karar Destek Sistemlerini elde etmek için hastaların Elektronik Sağlık Kayıtlarından (EHR) olan büyük miktardaki bilgiden yararlanmak olacaktır. Yapay zekânın tıp alanında kullanımının artmasıyla birlikte birçok hastalığın erken öngörüsü ve tedavisi bu yöntemlerle sağlanmaktadır. Günümüzde makine öğrenimi yöntemlerinin en popüler algoritmalarının başında gelen yapay sinir ağları, akıl yürütme ve karar verme yeteneklerini göz önünde bulundurularak sepsis hastalığının erken evrelerinin tahmininde ve sepsis düzeyinin belirlenmesinde kullanılacağı çıkarımda bulunulmaktadır.

Hastalarda sepsisi tespit etmek, erken tahmin modelleri geliştirmek için çeşitli çalışmalar yapılmıştır. Makine öğrenimi ve yapay zekâdaki gelişmeler nedeniyle, bu alanlar tıp alanında da harika bir uygulamaya sahiptir. Makine öğrenme algoritmaları, sepsisi önceden tahmin etmek ve insanlara uygun ilaçları almalarına yardımcı olmak için kullanılır. Farklı makine öğrenme algoritmaları sepsisi teşhis edebilir veya tahmin edebilir ve böylece sepsisin ilerlemesini önleyebilir.

Covid-19 sürecinde öneminin bir kez daha ortaya çıktığı yapay zekâ, kısaca, insan zekâsına ihtiyaç duyulan işleri bilgisayarlara yaptıran teknoloji olarak tanımlanabilir. Yapay zekâ teknolojisinin günümüzde geldiği nokta göz önüne alındığında; tıbbi tanı koyabilmekte, kişiye özgü sağlık çözümleri sunabilmekte, yaşantımızın birçok alanında pratik çözümler sunabilmektedir. Covid-19 pandemisinin hızlandırdığı dijitalleşme sürecinde birçok sektörde üstün başarılarını gördüğümüz yapay zekâ teknolojilerinin odak noktalarının başında da sağlık teknolojileri gelmektedir.

Bu araştırmanın genel amacı, sepsisin erken tespiti için yapay zeka (AI) tabanlı bir teknik tasarlamak ve elde edilen başarılı sonucu Hastane Bilgi Yönetim Sistemi'nin (HBYS) yoğun bakım modülüne entegrasyonu yaparak sepsise bağlı ölüm oranlarını azaltabilmek için, uygulamaya alabilmek.

2. MATERYAL METOT

2.1. Literatür Çalışması

Literatürdeki birçok çalışma, farklı veri setleri ve pediatrik hastalar [10-18] kullanılarak sepsisin saptanması için farklı AI tekniklerinin performansını karşılaştırmıştır. Bu çalışmalar aynı veri setleri kullanılarak yapılmadığından, performansı doğrudan çalışmamızla karşılaştırmak doğru değildir. Bu nedenle literatürde benzer çalışmalara ağırlık verilmiştir.

Literatürde sepsiste makine öğrenmesi yöntemlerinin kullanımına ilişkin bazı çalışmalar şu şekilde özetlenebilir:

Tıp bilimindeki en zor sendromlardan biri olan sepsis, enfeksiyondan kaynaklanan ve insan vücudundan gelen düzensiz bir yanıtın sonucu olarak biyokimyasal, fizyolojik ve patolojik anormalliklerle ilişkili bir sendromdur. Sepsis, özellikle hemen tanınmaz ve tedavi edilmezse, enfeksiyondan ölümün birincil nedenidir ve tanınması acil dikkat gerektirir [1]. Basit bir ifadeyle sepsis, vücudun bir enfeksiyona verdiği yanıtın kendi doku ve organlarına zarar vermesiyle ortaya çıkan yaşamı tehdit eden bir durumdur.

Umut Kaya ve arkadaşları 2018 yılında yaptıkları çalışmada, sepsis teşhisine yardımcı olan çok katmanlı yapay sinir ağlarını kullanan bir model önerdi. Yapay sinir ağı modellerinin inşası için ileri beslemeli geri yayılım ağ yapısı ve Levenberg Marquardt eğitim algoritması kullandılar. Modelin girdi ve çıktı değişkenleri, doktorların sepsis hastalığını teşhis etmek ve sepsis seviyesini belirlemek için kullandıkları parametrelerdi. Kullanılan yöntem, sepsis hastalığının erken teşhisi için alternatif bir tahmin modeli sağladı. İstanbul'da 18-65 yaş arası yoğun bakım hastalarına ait veriler kullanılarak, yapay sinir ağları yardımıyla sepsise yakalanma riski tahmin edilmeye çalışıldı. Yoğun bakım ünitesine yerleştirilen ve sepsis tanısı alan hastaların girdi ve çıktıları 2017 yılında sepsis tanımı ile oluşturulmuş ve doktorların kullandığı algoritmalar ve değişkenler incelenerek yapay sinir ağının modellenmesinde kullanılmıştır. Sepsisin erken teşhisi için modellenmiş girdiler, doktorlar tarafından sepsis tanısı koymak ve sepsis seviyesini belirlemek için kullanılan parametrelerdi. Bu modelde% 99 eğitim, test ve doğruluk değerleri elde edilmiştir [3].

Gültepe ve meslektaşları 2014 yılında yaptıkları bu çalışmada, rutin olarak ölçülen yaşamsal belirtilere ve laboratuvar çalışmalarına dayalı olarak yüksek hiperlaktatemi riski taşıyan hastaları belirlemek için bir karar destek sistemi geliştirmeyi amaçlamışlardır. Kaliforniya Üniversitesi Davis Sağlık Sisteminde bulunan, en az iki sistemik inflamatuar yanıt sendromu kriterini karşılayan 741 yetişkin hastanın elektronik sağlık kayıtları, hastaların yaşamsal belirtileri olan beyaz kan hücresi sayımı (WBC) ile sepsis oluşumu ve ölüm oranını ilişkilendirmek için kullanıldı. Üretken ve ayırt edici sınıflandırma (Naive Bayes, destek vektör makineleri (SVM), Gauss karışım modelleri, gizli Markov modelleri), heterojen hasta verilerini entegre etmek ve laktat seviyesi ve ölüm riski çıkarımı için bir öngörü aracı oluşturmak için kullanıldı. Yaptıkları bu çalışmada, Hasta verilerinin geçici yönü ve değişkenliği göz önüne alındığında, birkaç klinik değişkenle laktat seviyeleri ve ölüm riskinin etkili tahminleri sağlanabilir sonucuna ulaştılar [4].

Desautels ve arkadaşları 2016 yılında yaptıkları çalışmada, Retrospektif verilerdeki yeni Sepsis-3 tanımları için bir sepsis tahmin yöntemini incelemek ve doğrulamak için InSight, elektronik sağlık kaydı verilerinden minimum bir değişken seti kullanarak tahminler yapıldı, bu yaklaşımın performansını mevcut puanlama sistemleriyle karşılaştırdılar ve Veri seyrekliğinin InSight performansı üzerindeki etkilerini araştırdılar.

15 yaş ve üzeri yoğun bakım ünitesi (YBÜ) hastalarıyla sınırlı 3 farklı veri seti ile retrospektif çok parametreli akıllı yoğun bakımda izleme (Multiparameter Intelligent Monitoring in Intensive Care - MIMIC) kullanarak sepsisi tahmin etmek için kolayca elde edilen hasta verilerinin (hayati değerler, periferik kapiller oksijen satürasyonu, Glasgow Koma Skoru ve yaş) çok değişkenli kombinasyonlarını kullanan bir makine öğrenimi sınıflandırma sistemi olan InSight'ı uyguladılar. Sepsis sendromunun Sepsis-3 tanımlarını takiben, InSight'ın sınıflandırma performansını hızlı sıralı organ yetmezliği değerlendirmesi (qSOFA), modifiye erken uyarı skoru (MEWS), sistemik inflamatuar yanıt sendromu (SIRS), basitleştirilmiş akut fizyoloji skoru (SAPS) ile karşılaştırdılar. Ayrıca InSight sisteminin sağlamlığını, bireysel girdi gözlemlerinin rastgele silinmesine karşı test ettiler.

Yaptıkları bu çalışmada; hayati değerlerden biraz daha fazlasını kullanmalarına rağmen, 'InSight sepsis başlangıcını tahmin etmek için etkili bir araçtır ve rastgele eksik verilerle bile iyi performans gösterir' sonucuna ulaştılar [5].

Fleuren ve arkadaşları 2020 yılında yaptıkları çalışmada sepsisi tahmin etmek için gerçek zamanlı modellerin performanslarını sistematik bir inceleme ve meta-analiz yaparak değerlendirmişlerdir. Kanıt kalitesi, önyargı riskini değerlendirmek için özel bir tanısal doğruluk çalışmalarının kalite değerlendirmesi (Quality Assessment of Diagnostic Accuracy Studies - QUADAS-2) kontrol listesiyle birlikte önerilerin değerlendirilmesi, geliştirilmesi ve değerlendirilmesi (Grading of Recommendations Assessment, Development and Evaluation - GRADE) metodolojisi kullanılarak değerlendirildi. Alıcı çalışma özelliği (AUROC) metriğinin eğrisi altında rapor edilen bir alana sahip modeller, model performansına en güçlü katkıları belirlemek için meta-analiz edildi. Bu çalışmada sistematik inceleme ve meta-analiz, geriye dönük verilere göre, bireysel makine öğrenimi modellerinin sepsis başlangıcını vaktinden önce doğru bir şekilde tahmin edebildiğini gösterdiği sonucuna ulaşılmıştır [6].

Gültepe ve arkadaşları, sepsis'in enfeksiyona karşı düzensiz bir immün yanıttan kaynaklanan ciddi bir tıbbi durum olduğunu 2012 yılında yaptıkları çalışmada bildirmektedirler. Sepsis semptomlarının erken tanısı, hastalığın her dört kişiden birinin ölümüyle sonuçlanan daha ağır aşamalara ilerlemesini önlemek için, 233 sepsis vakası içeren 1492 hastayı içeren bir EMR veritabanını kullanarak, sepsisin göstergesi olan ve bayes çıkarım ağının eğitimi için daha fazla kullanılabilecek özellikleri belirlemek için bir kümeleme analizi kullandılar. Bayes ağı, sepsis hastaları için sistemik enflamatuar yanıt sendromu kriterleri, ortalama arter basıncı ve laktat seviyeleri kullanılarak oluşturulmuştur. Ortaya çıkan ağ, laktat seviyeleri ile sepsis arasında açık bir korelasyon ortaya koymaktadır. Ağ yapısının daha fazla modellenebilmesi, laktat seviyelerinin organ disfonksiyonunun varlığından bağımsız olarak mortaliteyi tahmin edebildiğini gösterdi. Sonuç olarak, Bayes ağlarının; sepsis hastalarının laktat düzeyleri ile sepsis arasındaki ilişkiyi yakalayabildiği gösterilmiştir [7].

Stanculescu ve arkadaşları 2014 yılında, yeni-doğan sepsisinin erken teşhisi için uygulanan otoregressif gizli markov (autoregressive hidden markov) modelleri üzerinde yaptıkları çalışmada, prematüre bebeklerin yoğun bakım altında olduğu dönemde geç başlayan prematüre sepsisi en önemli klinik kaygılardan biri olarak değerlendirdiler. Uyguladıkları model, enfeksiyonun başlangıcı hakkında gerçek zamanlı tahminler üretebilir ve eksik verilerle çalışabilir. Sepsis tanısı için otoregressif

gizli markov modelinin etkinliğini Edinburgh Kraliyet Kliniği'ndeki yeni-doğan yoğun bakım ünitesinden gelen verilerle değerlendirdiler [8].

Guillen ve meslektaşları, yoğun bakımdaki yetişkin hastalardan toplanan klinik laboratuvar değerlerini ve yaşamsal bulguları kullanarak şiddetli sepsisin tahmin edilmesi için yeni bir çerçeve belirlemeyi ve erken tahmincileri belirlemeyi amaçladıklarını söyledi. Yaşamsal bulguları, laboratuvar değerlerini veya yaşamsal bulguların ve laboratuvar değerlerinin bir kombinasyonunu kullanarak lojistik gerileme modellerini keşfetmeye, vektör makinelerini ve lojistik model ağaçları yöntemlerini desteklemeye çalıştılar. Yaşamsal ve laboratuvar bulgularının kullanıldığı tahmine dayalı destek vektör makinesi modeli yoğun bakım hastalarının retrospektif kohortlarına uygulandığında, 3446 hastadan 339(e) hastada ciddi sepsis geliştiğine dair doğru bir tahminle diğer iki modele göre daha yüksek tahmin gücü göstermiştir. Ancak, bu model diğer iki modelin klinik şeffaflığına ve yorumlanabilirliğine sahip değildir. Bu çerçeveye dayalı ve geliştirilen modellerin yoğun bakımda ve yoğun bakım dışı ortamda klinik karar desteği için önerilebileceğini göstermişlerdir [9].

Bu çalışmalar incelendiğinde uluslararası veri tabanından alınan farklı veri setleri üzerinde makine öğrenmesi yöntemleri kullanılmasına rağmen önerdiğimiz kadar çok algoritma denenmemiştir. Ayrıca literatürde yer alan bu çalışmaların performans parametreleri incelendiğinde tez çalışmamızın sonuçları açısından bir iyileşme olduğu görülmektedir. Bu yönüyle literatüre katkı sağladığı anlaşılmaktadır.

2.2. Veritabanı Tanıtımı

Tezin bu bölümünde ise, kullanılan veri kümesinin temel bileşenlerinin ayrıntılarını, açıklamasını ve çeşitli makine öğrenimi algoritmalarına ve önerilen yaklaşımlara genel bir bakış sunacaktır.

Sepsisin meydana gelmesinde birçok enfeksiyon türü sebep olabilir. Bunların en sık görülenleri menenjit, idrar yolu enfeksiyonları, zatürree, solunum yolu enfeksiyonları, cilt enfeksiyonları, karın içi enfeksiyonlar, yaralar ve yanıklardır. 1 yaş altı çocuklar, hamileler, kronik hastalığı olanlar ve yaşlılar sepsis için risk grubundadır. Sepsis insidansı, kalp, kanser ve inme gibi bilinen diğer hastalıkların insidansından

daha yüksektir. Kolon ve meme kanserinden ölümlerin toplamından daha fazla insan sepsisten ölüyor.

Yakın zamanda yayınlanan Sepsis-3 kriterlerine göre, resmi olarak, şüpheli veya kanıtlanmış bir enfeksiyon nedeniyle sıralı organ yetmezliği değerlendirmesi (SOFA) skorunda ≥2 puanlık akut artış olarak tanımlanmaktadır [1]. Laboratuvar ve klinik parametrelerin izlenmesi yoluyla saptanan SOFA skorundaki artışlar, kardiyovasküler, solunum, böbrek, karaciğer, pıhtılaşma ve nörolojik sistemlerdeki olası bozuklukları yansıtır [32]. Genel olarak, yeni tanım, yoğun bakım ünitelerinde (YBÜ) yüksek ölüm riski taşıyan septik hastaları belirlemek için öncekinden daha iyi performans sunar [1]. Sepsis-3 ayrıca en kötü sepsis sonuçları olan hızlı SOFA (qSOFA) hastalarını taramak için kullanılır. Septik şok, altta yatan dolaşım ve hücresel/metabolik anormalliklerin mortaliteyi önemli ölçüde artıracak kadar derin olduğu sepsisin alt kümesidir [19].

Şiddetli sepsis ve septik şok, yoğun bakım ünitelerinde (YBÜ) hala önde gelen ölüm nedenleridir ve zamanında teşhis, tedavi sonuçları için çok önemlidir. Sepsis yönetiminin zamana çok duyarlı olduğu düşünüldüğünde, sepsisin erken tahmin edilmesi mortalitenin (ölüm hızı) önlenmesinde çok önemlidir. Elektronik tıbbi kayıtların (EMR) ilerlemesi, tıpta yapay zekanın (AI) gelişimini kolaylaştırabilecek büyük hacimli klinik verileri depolama imkanı sunar.

Yüksek riskli sepsis grubu; 65 yaş ve üstü yetişkinleri, diyabet, kanser, akciğer, karaciğer ve böbrek hastalığı gibi kronik hastalıkları olan kişileri, bağışıklık sistemi zayıf olanları, daha önce sepsis geçirmiş kişileri ve bir yaşından küçük çocukları içerir [20].

Tez çalışmasında kullandığımız, sepsis tanısında kullanılan klinik bulgular aşağıdaki gibidir.

1. Ateş (> 38.3C)
2. Hipotermi (düşük vücut ısısı) (<36C)
3. Kalp atış hızı >90
4. Takipne (hızlı solunum) (>30)
5. Mental durum değişikliği
6. Diyabet yokluğunda hiperglisemi (artmış glukoz) (>140 mg/dL veya 7.7 mmol/L)

7. Lökositoz (Artan Lökosit) >12.000/µL
8. Lökopeni (Lökositlerin Azaltılması) <4,000µL
9. İmmatür form beyaz kan hücresi sayımı Lökosit: N, Bant> %10
10. Plazma C-reaktif protein (CRP) normal değerden >2 SD
11. Plazma prokalsitonin normal değerden >2 SD
12. Hipotansiyon (SAB<90 mmHg, MAP<70 mmHg, SAB düşüşü 40 mmHg'den fazla)
13. Hipoksemi (Oksijen azalması) (PaO_2/FiO_2<300)
14. Saturasyon
15. Kreatinin artışı (>0.5 mg veya 44.2 µmol/L)
16. Pıhtılaşma bozukluğu (INR>1.5, aPTT>60 sn)
17. Trombositopeni (Azalmış Trombosit) (<100.000)
18. Hiperbilirubinemi (T.Bil>4mg veya 70 µmol/L)
19. Laktat yüksekliği (laktat>1 mmol/L)

Burada yer alan ifadelerden SBP; sistolik kan basıncı, MAP; ortalama kan basıncı, INR; uluslararası normalleştirilmiş oran, aPTT; aktive parsiyel tromboplastin zamanı, SD; Standart sapma, PaO2; Kısmi arteriyel oksijen basıncını ifade etmektedir. [21].

Veri setindeki 18 özellik sayısaldır. Sadece mental durum sayısal değildir. İkili kategorik (ikili) özelliklerde "0" zihinsel durumun kapalı olduğunu, "1" ise zihinsel durumun açık olduğunu gösterir. Zihinsel durum, sınıflandırma görevi için kullanılan veri kümesini yapmak için ikili olarak dönüştürülmüştür.

2.3. Kullanılan Yöntemler

2.3.1. Makine öğrenimi algoritmaları

Makine öğrenmesi, günümüzde yapay zekâ biliminin en yaygın kullanılan alt dallarından biridir. Yapay zekâ tabanlı sistemler, sürekli olarak insanlar gibi davranmayı öğrenir ve eğitir ve kendi kendine muhakeme ve problem çözme

yetenekleri geliştirir. Yapay zekâ yaklaşımını kullanan uygun şekilde eğitilmiş sistemler, daha yüksek derecede matematiksel manipülasyonlar olmadan karmaşık sorunları sorunsuz bir şekilde çözebilir. Bu nedenle, Yapay zekâ yaklaşımı, çeşitli geleneksel problem çözme tekniklerine umut verici bir alternatif haline geldi. Makine öğrenmesi, makinelerin yapay zekâ algoritmaları ile insan zekâsının düşüncelerine benzer kalıpları tanıma ve karar verme eylemlerini öğrenme ve gerçekleştirme yeteneğidir. Algoritmalar, öğrenme için mevcut örneklerin sayısı arttıkça performanslarını uyarlanabilir bir şekilde geliştirir [22]. Makine öğrenimi temel olarak geçmiş deneyimlerden geleceği tahmin etmektir [23]. Sınıflandırma, tahmin, tahmin oluşturma gibi pek çok amaç için kullanılan, etkin ve hatasız tahminler yapabilen algoritmalar, kuralları verilerden öğrenebilen, değişikliklere uyum sağlayabilen ve tecrübe ile performansını geliştirebilen yazılım tasarımından oluşmaktadır [24]. Makine öğrenimi alanı, deneyimle otomatik olarak gelişen bilgisayar programlarının nasıl oluşturulacağıyla ilgilenerek ve performansı optimize etmek için örnek verileri veya geçmiş deneyimleri kullanarak bilgisayar programlamadır [25]. Bir makine, deneyimler yoluyla performansını geliştirdiğinde, makinenin öğrenmiş olduğu kabul edilir; burada bahsedilen öğrenme, verileri tutan ve faydalı örnekleri ortaya çıkaran bilgisayar modellerini gerektirir [26].

Karar ağacı sınıflandırıcı algoritması, naive bayes sınıflandırıcı, yapay sinir ağları, destek vektör makinesi ve k-en yakın komşu algoritması gibi makine öğrenmesinde sınıflandırma için literatürde kullanılan çok sayıda algoritma vardır. Örneklerden çıkarım olarak da bilinen sınıflandırmanın amacı, daha önce algoritmaya dahil edilmemiş örnekleri, kavram tanımı elde edildikten sonra en yüksek doğrulukla elde edecek bir sınıflandırıcı geliştirmektir [27].

Yapay zekâ ve makine öğrenmesi uygulamalarının sağlık alanında kullanımı, tıbbi teşhis ve hastalık takibi, maliyet tahmini, görüntüleme analizi, kaynak planlama ve acil durum yönetimi, yapılandırılmamış verilerin işlenmesi gibi birçok alt faaliyet alanında gerçekleştirilmektedir. Makine öğrenimi, hem sağlık hizmetlerinin artan maliyetini azaltmak hem de daha iyi bir hasta-doktor ilişkisi kurmaya yardımcı olmak için bir çözüm sunar. Son zamanlarda, sağlık alanında büyük miktarda veri mevcut hale geldi. Bu, yapılandırılmış veya yapılandırılmamış verilere sahip EMR'leri içerir. Yapılandırılmış sağlık verileri, bir veri tabanında kategorize edilmesi kolay bilgilerdir; hasta ağırlıkları, vücut sıcaklıkları, baş ağrısı, karın ağrısı vb. tıbbi verilerin

çoğunluğunun çeşitli notlar, raporlar, resimler, ses ve video kaydı. Ancak yapay zekânın sağlık alanında kullanımına ilişkin olarak; Veri yönetimi, klinik verilerin doğruluğu, veri koruma ile ilgili etik ve yasal süreçler, yapay zekanın sağlık alanında kullanımını sınırlandırmaktadır.

Makine öğrenme algoritmaları, veri setinden üzerindeki öğrenme modeline göre:

- Denetimli öğrenme (supervised learning),
- Denetimsiz öğrenme (unsupervised learning),
- Pekiştirmeli (takviyeli) öğrenme (reinforcement learning)

olarak sınıflandırılır. Şekil 2.1'de görüldüğü gibi derim öğrenme algoritmaları denetimli, denetimsiz ve pekiştirmeli öğrenme tekniklerinin üçü ile de uygulanabilir.



Şekil 2.1. Yapay zekâ, makine öğrenmesi ve öğrenme teknikleri arasındaki ilişki

2.3.1.1. Denetimli öğrenme (Supervised Learning)

Denetimli öğrenme etiket adı da verilen cevap değişkeninin bulunduğu veri setlerine uygulanır. Burada cevap değişkeni sürekli veya kategorik olabilir. Denetimli öğrenme bir dizi eğitim örneğinden oluşan bu etiketli eğitim verilerinden bir fonksiyon çıkarımı yapar [46]. Örneğin bir dizi hastaya ait olan veri setinde cevap değişkenimiz her bir hastanın kanser olup olmadığı olabilir veya verilen bir ülke veya şehirdeki ev fiyatlarını veren veri kümesinde cevap değişkenimiz ev fiyatları olabilir [47]. Denetimli öğrenme örnek girdi cevap çiftlerine göre bir girdiyi bir cevaba eşleyen fonksiyonu öğrenerek eğitimde kullanılmayan yeni örnekler için kullanılabilecek fonksiyon çıkarımını üretir [48].

2.3.1.2. Denetimsiz öğrenme (Unsupervised Learning)

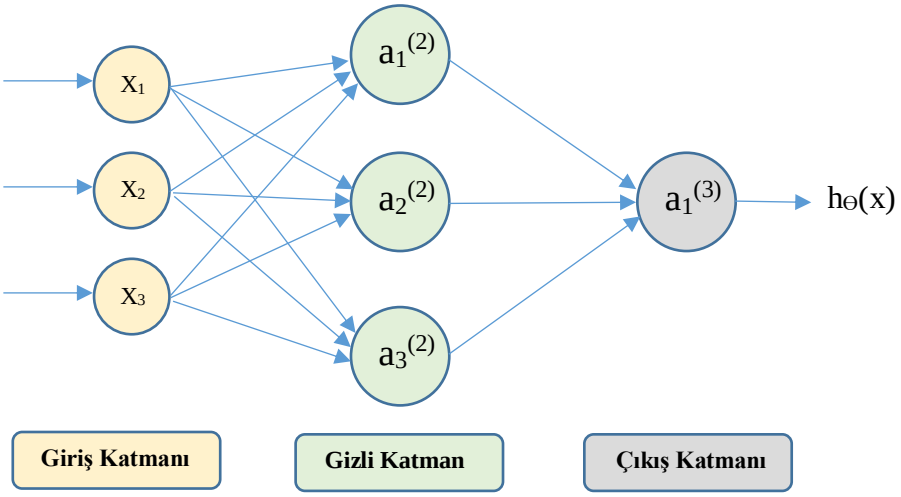
Cevap değişkenlerinin diğer adıyla etiketli verilerin olmadığı veri setlerinde öğrenme denetimsiz olarak gerçekleştirilir. Denetimsiz öğrenme veri kümesi içerisinde bazı benzerlik veya mesafe ölçümlerine göre gerçekleşir. Denetimsiz öğrenmede en çok kullanılan teknik kümelemedir. Denetimsiz öğrenmeye örnek olarak sosyal ağ analizi, görüntü segmentasyonu, klimatoloji, pazarlamada müşteri segmentasyonu ve daha birçok uygulama sayılabilir [47].

2.3.1.3. Pekiştirmeli öğrenme (Reinforcement Learning)

En iyi sonucu üretebilmek için denetimli ve denetimsiz öğrenme algoritmalarının her ikisi de temiz ve doğru veriye ihtiyaç duyar. Algoritmanın görmediği örnekler üzerinde de çalışabilmesi için eğitim setinde kullanılan verinin kapsayıcı olması gerekir. Pekiştirmeli öğrenme, yalnızca verilerin başlangıç durumunun bir girdi olarak mevcut olduğu ve tek bir olası cevabın olmadığı ancak çok fazla sonucun mümkün olduğu durumlarda ideal bir seçimdir [47]. Robotik ve sürücüsüz araçlarda uygulanan makine öğrenme algoritmaları pekiştirmeli öğrenme algoritmaları sınıfındandır. Pekiştirmeli öğrenme algoritmaları sürekli olarak çevresinden öğrenen algoritmalardır.

2.3.2. Yapay Sinir Ağları (YSA)

Makine öğrenmesi yöntemlerinden biri olan YSA, nöron adı verilen basit bilgi işlem birimlerinin bir araya gelmesiyle oluşan sistemlerdir. YSA'lar, değişkenler arasındaki doğrusal olmayan ilişkileri öğrenme ve yüksek dereceli ilişkileri tanıma konusunda oldukça yeteneklidir. YSA'ların karmaşık ilişkileri modelleme gücü, karmaşık matematiksel modellerde değil, çok sayıda basit nöronun etkileşimli birleşiminde yatmaktadır. YSA'lar, denetimli, denetimsiz ve pekiştirmeli öğrenme algoritmalarına tam olarak uygulanabilen modellerdir.



Şekil 2.2. Tek gizli katmanlı YSA

2.3.3. Destek Vektör Makinesi (SVM)

Destek vektör makinesi başlangıçta iki sınıfı ayırmak için kullanılırken, zamanla geliştirilmiş ve doğrusal olmayan sistemlerle regresyon, sınıflandırma ve aykırı değer algılama problemlerinde başarıyla kullanılmıştır. İstatistiksel öğrenme teorisine dayanan denetimli bir parametrik makine öğrenme algoritmasıdır. Teorik olarak, bir SVM, bir tümör numunesinden veya periferik sıvıdan türetilen gen ekspresyon profilini inceleyebilir ve bir teşhis veya prognoza ulaşabilir. SVM algoritmasında iki sınıfı ayırmak için sınıfları oluşturan veriler arasına paralel bir çizgi/hiper düzlem çizilir. Sınıfları ayırmak için kullanılan yapı, iki boyutlu uzayda bir doğru, üç boyutlu uzayda bir düzlem olarak temsil edilir. Hiper düzleme en yakın verilere destek vektörleri denir. Zıt sınıfların destek vektörleri arasındaki marj maksimize edilerek daha dayanıklı hale getirilir [28].

2.3.4. k-En Yakın Komşu Algoritması (KNN)

K-NN, literatürdeki makine öğrenmesi algoritmaları arasında en basit ve en eski parametrik olmayan denetimli sınıflandırma yaklaşımlarından biri olarak bilinmektedir. Toplam veri setinde özel bir k sayısı tanımlanarak en yakın komşuların ortalama/mod sınıfları elde edilir ve yeni nesne komşularına en yakın sınıfa atanır. yeni nesnenin komşularına olan uzaklıkları euclid, manhattan, minkowski ve chebyshev gibi fonksiyonlarla hesaplanabilir. k sayısının yeterince büyük olması koşuluyla eğitim verilerine karşı sağlam bir yapıya sahiptir. Veri seti ve k boyutu arttığında işlem süresi oldukça artar ve bu yaklaşımda tüm bu mesafe hesaplamalarının hafızada tutulması gerekir [29]. Bu nedenle k değerinin seçimi son derece önemlidir.

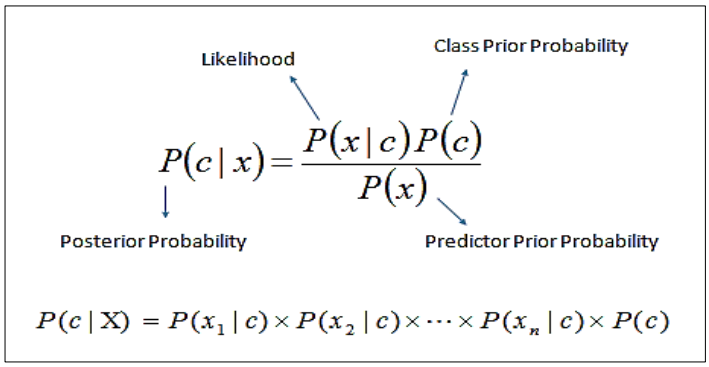
2.3.5. Karar Ağacı Sınıflandırıcı Algoritması

Karar ağacı algoritması, denetimli öğrenme kategorisine girer. Hem regresyon hem de sınıflandırma problemlerini çözmek için kullanılırlar. Karar ağacı, her yaprak düğümün bir sınıf etiketine karşılık geldiği ve özniteliklerin ağacın iç düğümünde temsil edildiği sorunu çözmek için ağaç temsilini kullanır. Hedef değişkenin ayrık değerler alabildiği ağaç modellerine sınıflandırma ağaçları denir; bu ağaç yapılarında yapraklar sınıf etiketlerini temsil eder ve dallar bu sınıf etiketlerini ortaya çıkaran özelliklerin kombinasyonlarını temsil eder. Hedef değişkenin sürekli değerler alabileceği karar ağaçlarına regresyon ağaçları denir. Karar analizinde, kararları görsel ve açık bir şekilde temsil etmek için bir karar ağacı kullanılabilir. Karar ağacı yöntemi, girdi verilerinin bir algoritma yardımıyla gruplara ayrılarak tüm elemanların aynı sınıf etiketine sahip olması için yapılan sınıflandırma işlemidir. İlk olarak, veri seti için entropi Denklem 2.1'de gösterilen H yöntemi ile hesaplanır. H(S) ile sembolize edilen entropi, veri setindeki belirsizlik miktarının ölçüsüdür. S, entropiyi hesaplamak için kullanılan mevcut kümedir. S'de, C = {True, False}, sınıflar kümesidir. P fonksiyonu, "c" sınıfındaki eleman sayısının kümedeki eleman sayısına oranıdır [30].

$$H(S) = \sum_{c \in C} -\mathrm{p(c)}\log_2 \mathrm{p(c)} \qquad (2.1)$$

2.3.6. Naive Bayes

Naive bayes, Denklem 2.2'de görüldüğü gibi bayes teoremine dayanan istatistiksel bir yaklaşım olan bir sınıflandırma modelidir, NB sınıflandırıcısı, sınıftaki belirli bir özelliğin etkisinin diğer niteliklerden bağımsız olduğunu varsayar. Gerçekten de, bu nitelikler bağlantılıysa, bağımsız olarak tahmin edilirler. Bu varsayım, hesaplamayı kolaylaştırır ve bu nedenle naive olarak adlandırılır. Bu varsayım aynı zamanda sınıf koşullu bağımsızlığı olarak da adlandırılır [31].



(2.2)

2.3.7. Lojistik regresyon

Regresyon analizi, iki veya daha fazla değişken arasındaki ilişkiyi ölçmek amacı ile kullanılmakta ve tanımlayıcı çıkarımsal istatistik sağlamaktadır. Hedeflenen, bağımlı ve bağımsız değişken arasındaki ilişkiyi en az sayıda değişken ile en iyi uyuma sahip olacak şekilde oluşturabilecek en iyi modeli kurmaktır [49].

Yayınlanmış eğilim puanı analizlerinin büyük çoğunluğu, puanları tahmin etmek için lojistik regresyon kullanır. Ayrıca, lojistik regresyon, çeşitli disiplinlerdeki araştırmacılar için tanıdık ve makul ölçüde iyi anlaşılmış bir araçtır ve çoğu istatistiksel pakette (örneğin, SAS, STATA, R) uygulanması kolaydır.

Literatürdeki araştırmalar incelendiğinde, hastalık teşhisi, genetik araştırmalar, deprem tahmini, stok tahmini, metin sınıflandırması için sınıflandırmaya yönelik makine öğrenmesi kullanıldığını ortaya koymaktadır. Bu alanlardan sağlık, makine öğrenmesi çalışmalarının yapıldığı önemli bir sektördür ve başarılı sonuçlara ulaşılması durumunda hem doktorlara hem de hastalara fayda sağlayacağı düşünülmektedir. Bu tez çalışmasında bazı kan testi değerleri (PCT, PLT, WBC, Glikoz (açlık), kreatinin (kan), Laktik Dehidrojenaz (LDH), INR, CRP, Protrombin Zamanı, Bilirubin, PCO2) ve hayati bulgular (tansiyon, nabız, ateş, satürasyon vb.)

gibi özelliklere bakarak sepsisin erken teşhisini tahmin etmek için MATLAB programlama dilinin makine öğrenimi araç kutusu kullanılarak farklı yapılara sahip sınıflandırıcı modeller geliştirilmiştir. Bu modeller, aşağıda detayları verilen çok katmanlı sinir ağı, karar ağacı sınıflandırıcı algoritması, naive bayes sınıflandırıcı, yapay sinir ağları (YSA), destek vektör makinesi (SVM) ve k-en yakın komşu algoritması (KNN) yöntemleridir. Tüm sayısal sonuçlar, Windows 10 işletim sistemi altında bir Intel işlemci üzerinde MATLAB R2021b kullanılarak elde edildi.

2.4. Başarım Metriklerinin Tanıtımı

İkili sınıflandırma problemleri için, sınıflandırma eğitimi sırasında en iyi (optimal) çözümün ayrım değerlendirmesi Tablo 2.1'de gösterildiği gibi karışıklık matrisine dayalı olarak tanımlanabilir. Tablonun satırı tahmin edilen sınıfı, sütun ise gerçek sınıfı temsil eder. Bu karışıklık matrisinden, doğru pozitif (DP) ve doğru negatif (DN) doğru şekilde sınıflandırılan pozitif ve negatif örneklerin sayısını gösterir. Bu arada yanlış negatif (YN) ve yanlış pozitif (YP) sırasıyla yanlış sınıflandırılan negatif ve pozitif örneklerin sayısını gösterir. Tablo 2.1'den, farklı değerlendirme odaklarıyla sınıflandırıcının performansını değerlendirmek için Tablo 2.2'de gösterildiği gibi yaygın olarak kullanılan birkaç metrik oluşturulabilir. Çok sınıflı problemler nedeniyle, Tablo 2.2'de listelenen metriklerden birkaçı, çok sınıflı sınıflandırma değerlendirmeleri için genişletilmiştir.

Tablo 2.1. İkili sınıflandırma için karışıklık matrisi

	Gerçek Pozitif Sınıf	**Gerçek Negatif Sınıf**
Öngörülen Pozitif Sınıf	Doğru pozitif (dp)	Yanlış negatif (yn)
Öngörülen Negatif Sınıf	Yanlış pozitif (yp)	Doğru negatif (dn)

Doğruluk pratikte hem ikili hem de çok sınıflı sınıflandırma problemleri için en çok kullanılan değerlendirme metriğidir. Doğruluk sayesinde, üretilen çözümün kalitesi, toplam örnekler üzerinden doğru tahminlerin yüzdesine dayalı olarak değerlendirilir [50]. Tamamlayıcı doğruluk metriği, üretilen çözümü yanlış tahmin yüzdesine göre değerlendiren hata oranıdır. Bu metriklerin her ikisi de pratikte, optimal çözümü ayırt etmek ve seçmek için araştırmacılar tarafından yaygın olarak kullanılmıştır.

Doğruluk veya hata oranının avantajları, bu metriğin daha az karmaşıklıkla hesaplanmasının kolay olmasıdır; çok sınıflı ve çok etiketli problemler için geçerlidir; kullanımı kolay puanlama ve insan tarafından anlaşılması kolaydır. Pek çok çalışmanın işaret ettiği gibi, doğruluk metriğinin değerlendirme ve ayırt etme süreçlerinde sınırlamaları vardır. Doğruluğun ana sınırlamalarından biri, daha az ayırt edici ve daha az ayırt edilebilir değerler üretmesidir [51, 52]. Sonuç olarak, optimal sınıflandırıcıyı seçme ve belirlemede doğruluk için daha az ayırt edici güce yol açar.

Tablo 2.2. Sınıflandırma değerlendirmeleri için eşik metrikleri

Metrik	Formül	Değerlendirme
Doğruluk (acc)	$Doğruluk = \frac{DP + DN}{DP + DN + YP + YN}$	Genel olarak, doğruluk metriği, doğru tahminlerin değerlendirilen toplam örnek sayısına oranını ölçer.
Hata Oranı (err)	$Hata\ oranı = \frac{YP + YN}{DP + DN + YP + YN}$	Yanlış sınıflandırma hatası, yanlış tahminlerin değerlendirilen toplam örnek sayısına oranını ölçer.
Duyarlılık	$Duyarlılık = Anma = DPO = \frac{DP}{DP + YN}$	Bu metrik, doğru sınıflandırılmış pozitif kalıpların kesrini ölçmek için kullanılır
Özgüllük	$Özgüllük = DNO = \frac{DN}{DN + YP}$	Bu metrik, doğru şekilde sınıflandırılan negatif kalıpların oranını ölçmek için kullanılır.
Kesinlik	$Kesinlik = Pozitif\ Tahmin\ Değeri = = \frac{DP}{DP + YP}$	Kesinlik, pozitif bir sınıfta toplam tahmin edilen kalıplardan doğru tahmin edilen pozitif kalıpları ölçmek için kullanılır.
F-Ölçü (FM)	$f_\beta - ölçütü = \frac{(1 + \beta^2)x(kesinlik * anma)}{\beta^2 + kesinlik * anma}$	Bu metrik, kesinlik ve anma değerlerinin harmonik ortalaması alınarak elde edilmektedir.

Alıcı işlem karakteristikleri eğrileri altında kalan alan (AUC); popüler sıralama türü metriklerinden biridir. AUC, optimize edilmiş bir öğrenme modeli oluşturmak ve ayrıca öğrenme algoritmalarını [52, 53] karşılaştırmak için kullanıldı. Eşik ve olasılık ölçümlerinden farklı olarak AUC değeri, bir sınıflandırıcının genel sıralama performansını yansıtır.

AUC'nin sınıflandırma eğitimi sırasında sınıflandırıcı performansını değerlendirmek ve optimal bir çözümü ayırt etmek için doğruluk metriğinden [53] teorik ve ampirik olarak daha iyi olduğu kanıtlanmıştır. AUC'nin performansı değerlendirme ve ayırt etme süreçleri için mükemmel olmasına rağmen, AUC'nin

hesaplama maliyeti, özellikle çok sınıflı problemlerin üretilen çözümlerinin bir hacmini ayırt etmek için yüksektir.

2.5. Özellik Mühendisliği

Makine öğrenimi algoritmaları, sorunu temsil eden girdi verilerinden öğrenir. Belirlenen bir problemin çözümü için algoritmanın doğru verilerle beslenmesi çok önemlidir. Veriler gerekli ve dikkatli bir şekilde toplanmış olsa bile, önemli özelliklerin belirli bir format ve hassasiyette sisteme dahil edilmesi gerekir. Bu önemli veri biçimini elde etmek için:

1. Veri seçimi
2. Veri ön işleme
3. Veri dönüşümü

işlemleri uygulanır.

Özellik mühendisliği, ham verileri bir modelin daha iyi performans göstermesine yardımcı olacak bir forma dönüştürme uygulamasıdır. Veri dönüştürme adımlarına özellik mühendisliği de denir.

Mevcut tüm veri kümesinden belirli bir alt küme seçilerek "Veri Seçimi" işlemi yapılır. Genel olarak var olan veri kümesinin tamamının "çok daha iyi" felsefesi ile seçilmesi isteği olmasına karşın bazı durumlarda bunun doğruluğu tesis edilemeyebilir. Bunun için çözülmesi planlana problem üzerinde hangi verilerin fiilen etkili olduğunu bilmek gerekir. Önemli veriler için veri seçimi varsayımlar üzerinden ve daha sonra varsayımı doğrulayacak şekilde yapılabilir. Bu çalışmada geleneksel veri seçim yöntemi olan "traditional validation" yöntemi ve k-kat çaprazlama yöntemleri ayrı ayrı kullanılmıştır. Geleneksel bölütlemede verilerin rastgele olarak seçilmiş %70'i eğitim işleminde kullanılmış olup %30'luk kısmı ise performansın ölçülebilmesi için test işlemlerinde kullanılmıştır. K-kat çaprazlama yönteminde veri seti 3, 5 ve 10 bileşene bölütlenerek eğitim ve test setleri oluşturulmuştur.

Literatürde yaygın olarak kullanılan 3 adet veri ön işleme adımı vardır: temizleme, biçimlendirme ve örnekleme. Bu işlemler sırasıyla şu şekilde gerçekleştirilir: ilk olarak gereksiz ve eksik veriler verilerden temizlenir. Ardından, temizlenen veriler üzerinde sayısal değerler ve normalizasyon olarak bilinen bir

biçimlendirme işlemi gerçekleştirilir. Makine öğrenimi yöntemleri metin verileri üzerinde çalışamaz. Ağa verilen girdiler sayısal değerlere dönüştürülür. Bu dönüşüme örnek olarak cinsiyet verilebilir (erkek = 1, kadın = 2 olarak sayısallaştırılmıştır). Ardından ağın girişlerine uygulanan özelliklerin ağ üzerinde yanlı bir etki yaratmasını önlemek için normalizasyon işlemi gerçekleştirilir. Tüm girdilerin ağ üzerinde homojen bir etkiye sahip olması sağlandıktan sonra örnekleme işlemi gerçekleştirilir. Burada kullanılan iki yaygın yöntem: geleneksel veri segmentasyonu ve k-katlı çapraz doğrulama.

Son adım ise kullanılan algoritmanın problem alanını doğrudan etkileyen "veri dönüştürme" işlemidir. Birçok çalışmada birden fazla veri dönüştürme işlemine ihtiyaç duyulabilir. Ölçekleme, özellik ayrıştırma, birleştirme olmak üzere üç genel veri dönüştürme yöntemi vardır. İncelendiğinde birçok problemin ön işleme adımında birleştirme, ayrıştırma veya ölçekleme işlemlerinin birlikte yapıldığı görülmektedir.

Özelliklerin kendi içinde sınıflandırılması ve özelliklerinin belirlenmesi noktasında ölçeklendirme işlemi ile başlandıktan sonra özelliklerin ağ üzerindeki etkilerinin araştırılması işlemidir. Temel bileşen analizi (PCA) yöntemi burada en yaygın olarak kullanılan yöntemlerin başında gelmektedir. PCA yöntemi aynı zamanda verilerin normal dağılıma sahip olduğu varsayımı altında iyi sonuçlar veren bir boyut küçültme ve öznitelik çıkarma yöntemidir. Birleştirme işlemi, özellik sırasında benzer özellikleri içeren girişlerin birleştirilmesi işlemine dayanır. Derin öğrenme gibi algoritmalar bu veri dönüştürme işlemlerini kendi içlerinde kapalı bir kutu olarak gerçekleştirir ancak bu işlemin makine öğrenmesi algoritmalarında uzman bir kişi tarafından yapılması gerekir.

Bu tez çalışmasında, özellik mühendisliğindeki temel işlemleri kullanıldı. Veritabanı sahibi tarafından toplanan ham veriseti üzerinde veri normalizasyon gibi işlemler gerçekleştirilmiştir. Daha sonra, veri seti eğitim ve test için bölünmüştür. Bu sayede olası istatistiksel hataların önüne geçilmiş olur.

Veri seçimi adımında, sırasıyla rastgele belirlenen geleneksel doğrulama yöntemi ve k-katlama (3-kat, 5-kat, 10-kat) çapraz doğrulama yöntemi kullanılarak veri seçim işlemi gerçekleştirilmiştir. Bu çalışmada, ağın eğitim süreci için geleneksel değer yöntemi için 691 örnek eğitim ve 296 adet örnek ise test işlemi için kullanılmıştır. 3-kat çapraz doğrulama yöntemi için veriseti 3 parçaya (a, b ve c)

bölünür ve her adımda 658 (a+b, 329+329) adet örnek eğitim, 329 (c) adet örnek test işlemleri için kullanılmıştır. 5-kat çapraz doğrulama yöntemi için veriseti 5 parçaya (a,b,c,d ve e) bölünür ve her adımda 788 (a+b+c+d, 197+197+197+197) örnek eğitim işlemleri için 197 (e) örnek de test işlemleri için kullanılmıştır. Son olarak 10-kat çapraz doğrulama yöntemi ile veri seti 10 parçaya bölünmüş olup her adımda farklı kombinasyonlarda 888 adet hasta verisi eğitim ve 99 hasta verisi de test işlemi için kullanılmıştır.

İkinci adımda, kullanılan veri seti üzerinde tüm veri ön işleme adımları yazılım arayüzü kullanılarak otomatik olarak gerçekleştirilmiştir. Ayrıca veri dönüştürme işlemi olarak ölçeklendirme işlemi, veri seti oluşturulurken üzerinde gerçekleştirilmiş ve bu adımda herhangi bir işlem yapılmamıştır. Ancak üçüncü aşamada öznitelik belirlemede PCA yöntemi kullanılmıştır. Bu amaçla veri setindeki 18 girdi verisi için temel bileşen analizi sırasında 1, 2 ve 3 öznitelik boyut küçültme kombinasyonları denenmiş ve PCA işlemleri sonucunda en iyi boyut küçültme performansları sonuçlarda verilmiştir. Bu işlemler sırasında en doğru ağ parametrelerini bulmak için literatürdeki genel bilgiler kullanılmıştır. Örneğin, gizli katman nöronlarının sayısı, giriş katmanı sayısının yaklaşık 2 katı olarak belirlenmiş ve toplam hatanın yerel minimumlarını bulmak için ortalama kare hatası yöntemi kullanılmıştır. Çalışmada çok fazla ağ kullanıldığı için her bir yöntemin parametreleri ve karışıklık matrisi verilmemiştir. Ancak, tüm farklı makine öğrenme algoritmaları için aynı özellik mühendisliği yöntemleri uygulanarak tüm ağların performansı paralel olarak incelenmiştir.

3. SEPSİS TANIMI

Sepsis tıp biliminde, enfeksiyonun neden olduğu ve insan vücudundan gelen düzensiz bir yanıtın bir sonucu olarak biyokimyasal, fizyolojik ve patolojik anormalliklerle ilişkili bir sendrom olan en zor sendromlardan biri olmaya devam etmektedir. Sepsis özellikle tanınmazsa ve hemen tedavi edilmezse, enfeksiyondan kaynaklanan ölümlerin birincil nedenidir. Tanınması acil dikkat gerektirir [1]. Basit bir ifadeyle sepsis vücudun bir enfeksiyona tepkisi esnasında kendi doku ve organlarını yaraladığında ortaya çıkan yaşamı tehdit eden bir durumdur.

Sepsis oluşumunda birçok enfeksiyon türü rol oynayabilmektedir. Bunlardan en yaygınları idrar yolu enfeksiyonları, solunum yolu enfeksiyonları, zatürre, karın içi enfeksiyonlar, yara ve yanıklar, menenjit ve deri enfeksiyonlarıdır. 1 yaş altı çocuklar, hamile kadınlar, kronik hastalığı bulunan kişiler ve yaşlı bireyler sepsis açısından risk grubunda yer almaktadır.

1992 ve 2016 yılları arasında, sepsis tanımları, Sepsis-1 ve Sepsis-2, sepsisi tanımlamak için sistemik inflamatuar yanıt sendromu (SIRS) kriterlerine dayanıyordu. Sistemik inflamatuar yanıt sendromu (SIRS), vücudun enfeksiyon ve enfeksiyon dışı nedenlere (yanık, multitravma, pankreatit, hemorajik şok, iskemi, otoimmün hastalık) bağlı olarak vermiş olduğu seyri klinik olarak değişken olabilen inflamatuar yanıt olarak tanımlanmıştır [33]. 2008 ve 2012 yılında yayınlanan SSC (surviving sepsis campaign-sepsiste sağkalım kampanyası) kılavuzlarında da SIRS temelli tanımların yetersiz olduğu bilinse de SIRS - sepsis - ağır sepsis - septik şok temelli tanımlamalar yapılmıştır [35, 36]. 2016 yılında, tanı kriterlerinin öngörücü geçerliliğini en üst düzeye çıkarmak için, SIRS üzerindeki vurguyu sıralı [Sepsisle İlişkili] organ yetmezliği değerlendirmesinin kullanılması lehine terk eden yeni tanım Sepsis-3 önerildi. En güncel olanı ise Sepsis-3 kılavuzu (2016); [1], sepsisten şiddetli sepsis yoluyla septik şoka doğru sepsis ilerlemesinin geleneksel üçlü sınıflandırmasını ortadan kaldırır ve bunun yerine, ölüm olasılığındaki artışlara bağlı iki aşamalı bir tanımlama sistemini kullanır. Yeni tanıma göre, "sepsis" terimi, enfeksiyona düzensiz bir konukçu tepkisinin neden olduğu "yaşamı tehdit eden bir organ işlev bozukluğu" olarak tanımlanmaktadır. Sepsis-3 ayrıca, sepsis sonuçlarının en kötü olduğu yoğun bakım ünitesi (YBÜ) hastalarını taramak için hızlı SOFA (qSOFA) skorunu da getirmiştir [1]. Altta yatan dolaşım ve hücresel / metabolik anormalliklerin mortaliteyi

önemli ölçüde artıracak kadar derin olduğu sepsis alt kümesine Septik Şok denir[19]. Bu kılavuza göre değişen terminoloji Tablo 3'de verilmiştir.

Tablo 3.1. 2016 kılavuzuna göre yenilenmiş sepsis ve septik şok tanımı

	Önceki Tanımlar	2016 Tanımı
Sepsis	SIRS + Enfeksiyon Şüphesi	Şüpheli veya dökümente edilmiş Enfeksiyon + SOFA skoru ≥ 2 puan (organ disfonksiyonu olarak kabul edilir)
Ağır Sepsis	Sepsis + Organ Disfonksiyon Bulguları + Sıvı tedavisine yanıt veren hipotansiyon	Tanımı kaldırıldı
Septik Şok	Sepsis + inotrop gernektiren hipotansiyon (hipovoleminin olmadığı durumlarda)	Sepsis + OAB ≥ 65mmHg olması için inotrop ihtiyacı + Laktat > 2mmol/L

Yoğun bakım ünitelerinde (YBÜ) şiddetli sepsis ve septik şok hala önde gelen ölüm nedenleridir ve tedavi sonuçları için zamanında teşhis çok önemlidir. Sepsis yönetiminin zamana oldukça duyarlı olduğu düşünüldüğünde, sepsisin erken tahmini, mortalitenin (ölüm oranı) önlenmesinde çok önemlidir. Elektronik tıbbi kayıtların (EMR) ilerlemesi, tıpta yapay zekanın (AI) gelişimini kolaylaştırabilecek büyük miktarda klinik veriyi saklama imkanı sunar.

3.1. Ağır Sepsis

Sepsisle ilişkili hipotansiyon (sistolik kan basıncı < 90 mmHg, ortalama arteryel basınç < 60 mmHg veya sistolik kan basıncında bazal değerlerden 40 mmHg daha fazla düşme) gelişmesi veya sepsis ile ilişkili olan organ hipoperfüzyonu gelişmesi durumunda bu klinik ağır sepsis olarak tanımlanmaktaydı [34]. Sepsisle ilişkili organ hipoperfüzyon bulguları Tablo 3.2'de gösterildi.

Tablo 3.2. Sepsisle ilgili organ hipoperfüzyon bulguları

Değişen bilinç durumu
Laktat yüksekliği > 2 mmol/L
Yeterli sıvı resüsitasyonuna rağmen akut olgüri (0.5 ml/kg/saat)
Arteryel hipoksemi (PaO_2/FiO_2<300)
Kreatinin artışı >0.5 mg/dL veya Kreatinin yüksekliği >2.0 mg/dL
Koagülasyon bozukluğu (INR>1.5 ya da aPTZ>60sn)
Trombositopeni (<100.000/μL)
Hiperbilirubinemi (plazma total bilirubin >4mg/dL)

3.2. Septik Şok

Sepsisle ilişkili olarak ve klinik olarak hipovoleminin olmadığı durumlarda gelişen yani sıvı resüsitasyonuna rağmen mevcut olan hipotansiyon (sistolik kan basıncı < 90 mmHg, ortalama arteryel basınç < 60 mmHg veya sistolik kan basıncında bazal değerlerden 40 mmHg daha fazla düşme gelişmesi) ile birlikte organ perfüzyon bulgularının (Tablo 3.2. Organ Perfüzyon Bulguları) gelişmesi durumu septik şok olarak olarak tanımlanmaktadır [33]. Bu tanımlama daha sonra 2016 yılında yayınlanan kılavuzla birlikte değişmiş olmakla birlikte SOFA skorunun 2 puan ve üstü olması durumunda OAB ≥ 65 mmHg üstünde tutmak için inotrop ihtiyacı gelişmesi ve laktak yüksekliği (> 2 mmol/L) olması durumunda septik şoktan bahsedilmektedir. Bu tanıma göre septik şok tanısı alan hastaların hastane mortalitesi % 40'dan yüksek olduğu saptanmıştır [1].

3.3. Multiple Organ Disfonksiyonu Sendromu

Multiple organ disfonksiyonu sendromu akut hastalığı olan hastalarda görülen hızlı ve progresif seyreden ve vücut hemostazının sağlayabilmek için dışardan müdahele gerektiren organ yetmezliğidir [37]. Hem enfeksiyon kaynaklı (sepsis veya septik şok) hem de enfeksiyon dışı sebeplerle (pankratite bağlı SIRS) primer ve sekonder olarak ikiye ayrılmaktadır.

MODS'da organ disfonksiyonunu belirlemek için kabul edilmiş evrensel bir tanı kriteri yoktur. Ancak aşağıda belirtilen organ parametreleri MODS'a gidişi belirlemede ve yoğun bakım mortalitesini tahmin etmede belirli skorlama sistemlerinde (SOFA ve LODS gibi) sıkça kullanılmaktadır. Bu organ parametreleri Tablo 3.3'de verilmiştir [38, 9].

Tablo 3.3. MODS organ disfonksiyon parametreleri

Parametreler	
Solunum	PaO_2/FiO_2
Hematoloji	Trombosit Sayısı
Karaciğer	Serum Bilirubin
Böbrek	Serum Kreatinin
Beyin	GKS
Kardiyovasküler	Hipotansiyon ve İnotrop ihtiyacı

3.4. Sepsis Risk Faktörleri ve Klinik Bulguları

CDC verilerine göre sepsis için yüksek risk grubunda; 65 yaş ve üstü yetişkinler, diyabet, kanser, akciğer, karaciğer ve böbrek hastalığı gibi kronik hastalıkları olan kişiler, bağışıklık sistemi zayıf olanlar, daha önce sepsis geçiren kişiler ve bir yaşından küçük çocuklar yer almaktadır [40].

3.4.1. Sepsis için Risk Faktörleri

- Çok genç (1 yaş altı), yaşlı (75 yaş üstü) veya çok zayıf insanlar,
- Son 6 hafta içinde geçirilmiş travma, ameliyat veya invaziv prosedür,
- Hastalık (diyabet) veya ilaçlar (uzun süreli steroid, kemoterapi veya immünosupresif tedavi alan kişiler) nedeniyle bozulmuş bağışıklık,
- Kalıcı sonda ve kateter varlığı, intravenöz uyuşturucu ilaç kullanımı, cilt bütünlüğünün bozulması (kesik, yanık veya cilt enfeksiyonu),
- Nötropenik hastalar,
- Hamile olan, son 6 hafta içinde doğum yapmış ve düşük yapmış kadınlar için ek risk faktörleri:
    - ✓ Gebelik diyabeti, diyabet ve diğer komorbiditeler,
    - ✓ Sezaryen, forsepsle doğum ve invaziv prosedürler,
    - ✓ Uzun süreli membran rüptürü,
    - ✓ A grubu streptokok enfeksiyonu olan biriyle yakın temas,
    - ✓ Vajinal kanamanın devam etmesi ve kötü kokulu vajinal akıntı varlığıdır [41].

3.4.2. Sepsis klinik bulguları

Sepsis semptomları, mikrobiyal toksinlerden ve enflamatuvar hücrelerin ürettiği sitokinlerden kaynaklanmaktadır [42]. Enfeksiyon şüphesi olan hastada aşağıdaki bulguların bazılarının olması sepsis şüphesi lehine düşünülmektedir [36].

Genel Parametreler

- Ateş (> 38.3^{o}C)
- Hipotermi (düşük vücut sıcaklığı) (<36^{o}C)

- Kalp atım hızı >90
- Takipne (hızlı solunum) (>30)
- Mental durum değişikliği
- Belirgin ödem veya pozitif sıvı dengesi (>20 mL/kg/gün)
- Diabet yokluğunda hiperglisemi (Glukozun yükselmesi) (>140 mg/dL veya 7.7 mmol/L)

İnflamatuar Parametreler

- Lökositoz (Lökositin artması) >12.000/µL
- Lökopeni (Lökositin azalması) <4.000µL
- İmmatür form beyaz küre sayısının Lökosit: N, Bant>%10
- Plazma C-reaktif protein (CRP) normal değerden >2 SD
- Plazma prokalsitonin normal değerden >2 SD

Hemodinamik Parametreler

- Hipotansiyon (SAB<90 mmHg, OAB<70 mmHg, SAB'de 40 mmHg'den fazla azalma)

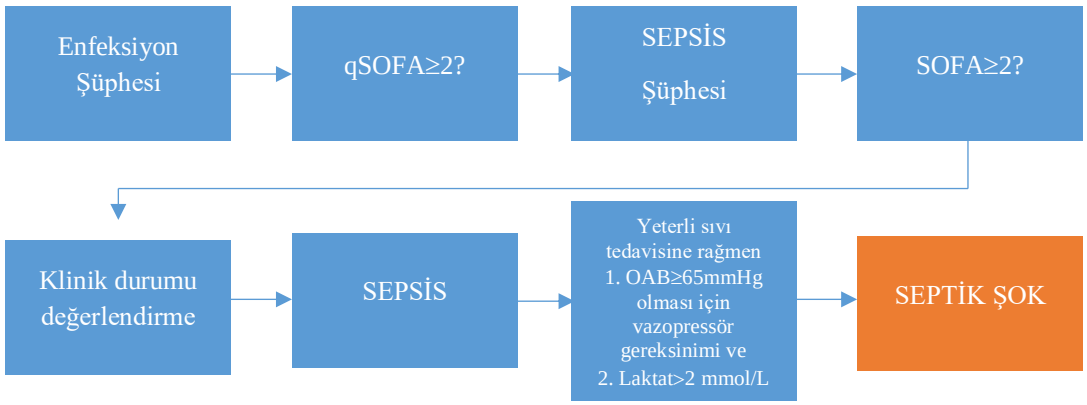
Organ Disfonksiyonu Parametreleri

- Hipoksemi (Oksijenin azalması) (PaO2/FiO2<300)
- Akut oligüri idrar çıkışı < 0.5 ml/kg/saat-en az 2 saat- yeterli sıvı tedavisine rağmen)
- Kreatinin artışı (>%0.5mg veya 44.2 µmol/L)
- Koagulasyon bozukluğu (INR>1.5, aPTT>60 sn)
- Trombositopeni (Trombositin azalması) (<100.000)
- İleus (Barsak sesi yok)
- Hiperbilurubinemi (T.Bil>%4mg veya 70 µmol/L)

Doku Perfüzyon Parametreleri

- Laktat yüksekliği (laktat>1mmol/L)
- Azalmış kapiller geridönüşü veya benekli görünüm
- SKB (sistolik kan basıncı)
- OAB (ortalama kan basıncı)
- INR (international normalized ratio)
- aPTT (aktive parsiyel tromboblastin zamanı)
- SD (standart deviasyon)
- PaO2 (Parsiyel arteryel oksijen basıncı) [22].

Şekil 3.1'de görüldüğü üzere sepsis tanısı için oluşturulan algoritmada enfeksiyon şüphesi olan hastaların qSOFA parametresinin 2 üstü olması beklenmektedir. Sepsis şüphesi olan hastaların ise SOFA parametresinin 2 üstü olması gerekmektedir. Akabinde yapılan klinik değerlendirme sonucu yeterli sıvı tedavisi yapılmasına rağmen OAB parametresinin 65 mmHg üstü olması için vazopressör gereksinimi ve Laktak değerinin 2 mmol/L'den yüksek olması beklenmektedir. Bu durumda hastaya Septik Şok tanısı konulabilmektedir.



Şekil 3.1. Sepsis tanı algoritması (Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Critical care medicine 2017;45(3):486-552'den uyarlanmıştır.)

3.5. SOFA

Sepsis, YBÜ hastalarında ana ölüm nedeni olan eş zamanlı çoklu organ fonksiyon bozukluğu sendromu ile ilişkilidir. SOFA, organ fonksiyon bozukluğu, organ yetmezliği, mortalite ve morbidite değerlendirilmesinde kullanılan bir skorlama sistemidir. Skor sıfır ile 24 arasında bir değer alır [44]. SOFA skorunda 2 veya daha fazla bir değişiklik sepsis sendromunun tanımlayıcı bir özelliğidir. Hastanın enfeksiyon başlangıcından önce önceden var olan (akut veya kronik) organ disfonksiyonuna sahip olduğu bilinmediği sürece, temel SOFA skorunun sıfır olduğu varsayılmalıdır. SOFA skoru 2 veya daha fazla olan hastalar, varsayılan enfeksiyonu olan genel bir hastane popülasyonunda yaklaşık % 10'luk bir genel mortalite riskine sahiptir. Organ fonksiyon bozukluğu, enfeksiyona bağlı olarak toplam SOFA skorunda ≥2 puanlık ani bir değişiklik olarak tanımlanabilir [19].

Seymour CW ve arkadaşlarının yapmış olduğu bir çalışmada ise ΔSOFA skoru 2 ve üzerinde olan hastalarda ölüm riskinin 2-25 kat arttığı bulunmuştur, bu sebeple ΔSOFA ≥2 olan skorlar organ fonksiyon bozukluğunu temsil etmektedir [1,19].

Tablo 3.4. SOFA kriteri

Puan	0	1	2	3	4
Sistem					
Solunum PaO_2/FiO_2, mmHg (kPa)	≥400 (53.3)	<400 (53.3)	<300(40)	<200 (26.77)	<100 (13.3)
Koagülasyon					
Platelet sayısı ×103µL	≥150	<150	<100	<50	<20
Karaciğer					
Bilirubin, mg/dL	<1.2	1.2-1.9	2.0-5.9	6.0-11.9	>12.0
Kardiyovasküler	OAB≥70mm Hg	OAB<70 mm Hg	Dopamin <5 veya dobutamin (tüm dozlar)	Dopamin 5.1-15 veya epinefrin ≤0.1	Dopamin >15 veya epinefrin >0.1 veya norepinefrin >0.1
Santral S.S.					
Glaskow Koma Skalası	15	13-14	10-12	6-9	<6
Renal					
Kreatin, mg/dL	<1.2	1.2-1.9	2.0-3.4	3.5-4.9	>5.0
İdrar Çıkışı				<500	<200

Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315(8):801-10'den uyarlanmıştır.

3.6. Quick SOFA

Hızlı Sepsis İlişkili Organ Yetmezliği Değerlendirme Skoru olarak türkçeleştirilen qSOFA skoru SCM/ESICM tarafından yayınlanan Sepsis-3 kılavuzunda acil servis ve hastane öncesi hizmetlerde yoğun bakım dışı kullanımı için önerilmiştir. Çünkü SOFA skorunun yatakbaşı hızlı uygulanamaması ve çok fazla laboratuvar testi gerektirmesi nedeniyle bir erken uyarı sistemi olarak kullanılması uygun değildir. Ayrıca bu kılavuzun temelini oluşturan büyük kohort çalışmasında SOFA skoru sadece yoğun bakım hastalarında hesaplanmıştır [1]. Hastane dışı, acil servis veya genel hastane koşullarında şüpheli enfeksiyonu olan erişkin hastalarda yatak başı sepsis ilişkili prognozun belirlenmesi açısından kullanılabilmektedir. Uzun süreli YBÜ'de kalması veya hastanede ölmesi muhtemel enfeksiyon şüphesi olan hastalar, yatak başında hemen qSOFA ile tanımlanabilir, yani zihinsel durumda değişiklik, sistolik kan basıncı ≥100 mm Hg veya solunum hızı ≥22 / min.

Tablo 3.5. qSOFA kriterleri

Kriterler	**Puan**
Solunum Sayısı ≥22/dk	1
Sistolik Kan Basıncı ≤100 mmHg	1
Değişen Bilinç Durumu (GKS<15)	1

Yakın bir tarihe kadar sistemik inflamatuar yanıt sendromu SIRS (systemic inflammatory response syndrome: sistemik inflamatuar yanıt sendromu) kriterleri, sepsis tanısı koyabilmek için kullanılan kabul görmüş bir skorlama sistemi idi [34,45]. Sistemik inflamasyona sebep olabilen enfeksiyon dışı sebepler de (pankreatit, yanık, hemorajik şok, iskemi, çoklu travma) SIRS'ı pozitifleştirdiğinden ve basit enfeksiyonların da sepsis olarak kabul edilmesine sebep olduğundan, bu kriterlerin kullanımı sorgulanmaya başlanmıştır ve yerine SOFA (Sequential-Sepsis Related Organ Failure Assessement: Tekrarlanan-Sepsis ilişkili Organ Yetmezliği Değerlendirmesi), hızlı SOFA (qSOFA), NEWS (National Early Warning Score: Ulusal Erken Uyarı Skoru) gibi skorlama sistemleri kullanılarak, sepsis hastalarını erken dönemde tanımak ve tedavi etmek amaçlanmıştır [43].

4. DENEYSEL ÇALIŞMALAR

Bu bölüm, önerilen makine öğrenimi algoritmalarının erken algılama tahmin performansını gösterir. Erken tahmin performansının değerlendirilmesi, sepsisli hastaların veri setinde gerçekleştirilir.

Bu bölümde yapay sinir ağı sonuçlarını sınıflandırma doğruluğuna göre değerlendirerek Sepsis hastalarının erken teşhisini tahmin ediyoruz. Bu amaçla 19 öznitelik (18 Girdi - 1 Çıktı) ve 977 örnek içeren bir veri seti kullanılmıştır. Önerilen algoritmaların performansını değerlendirmek için geleneksel doğrulama ve k-kat çapraz doğrulama yaklaşımı kullanılmıştır. Sepsis tanı verileri, çalışmanın başarısını göstermek için birçok farklı makine öğrenme tekniğiyle test edildi. Bu amaçla fine tree, medium tree, coarse tree, linear discriminant, quadratic discriminant, logistic regression, gaussian naive bayes, kernel naive bayes, linear SVM, quadratic SVM, cubic SVM, fine gaussian SVM, medium gaussian SVM, coarse gaussian SVM, fine KNN, medium KNN, coarse KNN, cosine KNN, cubic KNN, weighted KNN, boosted trees, bagged trees, subspace discriminant, subspace KNN, rusboosted trees, narrow neural network, medium neural network, wide neural network bilayer neural network, trialayer neural network, SVM kernel yöntemleri test edilmiş ve tüm öznitelikleri (PCA'lı veya PCA'sız) kullanan farklı sınıflandırıcılar ile elde edilen sınıflandırma performans sonuçları Tablo 4.1'de gösterilmiştir.

Tablo 4.1. Karşılaştırma için makine öğrenmesi teknikleri (%)

Machine Learning Models	TV		3-Fold CV		5-Fold CV		10-Fold CV	
	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable
1.1 Fine Tree	96.1	96.5	82.2	71.5	81.7	75.9	81.5	75.6
1.2 Medium Tree	87.5	82.3	81.5	70.7	82.2	74.2	80.7	73.0
1.3 Coarse Tree	78.4	73.9	75.1	72.6	76.8	70.8	77.8	71.8
1.4 Linear Discriminant	75.4	74.0	73.1	73.4	74.4	72.7	74.3	73.2
1.5 Quadratic Discriminant	76.4	73.7	75.6	72.5	75.4	72.4	74.9	72.2
1.6 Logistic Regression	74.7	73.5	73.0	72.5	74.2	72.3	74.3	72.6
1.7 Gaussian Naive Bayes	70.2	69.6	68.9	68.9	69.2	69.9	69.4	68.5
1.8 Kernel Naive Bayes	85.5	80.5	77.7	74.3	80.2	74.7	79.4	75.3
1.9 Linear SVM	74.0	73.2	73.0	73.4	73.0	72.4	73.2	71.7
1.10 Quadratic SVM	91.3	87.6	83.2	82.4	85.2	82.7	86.1	82.9
1.11 Cubic SVM	**100**	98.5	**87.8**	**85.7**	89.0	86.0	88.4	86.7

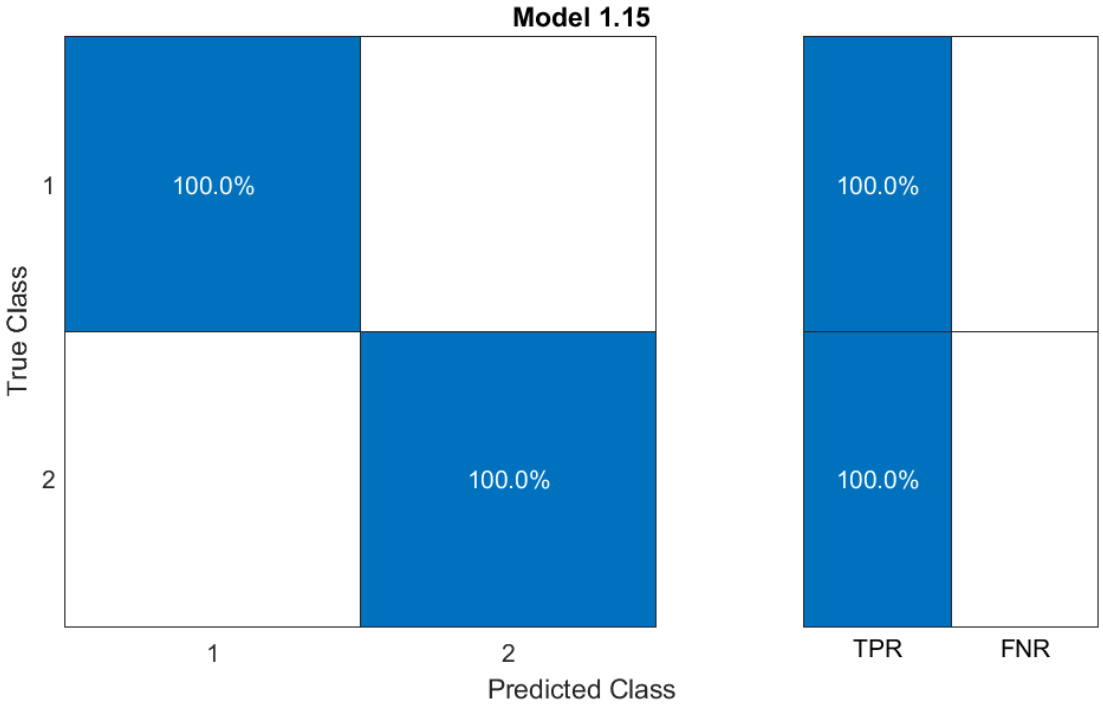
Tablo 4.1. Devamı

Machine Learning Models	TV		3-Fold CV		5-Fold CV		10-Fold CV	
	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable
1.12 Fine Gaussian SVM	**100**	99.6	75.8	77.3	78.9	80.4	79.8	81.6
1.13 Medium Gaussian SVM	90.8	87.7	85.3	82.4	86.6	83.5	86.4	83.8
1.14 Coarse Gaussian SVM	73.4	73.5	71.9	70.4	72.8	72.5	72.7	73.0
1.15 Fine KNN	**100**	**100**	87.4	83.2	**93**	85.9	90.5	85.7
1.16 Medium KNN	89.5	88.2	83.3	83.4	84.7	84	84.3	84.4
1.17 Coarse KNN	77.4	76.3	73.7	72.3	74.5	74.1	76.2	75.0
1.18 Cosine KNN	89.7	87.4	83.5	83.2	84	83.4	85.3	83.2
1.19 Cubic KNN	88.8	86.8	80.6	82.5	81.8	82.8	82.9	83.8
1.20 Weighted KNN	**100**	**100**	85.9	84.4	87.8	86.7	89.3	87.4
1.21 Boosted Trees	**100**	94.2	86.7	79.2	88.2	78.8	88.0	78.9
1.22 Bagged Trees	**100**	99.9	87.1	81.5	89.7	81.6	**92.7**	82.9
1.23 Subspace Discriminant	73.4	73.0	73.0	72.5	73.1	72.5	73.2	72.0
1.24 Subspace KNN	**100**	**100**	88.6	86	90.5	**87.5**	90.8	**87.6**
1.25 gtnvn gv	89.7	88.0	85.7	78.2	83.9	75.5	86.0	78.4
1.26 Narrow Neural Network	**100**	92.5	80.7	78.8	85.1	79.3	83.4	79.9
1.27 Medium Neural Network	**100**	**100**	84.1	83.2	86.6	83.5	85.6	81.9
1.28 Wide Neural Network	**100**	**100**	85.5	83.1	87.3	84.5	88.9	84.1
1.29 Bilayerde Neural Network	**100**	**100**	83.1	79.5	83.0	77.9	84.0	77.9
1.30 Trialayerde Neural Network	**100**	**100**	81.7	77.6	82.5	79.4	84.6	81.1
1.31 SVM Kernel	93.9	90.8	83.0	79.4	84.4	82.2	85.6	81.2
1.32 Logistic Regression Kernel	87.5	86.4	78.6	77.2	80.5	77.9	81.0	81.0

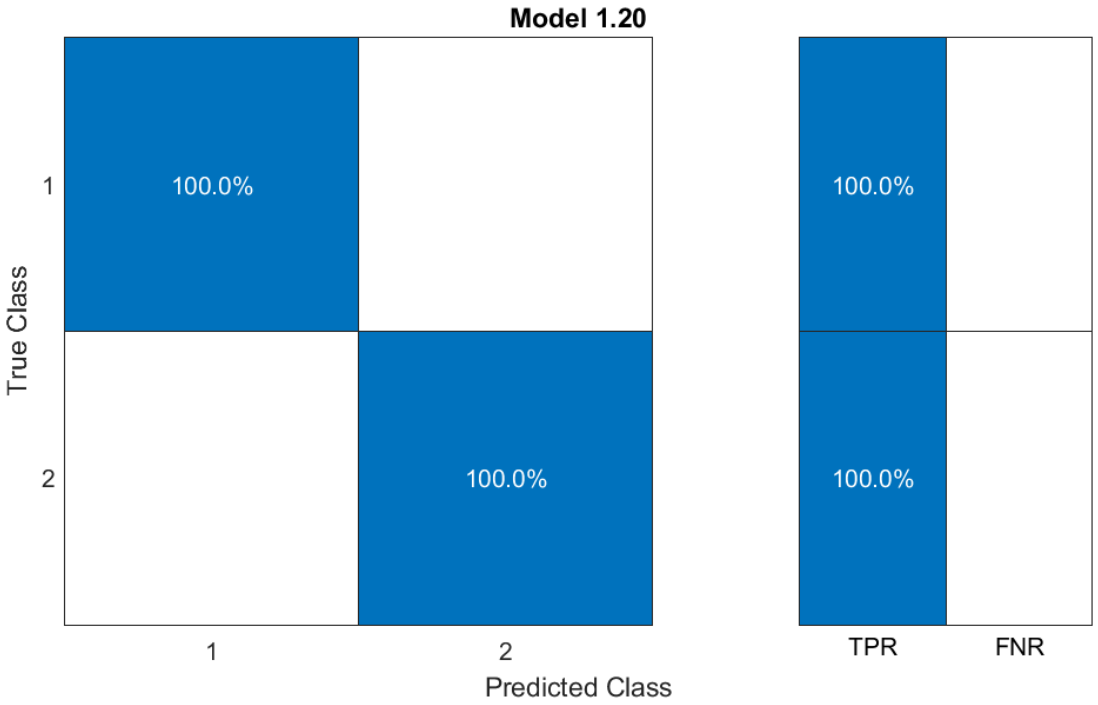
Aynı veri seti için bu çalışmada önerilen yöntemlerin doğruluk performansı literatürdeki diğer makine öğrenme yöntemlerine göre daha başarılı olmuş ve sepsisli hastaların erken teşhisini başarıyla belirleyebilecek etkili bir karar destek sistemi tasarlanmıştır. Veri setindeki dengesizliğe rağmen, risk tahmininde en iyi performansı gösteren cubic SVM, fine gaussian algorithms SVM, fine KNN, boosted trees, bagged trees, subspace KNN, narrow neural network, medium neural network, wide neural network bilayer neural network ve trialayer neural network yöntemleri yüksek bir performans ile (%100) test işlemlerini gerçekleştirmiştir. Ayrıca, PCA gibi farklı parametre ve optimizasyon yöntemleri ile düşük performanslı algoritmaların başarımı artırılmaya çalışılmıştır. Bu çalışmalar sonucunda elde edilen en iyi sonuçlar tabloda verilmiştir. Bu sonuçlar incelediğinde, önemli bir performans etkisi iyileşmesinin elde edilemediği sonucuna ulaşılmıştır. Bunun nedeni ise örneklem setinde belirlenen girişlerin zaten yüksek bir eğitim ve test sonucu elde etmesi ve probleme uygun olarak

belirlenmiş olması, ön işlemlerin bu performans artışını yeterli seviyede gerçekleştirmesidir.

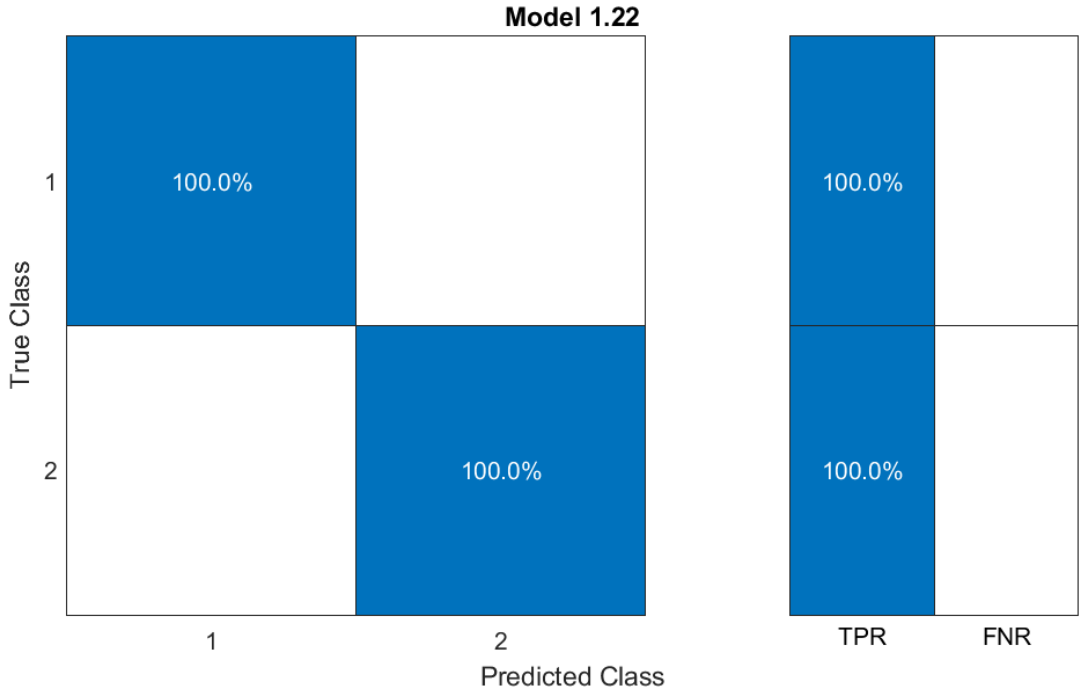
Uygulanan algoritmalar arasında en başarılı algoritmaya ait karışıklık matrisleri aşağıda verilmiştir. Makine öğrenimi, kullanılan sınıflandırma modelinin performansını yorumlamak için bir karışıklık matrisi kullanır. Şekil 5-16, Hedef özniteliğin tahminlerini ve gerçek değerlerini karşılaştıran karışıklık matrisi verilmiştir.



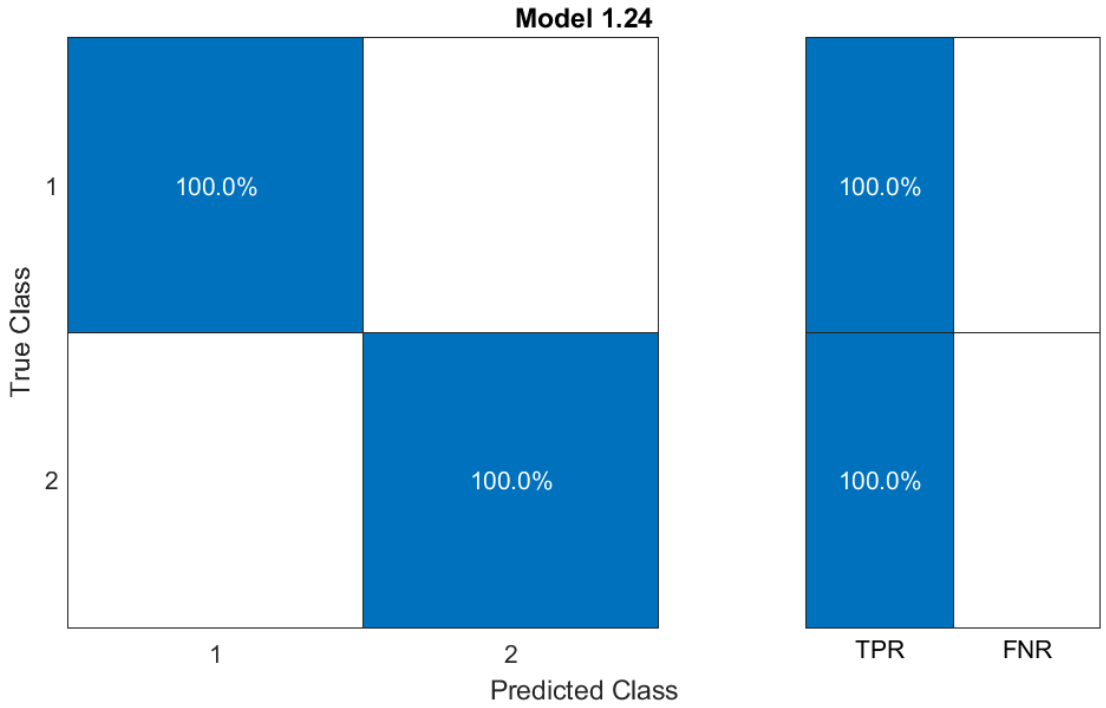
Şekil 4.1. En başarılı fine KNN algoritması için karışıklık matrisi.



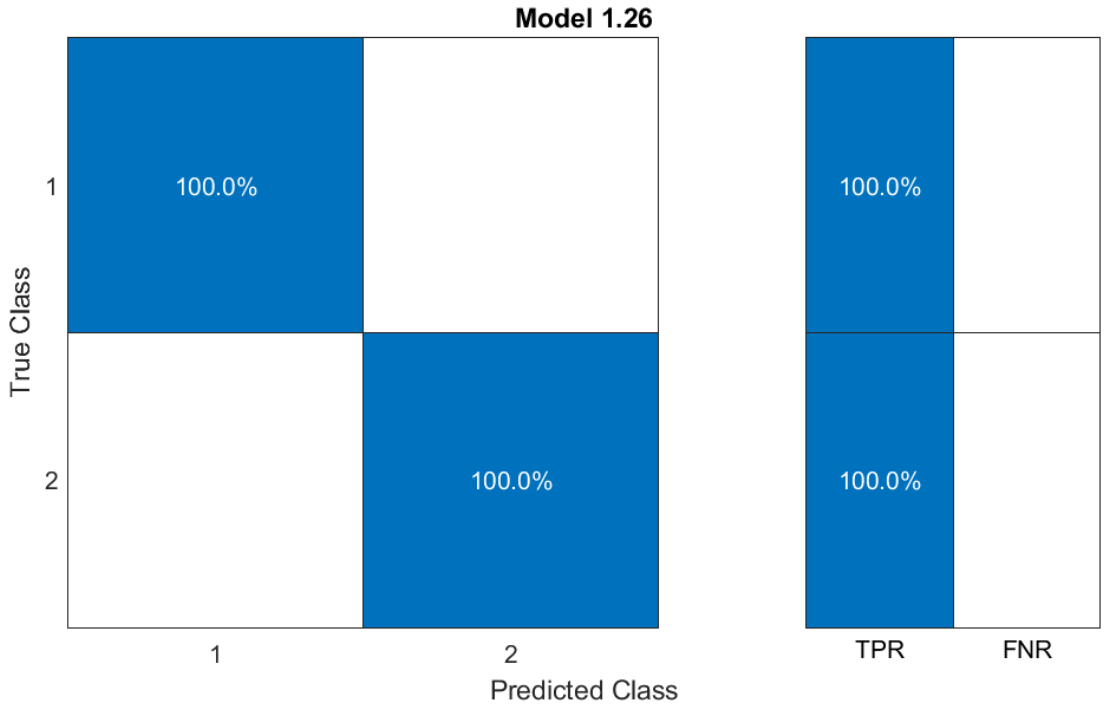
Şekil 4.1. En başarılı weighted KNN algoritması için karışıklık matrisi



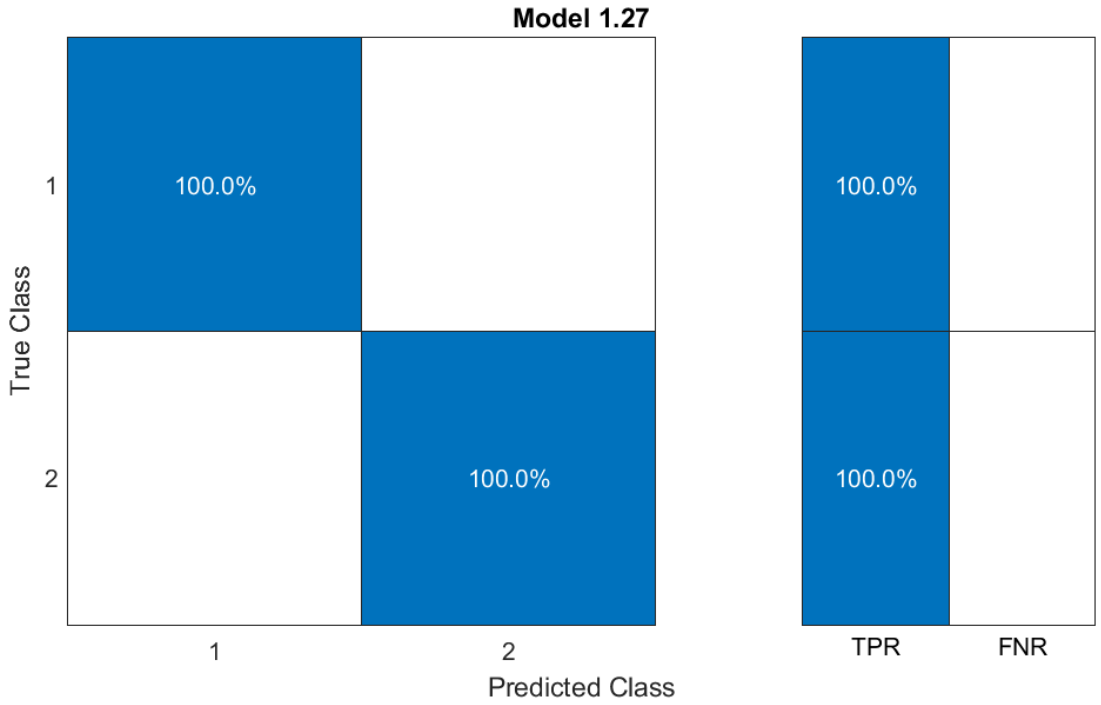
Şekil 4.2. En başarılı bagged trees algoritması için karışıklık matrisi.



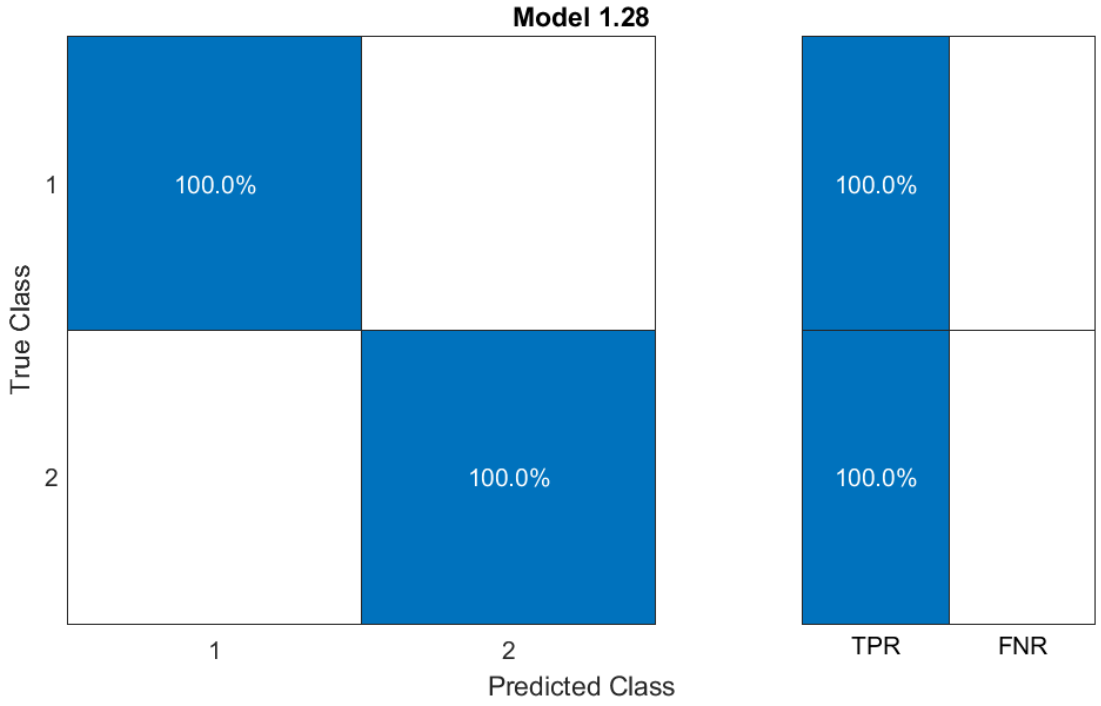
Şekil 4.3. En başarılı subspace KNN algoritması için karışıklık matrisi.



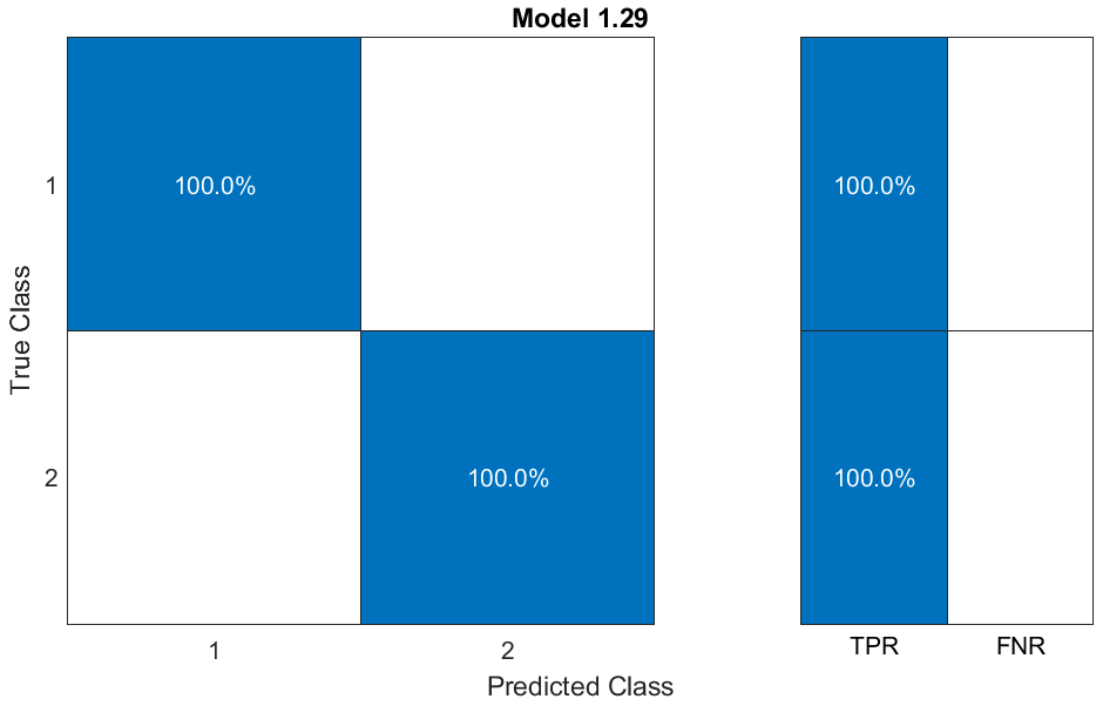
Şekil 4.4. En başarılı narrow neural network algoritması için karışıklık matrisi.



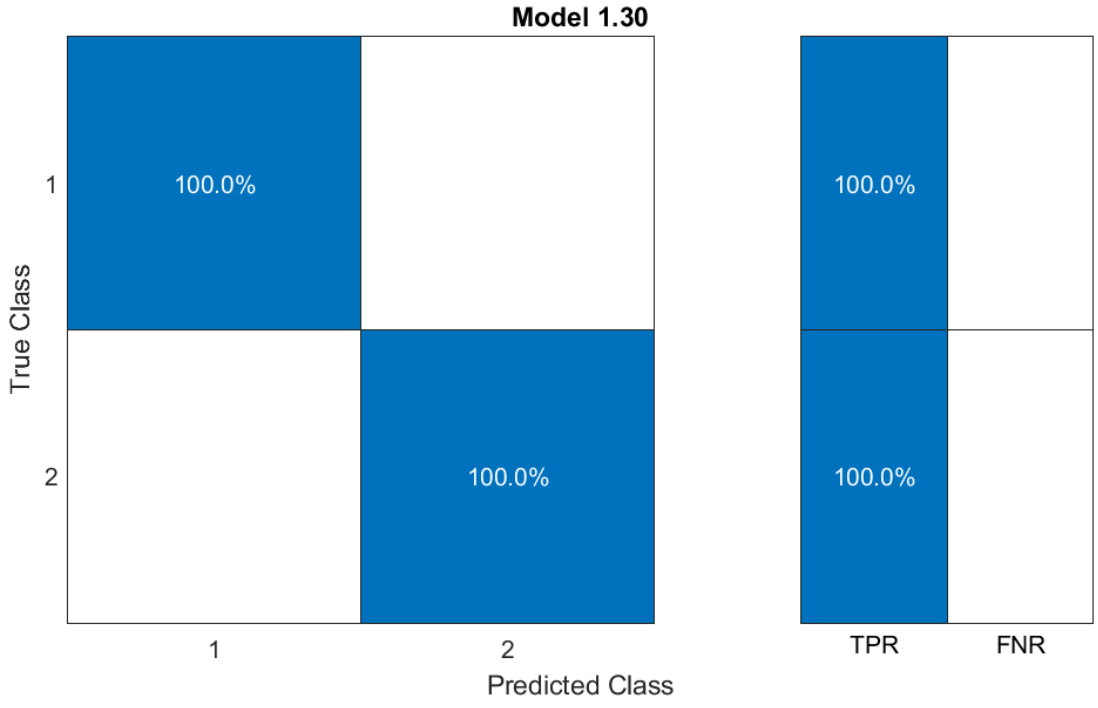
Şekil 4.5. En başarılı medium neural network algoritması için karışıklık matrisi.



Şekil 4.6. En başarılı wide neural network algoritması için karışıklık matrisi.



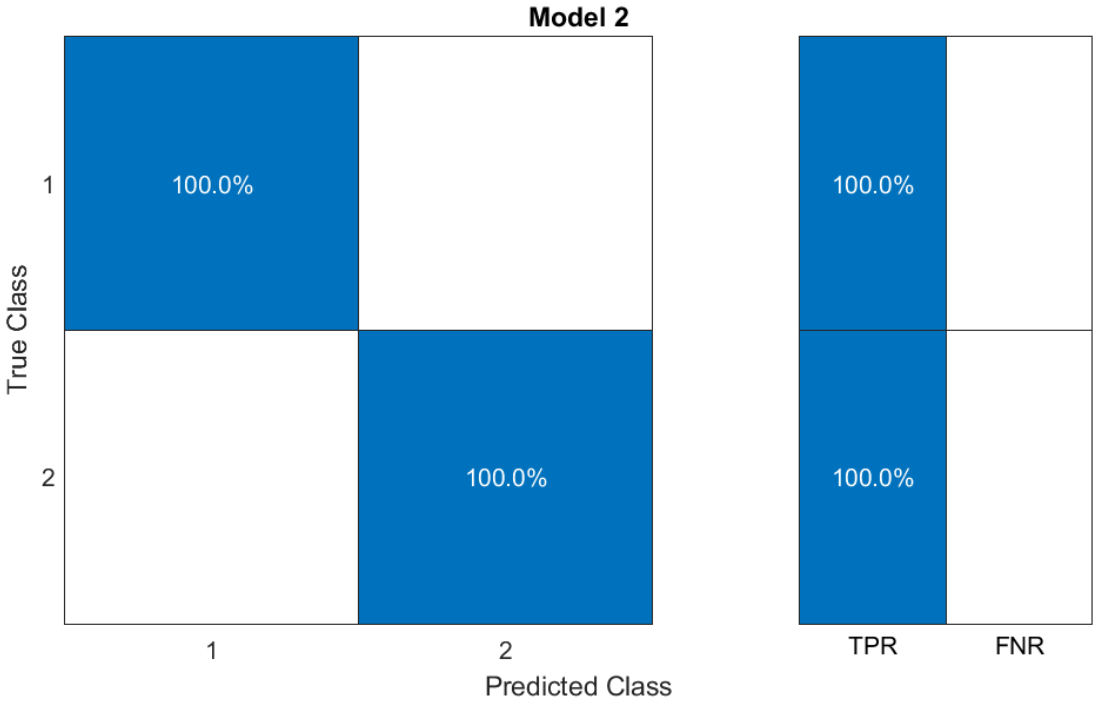
Şekil 4.7. En başarılı bilayered neural network algoritması için karışıklık matrisi.



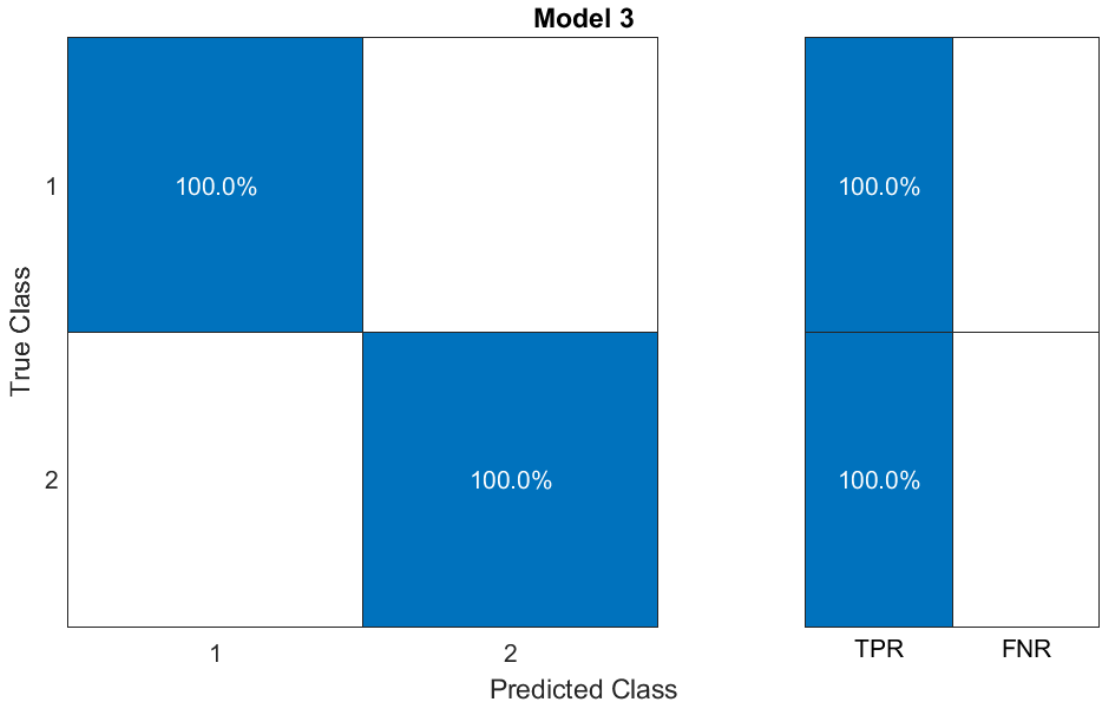
Şekil 4.8. En başarılı trilayered neural network algoritması için karışıklık matrisi.

Arıca yapılan çalışmada 3 farklı algoritma üzerinde iyileştirmeler yapılmış çıkan sonuçlar %100 doğruluğa yükseltilmiştir. Bunlar; cubic SVM algoritması için kutu

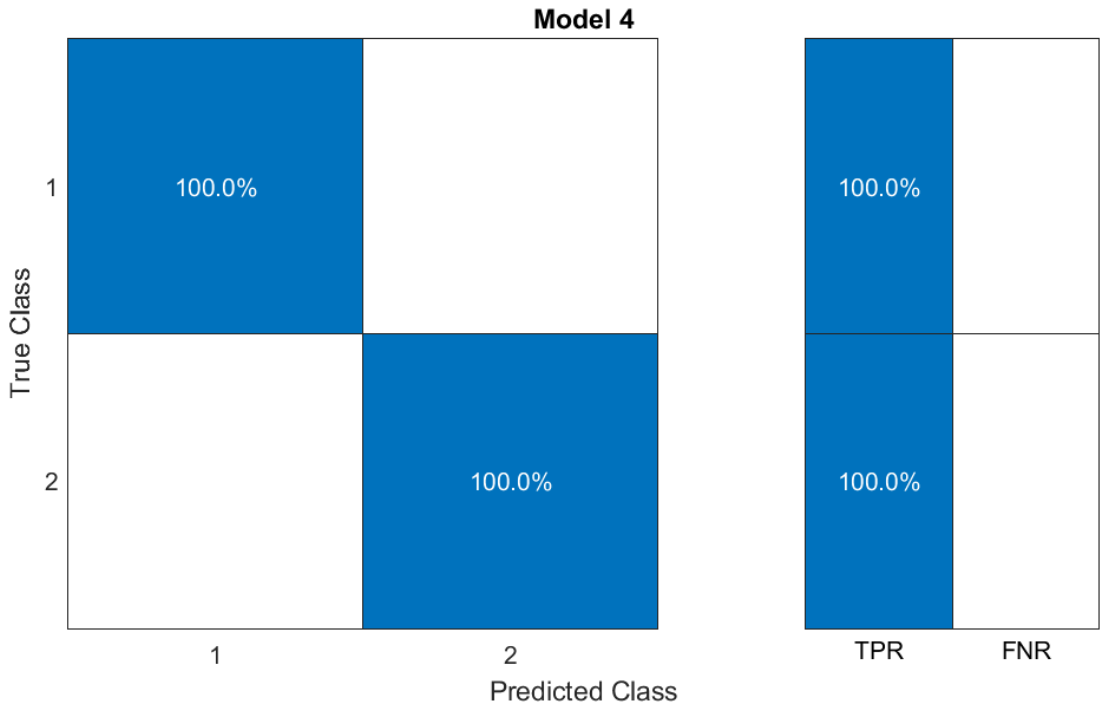
kontras seviyesi parametresi (box constraint level) 4 ve manuel çekirdek tarama parametresi (kernel scan manuel) ayarlanarak yapılan denemede sonuç %100 bulunmuştur. Fine gaussian SVM algoritması için box constraint level 2 yapılarak algoritma çalıştırıldığında sonuç aşağıda görüldüğü gibi %100 olmuştur. Son olarak ise boosted trees algoritmasında max number of splits 50, number of learners 50 ve learning rate 0,11 yapılarak advanced modda çalıştırıldığında %100 sonucuna ulaşılmıştır. İlgili algoritmaların karışıklık matrisleri aşağıda gösterilmiştir.



Şekil 4.9. Revize edilen başarılı cubic SVM algoritması için karışıklık matrisi.

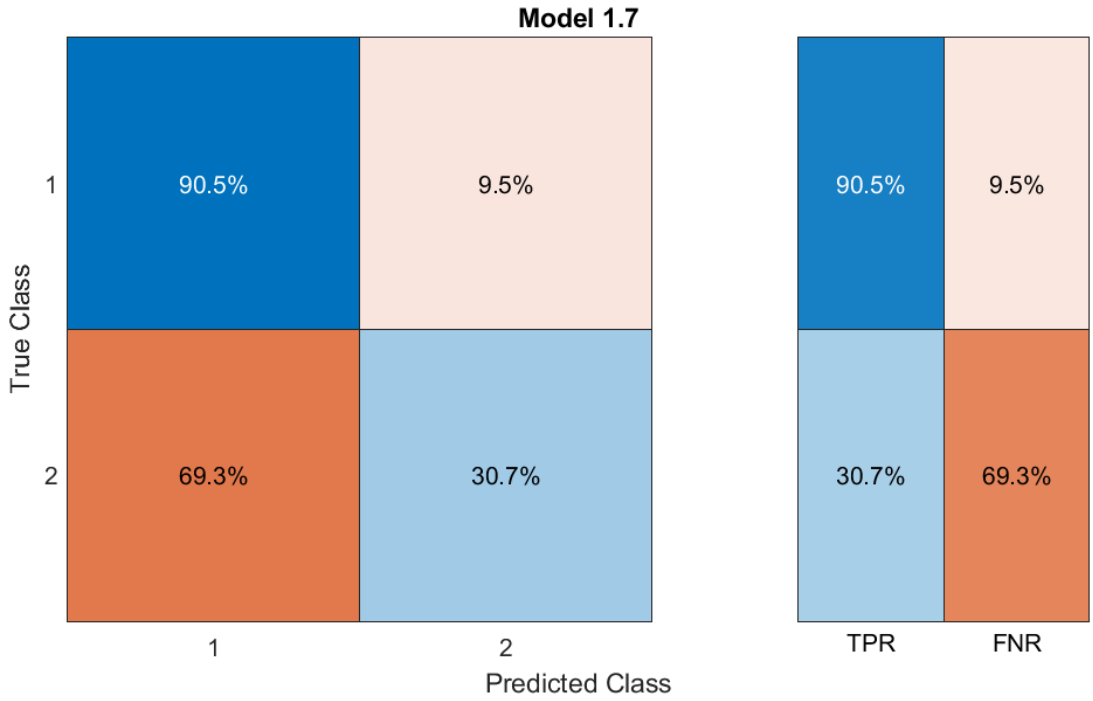


Şekil 4.10. Revize edilen başarılı fine gaussian SVM algoritması için karışıklık matrisi.



Şekil 4.11. Revize edilen başarılı boosted trees algoritması için karışıklık matrisi.

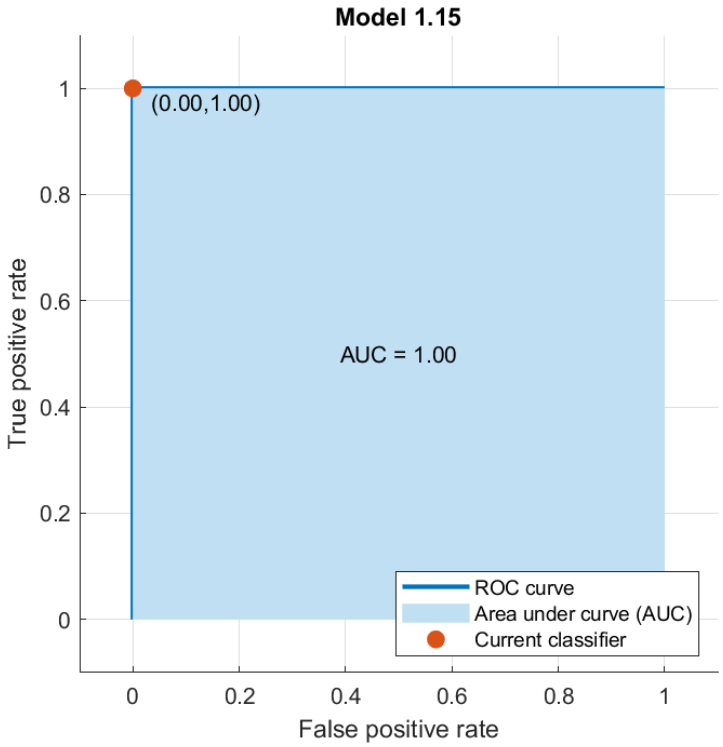
Uygulanan algoritmalar arasında en başarısız algoritmaya ait karışıklık matrisleri aşağıda verilmiştir. Şekil 4.12'de, Hedef özniteliğin tahminlerini ve gerçek değerlerini karşılaştıran karışıklık matrisi verilmiştir.



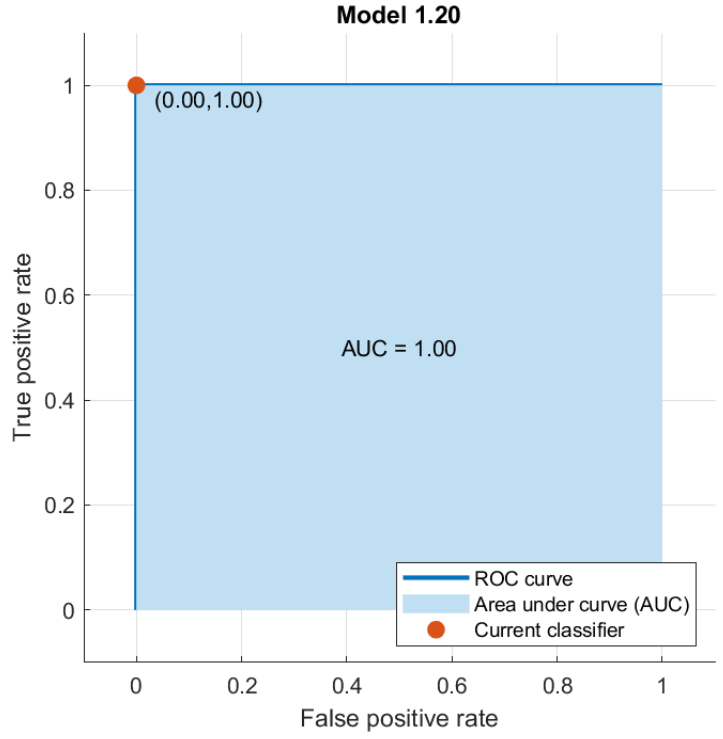
Şekil 4.12. En kötü başarılı gaussian naive bayes algoritması için karışıklık matrisi.

Şekil 4.1 ile Şekil 4.11 arasındaki algoritmalarda görüldüğü gibi, karışıklık matrisi, 977 örnekten 977 sepsis hastasının gerçekten yaşayacağı ve 0'ın yaşamayacağı belirlendi. Şekil 4.12'de görüldüğü gibi, gaussian naive bayes algoritması karışıklık matrisi, 977 örneğin 678'inin sepsis hastasında fiilen hayatta kalacağı ve 299'unun hayatta kalmayacağı belirlendi.

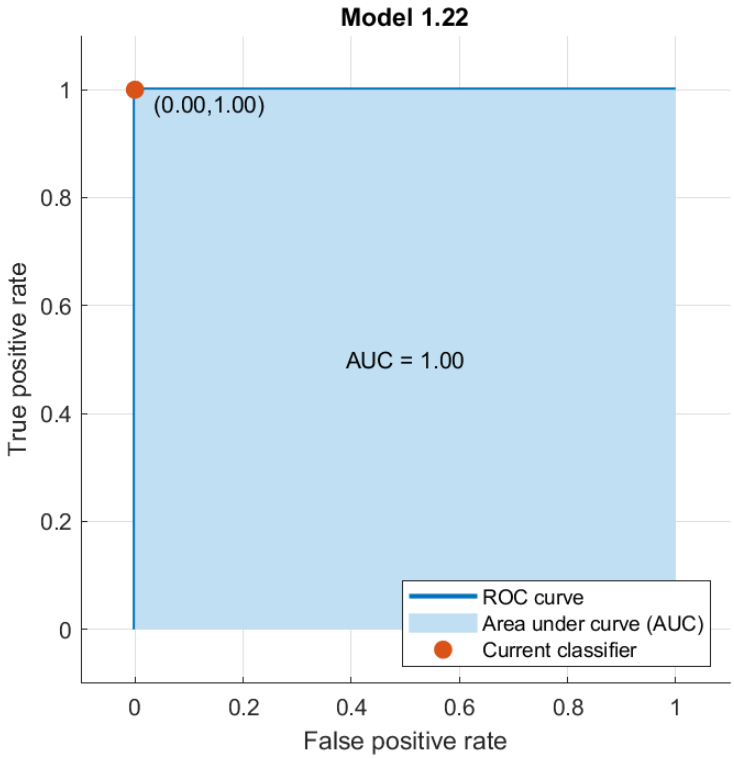
Alıcı işletim karakteristiği (ROC) eğrisi, bir ikili sınıflandırıcı sisteminin ayrım eşiği değiştikçe sınıflandırma yeteneğini gösteren bir grafiktir. Algoritma performanslarının belirlenmesi sonucunda elde edilen ROC eğrisi en iyi performans gösteren 9 algoritma ayrıca revize edilen 3 algoritma ile toplam 12 algoritmanın başarılı ROC eğrileri sırasıyla paylaşılmıştır.



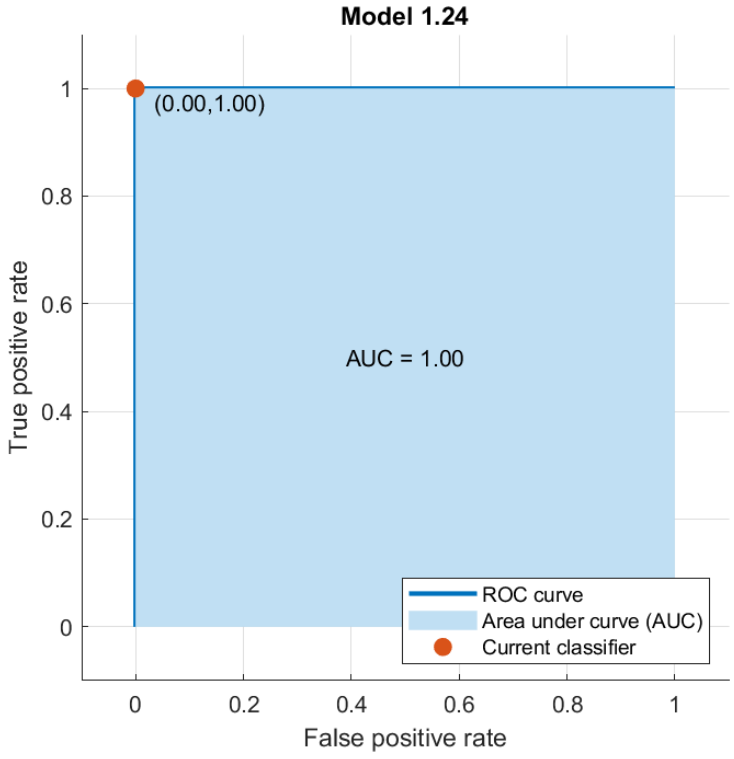
Şekil 4.13. En iyi performans gösteren fine KNN algoritması için ROC eğrisi



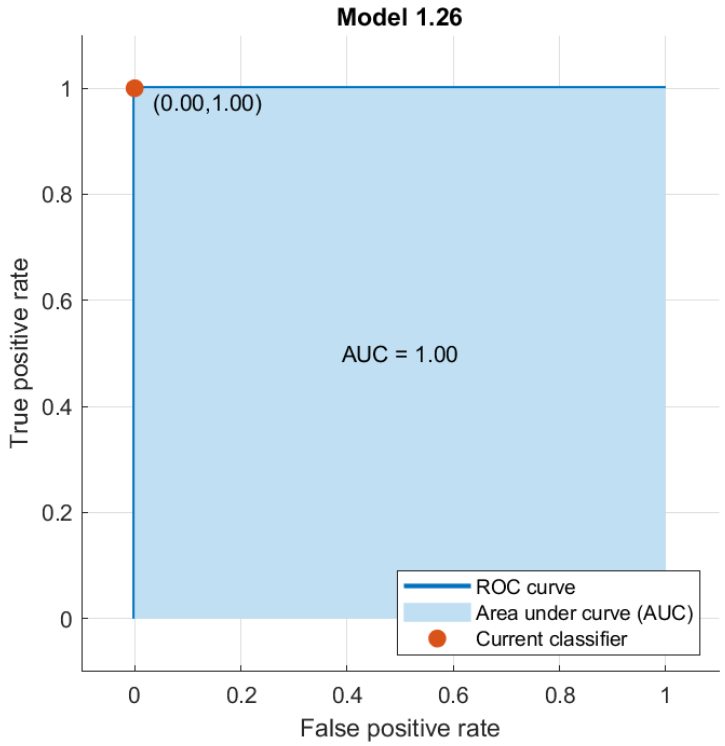
Şekil 4.14. En iyi performans gösteren weighted KNN algoritması için ROC eğrisi



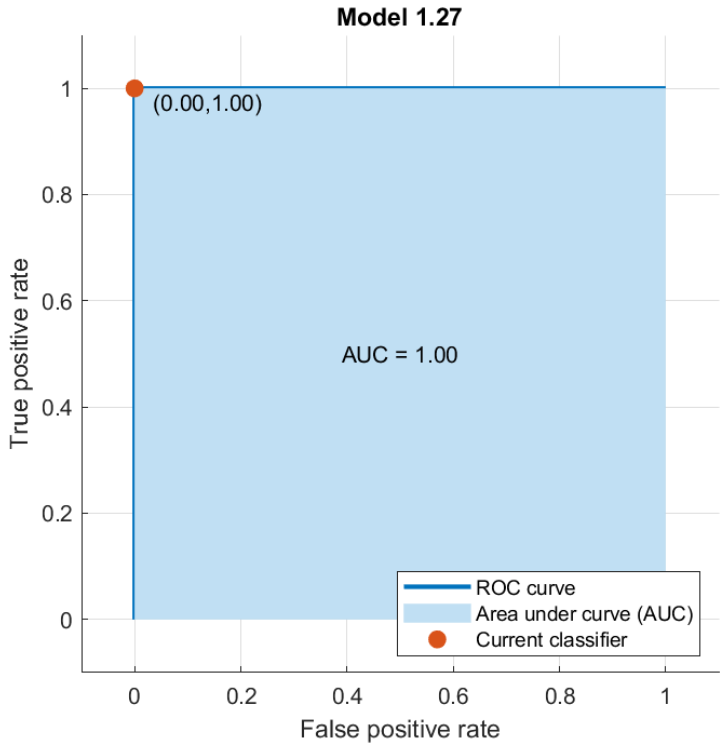
Şekil 4.15. En iyi performans gösteren bagged trees algoritması için ROC eğrisi



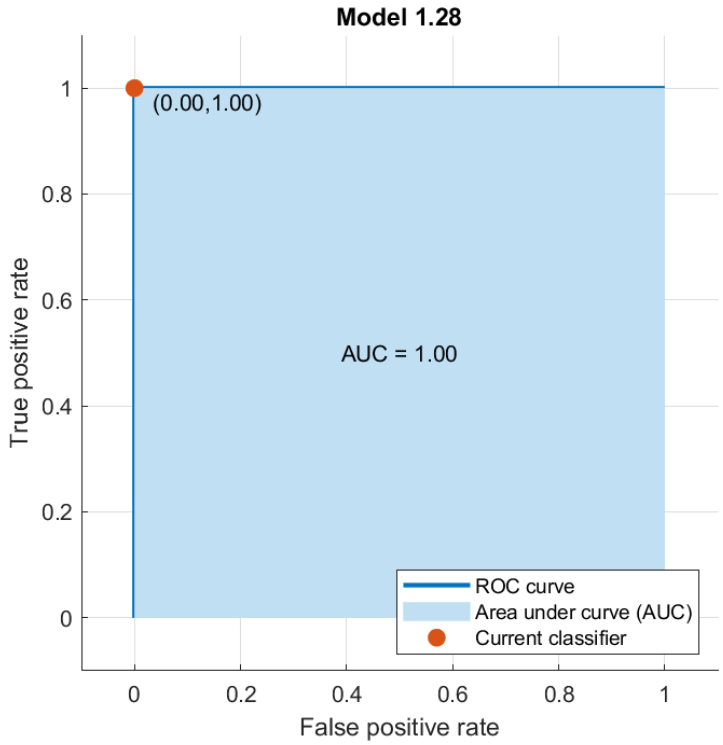
Şekil 4.16. En iyi performans gösteren subspace KNN algoritması için ROC eğrisi



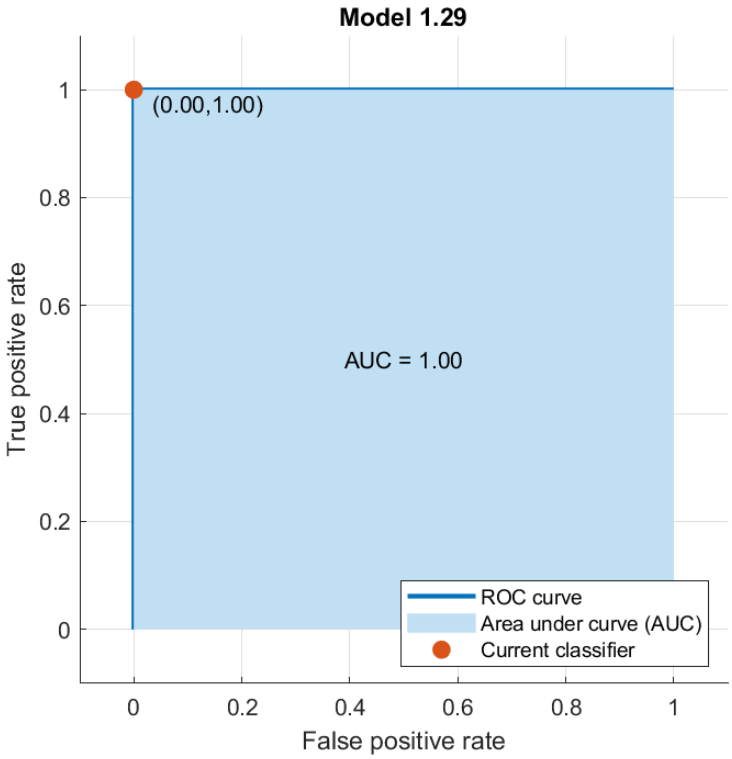
Şekil 4.17. En iyi performans gösteren narrow neural network algoritması için ROC eğrisi



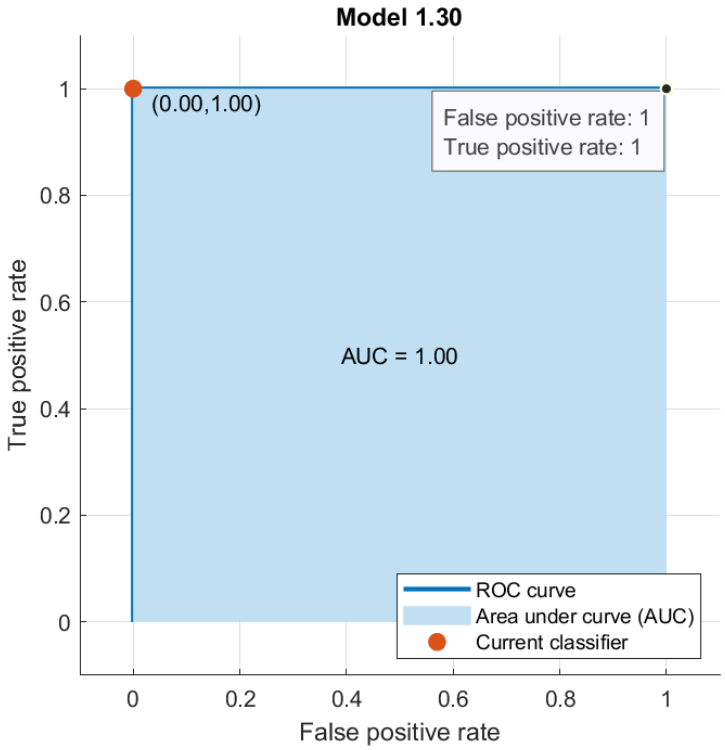
Şekil 4.18. En iyi performans gösteren medium neural network algoritması için ROC eğrisi.



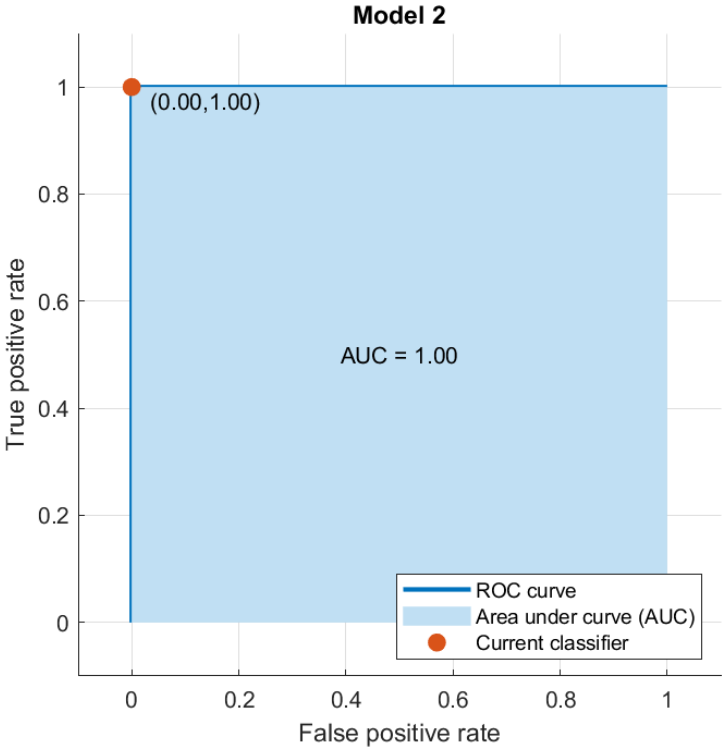
Şekil 4.19. En iyi performans gösteren wide neural network algoritması için ROC eğrisi.



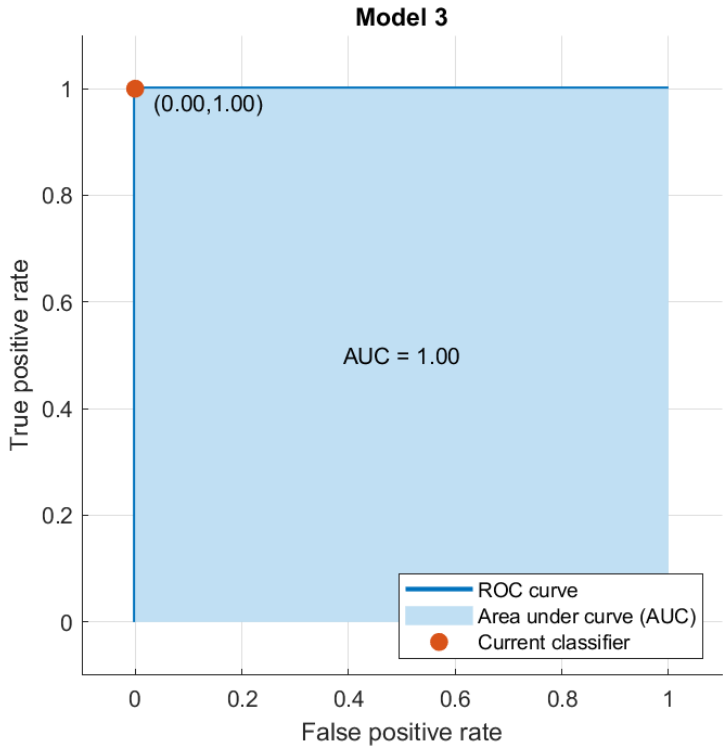
Şekil 4.20. En iyi performans gösteren bilayered neural network algoritması için ROC eğrisi.



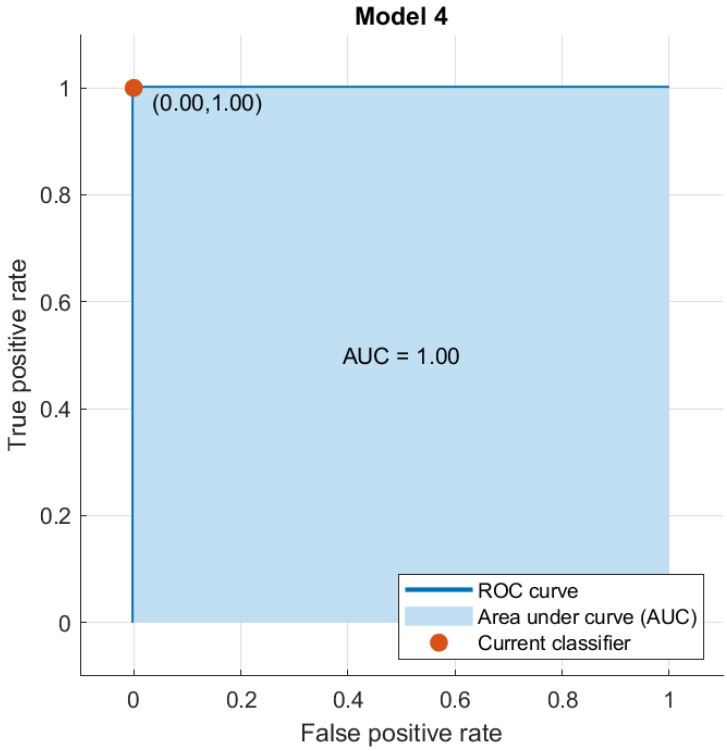
Şekil 4.21. En iyi performans gösteren trilayered neural network algoritması için ROC eğrisi.



Şekil 4.22. Revize edilen başarılı cubic SVM algoritması için ROC eğrisi.

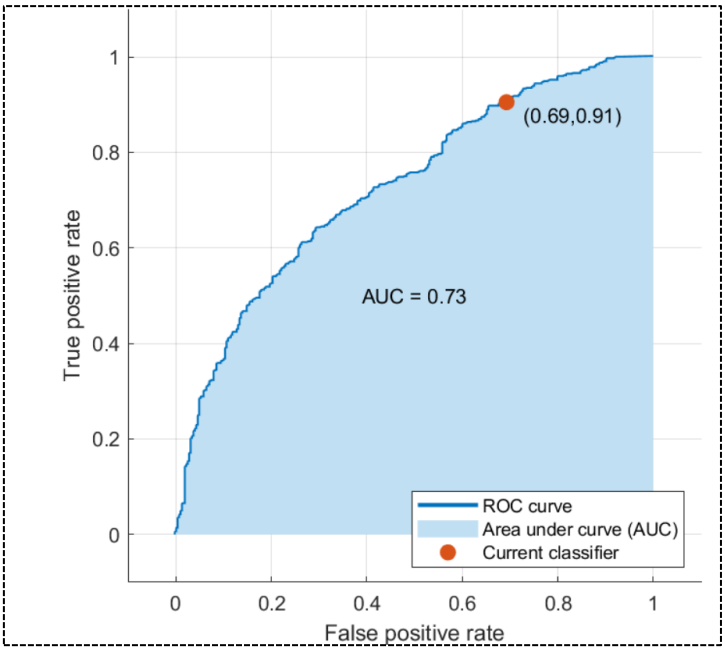


Şekil 4.23. Revize edilen başarılı fine gaussian SVM algoritması için ROC eğrisi.



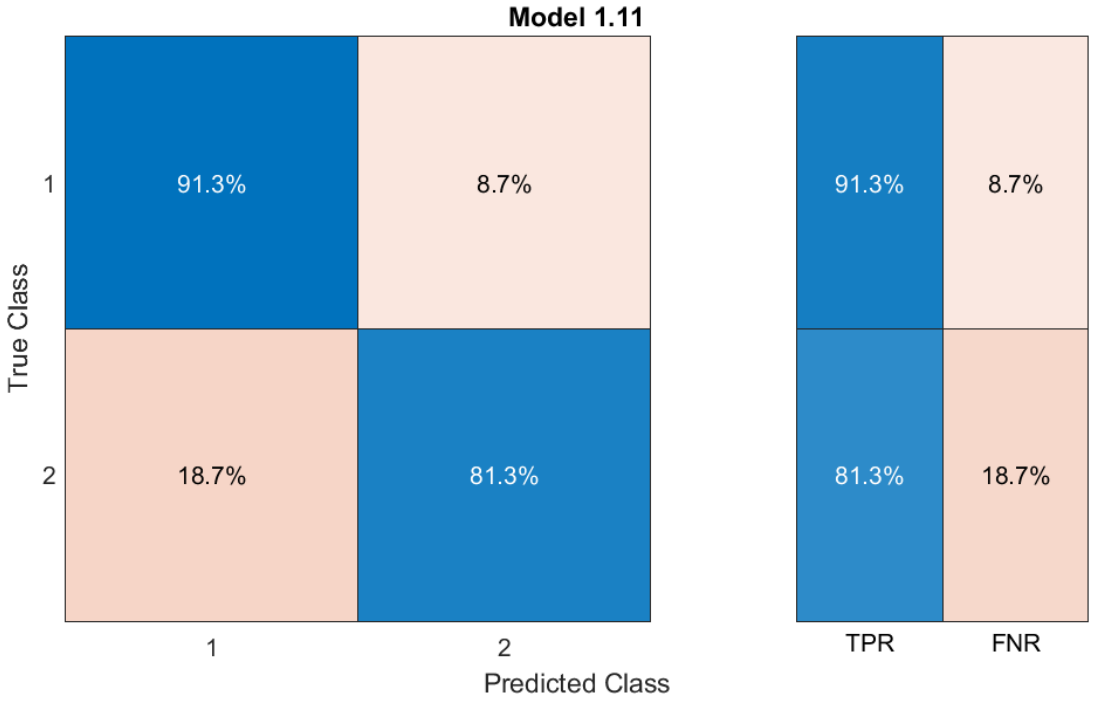
Şekil 4.24. Revize edilen başarılı boosted trees algoritması için ROC eğrisi.

Algoritma performanslarının belirlenmesi sonucunda elde edilen ROC eğrilerinde en kötü performans gösteren gaussian naive bayes algoritması için ROC eğrisi Şekil 4.25'de gösterilmiştir.

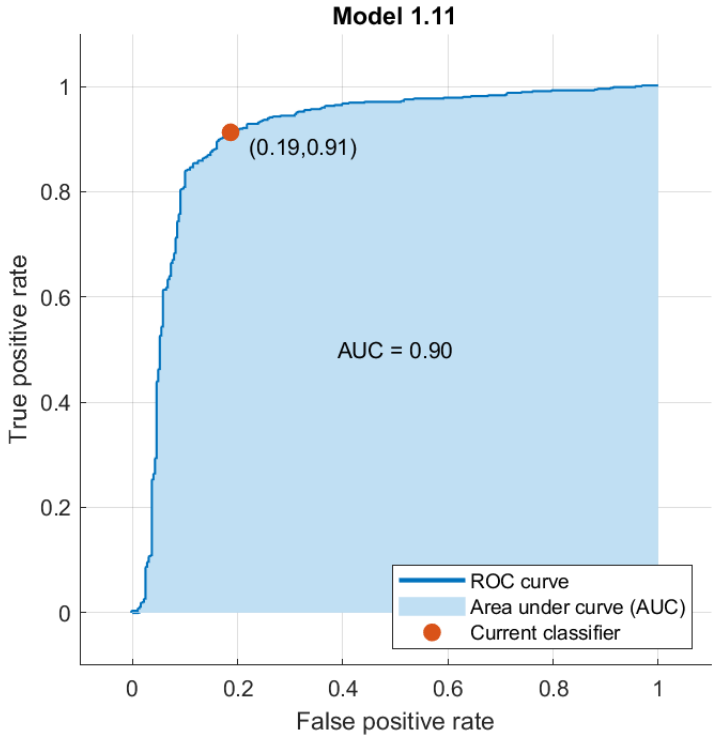


Şekil 4.25. En kötü performans gösteren gauss naive bayes algoritması için ROC eğrisi.

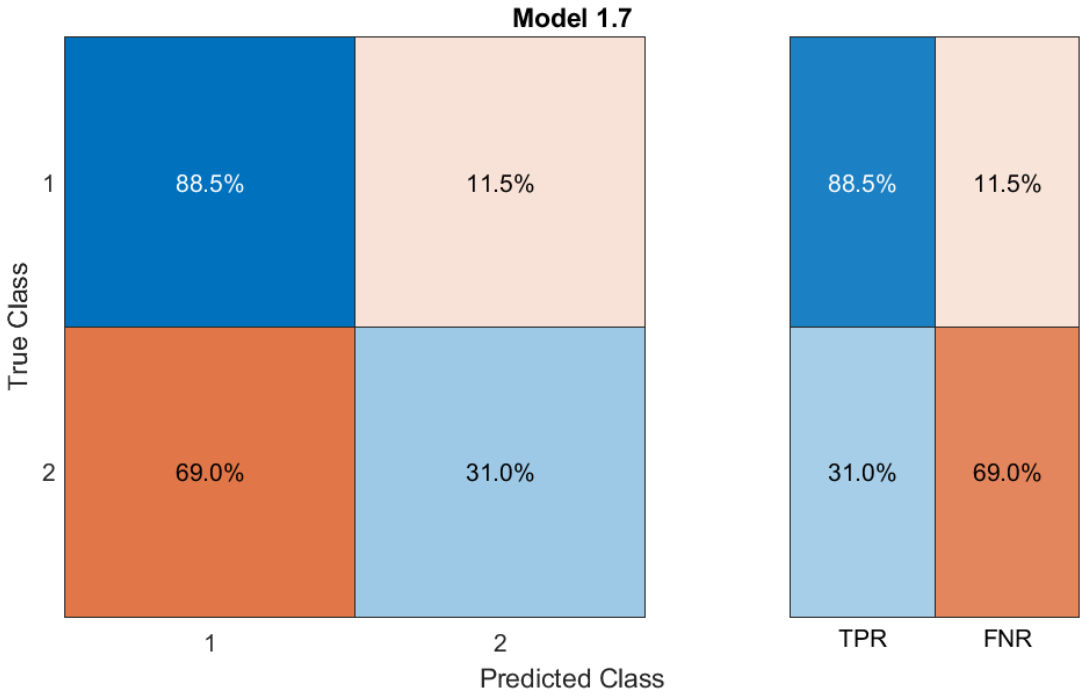
Veri seçimi adımında, sırasıyla rastgele belirlenen geleneksel doğrulama yöntemi yanında k-katlama (3-k Fold Cross Validation, 5-CV,10-CV) çapraz doğrulama yöntemi de kullanılarak veri seçim işlemi gerçekleştirilmiştir. K-katlama yöntemlerine ait en başarılı ve en başarısız Karışıklık Matrisi ile ROC eğrileri aşağıda gösterilmiştir.



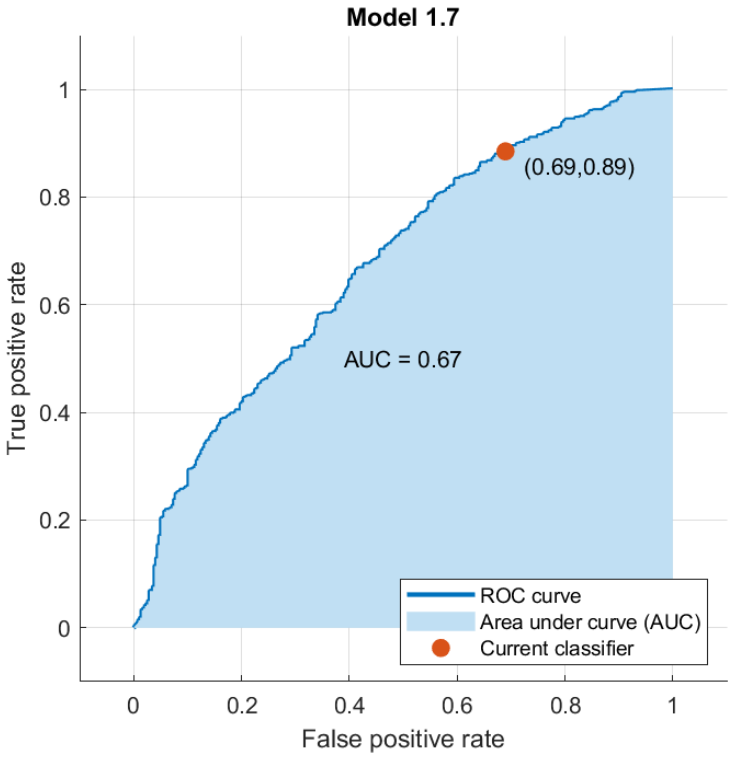
Şekil 4.26. 3-CV en başarılı Cubic SVM algoritma için Karışıklık Matrisi



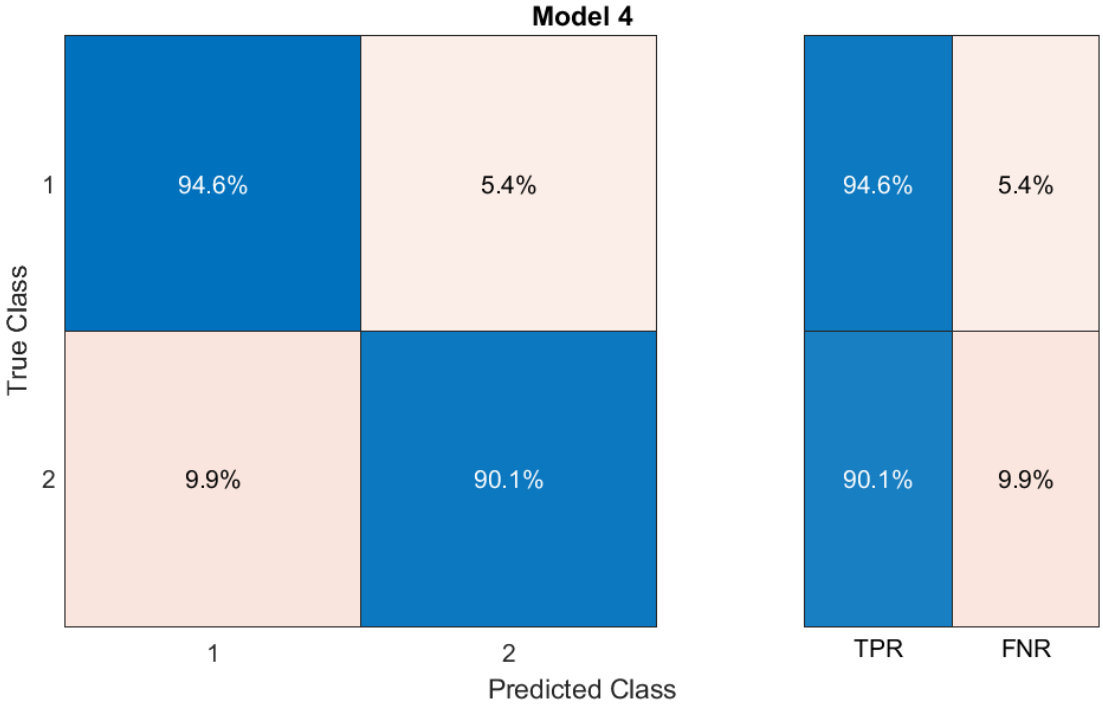
Şekil 4.27. 3-CV en başarılı Cubic SVM algoritması için ROC eğrisi



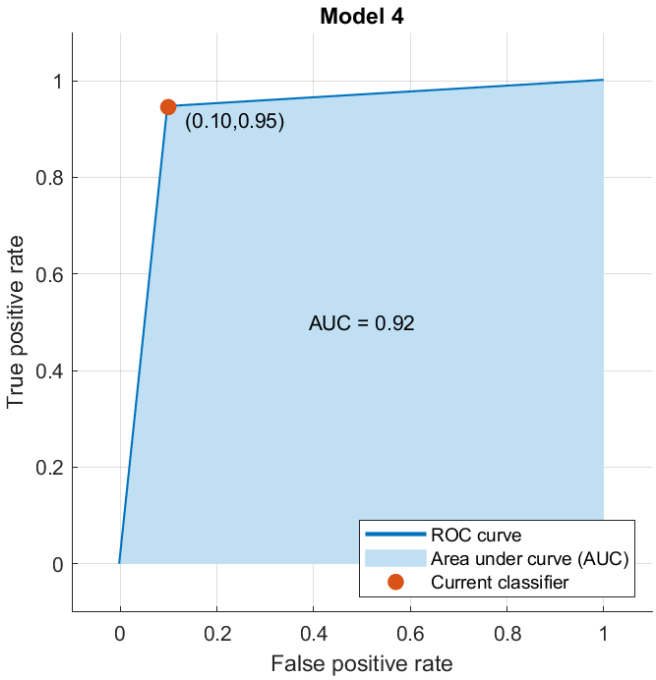
Şekil 4.28. 3-CV en başarısız Gaussian Naive Bayes algoritma için Karışıklık Matrisi



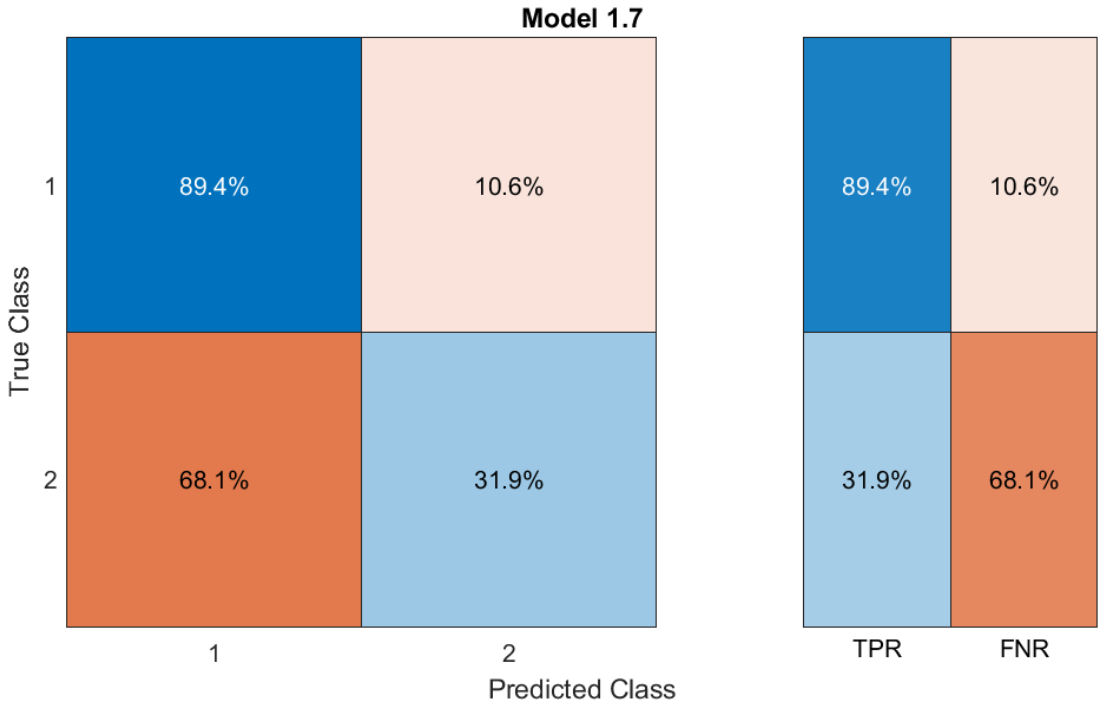
Şekil 4.29. 3-CV en başarısız Gaussian Naive Bayes algoritma için ROC eğrisi



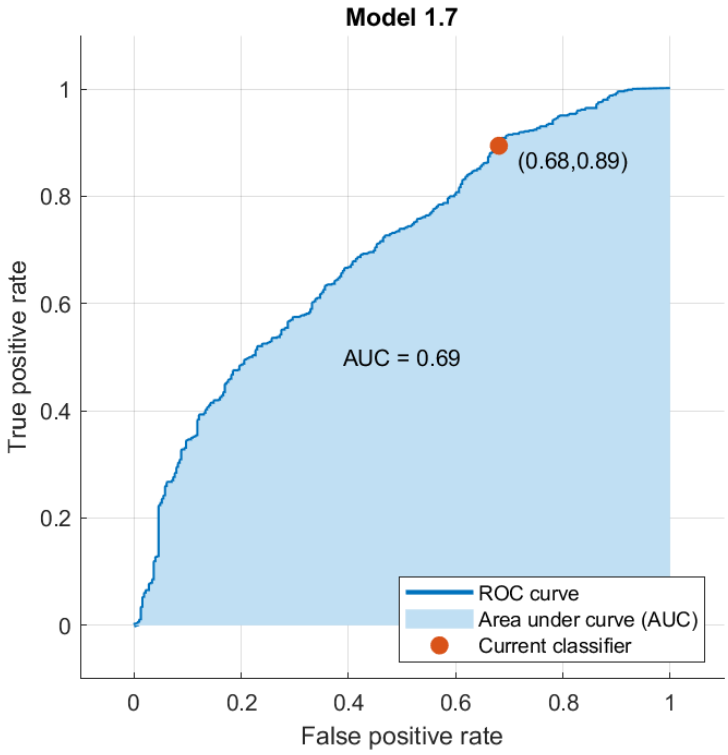
Şekil 4.30. 5-CV en başarılı Fine KNN algoritma için Karışıklık Matrisi



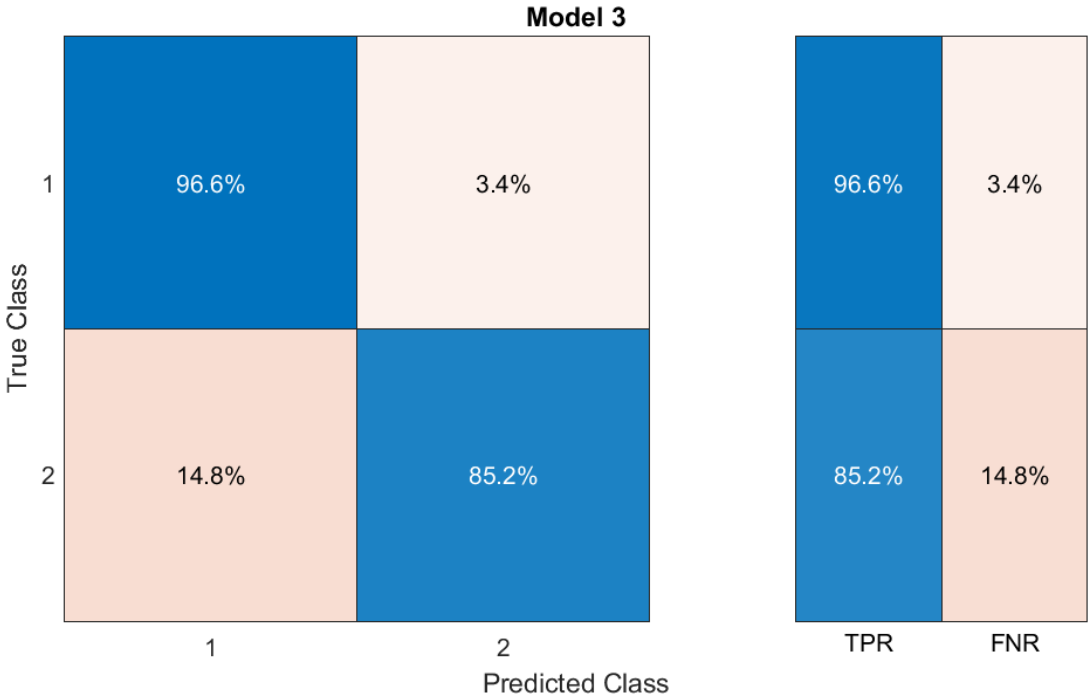
Şekil 4.31. 5-CV en başarılı Fine KNN algoritması için ROC eğrisi



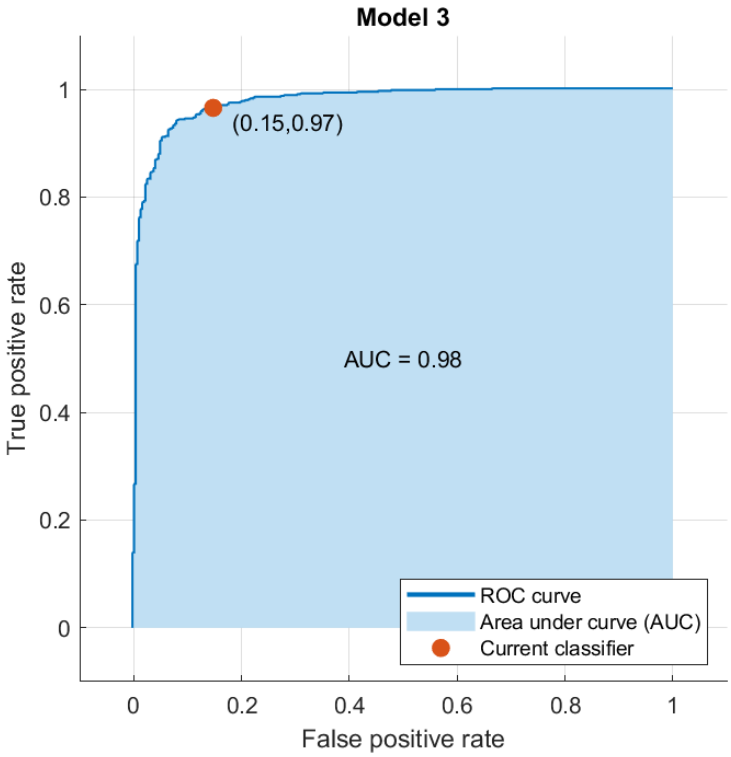
Şekil 4.32. 5-CV en başarısız Gaussian Naive Bayes algoritma için Karışıklık Matrisi



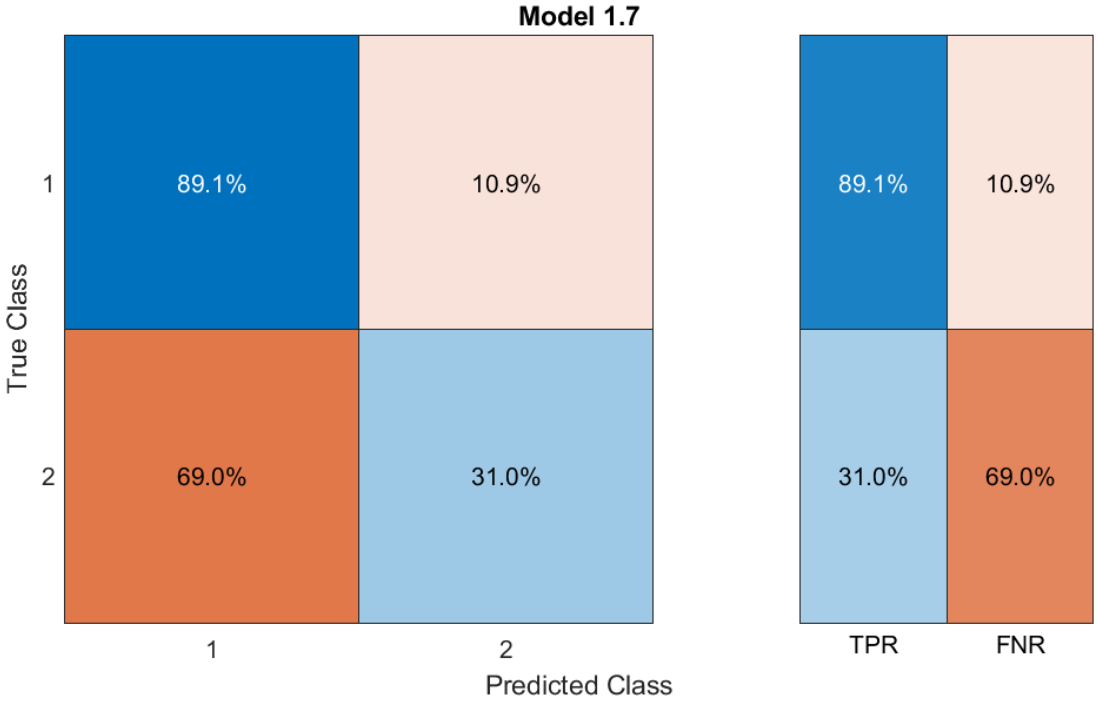
Şekil 4.33. 5-CV en başarısız Gaussian Naive Bayes algoritma için ROC eğrisi



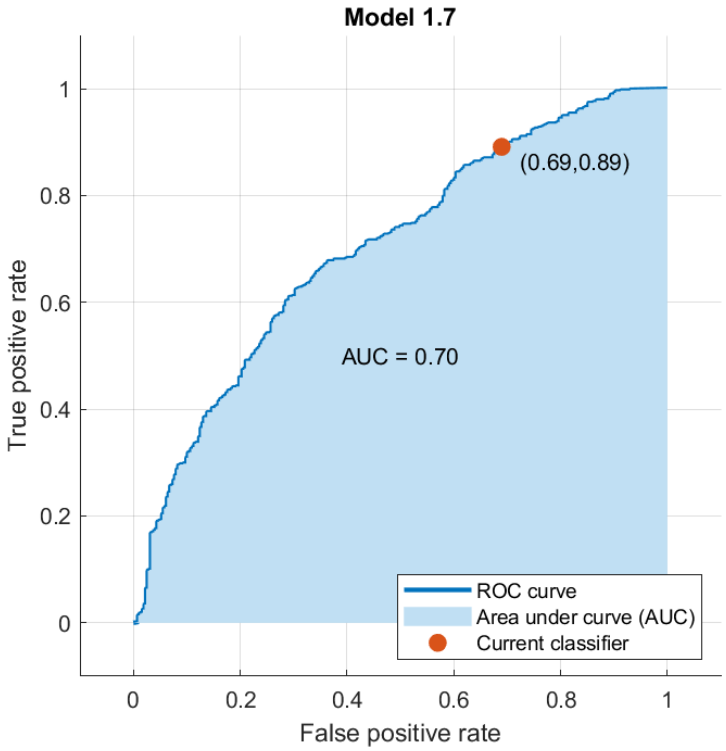
Şekil 4.34. 10-CV en başarılı Bagged Trees algoritma için Karışıklık Matrisi



Şekil 4.35. 10-CV en başarılı Bagged Trees algoritma için ROC eğrisi



Şekil 4.36. 10-CV en başarısız Gaussian Naive Bayes algoritma için Karışıklık Matrisi



Şekil 4.37. 10-CV en başarısız Gaussian Naive Bayes algoritma için ROC eğrisi

Elde edilen sonuçlar incelendiğinde, yukarıda bahsedilen yapay sinir ağı yöntemlerinin, sepsis hastasının erken teşhisini öngörmede olağanüstü bir başarıya sahip olduğu anlaşılmaktadır. Böylece mevcut durumla yapılan çalışmanın başarılı olduğu ve hastaların erken teşhisini öngörmek için makine öğrenmesi algoritmalarının kullanılmasının uygun olduğu sonucuna varılmıştır. Ayrıca bu çalışma, sepsis hastasının erken teşhisinin öngörülmesinde hekimler için yeni bir destek aracı olarak kullanılma potansiyeline sahiptir.

Sonuç olarak, sepsisin erken tahmini için, 32 farklı makine öğrenme tekniği uygulanmış ve 12 farklı yöntemin %100 doğruluk performansı elde ettiği görülmüştür. Diğer yöntemlerin başarı oranları %68,9 ile %96,1 arasında değişmektedir ve küçümsenemez. Bu performanslar incelendiğinde makine öğrenmesi algoritmalarının sepsis karar destek sistemleri olarak uygulamalara entegre edilebileceği görülmektedir. Yüksek doğrulukta tahmin yetenekleri nedeniyle sepsis ile ilişkili ölüm-hayatta kalma tahmininde yapay zeka yöntemlerinin kullanılması önerilir.

5. SONUÇLAR VE TARTIŞMA

Sepsis hastaları çok yüksek bir ölüm oranına sahiptir ve doktorların erken teşhis koyabilmek için güvenilir prognostik tahminlere ihtiyacı vardır ve bilinçli kararlar vermek için ilaçları doğru şekilde uygulamak gerekir. Bilişim Bilimlerinin sağlık sektöründeki uygulaması olarak değerlendirilen özgün araştırmamızda sepsis tanısında kullanılan veri seti üzerinde kestirimci öğrenme modelleri ve verilerden öğrenen karar destek sistemleri üretilmiştir. Bu modeller hastalığın erken tahminini değerlendirerek bilime katkı sağlamayı amaçlamaktadır.

Sepsis tanısı için oluşturulan veritabanı ile farklı yapay zeka yöntemleri üzerine deneysel bir tez çalışması yapılmıştır. Sonuçlar Matlab ortamında denenmiş olup farklı modeller üzerinde deneysel çalışmalar ve özellik mühendisliği yöntemleri denenmiştir. Sonuçlar incelendiğinde birçok farklı yöntem ile %100'e yakın bir sınıflandırma başarı oranına ulaşıldığı görülmüştür (Accuracy).

Bu çalışmada, veri setindeki tüm özellikler makine öğrenmesi yöntemleriyle sınıflandırılmış ve Sepsis'in erken tahmini açısından kontrollü bir sınıflandırma problemi için %100 doğrulukla tahmin edilmiştir. Buna göre çalışmamız, Sepsis hastalarının sağlık kayıtlarının ikili sınıflandırmasında makine öğreniminin etkin bir şekilde kullanılabileceğini göstermiştir.

Yüksek özellikli veri kümelerinde PCA kullanılması önerilir. PCA, verilerdeki gerekli bilgileri ortaya çıkarmak için oldukça etkili bir yöntemdir. PCA'nın arkasındaki temel mantık, verilerdeki temel özellikleri yakalayarak çok boyutlu bir veriyi daha az değişkenle temsil etmektir. Veri setindeki küçük veriler için k-katlama çapraz doğrulama yöntemi önerilir. Bu yöntemin yanında PCA önerilmez. Bunun yerine parametre optimizasyon yöntemi denenir. Bu tez çalışmasından bu sonuçları elde ediyoruz.

Karşılaştırma için farklı türlerde makine öğrenmesi yöntemleri farklı varyantlarla denenmiş ve %69 ile %100 arasında sınıflandırma doğruluğu elde edilmiştir. Cubic SVM, Fine Gaussian SVM, Fine KNN, Boosted Trees, Baggged Trees, Subspace KNN, Dar Sinir Ağı, Orta Sinir Ağı, Geniş Sinir Ağı Çift Katmanlı Sinir Ağı ve Trialayer Sinir Ağı yöntemlerinin performansının maksimum olduğu gözlemlenmiştir. Mevcut çalışmanın bir kısıtlılığı olarak, veri setinin (977 hasta) küçük ve sınıflandırma için kararsız olduğu göz ardı edilmemelidir.

Sepsise bağlı hastane ölümlerini azaltmak için yapay zeka yöntemlerine dayalı yaptığımız çalışmada, daha hızlı tanı konularak Sepsisin ilerlemesinde önlem alınma imkanı ile kişileri hayatta tutmak için erken teşhis konusunda hekimlere destek olacak bir karar destek yazılımı oluşturma imkanı kazandırmıştır.

Ayrıca başarı oranı en yüksek olan yapay zeka yöntemlerinden HBYS(Hastane Bilgi Yönetim Sistemi) için en uygun yöntem seçilerek, HBYS'nin yoğun bakım modülüne entegrasyonunu yaparak, Sepsise bağlı ölüm oranlarını azaltabilmek için, Hekimlere destek olacak karar destek sistemlerini uygulamaya alabilmektir.

7. KAYNAKÇA

[1] Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M, vd. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 23 Şubat; 315 (8): 801-810.

[2] Global Sepsis Alliance Web Site (Last Access: 03, March 2022), https://www.global-sepsis-alliance.org/news/2020/1/16/the-lancet-sepsis-associated-with-1-in-5-deaths-worldwide-double-prior-estimates-children-and-poor-regions-hit-hardest-global-burden-disease-study-kristina-rudd.

[3] Kaya, U., Yilmaz, A. ve Díkmen, Y. (2018). Prediction of sepsis disease by Artificial Neural Networks. Journal of Selcuk-Technic.Special Issue 2018 (ICENTE'18):107-31.

[4] Gultepe, E., Green, J., Nguyen, H., Adams, J., Albertson, T. ve Tagkopoulos I. (2014). From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system J Am Med Inform Doç. 21 (2): 315–25. DOI: 10.1136.

[5] Desautels T., Calvert, J., Hoffman, J., et al. (2016). Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform 4:e28.

[6] Fleuren, L.M., Klausch, T.L.T., Zwager, C.L., Schoonmade, L.J., Guo, T., Roggeveen, L.F., Swart, E.L., Girbes, A.R.J., Thoral, P., Ercole, A., Hoogendoorn, M. ve Elbers, PWG. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy Yoğun Bakım Med. Mart; 46 (3): 383–400. doi: 10.1007.

[7] Gültepe, E. Nguyen, H. Albertson, T. et al. (2012). A Bayesian network for early diagnosis of sepsis patients: a basis for a clinical decision support system 2. IEEE Uluslararası Biyo ve Tıp Bilimlerinde Hesaplamalı Gelişmeler Konferansı (ICCABS); Las Vegas, NV: 23–25, 1-5.

[8] Stanculescu, I., Williams, C.K.I., Y. ve Freer, Y. (2014). Autoregressive Hidden Markov Models for the Early Detection of Neonatal Sepsis, in IEEE Journal of Biomedical and Health Informatics, Sept. 18(5); 1560-1570.

[9] Guillén, J. et al. (2015). Predictive models for severe sepsis in adult ICU patients, 2015 Systems and Information Engineering Design Symposium, Charlottesville, VA, 182-187.

[10] Mani, S., Ozdas, A., Aliferis, C., Varol, H.A., Chen, Q., Carnevale, R., et al. (2014). Medical decision support using machine learning for early detection of late-onset neonatal sepsis. J. Am. Med. Inform. Assoc. 21, 326–36. 10.1136/amiajnl-2013-001854.

[11] Horng, S., Sontag, D.A., Halpern, Y., Jernite, Y., Shapiro, N.I. ve Nathanson, L.A. (2017). Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS ONE. 12:e0174708. 10.1371/journal.pone.0174708.

[12] Lamping, F., Jack, T., Rübsamen, N., Sasse, M., Beerbaum, P., Mikolajczyk, R.T., et al. (2018). Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children – a data-driven approach using machine-learning algorithms. BMC Pediatr. 18:112. 10.1186/s12887-018-1082-2.

[13] Kamaleswaran, R., Akbilgic, O., Hallman, M.A., West, A.N., Davis, R.L. ve Shah, S.H. (2018). Applying artificial intelligence to identify physiomarkers predicting severe sepsis in the PICU. Pediatr. Crit. Care Med. 19:e495–503. 10.1097/PCC.0000000000001666

[14] Calvert, J., Saber, N., Hoffman, J. ve Das, R. (2019). Machine-learning-based laboratory-d test for theeveloped diagnosis of sepsis in high-risk patients. Diagnostics. 9:20. 10.3390/diagnostics9010020.

[15] Masino, A.J., Harris, M.C., Forsyth, D., Ostapenko, S., Srinivasan, L, Bonafide, C.P., et al. (2019). Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data.

[16] Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., et al. (2019). Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. Comput. Biol. Med. 109:79–84. 10.1016/j.compbiomed.2019.04.027.

[17] Le, S., Hoffman, J., Barton, C., Fitzgerald, J.C., Allen, A., Pellegrini, E., et al. (2019). Pediatric severe sepsis prediction using machine learning. Front. Pediatr. 7, 413. 10.3389/fped.2019.00413.

[18] Yee, C.R., Narain, N.R., Akmaev, V.R. ve Vamulapalli, V. (2019). Yoğun bakım ünitesinde septik şoku tahmin etmeye yönelik veriye dayalı bir yaklaşım Biyomedya. Bilgi vermek. Analizler. 11 :1178222619885147. 10.1177/1178222619885147.

[19] Seymour, C.W., Liu, V.X., Iwashyna, T.J., Brunkhorst, F.M., Rea, T.D., Scherag, A., ... ve Deutschman, C.S. (2016). Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Jama, 315(8), 762-774.

[20] Centers for Disease Control and Prevention (CDC). (2018). Hospital Toolkit for Adult Sepsis Surveillance.

[21] Dellinger, R.P., Levy, M.M., Rhodes, A., Annane, D., Gerlach, H., Opal, S.M., ... ve Osborn, T. M. (2013). Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive care medicine, 39(2).

[22] Yang, M. ve De Loera, J.A. (2018). A machine learning approach to evaluate Beijing air quality.

[23] Hal Daumé III (2012). A Course in Machine Learning, http://ciml.info/dl/v0_8/ciml-v0_8-ch08.pdf, Accessed 22 Sep 2014.

[24] Mohri, M., Rostamizadeh, A. ve Talwalkar, A. (2012). Foundations of Machine Learning, the MIT Press, ISBN: 0-262-01825-X.

[25] Alpaydın, E. (2014). Introduction to Machine Learning, MIT Press, ISBN: 978-0-262-02818-9.

[26] Harrington, P. (2012). Machine Learning in Action, 1st Edition, Manning Publications Shelter Island, NY, ISBN: 978-1-61729-018-3.

[27] Kartal, E. (2015). Sınıflandırmaya dayalı makine öğrenmesi teknikleri ve kardiyolojik risk değerlendirmesine ilişkin bir uygulama, PhD. Thesis, İstanbul University.

[28] Zendehboudi Alireza, Baseer M.A. ve Saidur R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: a review. J Clean Prod. 199, 272–85.

[29] Tasçı E. ve Onan, A. (2016). Investigation of the effect of K-nearest neighbor algorithm parameters on classification performance. In: Turkish) Akademik bilişim.

[30] Gürfidan, R. ve Ersoy, M. (2021). Classification of Death Related to Heart Failure by Machine Learning Algorithms, Advances in Artificial Intelligence Research, 1(1), 13-18.

[31] Yang, F. (2018). An implementation of naive Bayes classifier,"Computational Science and Computational Intelligence (CSCI) 301-306.

[32] Vincent, J.L., Moreno, R., Takala, J., Willatts, S., De Mendonça, A., Bruining, H, et al. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 22, 707–10. DOI: 10.1007/BF01709751.

[33] Bone, R.C., Bak, R.A., Cerra, F.B., Dellinger, R.P., Fein, A.M., Knaus, W.A., … ve Sibbald, W.J. (1992). Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. Chest. Jun;101(6), 1644-55.

[34] Levy, M.M., Fink, M.P., Marshall, J.C., Abraham, E., Angus, D., Cook, D … and Ramsay, G. (2003). Crit Care Med. Apr;31(4):1250-6. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference.

[35] Dellinger, R.P., Levy, M.M., Carlet, J.M., Bion, J., Parker, M.M., Jaeschke, R, … and Vincent, J.L.(2008). Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. Crit Care Med. 36, 296-327.

[36] Dellinger, R.P., Levy, M.M., Rhodes, A., Annane, D., Gerlach, H., Opal, S.M … ve Moreno, R. (2013). Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. Crit Care Med. 41, 580-637.

[37] Pinsky, M.R. ve Matuschak, G.M. (1989). Multiple systems organ failure: failure of host defense homeostasis. Crit Care Clin. 5(2), 199-220.

[38] Marshall, J.C., Cook, D.J., Christou, N.V., Bernard, G.R., Sprung, C.L. ve Sibbald, W.J. (1995). Multiple organ dysfunction scores: a reliable descriptor of a complex clinical outcome. Crit Care Med. 23(10), 1638-52. 156.

[39] Le Gall, J.R., Klar, J., Lemeshow, S., Saulnier, F., Alberti, C., Artigas, A. ve Teres, D. (1996). The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. JAMA. 11;276(10), 802-10.

[40] Centers for Disease Control and Prevention (CDC). (2018). Hospital Toolkit for Adult Sepsis Surveillance.

[41] National Institute for Health and Care Excellence. Sepsis: recognition, diagnosis, and early management: NICE (2017). Sepsis: recognition, diagnosis and early management. BJU international. 2018;121(4), 497-514.

[42] Parrillo, J.E. (1993). Pathogenetic mechanisms of septic shock. The New England journal of medicine. 328(20), 1471-7.

[43] Sarıer, M. (2018). Sepsis Tanı Kriterlerinin Karşılaştırılması. Tıpta Uzmanlık Tezi.

[44] Ferreira, F.Lb, Bota, D.P., Bross, A., Mélot, C. ve Vincent, J. L. (2001). Serial evaluation of the SOFA score to predict outcome in critically ill patients. Jama, 286(14), 1754-1758.

[45] Bone, R.C., Balk, R.A., Cerra, F.B., Dellinger, R.P., Fein, A.M., Knaus, W.A., ... ve Sibbald, W.J. (1992). Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Chest, 101(6), 1644-1655.

[46] Mohri, M., Rostamizadeh, A. ve Talwalkar, A., (2012). Foundations of Machine Learning, The MIT Press.

[47] Ramasubramanian, K. ve Singh, A. (2019). Machine Learning Using R, 2nd in., Apress.

[48] Russell, S.J. ve Norvig, P. (2010). Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall.

[49] Balaban, M.E. ve Kartal, E. (2015). Veri madenciliği ve makine öğrenmesi temel algoritmalar ve R dili ile uygulamaları. İstanbul.

[50] Chawla, N.V., Japkowicz, N. ve Kolcz, A. (2004). Editorial: Special issue on learning from imbalanced data sets, SIGKDD Explorations. 6, 1-6.

[51] Huang, J. ve Ling, C.X. (2007). Constructing new and better evaluation measures for machine learning", in R. Sangal, H. Mehta and R. K. Bagga (Eds.) Proc. of the 20th International Joint Conference on Artificial Intelligence (IJCAI), San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 859-864.

[52] Rakotomamonyj, A. (2004). Optimizing area under ROC with SVMs, in J. Hernandez-Orallo, C. Ferri, N. Lachiche and P. A. Flach (Eds.) 1st Int. Workshop on ROC Analysis in Artificial Intelligence (ROCAI), Valencia, Spain, 71-80.

[53] Provost, F. ve Domingos, P. (2003). Tree induction for probability-based ranking. Machine Learning, 52, 199-215.

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	**Pınar Kaya Aksoy**
Doğum Yeri ve Yılı	
E-Posta	
Yabancı Dil	**İngilizce**

Eğitim	
Önlisans	2002.Marmara Üniversitesi, Sağlık Hizmetleri M.Y.O. Anestezi
Önlisans	2008, Mersin Üniversitesi, Mersin Meslek Yüksekokulu/Bilgisayar Teknolojisi ve Programlama
Lisans	2008, Anadolu Üniversitesi, İktisat Fakültesi, Kamu Yönetimi
Lisans	2014, Yaşar Üniversitesi/Mühendislik Fakültesi/Bilgisayar Mühendisliği (İngilizce),(Tam Burslu)

Yayın Listesi	
Makale	Kaya Aksoy, P. (2022). A Decision Support System on Artificial Intelligence Based Early Diagnosis of Sepsis. Artificial Intelligence Theory and Applications (AITA), Cilt 2 (1), 14.

İŞ TECRÜBELERİ

- **İzmir İl Sağlık Müdürlüğü - Destek Hizmetleri Başkanlığı – Sağlık Bilgi Sistemleri Birimi-** Sistem ve Network Eylül 2019 – Halen
- **Sağlık Bakanlığı - Sağlık Bilgi Sistemleri Genel Müdürlüğü - Sistem Yönetimi ve Bilgi Güvenliği Daire Başkanlığı** SOME Birimi Ocak 2017 – Ağustos 2019
- **Ankara DR.A.Y. Onkoloji Hastanesi** 2015 - Ocak 2017 Bilgi İşlem Sorumlusu
- **İzmir Bozyaka Eğitim ve Araştırma Hastanesi** 2007 – 2015 Anestezi Teknikeri
- **Manisa Saruhanlı Devlet Hastanesi** 2004 – 2007 Anestezi Teknikeri