

Makine Öğrenmesi Yöntemlerinin Felç Riskinin Belirlenmesinde Performansı: Karşılaştırmalı bir çalışma

Performance of Machine Learning Methods in Determining Stroke Risk: A Comparative Study

Özer Oğuz¹ , Suat Bayır¹ , Hasan Badem² 

¹ Enformatik Ana Bilim Dalı, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

² Bilgisayar Mühendisliği Bölümü, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

(ozersunayoguz@gmail.com, suatbayir1@gmail.com, hbadem@ksu.edu.tr)

Received: Sep.3, 2021

Accepted: Sep.16, 2021

Published: Oct.20, 2021

Özetçe— Felç (inme), beyin ya da kalbin belli bir bölgesinde kan akışının azalması ya da kesilmesi sonucunda gerçekleşen ani krizlerdir. Dünya genelinde ölüme en çok neden olan rahatsızlıklardan biri olan felcin kalıcı sakatlanmalara da neden olduğu bilinmektedir. Bu nedenle felç riskinin önceden belirlenmesi ölüm ya da kalıcı sakatlık riskinin azaltılması için oldukça önemlidir. Bu çalışmada felcin erken teşhisi ve risk sınıflandırması için 13 farklı makine öğrenme yöntemi kullanılmış ve deneysel sonuçlar elde edilmiştir. Elde edilen deneysel sonuçlar çeşitli başarı karşılaştırma ölçütlerine göre değerlendirilerek en başarılı makine öğrenme modeli belirlenmiştir. Elde edilen deneysel sonuçlarda Rastgele Orman Sınıflandırıcısı 99.425% doğruluk değeri ile en başarılı yöntem olduğu görülmüştür.

Anahtar Kelimeler: Felç, makine öğrenmesi, hastalık teşhisi, sınıflandırma, rastgele orman sınıflandırıcısı

Abstract— Stroke is a sudden crisis that occurs as a result of reduced or interrupted blood flow in a certain region of the brain or heart. The stroke, which is one of the most common causes of death in the world, also causes permanent disability is known. Therefore, predetermining the risk of stroke is very important to reduce the risk of death or permanent disability. In this study, 13 different machine learning methods have been used for early diagnosis and risk classification of stroke and experimental results have been obtained. The obtained experimental results have been evaluated on various comparison criteria. In the obtained experimental results, Random Forest Classifier has been found to be the most successful method with accuracy rate with 99.425%.

Keywords: Stroke, machine learning, disease diagnosis, classification, random forest

1. Giriş

Dünya Sağlık Örgütü'nün (WHO) tanımına göre felç(inme), hızla gelişen semptomlarla birlikte fokal serebral fonksiyonun bozukluğunun vasküler klinik belirtileri dışında belirlenebilir bir neden olmaksızın 24 saat veya daha uzun sürebilen ve aynı zamanda ölüme yol açabilen hastalıktır (Truelson

ve diğerleri.,2006). Felç, beynin ya da kalbin belirli bir bölgesinde meydana gelen kan akışının azalması veya tamamen kesintiye uğramasıyla beyin dokusunda oksijen ve besin yetersizliğine bağlı gerçekleşen ani krizlerdir (Louis ve Caplan, 2016). WHO 2019 yılı verilerine göre ani gerçekleşen felç, dünya genelinde ölüme en çok neden olan ikinci sıradaki rahatsızlıktır. Aynı zamanda WHO, inmenin dünya genelinde sakatlığa en çok neden olan dördüncü rahatsızlık olduğunu belirtmektedir (WHO, 2019). Türkiye’de ise felcin, en çok ölüme neden olan ikinci rahatsızlık olduğu rapor edilmektedir (WHO, 2020). Felç, bireyin yaşam standartlarını önemli ölçüde düşüren önemli sağlık sorunlarından biridir (Emre ve diğerleri, 2019). Bu nedenlerle felç riskinin erken dönemde belirlenmesi ve felcin erken teşhisi oldukça önem arz etmektedir. Felç, erken müdahale edildiğinde ya da erken teşhis edilebildiğinde ölüm ya da kalıcı sakatlık riskinin yaklaşık %80 oranında engellenebildiği belirtilmektedir (Singh ve Choudhary, 2017). Bu sebeple felcin erken ve hızlı teşhisi için hastane ve uzman bilgi ve deneyimlerinin yanında farklı alternatif yöntemler keşfetmek oldukça önemlidir.

Makine öğrenmesi teknikleri sınıflandırma problemlerinde başarılı sonuçlar verdiği için, tıbbi alanlarda da hastalıkların tahmini ve sınıflandırılmasında sıklıkla kullanılmaktadır (Badem, 2019). Kalp krizi (Anusha ve diğerleri, 2021), kanser (Erkal ve diğerleri, 2020), Epilepsi ve Parkinson (Badem, 2019; Çalışkan ve Badem, 2021) gibi farklı sağlık problemlerinin tespiti ve sınıflandırılması işlemlerinde başarılı sonuçlara ulaşıldığı rapor edilmektedir. Makine öğrenme teknikleri ile hastalık teşhisi hızlı ve kolay olduğundan dolayı tıbbi alanda sıklıkla tercih edilmektedir. Felç belirtilerini anlamının oldukça zor olması, çok ani gerçekleşmesi ve acil müdahale gerekliliği felcin erken teşhisini önemli kılmaktadır. Erken teşhis edilmediği takdirde kalıcı hasarlar bırakabildiği gibi ölüme de sonuçlanabilmektedir. Dolayısıyla diğer sağlık problemlerinde olduğu gibi felç teşhisi ve risk sınıflandırılmasında da hızlı ve güvenilir teşhis için makine öğrenmesi tekniklerinin kullanılması insan hayatı açısından oldukça önemlidir.

Felç rahatsızlığının yaş, cinsiyet, sigara kullanımı, hipertansiyon vb. çeşitli risk faktörleri bulunmaktadır. Bunun yanında kişinin yaşam koşullarının da felç için risk oluşturduğu belirtilmektedir (Emon ve diğerleri, 2020). Bu çalışmada felç rahatsızlığının risk faktörlerine göre elde edilen veri setinden yararlanarak hastalık riskinin belirlenmesinde en başarılı makine öğrenmesi tekniğini tespit etmek amaçlanmıştır. Amacımızı gerçekleştirmek için Lojistik Regresyon (LR, Logistic Regression), Karar Ağaçları (CART, Decision Tree Classifier), Rastgele Orman Sınıflandırıcısı (RFC, Random Forest Classifier), Destek Vektör Makinesi (SVM, Support Vector Machine), En Yakın Komşu Algoritması (KNN, K Neighbors Classifier), Bernoulli Naive Bayes (BNB), Gaussian Naive Bayes (GNB) Çok Katmanlı Algılayıcı Sınıflandırıcı (MLP, Multi-Layer Perceptron Classifier), Gradyan Artırıcı Sınıflandırıcı (GBM, Gradient Boosting Classifier), Adaboost Sınıflandırıcı (AB, Adaboost Classifier), LGBM Sınıflandırıcı (LGBM, Light Gradient Boosting Classifier), Ekstra Ağaç Sınıflandırıcısı (ET, Extra Trees Classifier) ve Derin Sinir Ağları (DNN, Deep Neural Networks) olmak üzere 13 farklı sınıflandırma algoritmasının performansları karşılaştırmalı olarak değerlendirilmiştir.

Çalışmanın ikinci bölümünde literatür özeti ve kullanılan yöntemler sunulmuştur. Üçüncü bölümünde ise, veri seti, karşılaştırma ölçütleri ve önerilen yöntem yer verilmiştir. Dördüncü bölümde ise sınıflandırma algoritmalarından elde edilen deneysel sonuçlar, literatür ile karşılaştırmalı olarak yer verilmiştir. Çalışma sonuç ve öneriler bölümleriyle sonlandırılmıştır.

2. Literatür Özeti ve Genel Bilgiler

2.1. Literatür Özeti:

Literatürde sınırlı araştırmada felcin teşhis ve tespit edilebilmesi için makine öğrenme tekniklerini kullandığı görülmektedir. Literatürde RFC, CART, NB, SVM vb. farklı birçok makine öğrenmesi tekniği önerilerek başarılı sonuçlar elde edildiği rapor edilmiştir. Emon ve ark. (2020), yaptıkları çalışmada felç hastalığı tahmini için 10 farklı makine öğrenmesi tekniğini kullanarak sonuçları karşılaştırmışlardır. Çalışmada önerilen modeller arasında Ağırlıklı Oylama Algoritmasının (Weighted Voting Algorithm) felci %97 doğruluk oranıyla daha başarılı bir şekilde tahmin ettiği rapor edilmiştir.

Singh ve ark. (2017), yapay zekâ kullanılarak felç riskinin tahmini için gerçekleştirdikleri çalışmada kullandıkları veri setinden öznitelik seçim işleminde CART, boyut azaltma işleminde Temel Bileşen Analizi (Principal Component Analysis-PCA) ve sınıflandırma işlemini gerçekleştirmek için Geri Yayılımlı Sinir Ağından (Back Propagation Neural Network) oluşan bir model önermişlerdir. Çalışma sonunda önerilen model ile %97,7 doğruluk değerini elde edilmiştir.

Sevli (2020), felç riskini belirlemek için gerçekleştirdiği çalışmada RFC makine öğrenmesi tekniğini kullanmıştır. Çalışmada kullanılan veri setinin örneklemelerinden kaynaklanan bazı veri problemlerinden dolayı yeniden örnekleme tekniği kullanılarak, sınıflandırma işlemi gerçekleştirilmiştir. Gerçekleştirilen sınıflandırma işlemi sonucunda %98,84 doğruluk değerine ulaşıldığı rapor edilmiştir. Ayrıca çalışmada felç riskinin tahmini için veri setindeki özniteliklerin etkisi de araştırılmıştır. Çalışma sonucunda en etkili özniteliklerin yaş, vücut kitle indeksi ve kandaki glukoz seviyesi olduğu rapor edilmiştir.

Revanth ve ark. (2020), felç riski tahmini için SVM, CART, RFC ve MLP makine öğrenmesi modellerini kullanarak karşılaştırmalı bir çalışma gerçekleştirmişlerdir. Bu amaçla geliştirdikleri web tabanlı uygulama üzerinden felç riski verilerini içeren veri seti ile modellerin eğitimini gerçekleştirerek sonuçları karşılaştırmışlardır. Çalışma sonucunda, önerilen SVM modelinin %98.99 doğruluk oranı ile önerilen diğer modellerden daha başarılı olduğu sonucuna ulaşılmıştır.

Cheon ve ark. (2019), felci tespit edebilmek için sağlık davranışlarını ve medikal hizmet kullanımını temel alan bir çalışma gerçekleştirmişlerdir. Çalışmada 15.099 adet felçli hastadan topladıkları verilerle derin öğrenme yöntemi kullanılarak felç tespiti yapılmıştır. Veri ön işleme aşamasında PCA kullanılarak veriler üzerinde boyut azaltma işlemi gerçekleştirilmiştir. Derin öğrenme ile gerçekleştirilen tahmin işleminde %83,48 AUC değerine ulaşıldığı rapor edilmiştir.

2.2. Makine Öğrenme Modelleri:

Bu çalışmada felcin erken tespiti ve risk sınıflandırması için literatürde en sık kullanılan 13 adet makine öğrenmesi tekniği belirlenmiştir. Bu amaçla; LR, CART, RFC, SVM, KNN, BNB, GNB, MLP, GBM, AB, LGBM, ET ve DNN sınıflandırıcıları ile modellerin eğitimi gerçekleştirilmiş, felç erken teşhisi ve risk sınıflandırması için en başarılı sonucu veren model belirlenmiştir. Kullanılan makine öğrenme modellerine aşağıda sunulmuştur.

SVM, sınıflandırma ve regresyon problemlerinde kullanılabilen denetimli (Supervised) öğrenme gerçekleştiren bir makine öğrenmesi tekniğidir (Cortes ve Vapnik, 1995). Sınıflandırma işlemini bir düzlem üzerine yerleştirilen girdi noktalarına bir sınır çizgisi çizerek iki ayrı gruba ayırarak gerçekleştirir. Bu işlemi gerçekleştirirken iki sınıfı birbirinden ayıran en uygun doğrunun çizilmesi amaçlanır.

CART, sınıflandırma ve regresyon problemlerinde kullanılabilen, küçük veri setleri için de uygun olan makine öğrenme tekniklerinden biridir. CART, kök, düğüm ve yapraklardan oluşan bir ağaç tipinde sınıflandırıcıdır. Bir kök ve başlayan CART modeli, dallanarak bir ağaç yapısı meydana gelir. Bu modelde her bir adımda girdilerin değeri hesaplanır. Hesaplanan bu değere göre diğer adımdaki eylem seçilir. Bu eylemler hesaplama için bir başka girdinin seçilmesi, fonksiyona ait değerlerin çıkartılması ya da girdi-çıktı değerleri için farklı kombinasyonların denenmesi olabilir (Kalles ve Morris, 1994).

LR, hesaplamasının basit olması ve etkili performansı nedeniyle istatistikte ve makine öğreniminde ikili sınıflandırma problemlerinde sıklıkla tercih edilmektedir (Song ve diğerleri, 2021). LR, ikili (binary) bağımlı değişkenler ile bağımsız değişkenlerin ilişkisini ortaya çıkaran en iyi matematiksel modeli bulmaya dayanır.

NB, Bayes teoremine dayanan bir makine öğrenmesi tekniğidir [18]. Bu algoritma oluşturulması görece kolay olması ve karmaşık tahmin şemalarına ihtiyaç duymaması nedeniyle sıklıkla tercih edilmektedir. NB, veri setindeki girdinin olasılıksal olarak hangi sınıfa ait olduğunun belirlenmesini

sağlayan makine öğrenmesi tekniğidir (Rish, 2001; Badem, 2019). GNB ve BNB, NB temeli makine öğrenmesi modellerindedir.

RFC, birden çok karar ağaçlarının bir araya gelmesi ile oluşan sınıflandırma algoritmasıdır. RFC'de karar ağaçları arasından en yüksek doğruluğa ve uygunluğa sahip olanlar tercih edilir. RF'da ormandaki her bir ağaç eşit sayıda veri dağılıma sahip olmaktadır. Bu ağaçlar veri setindeki özniteliklere göre dallanarak ortalama bir sonuç üretmektedir (Breiman, 2001).

ABC, güçsüz algoritmaların bir araya getirilerek yeni ve daha güçlü bir algoritma oluşturulmuştur. ABC ile daha önce geliştirilmiş çeşitli makine öğrenme modellerinin birleştirilmesi veya karşılaştırılmasıyla daha başarılı sonuçlara ulaşılması amaçlanmaktadır (Freund ve Schapire, 1997).

KNN, makine öğrenmesi modelleri arasında iyi bilinen ve sıkça tercih edilen algoritmalar arasında yer almaktadır. KNN'de ilk olarak k değeri belirlenmektedir. k değeri, belirlenmiş bir noktaya en yakın konumlarda bulunan komşuların sayısını ifade etmektedir. Minkowski, Öklid, Chebyshev ve kosinüs gibi uzaklık fonksiyonlarının kullanılmasıyla belirlenen noktanın, eldeki verilere göre uzaklığı hesaplanmaktadır. Mevcut noktanın hesaplanan yeni özelliklerine göre bir sınıfa etiketlenmesi ile gerçekleştirilir (Dudani, 1976). KNN basit olması ve hızlı çalışması nedeniyle sıklıkla tercih edilse de büyük verilerle çalışıldığında büyük miktarda belleğe ihtiyaç duyulması KNN makine öğrenme modelinin dezavantajlarından biridir.

GBC, karar ağaçları yapılarını güçlendirmek amacıyla geliştirilen güçlü bir makine öğrenme modelidir. Bu modelde de her veri seti için ayrı bir karar ağacı yapısı oluşturulmaktadır. Dallanmalarda bir sonraki ağaca geçmeden önce önceki ağaçların hata miktarları hesaplanır ve değerlendirilir. Oluşan hata oranını en aza indirerek ağaç oluşturulmaya devam edilerek bir sınıfa atama gerçekleştirilir (Freidman, 2001).

ET, karar ağacı temelli makine öğrenme modeli yaklaşımlarından bir tanesidir. Bu modelde veri setinde öznitelikler ve düğümlerin kesim noktası belirlenirken rastgele seçim yapılır. Bu modelde oluşturulan ağaçlar da öğrenme çıktılarından bağımsız olarak rastgele oluşturulur. Rastgelelik çeşitli parametrelerin seçimi ve optimizasyonu ile güçlü şekilde işleyerek en iyi sonuca ulaşmak amaçlanmaktadır (Geurts ve diğerleri, 2006).

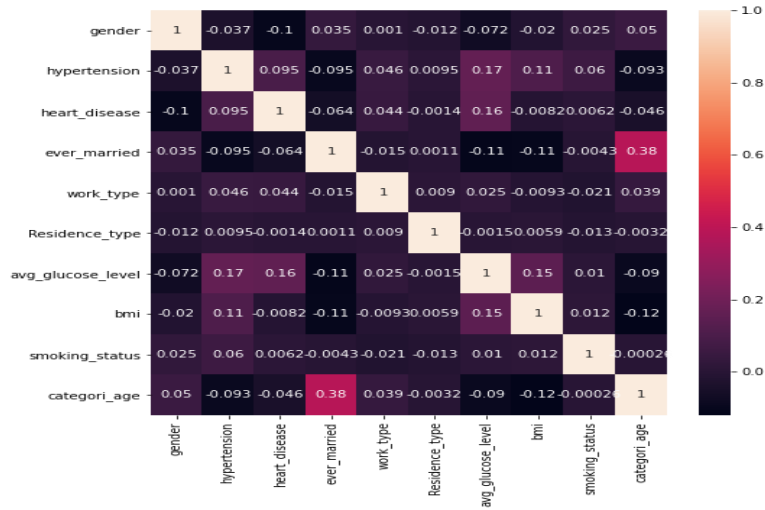
MLP, en yaygın yapay sinir ağı topolojilerinden bir tanesidir. MLP'de nöronlar arasında farklı ağırlık değerlerine sahip ileri beslemeli bağlantılar oluşturulur. Bu birbirine bağlı nöron bağlantıları ile girdi değerleri ve çıktı değerlerine göre katmanlar arasında hesaplamalar gerçekleştirilerek sonuçta bir hata oranı hesaplanır. Hata oranı en düşük seviyeye gelen kadar bu işlem ve hesaplama devam ederek eğitim tamamlanmış olur (Uzun ve diğerleri, 2018).

DNN, son yıllarda üstün sınıflandırma kabiliyeti nedeniyle karmaşık ve zor sınıflandırma problemlerinde sıklıkla tercih edilmektedir (Badem ve Baştürk, 2017; Yüksel ve diğerleri, 2018). DNN, bir girdi katmanı, çok sayıda gizli katman ve bir çıktı katmanında oluşan yapılardır. Bu katmanlarda nöronlar arasında karmaşık hesaplamalar gerçekleştirilerek sınıflama işlemi gerçekleştirilmektedir. Çalışmamızda 64 nöronlu dense tipinde 1 gizli katmandan oluşan, sigmoid aktivasyon ve adam optimizasyon yöntemi ile oluşturulmuş sıralı tam bağlı bir DNN modeli önerilmiştir.

3. Yöntem

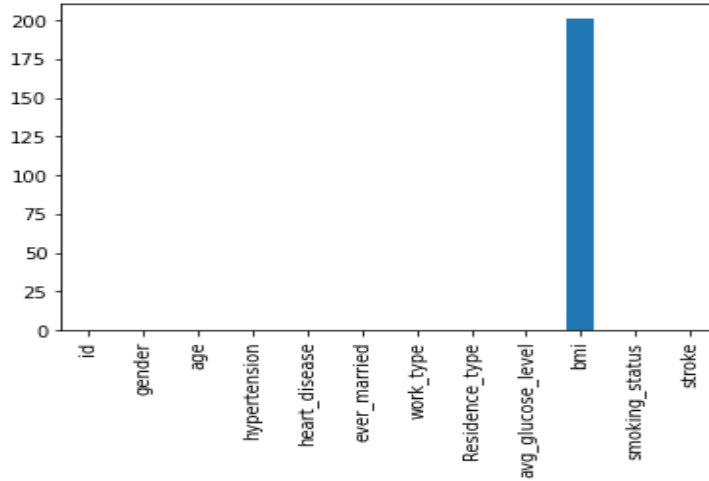
Makine öğrenmesi tekniklerinin sağlık, örüntü sınıflandırma ve sinyal sınıflandırma gibi birçok alanda başarılı sonuçlar verdiği bilinmektedir. Makine öğrenmesi teknikleri ile bir problemin giriş ve çıkış verileri arasındaki ilişki modellenmektedir. Makine öğrenme teknikleri ile giriş ve çıkış verileri arasındaki ilişkinin belirlenmesiyle çeşitli hastalıkların risk durumu tespitleri yapılabildiği gibi hastalıkların erken teşhisi de yapılabilmektedir.

Bu çalışmada öznitelik sayısı 11 olan ve 5110 örneklem içeren bir veri seti üzerinde 13 adet makine öğrenmesi tekniği ile felç riski sınıflandırılması gerçekleştirilmiştir. Öznitelikler arasındaki korelasyonu gösteren grafik Şekil 1'de görülmektedir.



Şekil 1. Korelasyon matrisi

Veri setinde bulunan kayıp veriler makine öğrenme modellerinin hatalı sonuçlar üretmesine ya da modellerin başarılarının düşük ya da yüksek bulunmasına neden olabileceğinden dolayı mevcut kayıp veriler veri setinden çıkarılmıştır. Şekil 2 incelendiğinde veri setinde sadece vücut kitle indeksi (BMI) verilerinde kayıp verilerin bulunduğu sonucuna ulaşılmaktadır.



Şekil 2. Kayıp veriler

Veriler arasındaki büyük değer farklarını gidermek için veri ön işleme aşamasında normalizasyon işlemi uygulanmıştır. Sınıflandırma etiketi verilerinin dengesiz dağılımının meydana getirebileceği olumsuz durumların önüne geçebilmek için sayıca az olan sınıflandırma verileri aşırı örnekleme (oversampling) yöntemiyle sayıları artırılarak dengeli bir sınıflandırma etiket dağılımı elde edilmiştir. Veri ön işleme aşamasından sonra makine öğrenme modelleri 10 kat çaprazlama tekniği ile koşturularak felç riski sınıflandırma işlemi gerçekleştirilmiştir. Modellerin başarımları, çeşitli ölçütlere göre karşılaştırılarak raporlanmıştır.

3.1. Veri Seti:

Bu çalışmada Kaggle veri deposunda yaş, cinsiyet, sigara içme durumları gibi özniteliklere göre 5110 adet bireyin kayıtlarını içeren "Stroke Prediction Dataset (2021)" veri seti kullanılmıştır. Veri seti felç geçirme risk sınıflandırması ve tahmini için kullanılan 10 adet girdi ve felç geçirme durum etiketlerini belirten 1 adet çıktı olmak üzere toplam 11 adet öznitelikten oluşmaktadır. Bu öznitelikler ve bu özniteliklere ait açıklamalara aşağıda yer verilmiştir.

Yaş: Bu öznitelik bir bireyin yaş bilgisinin içermektedir. Sayısal türde verilerden oluşmaktadır.

Cinsiyet: Bu öznitelik bir bireyin cinsiyeti bilgisini içermektedir. Kategorik türde verilerden oluşmaktadır.

Hipertansiyon: Bu öznitelik bir bireyin hipertansiyon rahatsızlığı olup olmadığı bilgisini içermektedir. Sayısal türde verilerden oluşmaktadır (1 – Hipertansiyon var, 0 – Hipertansiyon yok).

İş Türü: Bu öznitelik bireyin meslek bilgilerini içermektedir (Devlette çalışıyor, çalışmıyor, serbest meslek vb.). Kategorik türde verilerden oluşmaktadır.

Kalp Hastalığı: Bu öznitelik bir bireyin kalp hastalığı olup olmadığı ile ilgili bilgisini içermektedir (0 – Kalp hastalığı yok, 1 – kalp hastalığı var). Sayısal türde verilerden oluşmaktadır.

Evlilik Durumu: Bu öznitelik bir bireyin evlilik durumu bilgisini içermektedir (Evet-Hayır). Kategorik türde verilerden oluşmaktadır.

Yaşanılan Bölge: Bu öznitelik bir bireyin yaşadığı bölge bilgisini içermektedir (Kent-Kırsal). Kategorik türde verilerden oluşmaktadır.

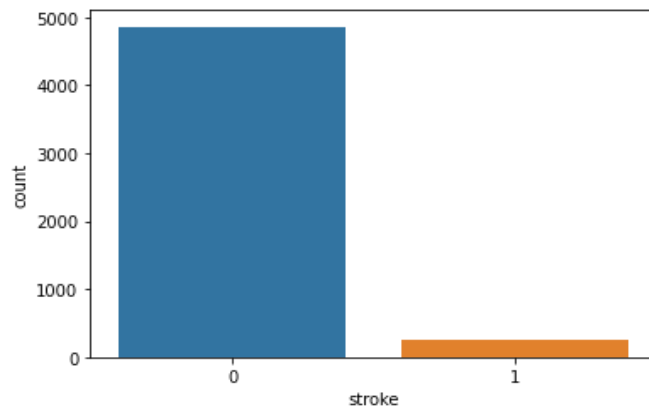
BMI: Bu öznitelik bir bireyin vücut kitle indeksi bilgisini içermektedir. Sayısal türde verilerden oluşmaktadır.

Ortalama Glikoz Seviyesi: Bu öznitelik bir bireyin kanındaki ortalama glikoz seviyesi bilgisini içermektedir. Sayısal türde verilerden oluşmaktadır.

Sigara Kullanma Durumu: Bireyin sigara kullanıp kullanmadığı bilgisini içeren verilerden oluşmaktadır (Daha önce sigara kullandı-Hala kullanıyor-Hiç kullanmamış). Kategorik türde verilerden oluşmaktadır.

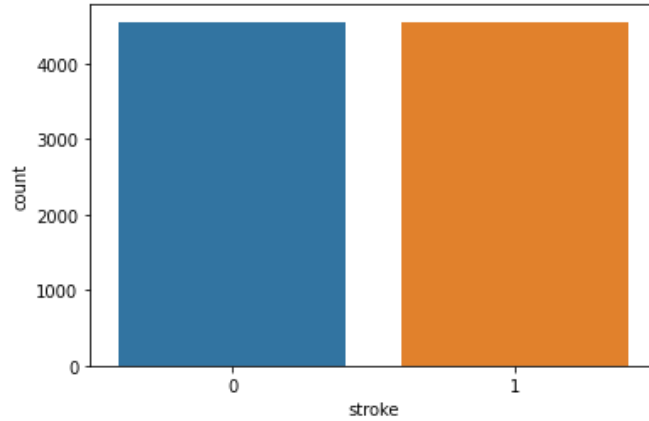
Felç Durumu: Bu öznitelik bir bireyin daha önce felç geçirip geçirmediği bilgisini içermektedir (1-Felç geçirdi, 0-Felç geçirmedi). Sayısal türde verilerden oluşmaktadır.

Veri seti incelediğinde sınıflandırma etiketi bilgisini içeren verilerin dengesiz dağılım gösterdiği görülmektedir. Veri setinin kayıp verilerden arındırılmasının ardından oluşan sınıf etiketi dengesiz dağılımı Şekil 3'te gösterilmiştir.



Şekil 3. Dengesiz dağılım grafiği

Dengesiz dağılımın meydana getirebileceği olumsuz durumlarla başa çıkabilmek için veri setine veri önileme adımında yeniden örnekleme yöntemlerinden olan aşırı örnekleme (oversampling) işlemi uygulanmıştır. Bu işlemle çeşitli performans ölçütlerine ait sonuçların daha dengeli ve birbirleriyle uyumlu olmalarını sağlamak amaçlanmıştır. Az sayıda olan sınıf etiketine ait toplam verilerin bir bölümü kopyalanarak bu sınıfa ait veri sayısı artırılarak etiket sayıları ve veriler dengelenmiştir. Şekil 4'te aşırı örnekleme işlemi sonucunda oluşan sınıf etiket dağılımları gösterilmiştir.



Şekil 4. Aşırı örnekleme sonucu oluşan dağılım grafiği

3.2. Karışıklık-Hata Matrisi (Confusion Matrix):

İkili sınıflandırma problemlerinde modelin performansını daha doğru bir şekilde değerlendirebilmek için karışıklık-hata matrisinden faydalanılmaktadır (Badem, 2019). Örnek bir karışıklık matrisi tablosu Tablo 1'de gösterilmektedir. Karışıklık matrisinde satırlar tahmin edilen sınıf değerlerini gösterirken sütunlar gerçek sınıf değerlerini göstermektedir. Karışıklık matrisinde ifade edilen TP ve TN doğru şekilde sınıflandırılan pozitif ve negatif değerlerin sayısını ifade etmektedir. FP ve FN ise yanlış şekilde sınıflandırılan pozitif ve negatif değerlerin sayısını göstermektedir. Karışıklık matrisindeki verilerden makine öğrenme tekniklerinin başarımlarını çeşitli başarımların ölçütleri kolayca hesaplanabilmektedir (Hossin ve Sulaiman, 2015; Badem, 2019).

Tablo 1. Örnek karışıklık-hata matrisi yapısı

	Tahmin edilen pozitif etiket (+)	Tahmin edilen negatif etiket (-)
Gerçek pozitif etiket (+)	TP	FN
Gerçek negatif etiket (-)	FP	TN

3.3. Performans Karşılaştırma Ölçütleri

Makine öğrenme tekniklerinin performanslarını birbirleriyle kıyaslayabilmek için farklı başarımların karşılaştırma ölçütlerinden yararlanılmaktadır. Bu çalışmada da makine öğrenme tekniklerinin sınıflandırma performanslarını karşılaştırabilmek için farklı ölçütlerden faydalanılmıştır. Doğruluk skoru (Accuracy-acc), kesinlik skoru (Precision-p), duyarlılık skoru (Recall-r) F Skoru (F measure - f) ve AUC (Area Under Curve-AUC) skoru olmak üzere 5 farklı karşılaştırma ölçütü performans kıyaslaması yapabilmek için belirlenmiştir. 10 kat çaprazlama tekniği ile eğitilen bütün modellerin ortalama skorları

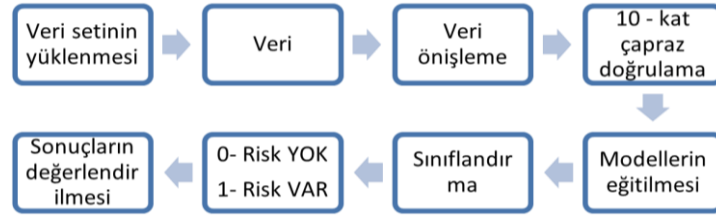
ve standart sapma deęerleri raporlanarak modellerin performanslarının deęerlendirilmesinde birer ölçüt olarak kullanılmıřtır. Tablo 2’de beř farklı performans karřılařtırma ölçütüne yönelik açıklamalara yer verilmektedir (Hossin ve Sulaiman, 2015; Badem, 2019).

Tablo 2. Performans karřılařtırma ölçütleri

Ölçüt	Açıklama	Formül
Doęruluk Skoru (Accuracy- acc)	Doęruluk skoru, doęru tahminlerin tahmin edilen tüm deęerlere oranını ifade etmektedir.	$\frac{TP + TN}{TP + TN + FP + FN}$
Kesinlik Skoru (Precision – p)	Kesinlik skoru, gerçek deęeri pozitifken pozitif olarak sınıflandırılan deęer sayısının bütün pozitif deęere sınıflandırılanların sayısına oranını ifade etmektedir.	$\frac{TP}{TP + FP}$
Duyarlılık Skoru (Recall – r)	Duyarlılık skoru, gerçek deęeri pozitifken pozitif olarak sınıflandırılan deęer sayısının bütün gerçek pozitif deęer sayısına oranını ifade etmektedir.	$\frac{TP}{TP + FN}$
F Skoru (F measure – f)	Kesinlik skoru ile duyarlılık skorunun harmonik ortalamasını ifade etmektedir. F skoru iki ölçütü birleřtirerek tek bir ölçüt olarak ifade etmekte kullanılmaktadır.	$2 \times \frac{p \times r}{p + r}$
AUC Skoru (Area Under Curve)	AUC, ROC eęrisinin altında kamıř olan alanı ifade etmekte kullanılmaktadır.	

3.4. Önerilen Yöntem

Bu çalıřmada felç riski sınıflandırması gerçekleřtirilmek amacıyla bir model önerilmektedir. Bu amaçla önerilen modelde “Stroke Prediction Dataset” veri setinde yer alan veriler ile LR, CART, RFC, SVM, KNN, NB, MLP, BNB, GBM, AB, LGBM, ET ve DNN makine öęrenmesi teknikleri ile sınıflandırma işlemleri gerçekleřtirilmiřtir. Her bir modelin sınıflandırma performansı çeřitli ölçütlere göre karřılařtırılmıřtır. Modellerin performanslarının karřılařtırılarak deęerlendirilmesinden sonra felç riskini en iyi sınıflandıran algoritma belirlenerek raporlanmıřtır. Çalıřmamızda önerilen model Şekil 5’te sunulmaktadır.



Şekil 5. Önerilen model

4. Deneysel Sonuçlar

Bu çalışmada LR, CART, RFC, SVM, KNN, BNB, GNB, MLP, GBM, ABC, LGBM, ET ve DNN olmak üzere 13 farklı sınıflandırıcı makine öğrenme modeli kullanılarak bireylerin sağlık verilerinden elde edilen değerlere göre felç geçirme riski belirlenmiştir. Tüm modeller 10 kat çapraz doğrulama tekniği ile eğitilerek acc, p, r, f, ve AUC skorlarına ait sonuçlar elde edilmiştir. Her bir modelin 10 kat çaprazlama ile eğitilmesinden elde edilen karşılaştırma ölçütlerine ait skorlarının standart sapmaları da elde edilerek raporlanmıştır. Makine öğrenmesi modellerine, ait sonuçlar, Google Colaboratory ortamında, 2.4 GHz Intel i5-6200U işlemciye ve 12 GB RAM belleğine sahip olan sistem üzerinden elde edilmiştir.

Deneysel sonuçlarda, felç riskini sınıflandırmak amacıyla 10 kat çapraz doğrulama tekniği ile eğitilen 13 farklı makine öğrenme modellerinden RFC sınıflandırıcısının 5 karşılaştırma ölçütünden tamamında diğer modellere göre daha üstün geldiği görülmektedir. Tablo 3'te tüm modellerin doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve AUC skoru değerleri görülmektedir. Tablo 3 incelendiğinde sırasıyla CART, ET, MLP, LGBM, KNN ve DNN sınıflandırıcılarının da felç riskini sınıflandırmada oldukça başarılı olduğu sonucuna ulaşılmaktadır.

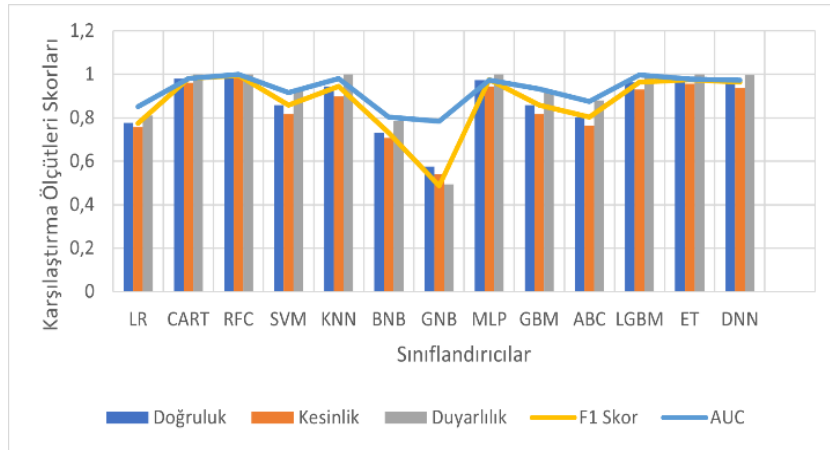
Tablo 3. Performans karşılaştırma ölçüt skorları

Model	acc	p	r	f	AUC
LR	0.77563	0.75811	0.81042	0.77530	0.85197
CART	0.98010	0.96098	1.0	0.98063	0.98053
RFC	0.99425	0.98804	1.0	0.99478	1.0
SVM	0.85882	0.81839	0.92276	0.85822	0.91544
KNN	0.94382	0.89924	1.0	0.94363	0.97957
BNB	0.73095	0.70818	0.78595	0.73008	0.80267
GNB	0.57574	0.54141	0.49340	0.48560	0.78392
MLP	0.97297	0.94371	1.0	0.97455	0.97415
GBM	0.85936	0.81721	0.92617	0.85870	0.93138
ABC	0.80319	0.76361	0.87872	0.80201	0.87461
LGBM	0.96265	0.93057	1.0	0.96260	0.99700
ET	0.97702	0.95749	1.0	0.97647	0.97872
DNN	0.96635	0.93664	0.99780	0.96460	0.97418

Deneysel sonuçlar felç riskini sınıflandırmak amacıyla 10 kat çapraz doğrulama tekniği ile eğitilen 13 farklı makine öğrenme modellerinden GNB sınıflandırıcısının beş karşılaştırma ölçütünden

dördünde diğer modellere göre daha başarısız olduğunu göstermektedir. GNB’de özellikle doğruluk, kesinlik ve F1 skoru değerleri incelendiğinde diğer sınıflandırıcılara göre oldukça geride sonuçlar elde edilmiştir. Bunun yanında LR ve BNB sınıflandırıcılarıyla GNB’ye göre nispeten daha başarılı sonuçlar elde edilmişse de diğer 10 sınıflandırıcıya göre GNB’nin felç riski sınıflandırma performansının en düşük olduğu sonucuna ulaşılmaktadır.

DeneySEL çalışmalarından elde edilen karşılaştırma ölçütlerine ait değerlerin daha kolay anlaşılabilmesi için Şekil 6 ‘da karşılaştırma ölçütlerinden elde edilen skorların çizgi-sütun grafiğine yer verilmiştir. Grafik incelendiğinde de felç riski sınıflandırmasında en başarılı modelin RFC olduğu ve en başarısız modelin ise GNB olduğu görülmektedir.



Şekil 6. Performans karşılaştırma ölçütleri skor grafiği

Tablo 4. Makine öğrenme modellerinin karşılaştırma ölçütlerine ait standart sapma değerleri

Model	acc_std	p_std	r_std	f_std	AUC_std
LR	0.01123	0.01621	0.01419	0.01137	0.00743
CART	0.00359	0.00552	0.0	0.00386	0.00306
RFC	0.00159	0.00436	0.0	0.00208	0.0
SVM	0.00807	0.01271	0.01051	0.00822	0.00571
KNN	0.00903	0.01459	0.0	0.00912	0.00440
BNB	0.01360	0.01301	0.02082	0.01364	0.00873
GNB	0.01521	0.00920	0.00200	0.02523	0.01018
MLP	0.00478	0.01124	0.0	0.00650	0.00751
GBM	0.00960	0.01254	0.01339	0.00971	0.00534
ABC	0.00788	0.01084	0.01401	0.00808	0.00675
LGBM	0.00464	0.00798	0.0	0.00467	0.00076
ET	0.00485	0.00843	0.0	0.00649	0.00393
DNN	0.00847	0.01411	0.00518	0.00864	0,00653

Sonuçların istatistiksel olarak daha anlamlı bir şekilde yorumlanabilmesi içinse Tablo 4’te 10 kat çaprazlama tekniği ile çalıştırılan her modele ait karşılaştırma ölçütlerinin standart sapma değerlerine

yer verilmiştir. İstatistikte standart sapma değerinin küçük olması verilerin ortalama değere daha yakın şekilde dağıldığı anlamına gelmektedir. Dolayısıyla bu çalışmada da standart sapma değerinin düşük olması 10 çapraz doğrulama tekniğinden elde edilen her bir karşılaştırma ölçütü değerinin ortalama değere daha yakın şekilde dağılım gösterdiği anlamına geldiğinden istenilen bir durumdur. Tablo 4 incelendiğinde RFC'nin beş adet karşılaştırma ölçütünün tamamında daha düşük standart sapma değeri elde ettiği görülmektedir.

Tablo 4 incelendiğinde GNB doğruluk, F1 skoru ve AUC skoru olmak üzere üç tanesinde, diğer modellere göre daha yüksek standart sapma değeri elde ettiği görülmektedir. BNB ve LR standart sapma değerleri GNB'ye göre nispeten daha iyi olsa da diğer on modele göre daha yüksek standart sapma değerlerine sahip oldukları görülmektedir.

Önerilen yöntemin sonuçlarının literatürden elde edilen aynı nitelikteki diğer çalışmalar ile karşılaştırılması önem arz etmektedir. Bu amaçla çalışmamızda Tablo 3 ve Tablo 4 birlikte değerlendirilerek felç riskini sınıflandırma işlemini 13 model arasından en iyi gerçekleştiren modelin RFC olduğu sonucuna ulaşılarak RFC modeli üzerinden karşılaştırmalar yapılmıştır. Tablo 5'te RFC modelinin sonuçlarıyla literatürden derlenmiş olan diğer modellerin sonuçları karşılaştırmalı olarak verilmiştir. Tablo 5 incelendiğinde önerilen RFC modelinin 0.99425 doğruluk değeriyle literatürdeki SVM, RFC, ANN, Weighted Voting, Two-class Boosted Decision Tree modellerine göre nispeten daha başarılı olduğu görülmektedir. Önerilen RFC'nin Tablo 5'te belirtilen diğer modellere göre ise oldukça başarılı olduğu görülmektedir.

Tablo 5. En başarılı modelin literatürdeki diğer modellerle karşılaştırılması

Sınıflandırıcı	Doğruluk	Açıklama
RFC (önerilen)	99.425%	10 – Kat Çapraz Doğrulama
SVM (Revanth ve diğerleri, 2020)	98.99%	Eğitim : %80, Test : %20
RFC (Sevli, 2021)	98.84%	10 – Kat Çapraz Doğrulama
Weighted Voting (Emon ve diğerleri, 2020)	97%	Eğitim ve Test verileri olarak ayrıldığı belirtilse de orana yer verilmemiştir.
XGB, GBC (Emon ve diğerleri, 2020)	96%	Eğitim ve Test verileri olarak ayrıldığı belirtilse de orana yer verilmemiştir.
Two-Class Boosted Decision Tree (Öznitelik Seçimi) (Ray ve diğerleri, 2020)	96.8%	Eğitim : %70, Test : %30
Two-Class Boosted Decision Tree) (Ray ve diğerleri, 2020)	97.6%	Eğitim : %70, Test : %30
Two-class Decision Jungle(Öznitelik Seçimi) (Ray ve diğerleri, 2020)	89%	Eğitim : %70, Test : %30
Two-class Decision Jungle) (Ray ve diğerleri, 2020)	96.7%	Eğitim : %70, Test : %30
ANN (Peng ve diğerleri, 2020)	98%	Eğitim : %70, Test : %15, Doğrulama : %15
RFC (Saleh ve diğerleri, 2019)	90%	10 – Kat Çapraz Doğrulama

5. Sonuç

Bu çalışmada felç riskinin belirlenmesinde makine öğrenme modellerinin başarımlarının analiz edilerek en iyi makine öğrenme modelini belirlemek amaçlanmıştır. Literatürde en sık kullanılan 13 farklı makine öğrenme modeli ile deneysel sonuçlara ulaşılarak modeller birbirleriyle farklı skorlar ve standart sapma değerleri üzerinden karşılaştırılmıştır. Deneysel sonuçlar neticesinde karar ağaçları temelli algoritmaların genelde başarılı sonuçlar verdiği gözlemlenmiştir. Modeller arasından en başarılı modelin RFC olduğu sonucuna ulaşılarak deneysel çalışma tamamlanmıştır. Ayrıca literatürde belirlenen aynı nitelikteki çalışmalarda yer alan rakip modellere göre çalışmada önerdiğimiz RFC modelinin daha başarılı olduğu belirlenmiştir.

Felç riskinin belirlenmesinde, veri seti üzerinde farklı ve daha detaylı ön işleme işlemleri gerçekleştirilerek yeni çalışmalar geliştirilebilir. Ayrıca gelecekte öznelik seçme, hiperparametre dönüşümü gibi işlemlerle modellerin başarımlarını artırmaya yönelik çalışmalar yapılabilir.

References

- Anusha M., Suresh K., Chandana M. (2021) Earlier Prediction on the heart disease based on supervised machine learning techniques. Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1696-1703. Madurai, India.
- Badem H. (2019) Parkinson Hastalığının Ses Sinyalleri Üzerinden Makine Öğrenmesi Teknikleri ile Tanımlanması. *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi* 8(2): 630-637.
- Badem H., Baştürk A., Çalışkan A. Yüksel M. E. (2017) A new efficient training strategy for deep neural networks by hybridization of artificial bee colony and limited-memory BFGS optimization algorithms. *Neurocomputing* 266: 506-526.
- Bayes T. (1763) An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions (1683-1775)* 53: 370-418.
- Berrar D. (2019) Cross-Validation. Encyclopedia of Bioinformatics and Computational Biology, Elsevier, pp. 542-545.
- Breiman L. (2001) Random Forests. *Machine Learning* 45: 5-32.
- Caplan L. R. (2016) Caplan's Stroke - A Clinical Approach, Cambridge University Press, p. 19.
- Cheon S., Kim J., Lim J. (2019) The Use of Deep Learning to Predict Stroke Patient Mortality. *International Journal of Environmental Research and Public Health* 16(11):1-12.
- Cortes C., Vapnik V. (1995) Support vector networks. *Machine Learning* 20(3): 273-297.
- Çalışkan M., Badem H. (2021) Makine Öğrenme Teknikleri Kullanılarak Epilepsi Teşhisi, M. Kalkancı, A. Günday (Ed.), Mühendislik Alanında Araştırma ve Değerlendirmeler, Gece Kitaplığı Yayınları, pp. 61-76.
- Dudani S. A. (1976) The Distance-Weighted k-Nearest-Neighbor Rule. *IEEE Transactions on Systems, Man, and Cybernetics* 6(4): 325-327.
- Emon M. U., Keya M.S., Meghla T. I., Rahman M. M., Al Mamun S., Kaiser M. S., (2020) Performance Analysis of Machine Learning. Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1464-1469. Coimbatore, India.
- Emre A., Çetiner M., Korkut Y. (2019) İnmeli Hastalarda Yaşam Kalitesi ve İlişkili Faktörler. *Turkish Journal of Family Medicine and Primary Care (TJFMPC)* 13(3): 103-111.

- Erkal B., Başak S., Ciloğlu A., Dede Şener D. (2020) Multiclass Classification of Brain Cancer with Machine Learning Algorithms. 2020 Medical Technologies Congress (TIPTEKNO), pp.1-4. Antalya, Turkey.
- Freidman J. H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29(5): 1189-1232.
- Freund Y., Schapire R. E. (1997) A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences* 55(1): 119-139.
- Geurts P., Ernst D., Wehenkel L. (2006) Extremely randomized trees. *Machine Learning* 63: 3-42.
- Hossin M., Sulaiman N. (2015) A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP)* 5(2): 1-11.
- Kaggle Stroke Prediction Dataset (2021). <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>. Erişim tarihi: 04 July 2021.
- Kaltes D., Morris T. (1994) Efficient Incremental Induction of Decision Trees. *Machine Learning* 24(3): 231-242.
- Peng C. C., Wang S. H., Liu S. J., Yang Y. K., Liao B. H. (2020) Artificial Neural Network Application to the Stroke Prediction. 2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), pp. 130-133.
- Ray S., Alshouli K., Roy A., AlGhamdi A., Agrawal D. P. (2020) Chi-Squared Based Feature Selection for Stroke Prediction using AzureML. 2020 Intermountain Engineering, Technology and Computing (IETC), pp. 1-6. Orem, UT, USA.
- Revanth S., Sanjay S., Sanjay N., Vijagayaganth V. (2020) Stroke Prediction using Machine Learning Algorithms. *International Journal of Disaster Recovery and Business Continuity* 11(1): 3081-3086.
- Rish I. (2001) An Empirical Study of the Naïve Bayes Classifier. IJCAI 2001 workshop on empirical methods 3: 41-46. IBM New York.
- Saleh H., Ghanny F.A., Younis E., Omran N., Abdelmgeid A. (2019) Stroke Prediction using Distributed Machine Learning Based on Apache Spark. *International Journal of Advanced Science and Technology (IJAST)* 28(15): 89-97.
- Sevli O. (2021) İnme (Felç) Riskinin Makine Öğrenmesi Kullanılarak Tespiti. 7. Uluslararası Mühendislik Mimarlık ve Tasarım Kongresi (7th International Congress on Engineering, Architecture and Design), pp. 661-667. İstanbul, Türkiye.
- Singh M. S., Choudhary P. (2017) Stroke Prediction using Artificial Intelligence. 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), pp. 158-161. Bangkok, Thailand.
- Song Y., Kong X., Huang S., Zhang C. (2021) Fast Training Logistic Regression via Adaptive Sampling. *Scientific Programming* 2021(2): 1-11.
- Truelsen T., Begg S., Mathers C. (2006) The Global Burden of Cerebrovascular Disease, Genova.
- Uzun R., ErKaymaz O., Şenyay Yapıcı İ. (2018) Comparison of Artificial Neural Network and Regression Models to Diagnose of Knee Disorder in Different Postures Using Surface Electromyography. *Journal of Science* 31(1): 100-110.
- WHO (2019) Global Health Estimates: Life expectancy and leading causes of death and disability, World Health Organization.
- WHO (2020) Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019., World Health Organization, Genova.

Yüksel M. E., Sarıkaya Baştürk N., Badem H., Çalışkan A., Baştürk A. (2018) Classification of high resolution hyperspectral remote sensing data using deep neural networks. *Journal of Intelligent & Fuzzy Systems* 34(4): 2273–2285.