ZAMAN SERİSİ VE SİNİR AĞLARI KULLANARAK MEME KANSERİ TEŞHİSİNDE MAKİNE ÖĞRENMESİ MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

YÜKSEK LİSANS TEZİ

Serkan TAŞPINAR

Danışman

Dr. Öğr. Üyesi Kerem GENCER

BİLGİSAYAR ANABİLİM DALI

Haziran 2025

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ZAMAN SERİSİ VE SİNİR AĞLARI KULLANARAK MEME KANSERİ TEŞHİSİNDE MAKİNE ÖĞRENMESİ MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

Serkan TAŞPINAR

Danışman
Dr. Öğr. Üyesi Kerem GENCER

BİLGİSAYAR ANABİLİM DALI

Haziran 2025

TEZ ONAY SAYFASI

Serkan TAŞPINAR tarafından hazırlanan "Zaman Serisi ve Sinir Ağları Kullanarak Meme Kanseri Teşhisinde Makine Öğrenmesi Modellerinin Karşılaştırılmalı Analizi" adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca 26 / 06 / 2025 tarihinde aşağıdaki jüri tarafından **oy birliği** ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **Bilgisayar Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman **:** Dr. Öğr. Üyesi Kerem GENCER

İmza

Başkan **:** Doç. Dr. Emre AVUÇLU
Aksaray Üniversitesi, Mühendislik Fakültesi …….

Üye **:** Dr. Öğr. Üyesi İnayet Hakkı ÇİZMECİ
Afyon Kocatepe Üniversitesi Mühendislik Fakültesi …….

Üye **:** Dr. Öğr. Üyesi Kerem GENCER
Afyon Kocatepe Üniversitesi Mühendislik Fakültesi …….

Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
..... /..... /..... tarih ve
……. sayılı kararıyla onaylanmıştır.

…….
Prof. Dr. Bekir YALÇIN
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

26 / 06 / 2025

İmza

Serkan TAŞPINAR

ÖZET

Yüksek Lisans Tezi

ZAMAN SERİSİ VE SİNİR AĞLARI KULLANARAK MEME KANSERİ TEŞHİSİNDE MAKİNE ÖĞRENMESİ MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

Serkan TAŞPINAR
Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Kerem GENCER

Meme kanseri, dünya genelinde kadınlar arasında en sık görülen kanser türlerinden biri olup, kansere bağlı ölümlerin başlıca nedenleri arasında yer almaktadır. Erken teşhis hem tedavi sürecinin etkinliği hem de hastaların yaşam süresi açısından kritik bir öneme sahiptir. Bu bağlamda, makine öğrenimi algoritmalarının tıbbi veri analizi ve teşhis süreçlerine entegre edilmesi, özellikle erken tanı açısından önemli katkılar sunmaktadır. Bu çalışma, meme kanseri teşhisinde kullanılan üç farklı makine öğrenimi modelinin Kolmogorov Arnold Network (KAN), Autoregressive Integrated Moving Average (ARIMA) ve Long Short-Term Memory (LSTM) performansını karşılaştırarak, bu alandaki önemli bir bilgi boşluğunu doldurmayı amaçlamaktadır. UCI Makine Öğrenimi Deposundan alınan Wisconsin Meme Kanseri ve Breast Cancer Coimbra veri setleri kullanılarak her bir modelin teşhis doğruluğu ve karmaşık veri yapılarını öğrenme yeteneği değerlendirilmiştir. KAN modeli, nöral ağların esnek yapısını temel fonksiyonlarla birleştirerek küçük ve karmaşık veri kümelerinde başarılı sonuçlar vermektedir. Özellikle biyomedikal veri setlerinde sunduğu esneklik, teşhis doğruluğunu artırma potansiyeline sahiptir. ARIMA modeli ise zaman serisi verilerinde yaygın olarak kullanılan bir yöntem olup, verilerin zamana bağlı değişkenliğini modelleyerek meme kanseri tanısında uygulanabilirlik sunmaktadır. LSTM modeli, uzun vadeli bağımlılıkları öğrenme yeteneği ile zaman serisi verilerinde öne çıkan bir derin öğrenme yöntemidir.

Elde edilen bulgular, KAN modelinin Mean Squared Error (MSE), Root Mean Squared Error (RMSE) ve Mean Absolute Error (MAE) gibi hata metrikleri açısından en düşük değerlere sahip olduğunu ve bu sayede en yüksek tahmin doğruluğunu sağladığını göstermiştir. LSTM modeli, mimarisinin karmaşıklığına rağmen bu çalışmada beklenen performansı gösterememiş; hata metrikleri açısından yüksek değerler üretmiş ve daha düşük bir başarı sergilemiştir. ARIMA modeli ise LSTM'ye kıyasla daha iyi performans göstermiş ancak KAN modelinin gerisinde kalmıştır. Sonuçlar, KAN modelinin meme kanseri teşhisinde güvenilir ve etkili bir araç olarak kullanılabileceğini ortaya koyarken, makine öğrenimi modellerinin dikkatli seçimi ve hiper parametre optimizasyonunun, tıbbi teşhis süreçlerinde kritik rol oynadığını da vurgulamaktadır. Gelecekte yapılacak çalışmaların, bu modellerin farklı veri setleri üzerinde test edilmesi, model parametrelerinin kapsamlı biçimde optimize edilmesi ve alternatif veri önişleme tekniklerinin kullanılmasıyla daha gelişmiş teşhis sistemlerinin geliştirilmesine katkı sağlaması beklenmektedir.

Anahtar Kelimeler: Meme kanseri teşhisi, Makine öğrenmesi modelleri , KAN, ARIMA, LSTM, Tahmin performansı.

ABSTRACT

M.Sc. Thesis

COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR BREAST CANCER DIAGNOSİS USING TIME SERIES AND NEURAL NETWORKS

Serkan TAŞPINAR

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Computer

Supervisor: Asst. Prof. Kerem GENCER

Breast cancer is one of the most common types of cancer among women worldwide and remains a leading cause of cancer-related deaths. Early diagnosis is critically important for both the effectiveness of treatment and patient survival. In this context, integrating machine learning algorithms into medical data analysis and diagnostic processes offers significant contributions, particularly in achieving early detection. This study aims to address a crucial gap in the literature by comparing the performance of three distinct machine learning models—Kolmogorov-Arnold Network (KAN), Autoregressive Integrated Moving Average (ARIMA), and Long Short-Term Memory (LSTM)—in the diagnosis of breast cancer. The models were evaluated using the Wisconsin Breast Cancer and Breast Cancer Coimbra datasets obtained from the UCI Machine Learning Repository, focusing on diagnostic accuracy and the ability to learn from complex data structures. The KAN model combines the flexible structure of neural networks with fundamental functions, demonstrating successful results on small and complex datasets. Its adaptability, especially in biomedical data analysis, offers potential for improving diagnostic accuracy. The ARIMA model, a widely used method for time series data, provides applicability in breast cancer diagnosis by modeling temporal variability in the data. The LSTM model, a deep learning approach known for its ability to learn long-term dependencies, stands out in time series analysis. The findings indicate that the KAN model achieved the lowest error values in terms of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), thereby delivering the

highest prediction accuracy. Despite its architectural complexity, the LSTM model did not perform as expected in this study; it produced higher error values and showed relatively lower success. The ARIMA model outperformed the LSTM but still fell short compared to the KAN model. These results suggest that the KAN model can be considered a reliable and effective tool in the diagnosis of breast cancer. Moreover, the study underscores the importance of carefully selecting machine learning models and optimizing hyperparameters, as they play a critical role in medical diagnostic processes. Future studies are expected to contribute to the development of more advanced diagnostic systems through testing these models on different datasets, comprehensive optimization of model parameters, and the application of alternative data preprocessing techniques.

Keywords: Breast cancer diagnosis, Machine learning models, Kolmogorov–Arnold Network (KAN), ARIMA, LSTM, Prediction performance.

TEŞEKKÜR

Bu araştırmanın yürütülmesinde, konunun belirlenmesinden sonuçların değerlendirilmesine ve tez metninin hazırlanmasına kadar her aşamada bilgi ve deneyimiyle bana yol gösteren, yönlendirmeleri ve değerli katkılarıyla çalışmama büyük destek veren tez danışmanım Dr. Öğr. Üyesi Kerem GENCER'e en içten teşekkürlerimi sunarım.

Ayrıca, eğitim hayatım boyunca bana her zaman destek olan, sabır ve anlayışlarıyla bu süreci kolaylaştıran sevgili eşime maddi ve manevi katkıları için sonsuz teşekkür ederim.

Serkan TAŞPINAR

Afyonkarahisar 2025

İÇİNDEKİLER DİZİNİ

Sayfa

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler	
KAN(x)	KAN giriş vektörüne karşılık gelen model çıktısı
f(x)	KAN kolmogorov temsiline göre çıktı fonksiyonu
$\phi_i, \psi_i,$	KAN Kolmogorov temsiline göre ara fonksiyonlar
y_t	LSTM çıkış değeri
W, b	LSTM Ağırlık matrisi ve bias terimi
x_t	LSTM zaman adımı t'deki giriş verisi
h_t	LSTM hücresindeki gizli durum (hidden state)
y_t	ARIMA zaman adımındaki gözlenen değer
ε	ARIMA hata terimi (white noise)
p,d,q	ARIMA model parametreleri (autoregressive, differencing, moving average)
C	ARIMA sabit terim
Kısaltmalar	
ANN	Artificial neural network (yapay sinir ağı)
AUC	Area under the curve (eğri altındaki alan)
ARIMA	AutoRegressive ıntegrated moving average
FN	False negative (yanlış negatif)
FP	False positive (yanlış pozitif)
KAN	Kolmogorov arnold network
LSTM	Long short-term memory (uzun kısa süreli bellek)
MAE	Mean absolute error (ortalama mutlak hata)
ML	Machine learning (makine öğrenmesi)
MSE	Mean squared error (ortalama kare hata)
RNN	Recurrent neural network (yinelemeli sinir ağı)
RMSE	Root mean squared error (kök ortalama kare hata)
ROC	Receiver operating characteristic (alıcı işlem eğrisi)
TN	True negative (doğru negatif)
TP	True positive (doğru pozitif)
CEC	Constant error carousel (sabit hata döngüsü)

ŞEKİLLER DİZİNİ

Sayfa

ÇİZELGELER DİZİNİ

Sayfa

1. GİRİŞ

Meme kanseri, dünya genelinde kadınlar arasında en sık rastlanan kanser türlerinden biri olup aynı zamanda kansere bağlı ölümlerin başlıca nedenleri arasında yer almaktadır (American Cancer Society 2020). Dünya Sağlık Örgütü'ne (DSÖ) göre her yıl yaklaşık 2.3 milyon kadına meme kanseri teşhisi konulmakta olup vaka sayıları artmaya devam etmektedir (WHO 2022). Kanser insidansı tahminlerine göre, meme kanseri akciğer kanserini geride bırakarak en yaygın kanser türü haline gelmiştir (Siegel vd. 2023). Erken evrede tespit edilen meme kanserinde tedavi başarısı %90 gibi yüksek oranlara ulaşmakta; orta evrelerde ise bu oran %50 ila %80 arasında değişmektedir. Geç evrede tanı konulan vakalarda ise tedavi süreci oldukça zorlaşmakta ve başarı oranı ciddi şekilde azalmaktadır. Bu nedenle, hastalığın zamanında teşhis edilmesi kritik bir öneme sahiptir (Verywell Health 2023).

Geleneksel olarak meme kanseri; fiziksel muayene, mamografi, ultrason ve biyopsi yöntemleri ile teşhis edilmektedir (American Cancer Society 2020). Mamografiyle birlikte uygulanan ultrason taramasının, özellikle yoğun meme dokusuna sahip kadınlarda, kanser tespit oranını artırdığı farklı çalışmalarda gösterilmiştir. Bu yöntemler, tanı sürecinde önemli rol oynamakta; ancak bazı sınırlamalara da sahiptir. Örneğin mamografi, genç kadınlarda yoğun meme dokusu nedeniyle yanlış negatif sonuçlara yol açabilmekte ve bu gibi durumlarda ultrason gibi tamamlayıcı yöntemlerin kullanımı tanı doğruluğunu artırabilmektedir (Skaane vd. 2008). Bununla birlikte, söz konusu tanı yöntemleri hem zaman alıcıdır hem de her zaman yeterli doğrulukta sonuç vermemektedir. Bu sebeple, tanı sürecinin hızlandırılması ve doğruluğunun artırılması amacıyla bilgisayar destekli teşhis sistemleri yaygın şekilde kullanılmaktadır (Doi 2007).

Makine öğrenmesi algoritmaları, büyük boyutlu veri setlerinden anlamlı örüntüler çıkararak, bu örüntüler üzerinden tahminler yapma yeteneğine sahiptir. Bu çalışmada meme kanseri teşhisine yönelik olarak üç farklı makine öğrenmesi modeli Kolmogorov Arnold Network (KAN), Autoregressive Integrated Moving Average (ARIMA) ve Long Short-Term Memory (LSTM) kullanılmıştır. KAN modeli, nöral ağların esnek yapısını çekirdek (kernel) yöntemlerle birleştirerek yüksek veri uyumu sağlamayı amaçlamaktadır

(Van Vaerenbergh vd. 2012).

Matematiksel olarak, KAN modeli herhangi bir sürekli fonksiyonu belirli bir doğruluk düzeyinde gösterip bir dizi fonksiyonun toplamı olarak ifade edebilmektedir. Bu özellik sayesinde, tıbbi verilerdeki karmaşık ilişkileri öğrenerek meme kanseri teşhisinde kullanılabilecek nitelikli sınıflandırmalar yapılabilmektedir. ARIMA modeli, zaman serisi analizlerinde yaygın olarak tercih edilen bir yöntem olup geçmiş verilere dayanarak gelecekteki değerlerin tahmini için kullanılmaktadır (Box vd. 2015). Bu yönüyle ARIMA, meme kanseri tanısında tıbbi verilerdeki zaman içerisindeki değişimleri takip ederek risk tahmini yapılmasına imkân sunmaktadır. LSTM modeli ise uzun vadeli bağımlılıkları öğrenme kapasitesiyle öne çıkan tekrarlayan bir sinir ağı yapısına sahiptir ve zaman serisi verilerinde başarılı sonuçlar vermektedir (Hochreiter ve Schmidhuber 1997).

LSTM modelleri, hastaların geçmiş tıbbi kayıtları ile tedavi sürecinde elde edilen görüntüleme verilerini analiz ederek, uzun dönemli örüntüleri tanımlayabilir; tümör hücrelerinin büyüme hızını tahmin edebilir ve daha önce fark edilemeyen anormal yapıları belirleyerek teşhis doğruluğunu artırabilir. Bu çalışmada kullanılan Wisconsin Meme Kanseri veri seti ile Breast Cancer Coimbra veri seti literatürde yaygın olarak kullanılan referans veri kümelerinden biridir (Wolberg vd. 1994, Patricio vd. 2018).

Araştırmamızda, KAN, ARIMA ve LSTM modelleri kullanılarak meme kanseri teşhisinde hangi modelin daha yüksek performans gösterdiği değerlendirilmiştir. Elde edilen bulgular doğrultusunda yapılan karşılaştırmalı analizler, makine öğrenimi tekniklerinin sağlık alanında daha etkin ve verimli biçimde kullanılabilmesi adına önemli çıkarımlar sunmaktadır (Topol 2019).

2. LİTERATÜR BİLGİLERİ

Literatürde, ARIMA ve LSTM modellerini karşılaştıran çok sayıda çalışma yer almaktadır. Bu çalışmaların temel amacı, her iki modelin farklı veri kümeleri üzerindeki performanslarını değerlendirerek hangi modelin, hangi problem türünde ve hangi veri setlerinde daha etkili olduğunu ortaya koymaktır (Gencer ve Başçiftçi 2021). Yapılan incelemeler, ARIMA ve LSTM modellerinin KAN ile birlikte değerlendirildiği herhangi bir çalışmaya rastlanmadığını göstermektedir. Bu eksiklik, KAN modelinin literatürde yeterince ele alınmadığını ortaya koymakta ve özellikle sağlık verileri üzerinde ARIMA ve LSTM gibi yaygın yöntemlerle karşılaştırılmadığını göstermektedir. Bu bölümde, ARIMA ve LSTM modellerini karşılaştıran çalışmalara yer verilmiş ve ilgili bulgular özetlenmiştir.

Martis vd. (2013), çalışmalarında LSTM ve ARIMA modellerini elektrokardiyogram (EKG) verilerindeki anormallikleri tahmin etmek amacıyla kullanmışlardır. LSTM modellerinin, EKG verilerindeki uzun vadeli bağımlılıkları daha başarılı şekilde tespit ederek, ARIMA modellerine kıyasla daha düşük hata oranları ile üstün performans sergilediği gözlemlenmiştir. Elde edilen bulgular, LSTM modelinin karmaşık ve uzun vadeli bağımlılıklar içeren veri türlerinde daha etkili olduğunu; özellikle EKG gibi klinik verilerin analizinde ARIMA'ya kıyasla daha başarılı sonuçlar verdiğini ortaya koymuştur. Ayrıca, LSTM modelinin anomali tespiti gibi zorlayıcı görevlerde ARIMA'ya göre daha üstün performans gösterdiği vurgulanmaktadır. Buna karşılık, ARIMA modelinin daha basit ve kısa vadeli tahmin gerektiren durumlarda etkili bir seçenek olabileceği belirtilmiştir. Dolayısıyla, EKG gibi karmaşık zaman serilerinde LSTM modelinin en yüksek performansı sergileyerek daha üstün bir alternatif olarak öne çıktığı söylenebilir.

Shahid vd. (2020) tarafından yürütülen çalışmada, hastanelerdeki hasta yoğunluğunu tahmin etmek amacıyla LSTM ve ARIMA modelleri karşılaştırılmıştır. LSTM modelinin, hasta sayısındaki ani değişimleri ve karmaşık yapıları daha etkili bir şekilde öğrenebildiği için ARIMA'ya kıyasla daha yüksek bir performans sergilediği rapor edilmiştir. Bulgular, LSTM'nin dinamik ve karmaşık veri içeren veri kümelerine daha iyi uyum sağladığını ve

bu sayede daha doğru tahminler üretebildiğini göstermektedir. Buna karşılık, ARIMA modeli daha sade yapılı ve doğrusal ilişkilerin baskın olduğu veri yapılarında etkili sonuçlar vermiştir. Dolayısıyla, bu çalışma LSTM modelinin karmaşık zaman serilerinde daha yüksek doğrulukla tahmin yapabildiğini; ARIMA modelinin ise daha basit ve doğrusal veriler için uygun olduğunu ortaya koymaktadır.

Chimmula ve Zhang (2020), COVID-19 vaka sayılarının tahmini amacıyla ARIMA ve LSTM modellerinin performanslarını karşılaştırmışlardır. Çalışmada, LSTM modellerinin daha düşük hata oranları ve daha yüksek doğruluk sağladığı tespit edilmiştir. Özellikle COVID-19 gibi belirsiz, karmaşık ve dinamik veri setlerinde, LSTM modelinin uzun vadeli bağımlılıkları öğrenme konusundaki yeteneğinin ARIMA'ya göre çok daha üstün olduğu ifade edilmiştir. Öte yandan, ARIMA modelinin daha düzenli ve kısa vadeli tahmin gerektiren veri setlerinde daha etkili olduğu belirtilmiştir. Bu bağlamda, LSTM modellerinin karmaşık ve öngörülemeyen veri türlerinde daha yüksek performans sergilediği söylenebilir.

Mougiakakou vd. (2010), diyabetli bireylerin kan şekeri düzeylerini tahmin etmek amacıyla ARIMA ve LSTM modellerini karşılaştırmışlardır. Çalışmada, LSTM modelinin ani değişimleri ve uzun vadeli bağımlılıkları daha iyi yakalayarak ARIMA'ya kıyasla daha yüksek doğruluk sunduğu belirlenmiştir. Elde edilen sonuçlar, LSTM modelinin kan şekeri düzeylerindeki karmaşık dalgalanmaları tahmin etmede daha etkili olduğunu ve bu bağlamda diyabetli bireylerin takibinde daha uygun bir yöntem olduğunu göstermektedir. Öte yandan, ARIMA modelinin daha düzenli ve sabit veri yapılarında anlamlı sonuçlar verdiği belirtilmiştir.

Kandemir-Cavas ve Cavas (2019), yoğun bakım ünitelerindeki (YBÜ) yatak doluluk oranlarını tahmin etmeye yönelik yaptıkları çalışmada, LSTM ve ARIMA modellerini karşılaştırmışlardır. Bulgular, LSTM modelinin karmaşık ve mevsimsel verileri daha iyi modelleyerek daha yüksek doğruluk oranlarına ulaştığını göstermektedir. Mevsimsel dalgalanmaları ve doğrusal olmayan yapıları daha iyi öğrenebilen LSTM modeli, bu bağlamda ARIMA modelinden üstün performans sergilemiştir. Buna karşın, ARIMA modeli kısa vadeli ve doğrusal yapıya sahip verilerde daha başarılı sonuçlar vermiştir.

Kırbaş vd. (2020), COVID-19 vaka sayılarını tahmin etmek amacıyla ARIMA, NARNN ve LSTM modellerini kullanarak Danimarka, Belçika, Almanya, Fransa, Birleşik Krallık, Finlandiya, İsviçre ve Türkiye gibi ülkelerden elde edilen veriler üzerinde bir karşılaştırma yapmışlardır. Çalışma bulgularına göre, LSTM modeli doğrusal olmayan yapıları daha etkili şekilde öğrenerek ARIMA'ya kıyasla daha yüksek performans sergilemiştir. Özellikle kısa vadeli tahminlerde LSTM modelinin daha başarılı olduğu, ARIMA modelinin ise daha sade ve doğrusal veri yapılarında etkili olduğu belirlenmiştir.

Literatürde ARIMA ve LSTM modellerini karşılaştıran çalışmaların bulguları, her iki modelin de farklı veri türleri ve yapılarında kendine özgü avantajlara ve sınırlılıklara sahip olduğunu ortaya koymaktadır. Ancak bu çalışmaların hiçbirinde ARIMA ve LSTM modellerinin KAN ile birlikte analiz edilip karşılaştırılmadığı görülmektedir. KAN modeli, doğrusal ve doğrusal olmayan fonksiyonları aynı yapı içerisinde temsil edebilme özelliği sayesinde, karmaşık veri yapılarında etkili sonuçlar sunabilmektedir. Bu nedenle, ARIMA ve LSTM modelleri ile birlikte KAN modelinin sağlık verileri üzerinde karşılaştırılması, literatürdeki önemli bir boşluğu dolduracak ve makine öğrenimi yaklaşımlarının performansını daha etkili ve kapsamlı biçimde değerlendirme imkânı sağlayacaktır.

3. MATERYAL ve METOT

3.1 Veri Seti

Wisconsin Meme Kanseri Teşhis (WBCD) veri seti, literatürde oldukça yaygın olarak kullanılan ve makine öğrenimi çalışmalarında referans ölçüt olarak kabul edilen kapsamlı bir veri setidir. Street vd. (1993) tarafından yapılan çalışmada bu veri seti kullanılarak çeşitli sınıflandırma algoritmalarının performansları karşılaştırılmıştır. Ayrıca Wolberg vd. (1994) de makine öğrenimi yöntemlerinin meme kanseri teşhisindeki etkinliğini değerlendirmek amacıyla bu veri setinden faydalanmıştır. UCI Makine Öğrenimi Deposunda yayımlanan bu veri seti, 569 örnekten oluşmakta ve her bir örnek, bir hastanın meme dokusundan alınan hücre bazlı ölçümlerine dayanmaktadır.

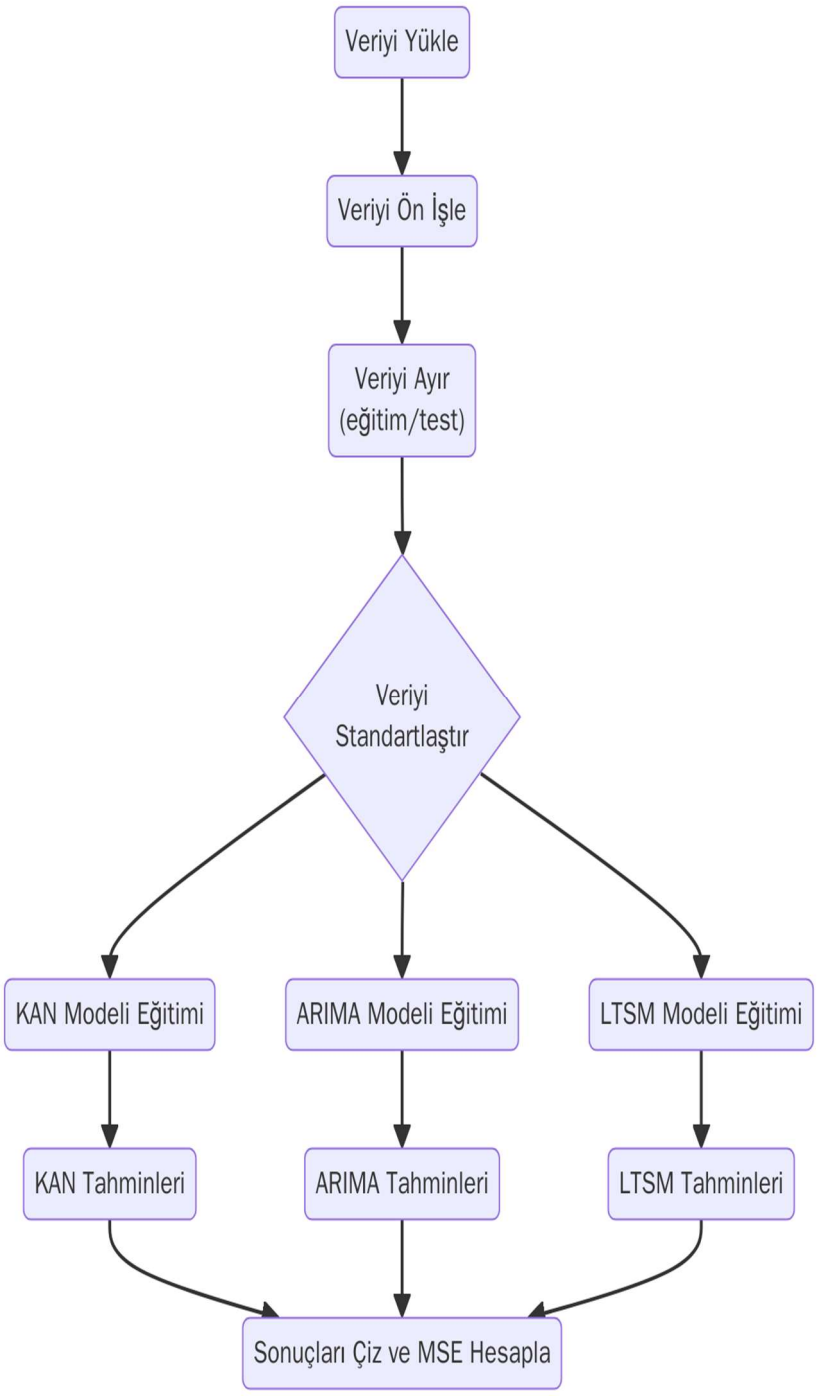
Veri setinde, her bir örnek iyi huylu (benign) ya da kötü huylu (malignant) olarak etiketlenmiş ve bu etiketler 'B' ve 'M' harfleri ile ifade edilmiştir. Özellikler arasında çap, doku, çevre, alan, düzgünlük, kompaktlık, oyukluk, oyuk noktalar, simetri ve fraktal boyut gibi parametreler yer almakta olup; bu parametreler ortalama, standart hata ve en kötü durum (worst) olmak üzere üç farklı şekilde ölçülmüştür. Bu çok boyutlu yaklaşım, hücrelerin kanserli olup olmadığını belirlemede yüksek doğruluk sağlar.

Breast Cancer Coimbra veri seti, Portekiz'deki Coimbra Üniversitesi'nden araştırmacılar tarafından oluşturulmuş ve UCI veritabanına dahil edilmiştir. Bu veri seti, geleneksel hücresel görüntü analizlerinden farklı olarak biyokimyasal göstergelere dayalı bir teşhis yaklaşımı sunar. Toplam 116 kadın hasta örneğini içeren veri setinde, yaş, vücut kitle indeksi (BMI), glikoz, insülin, leptin, adiponektin, resistin ve MCP-1 gibi biyobelirteç ölçümleri yer almaktadır.

Veri setinin hedef değişkeni, 0 (kanser değil) ve 1 (kanserli) olarak sınıflandırılmıştır. Küçük örneklem boyutuna rağmen, veri seti biyobelirteç tabanlı tanı sistemlerinin değerlendirilmesinde oldukça değerlidir ve özellikle daha karmaşık modellerin küçük veri üzerindeki genelleme yeteneğini test etmek için uygundur (Patricio vd. 2018).

Bu tez çalışmasında her iki veri seti, öncelikle eğitim ve test olmak üzere ikiye ayrılmış; %80 eğitim ve %20 test oranı kullanılmıştır. Özellikler, makine öğrenmesi modellerinin performansını artırmak amacıyla StandardScaler yöntemi ile normalize edilmiştir. Bu işlem, tüm özelliklerin aynı ölçekte olmasını sağlayarak öğrenme sürecini optimize eder. Elde edilen veriler sırasıyla KAN, ARIMA ve LSTM modelleri üzerinde eğitilmiş ve her veri seti için tahminler elde edilmiştir.

Şekil 3.1'de, veri kümesinin yüklenmesi, ön işlenmesi, eğitim ve test alt kümelerine ayrılması ile başlayan süreç; model eğitimi ve tahmin çıktılarının değerlendirilmesiyle devam eden genel iş akışı görselleştirilmiştir. Bu yapı, farklı veri setlerinin birbirine göre güçlü ve zayıf yönlerini analiz etmek ve model karşılaştırmalarını daha anlamlı kılmak için tasarlanmıştır. Veri setleri, meme kanseri teşhisi üzerine yapılan makine öğrenimi çalışmalarında oldukça faydalı bir kaynaktır ve farklı modellerin performanslarını karşılaştırmak için uygun bir temel sağlar (Street vd. 1993). Şekil 3.1'de, bu tez çalışmasında uygulanan veri işleme, model eğitimi ve tahmin sürecine ilişkin genel iş akışı görselleştirilmiştir. Süreç, veri kümesinin yüklenmesi ve ön işlenmesiyle başlamakta; ardından verinin eğitim ve test alt kümelerine ayrılmasıyla devam etmektedir. Bu veriler üzerinde sırasıyla KAN, ARIMA ve LSTM modelleri eğitilmekte ve tahminler elde edilmektedir.



Şekil 3.1 ARIMA, LSTM ve KAN modelleriyle tahmin süreci akışı

3.2 Kullanılan Modeller ve Eğitim Parametreleri

Bu çalışmada meme kanseri teşhisine yönelik sınıflandırma problemini çözmek amacıyla üç farklı model kullanılmıştır: Kolmogorov-Arnold Network (KAN), ARIMA ve LSTM. Her bir modelin yapısı, eğitim süreci ve parametre ayarları aşağıda detaylı şekilde

açıklanmıştır.

İlk olarak kullanılan KAN (Kolmogorov-Arnold Network) modeli, özellikle fonksiyon öğrenme ve ayrık sınıflandırma problemlerinde yüksek başarı gösteren yeni nesil bir yapay sinir ağı mimarisidir. Bu model PyTorch tabanlı kan kütüphanesi kullanılarak uygulanmıştır. Modelin yapısında giriş katmanında 30 nöron, gizli katmanda 5 nöron ve çıkış katmanında 1 nöron bulunmaktadır (width=[30, 5, 1]). Grid sayısı 5 olarak belirlenmiş, her bir düğümde 3 komşuya dayalı bir yapı (k=3) tercih edilmiştir. Optimizasyon işlemi, L-BFGS algoritmasıyla gerçekleştirilmiş; 50 adım boyunca eğitim yapılmıştır. Ayrıca, ağırlıkların kontrol altında tutulması amacıyla 0.01 değeriyle L2 cezası (lamb=0.01) ve 10 değeriyle entropi cezası (lamb_entropy=10) uygulanmıştır. Eğitim ve test verileri PyTorch tensörlerine dönüştürülerek modele aktarılmıştır.

İkinci olarak değerlendirilen ARIMA (AutoRegressive Integrated Moving Average) modeli ise zaman serisi analizi için sıklıkla kullanılan bir yaklaşımdır. Ancak bu çalışmada, sınıflandırma problemlerine kıyasla sınırlı bir kapasiteye sahip olması nedeniyle yalnızca karşılaştırma amacıyla kullanılmıştır. ARIMA modelinde sadece hedef değişken (etiket) zaman serisi olarak ele alınmış ve özellikler dikkate alınmamıştır. Modelin parametreleri (p=5, d=1, q=0) şeklinde belirlenmiş, yani 5 gecikmeli otoregresif terim, birinci dereceden fark alma ve sıfır hareketli ortalama bileşeni kullanılmıştır.

Üçüncü olarak kullanılan LSTM (Long Short-Term Memory) modeli, özellikle sıralı veriler üzerinde güçlü öğrenme yeteneğine sahip, geri beslemeli bir yapay sinir ağıdır. Model Keras kütüphanesi ile oluşturulmuş ve ardışık zaman adımlarında öğrenim yapacak şekilde yeniden şekillendirilmiştir. LSTM modeli 50 nöronlu tek bir LSTM katmanı ve ardından 1 nöronlu bir çıkış katmanından oluşmaktadır. Girdi boyutu (1, 30) olarak belirlenmiş, aktivasyon fonksiyonu olarak ‘ReLU’ seçilmiştir. Modelin çıkışında sigmoid aktivasyon fonksiyonu kullanılarak 0-1 aralığında olasılık tahmini yapılmıştır. Eğitim sürecinde Adam optimizasyon algoritması kullanılmış, kayıp fonksiyonu olarak binary_crossentropy tercih edilmiştir. Eğitim 50 epoch boyunca gerçekleştirilmiştir.

3.3 Kolmogorov Arnold Ağı (KAN)

Derin öğrenmenin temel yapı taşlarından birisi olan Çok Katmanlı Algılayıcılar (MLP), yapay zekânın gelişimi sürecinde büyük rol oynamıştır. MLP, özellikle derin öğrenmenin ilk yıllarında görüntü işleme ve çeşitli tahmin çalışmalarında oldukça başarılı olmuştur. MLP ve benzeri geleneksel yapay sinir ağları, öğrenme süreçlerinde birçok parametreyi optimize hale getirerek girdiler ile çıktılar arasındaki ilişkileri öğrenir. Ancak bu süreç, modelin karar mekanizmasını insan tarafından doğrudan yorumlanabilir hale getirmekte yetersiz kalmaktadır. Bu durum kara kutu problemi olarak adlandırılmaktadır (Kolmogorov 1961).

Kara kutu problemi tanım olarak modelin belirli bir girdiden nasıl bir çıktıya ulaştığını detaylı bir şekilde açıklayamaması nedeniyle ortaya çıkan bir durum olarak özetlenebilir. Geleneksel yapay sinir ağlarının açıklanabilirlik konusundaki sınırlılıklarını aşmak amacıyla geliştirilen KAN modeli, öğrenme süreçlerini daha şeffaf bir hale getirmektedir. KAN modelinde, klasik yapay sinir ağlarının aksine aktivasyon fonksiyonları nöronlar (düğümler) yerine kenarlar (bağlantılar) üzerinde tanımlanır. Bu yapı, her bir bağlantının özel bir dönüşüm fonksiyonu öğrenmesini sağlar. Bu sayede, her bir bağlantının çıktıya olan etkisi öğrenilebilir aktivasyon fonksiyonları aracılığıyla modellenmektedir. Geleneksel yapay sinir ağlarında, her katmanda sabit bir aktivasyon fonksiyonu kullanılırken, KAN modelinde ise her bir bağlantı kendi dinamik aktivasyon fonksiyonunu öğrenmektedir (Kolmogorov 1961).

Bu öğrenmenin sonucunda ise model daha esnek olurken, aynı zamanda karar mekanizmasının daha kolay ve net bir şekilde yorumlanmasına imkân sağlar. Bu sayede, hangi girdinin nasıl işlendiği ve hangi bağlantıların sonuca ne ölçüde katkı sağladığı açıkça izlenebildiğinden, modelin işleyişini anlamak ve yorumlamak çok daha kolay hale gelmektedir. Bu da modelin daha güvenilir olmasını sağlamaktadır. Aynı zamanda KAN'ı geleneksel yapay sinir ağlarındaki kara kutu problemine karşı güçlü bir alternatif haline getirmektedir. Özellikle modelin neden ve nasıl karar verdiğinin önemli olduğu durumlarda bizlere büyük bir avantaj sağlamaktadır.

KAN ağları, veriye uyum sağlama (data fitting) işlevlerinde gösterdikleri yüksek

performans sayesinde makine öğrenmesi mimarileri arasında dikkat çekmektedir. Veriye uyum sağlama, makine öğrenmesi modelinin eğitim verilerindeki hataları en aza indirerek girişlere karşılık en doğru çıktıları üretmeyi öğrenmesidir. Bu süreçte modelin temel amacı, eğitildiği örnekleri mümkün olduğunca doğru bir şekilde yeniden üretebilmektir. Böylelikle, sistemin öğrenmiş olduğu yapı sayesinde benzer diğer örneklerle karşılaşma durumunda başarılı tahminler yapması beklenir. Bu durum, özellikle tıbbi teşhislerde oldukça hayati bir önem taşır. Klasik sinir ağları, bu süreçte genellikle düğümlere sabit aktivasyon fonksiyonları atayarak öğrenir. Ancak bu yaklaşım, karmaşık ve yüksek boyutlu veri setlerinde doğruluk ve yorumlanabilirlik açısından sınırlayıcı olabilmektedir.

KAN modeli, tam da bu noktada geleneksel modellerden farklılaşarak daha esnek ve güçlü bir alternatif olarak karşımıza çıkmaktadır. KAN mimarisi, öğrenme sürecini düğümler yerine kenar bağlantıları üzerinden gerçekleştiren bir yapıya sahiptir. Bu yapı sayesinde her bağlantı, kendi öğrenilebilir aktivasyon fonksiyonu aracılığıyla giriş ile çıkış arasındaki ilişkiyi ayrı ayrı modelleyebilir. Bunu bir örnek bazında nasıl çalıştığını anlatmak gerekirse, meme kanseri teşhisine yönelik olarak geliştirilen bir model düşünelim. Bu model, geçmişe ait hasta kayıtlarından oluşan bir veri kümesi ile eğitilmektedir. Söz konusu bu veri setinde tümör büyüklüğü, hücre yoğunluğu ve hastanın yaşı gibi giriş değişkenleriyle birlikte bu kişilere konulan tanı bilgisi de yer almaktadır.

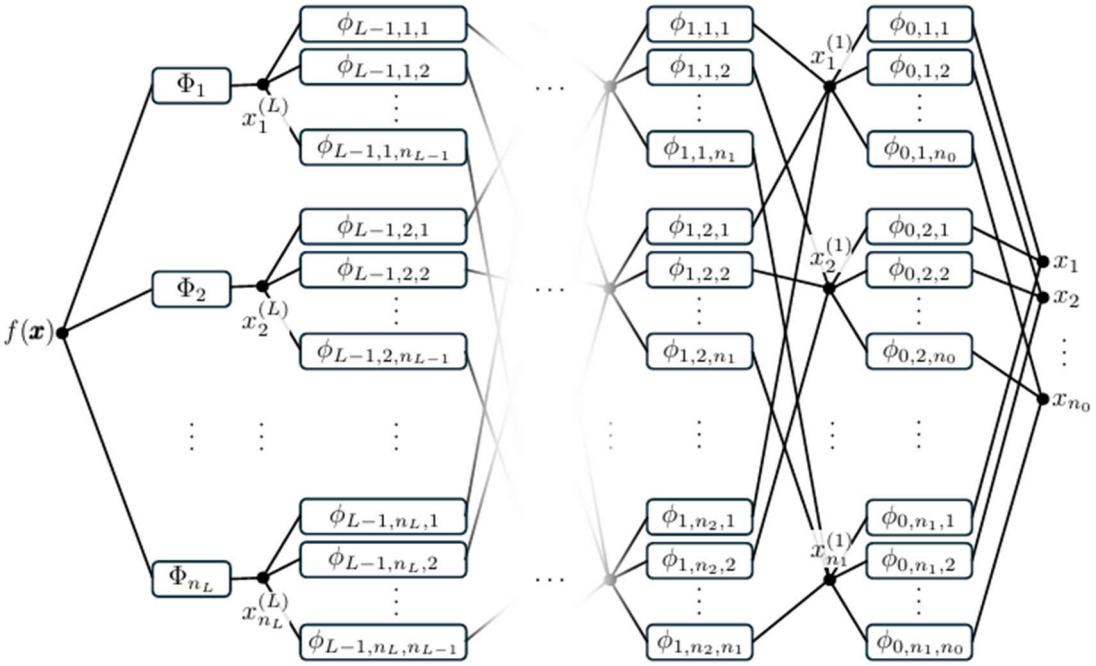
Burada tanı bilgisi, modelin doğru çıktıyı öğrenebilmesi için gerekli olan etiketi temsil etmektedir. Modelin eğitim aşamasındaki temel amacı, bu verileri öğrenerek, her bir hasta için doğru tanıyı yeniden tahmin edebilmektir. Örneğin, 2.1 cm çapında, hücre yoğunluğu yüksek ve 45 yaşındaki bir hastada kötü huylu bir tümör teşhisi/tanısı varsa, modelin bu örneği tekrar tekrar doğru bir şekilde tahmin etmesi beklenir. Model, eğitim verilerini doğru şekilde yeniden üretebilmenin yanında, benzer örnekler üzerinde de yüksek başarı göstererek genelleme yeteneğini ortaya koyar. Model, öğrenme sürecini tamamladıktan sonra, daha önce hiç görmediği fakat benzer özellikler taşıyan başka bir hasta verisiyle karşılaştığında, örnek olarak 2.5 cm tümör büyüklüğü, yüksek hücre yoğunluğu ve 48 yaşındaki bir hasta, doğru bir şekilde kötü huylu bir tümör teşhisi tanısı olarak tahmin edildiğinde modelin yalnızca ezber yapmadığını, aynı zamanda öğrenmiş olduğu yapıyı genelleştirebildiğini gösterir (Kolmogorov 1961).

KAN'ın bu özelliği sayesinde yalnızca yüksek doğruluk elde edilmez, aynı zamanda modelin karar mekanizması açıklanabilir ve yorumlanabilir hale gelir. Bu yönüyle KAN mimarisi, güçlü bir veriye uyum sağlama aracıdır. Kolmogorov Arnold Ağı (KAN), Kolmogorov Arnold temsil teoremi üzerine kurulmuş olan bir yapay sinir ağı modeldir. Bu teorem, 1957 yılında Andrey Kolmogorov tarafından gelişmiş olup, sürekli çok değişkenli fonksiyonların daha basit fonksiyonlarla ifade edilmesini sağlamayı amaçlar. Kolmogorov bu çalışmasında, her sürekli çok değişkenli fonksiyonun, sonlu sayıda sürekli tek değişkenli fonksiyonların bir birleşimi olarak ifade edilebileceğini öne sürmüştür (Kolmogorov 1961).

Kolmogorov'un öne sürmüş olduğu bu teoriden sonra Vladimir Arnold, Kolmogorov'un çalışmasını daha da genişletmiştir. Arnold 'un araştırması, Kolmogorov'un teoreminin evrenselliğini doğrulamış ve yüksek boyutlu fonksiyonların temsili konusundaki etkinliğini göstermiştir (Arnold 1957). KAN modeli, bu teorik temeller üzerine inşa edilmiş olup, her bir bağlantı üzerinde ayrı ayrı öğrenilen tek değişkenli fonksiyonlarla çalışır. Bu yapı sayesinde model, yüksek doğruluk elde etmenin yanı sıra, açıklanabilir ve esnek bir öğrenme süreci sağlar. KAN'ın esnek fonksiyon yapısı, zaman içerisinde değişen veri örüntülerine karşı duyarlılığını koruyabilmesini sağlar.

KAN modelleri, veri noktalarının etrafındaki yerel yapıları öğrenir ve bu bilgiyi nöral ağ üzerinden iletir (Liu vd. 2024). Bu modelin avantajlarından biri hem doğrusal olmayan hem de yüksek boyutlu veri setlerinde etkili olmasıdır. Çekirdek yöntemleri, veri noktalarını daha yüksek boyutlu bir uzaya yansıtarak bu tür verilerde etkili olmaktadır. Aynı zamanda uyarlanabilir öğrenme algoritmaları, zamanla değişen veri dağılımlarına hızlı bir şekilde uyum sağlamayı mümkün kılar (Chen vd. 2012). KAN modeli, çeşitli uygulama alanlarında başarıyla kullanılmaktadır. Özellikle zaman serisi analizi, biyomedikal sinyal işleme ve finansal veri analizi gibi alanlarda yüksek doğruluk ve genel performans göstermektedir (Zhao vd. 2019, Liu vd. 2024). Şekil 3.2'de, Kolmogorov-Arnold Ağı'nın (KAN) mimarisi şematik olarak sunulmaktadır. Bu yapıda, klasik yapay sinir ağlarından farklı olarak aktivasyon fonksiyonları düğümler (nöronlar) yerine kenar bağlantıları üzerinde tanımlanmıştır. Her bir bağlantı (edge), kendi öğrenilebilir

aktivasyon fonksiyonuna (ϕ) sahiptir ve bu özellik, giriş ve çıkış arasındaki ilişkilerin daha esnek ve özelleştirilebilir biçimde modellenmesini mümkün kılmaktadır (Liu vd. 2024).



Şekil 3.2 Kolmogorov Arnold Ağları

Kolmogorov teoremi, yukarıda da belirtildiği gibi herhangi bir sürekli çok değişkenli fonksiyonun, sonlu sayıda daha basit ve sürekli tek değişkenli fonksiyonların bileşimiyle ifade edilebileceğini ortaya koyar. Bu yaklaşım, karmaşık fonksiyonları daha yönetilebilir parçalara ayırarak analiz sürecini büyük ölçüde kolaylaştırır. Örnek olarak elimizde n-boyutlu sürekli bir fonksiyon olsun. Bu sürekli fonksiyon $f(x_1, x_2, \dots, x_{\Pi})$ için belirli bir yapıya sahip olan sürekli tek değişkenli fonksiyon kümeleri Φ_q ve φ_{qp} ile aşağıdaki gibi matematiksel olarak gösterilebilir:

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \varphi_{qp}(x_p) \right) \qquad (3.1)$$

Bu teoreme göre $f(x_1, x_2, \dots, x_n)$, n boyutlu çok değişkenli bir fonksiyon olup her bir giriş değişkeni (x_p) için tek değişkenli bir fonksiyon $\varphi_{qp}(x_p)$ kullanarak uygulanır. Bu

uygulama, her bir değişkenin bağımsız olarak analiz edilmesine imkân sağlar. Bununla beraber $\varphi_{qp}(x_p)$ fonksiyonları ise her bir giriş değişkenini işleyerek bu değişkenlerin çok boyutlu yapı içerisindeki etkilerini ayrıştırarak girdilerin aşağıdaki gibi bir ara toplam oluşturmasını sağlamaktadır:

$$\sum_{p=1}^{n} \varphi_{qp}\,(x_p) \tag{3.2}$$

Ara toplamlar, Φ_q olarak adlandırılan fonksiyonlara gönderildikten sonra her Φ_q sadece tek bir giriş alıp, aldığı bu giriş üzerinden aşağıdaki çıktıyı üretir:

$$\Phi_q\ \left(\sum_{p=1}^{n} \varphi_{qp}\,(x_p)\right) \tag{3.3}$$

Son olarak, bütün Φ_q fonksiyonlarının sonuçları toplanarak $f(x_1, x_2, \ldots, x_n)$ orijinal fonksiyonu elde edilir:

$$f(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q\ \left(\sum_{p=1}^{n} \varphi_{qp}\,(x_p)\right) \tag{3.4}$$

Liu vd. (2024) tarafından yapılan çalışmada aşağıda her bir Kolmogorov-Arnold Ağı (KAN) katmanı, giriş değişkenlerini farklı aktivasyon fonksiyonlarıyla eşlemiş olan bir fonksiyon kümesi genel formda

$$\Phi\ =\ \{\varphi_{q,p}\},\ \text{p = 1,2,…,}n_{in},\ \text{q= 1,2,…,}n_{out} \tag{3.5}$$

şeklinde tanımlanmaktadır. Buradaki $\varphi_{q,p}$ fonksiyonu giriş katmanındaki p. bileşenden çıkış katmanındaki q. bileşene giden dönüşüm fonksiyonlarını ifade eder (Liu vd. 2024). Kolmogorov Arnold teoremine göre, bir KAN katmanı iç katman olarak ele alındığında, giriş boyutu n_{in} = n, çıkış boyutu ise n_{out} = 2n+1 olacak şekilde tanımlanır. Dış katmanlarda ise bu oran n_{in} = 2n+1 ve n_{out} = 1 şeklindedir. Bu yapılar göz önüne alındığında, Kolmogorov-Arnold temsili, ardışık iki KAN katmanının fonksiyonel birleşiminde oluşur. Bu noktada daha derin bir temsile ulaşmak için yapılması gerekenin yalnızca katman sayısını artırmak olduğunu gösterir (Liu vd. 2024).

KAN mimarisinin yapısını tanımlarken, temel bir gösterim olarak şu tamsayı dizisi tanımlanır: $[n_0, n_1, \dots, n_L]$. Buradaki her n_i, i. katmanındaki nöron sayısını gösterir. Her bir nöron (l,i) biçiminde tanımlanır ve bu nöronun aktivasyon değeri $x_{l,i}$ ile ifade edilir. Ardışık iki katman olan l ve l+1 arasında $n_l n_{l+1}$ adet aktivasyon fonksiyonu bulunur. Bu fonksiyonlardan biri l. katmandaki i. nöron ile l+1. katmandaki j. nöron arasındaki bağlantıyı ifade eder ve aşağıdaki şekilde gösterilir (Liu vd. 2024).

$$\varphi_{l,j,i,} \quad \text{l=0, … , L-1,} \quad \text{i = 1, …, } n_l, \quad \text{j = 1, … , } n_{l+1} \tag{3.6}$$

$\varphi_{l,j,i}$ fonksiyonunun ön aktivasyonu doğrudan $x_{l,i}$' dir. Bu fonksiyonun çıktısı yani post-aktivasyonu ise $\tilde{x}_{l,j,i} \equiv \varphi_{l,j,i}(x_{l,i})$ olarak gösterilir. Bir j. nöronun l+1 katmandaki aktivasyon değeri, kendisine gelen tüm post aktivasyonların toplamı olarak hesaplanır(Liu vd. 2024).:

$$x_{l+1,j} = \sum_{i=1}^{n_l} \tilde{x}_{l,j,i} = \sum_{i=1}^{n_l} \varphi_{l,j,i}(x_{l,i}), \quad \text{j = 1, … ,} n_{l+1} \tag{3.7}$$

Bu denklem, matris formuna aşağıdaki gibi dönüştürülebilir:

$$x_{l+1} = \underbrace{\begin{bmatrix} \varphi_{l,1,1}(.) & \varphi_{l,1,2}(.) & \cdots & \varphi_{l,1,n_l(.)} \\ \vdots & & \ddots & \vdots \\ \varphi_{l,n_{l+1},1}(.) & \varphi_{l,n_{l+1},2}(.) & \cdots & \varphi_{l,n_{l+1},n_l(.)} \end{bmatrix}}_{\Phi_l} . x_l \tag{3.8}$$

Burada Φ_l, L. katmana ait bir fonksiyon matrisini ifade etmektedir. Genel anlamda bir KAN mimarisi, ardışık L katmanın bileşiminden oluşmaktadır. Girdi vektörü $x_0 \in \mathbb{R}^{n_0}$ olarak alındığında KAN'ın çıktısı şu şekilde ifade edilir (Liu vd. 2024):

$$\text{KAN(x)} = (\Phi_{L-1} \, o \, \Phi_{L-2} \, o \dots o \, \Phi_1 \, o \, \Phi_0) \tag{3.9}$$

Bu yapı, her katmanın çıktısını bir sonraki katmana girdi olarak aktarır ve fonksiyonel

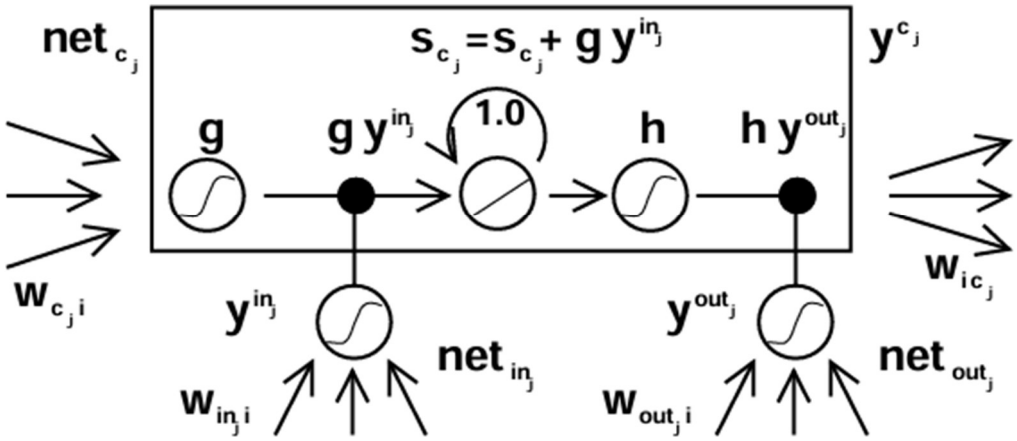
aktarım zinciri derinleştikçe ağın öğrenme kapasitesi de artar. Yukarıdaki ifadeyi klasik Kolmogorov temsiline daha benzer bir biçimde yazmak istersek, çıktı boyutunun $n_L = 1$ olduğu varsalılır ve f(x) = KAN(x) şeklinde tanımlama yapılır (Liu vd. 2024). Böylece:

$$\sum_{i_{L-1}=1}^{n_{L-1}} \varphi_{L-1,i_L,i_{L-1}} \left(\sum_{i_{L-2}=1}^{n_{L-2}} \cdots \left(\sum_{i_2=1}^{n_2} \varphi_{2,i_3,i_2} \left(\sum_{i_1=1}^{n_1} \varphi_{1,i_{2,i_1}} \left(\sum_{i_0=1}^{n_0} \varphi_{0,i_1,i_0} \right) \right) \right) \cdots \right) \tag{3.10}$$

şeklinde ifade edilebilir.

3.4 LSTM

LSTM, yinelemeli sinir ağları (RNN'ler), ardışık verilerle çalışmak üzere geliştirilmiş olan bir yapay sinir ağı türü olup geçmiş zamandaki verileri belleğinde tutarak, zaman serisi şeklinde ilerleyen veriler üzerinde anlamlı tahminlerde bulunma yeteneğine sahiptir. Ancak RNN mimarisinde, özellikle uzun dizilere sahip veri kümelerinde geçmiş bilgilerin taşınmasında gradyan sönmesi ya da kaybolan gradyan adı verilen literatürde vanishing gradient olarak geçen önemli bir problem ortaya çıkmaktadır. Bandar (2024), yapmış olduğu çalışmada LSTM ağları, geleneksel RNN'lerde sıklıkla karşılaşılan gradyan sönmesi problemini azaltmak amacıyla geliştirilmiş özel bir yinelemeli sinir ağı türüdür. LSTM'ler, bünyelerinde barındırdıkları ek bellek hücreleri ve kapı mekanizmaları sayesinde, bilgi saklama ya da unutma süreçlerinde gereken kararları alabilme yetkisine sahiptir. Bu özgün mimari yapı LSTM'leri ardışık verilerdeki uzun vadeli bağımlılıkları etkin bir şekilde yönetir (Bandar 2024). Şekil 3.3'te, Hochreiter ve Schmidhuber (1997) tarafından önerilen LSTM hücresinin temel yapısı gösterilmektedir. Bu yapıda, giriş kapısı (input gate), unutma kapısı (forget gate) ve çıkış kapısı (output gate) olmak üzere üç temel kapı mekanizması yer almaktadır. Bu kapılar, hücre durumundaki bilgilerin ne kadarının korunacağı, güncelleneceği ya da aktarılacağına karar verir. Hücre durumu s_c , geçmiş bilginin uzun süre saklanmasını sağlayarak geleneksel RNN'lerde karşılaşılan gradyan sönmesi (vanishing gradient) problemini önemli ölçüde azaltır. Şekilde gösterildiği üzere, her kapı bir sigmoid (σ) ya da tanh aktivasyonu ile kontrol edilmekte ve bu aktivasyonların çıktıları hücre durumuna etki etmektedir.



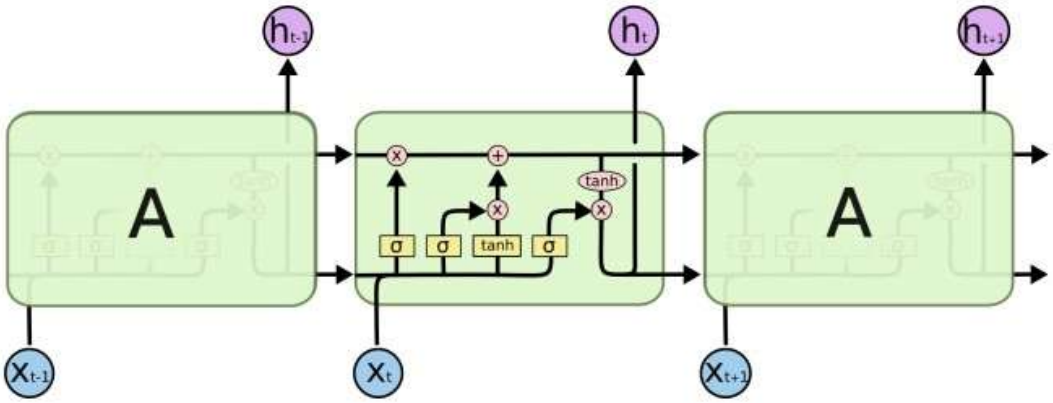
Şekil 3.3 LSTM hücresinin temel mimarisi

LSTM modeli, zaman serisi verilerinde uzun vadeli bağımlılıkları öğrenmede etkili performans gösteren özel bir tür RNN mimarisidir. LSTM modelleri, uzun süreli bağımlılıkları yakalama yetenekleri sayesinde; dil modelleme, konuşma tanıma ve finansal zaman serisi tahminleri gibi birçok alanda başarıyla kullanılmaktadır (Hochreiter ve Schmidhuber 1997).

LSTM modelinin temel yapı taşı olan LSTM hücresi, üç ana bileşen içerir: giriş kapısı (input gate), unutma kapısı (forget gate) ve çıkış kapısı (output gate). Bu kapılar, hücre durumu ile gizli durumu dinamik olarak güncelleyerek bilgilerin ne kadar saklanacağını veya unutulacağını belirler. Bu yapı, modelin önemli bilgileri uzun süre boyunca tutmasına olanak sağlarken, önemsiz bilgileri de etkin bir biçimde silebilmesini sağlar (Gers vd. 2000).

LSTM modeli, klasik RNN'lerin uzun vadeli bağımlılıkları öğrenmede karşılaştığı gradyan sönmesi (vanishing gradient) problemini aşmak amacıyla Hochreiter ve Schmidhuber tarafından 1997 yılında geliştirilmiştir. LSTM mimarisi, özellikle constant error flow (sabit hata akışı) mekanizması sayesinde, hata sinyalinin zaman boyunca bozulmadan hücre içinde aktarılmasını sağlar. Bu mekanizma, LSTM hücresindeki constant error carousel (CEC) olarak bilinen özel yapı aracılığıyla gerçekleştirilir. Böylece, uzun süreli bağımlılıkların etkili bir şekilde öğrenilmesi mümkün hale gelir.

Standart bir RNN'de tekrarlayan yapı yalnızca tek bir katmandan oluşurken, LSTM hücresi çok daha karmaşık bir yapıya sahiptir. LSTM'deki tekrarlayan modül, birbiriyle özel olarak etkileşim kuran dört alt bileşenden oluşur. Bu yapı, modelin daha karmaşık ilişkileri öğrenmesine olanak tanır. Söz konusu alt bileşenler ve hücrenin genel yapısı aşağıdaki şekilde şematik olarak gösterilmiştir.



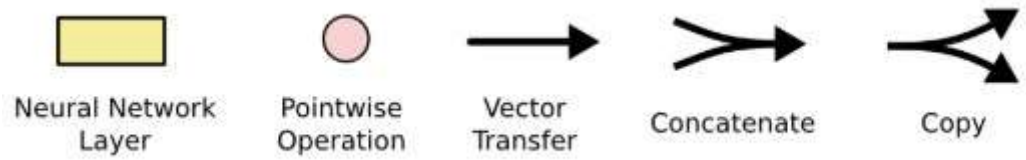
Şekil 3.4 LSTM'nin tekrarlayan yapıda iletişim halindeki dört etkileşimli katmanı

Şekil 3.4'te, LSTM hücresinin iç yapısını oluşturan ve birbirleriyle etkileşim hâlinde çalışan dört temel katman ayrıntılı olarak gösterilmektedir (Olah 2015). Bu yapı, klasik RNN mimarilerinden farklı olarak yalnızca ardışık bilgi akışına dayanan bir tekrar modülünden ibaret değildir. Bunun yerine, bilgi işleme, güncelleme ve aktarım süreçlerini birlikte yöneten bir mimari sunar.

Her LSTM hücresi; giriş kapısı, unutma kapısı, çıkış kapısı ve hücre durumu güncelleyici olmak üzere dört ana bileşeni içermektedir. Bu bileşenler sayesinde hücre, zaman boyunca taşıdığı bilgiyi koruma, gerektiğinde güncelleme veya silme yeteneğine sahiptir. Şekil 3.4'teki diyagram, bu bilgi akışını ve kapılar arası etkileşimi açık bir biçimde yansıtarak, LSTM'nin uzun vadeli bağımlılıkları öğrenme konusundaki başarısını görsel olarak ortaya koymaktadır.

Bu yapı, LSTM'nin geçmişe ait bilgiyi unutmadan uzun süre saklayabilmesini sağlayan kapı mekanizmalarının birlikte çalışmasını esas alır. Bu sayede model, dil işleme, zaman serisi tahmin gibi birçok alanda uzun vadeli bağlamı etkili şekilde

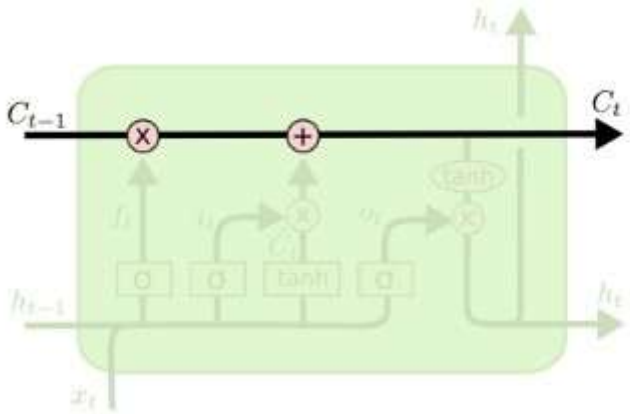
değerlendirebilmektedir.



Şekil 3.5 LSTM mimarisi içinde yer alan temel bileşenler ve sembollerin anlamları

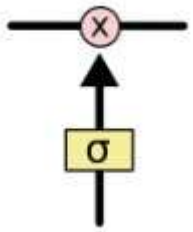
Şekil 3.5'te, LSTM ağlarında kullanılan temel yapısal semboller görsel olarak sunulmuştur (Olah 2015). Sarı dikdörtgenler, yapay sinir ağı katmanlarını; pembe daireler ise noktasal işlemleri (pointwise operations) temsil etmektedir. Vektör transferini gösteren sembollerdeki oklar, verinin akış yönünü belirtir. Kalın çizgili ok, bir düğümden çıkan veriyi; ince çizgili ok ise verinin başka bir düğüme girdiğini simgeler.

İki okun bir noktada birleştiği concatenate sembolü, iki farklı vektörün birleştirilerek aynı düğüme aktarımını ifade eder. Okların çatallandığı copy sembolü ise, bir verinin birden çok yere yönlendirilerek çoğaltıldığını gösterir. Bu semboller, LSTM mimarisinin hem yapısal işleyişini hem de hücre içi bilgi aktarım mekanizmasını kavramsal olarak sadeleştirmek için kullanılır. LSTM'nin bilgi işleme sürecinde kullandığı temel bileşenlerin diyagramlarda nasıl temsil edildiği açık biçimde anlaşılmakta ve bu yapı, özellikle hücre durumu (cell state) gibi kritik bileşenlerin zaman içerisindeki güncellenme süreçlerini tanımlamada kolaylık sağlamaktadır (Olah 2015).



Şekil 3.6 LSTM mimarisinde hücre durumunun ($\boldsymbol{C_t}$) bilgi akışındaki temel rolü

Şekil 3.6'da, LSTM hücresindeki hücre durumu bileşeninin C_t zaman boyunca bilgi taşıma sürecindeki rolü görselleştirilmiştir (Olah 2015). Bu yapı, bir tür taşıma bandı gibi düşünülebilir. Hücre durumu, zaman adımları boyunca büyük ölçüde doğrusal bir şekilde akar ve yalnızca sınırlı sayıdaki işlemden geçer. Bu özellik, bilgilerin bozulmadan aktarılmasına olanak sağlayarak uzun vadeli bağımlılıkların korunmasına katkıda bulunur. LSTM'nin en temel özelliklerinden biri, hücre durumu üzerinde bilgiyi ekleyebilme veya silebilme yeteneğidir. Bu işlemler, kapı (gate) adı verilen özel bileşenler tarafından kontrol edilmektedir. Giriş kapısı, unutma kapısı ve çıkış kapısı gibi mekanizmalar, hangi bilginin hücre durumuna yazılacağına, korunacağına veya silineceğine karar verir.



Şekil 3.7 LSTM hücresinde sigmoid fonksiyonunun çıktısının noktasal çarpım ile bilgi akışını denetlemesi

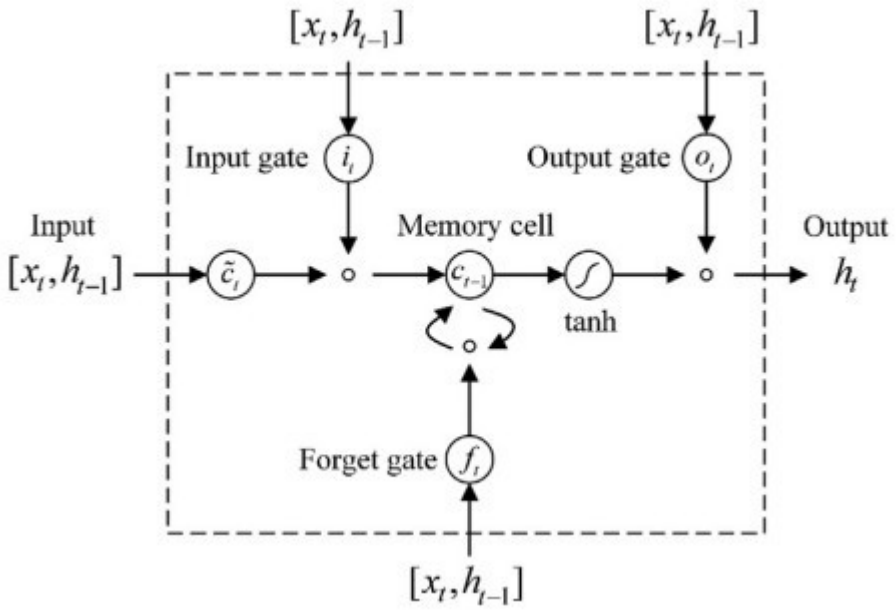
Şekil 3.7'de, LSTM hücresinde yer alan kapı (gate) yapılarında bilginin hücreye nasıl aktarıldığını denetleyen temel mekanizma görselleştirilmiştir (Olah 2015). Bu kapılar, hücreye hangi bilginin ekleneceğine, hangi bilginin tutulacağına ya da hangi bilginin silineceğine karar verir. Her kapı, bir sigmoid aktivasyon fonksiyonu tarafından kontrol edilen ve çıktı olarak [0, 1] aralığında değer üreten bir sinir ağı katmanıdır. Bu çıktı daha sonra noktasal çarpım (element-wise multiplication) işlemiyle hücre durumu veya başka bir vektörle çarpılır. Elde edilen sonuç, ilgili bilginin hücreye ne ölçüde aktarılacağını belirler. Sigmoid çıktısı 0'a yakınsa, bilgi aktarımı baskılanır (yeni bilgi geçmez). Sigmoid çıktısı 1'e yakınsa, bilgi tamamen geçer (izin verilir). LSTM hücresi içinde yer alan giriş, unutma ve çıkış kapıları, bu mekanizmayı kullanarak zaman adımları boyunca bilgi akışını kontrol eder.

- Giriş kapısı, yeni gelen bilginin hücreye aktarımına karar verir.

- Unutma kapısı, geçmiş bilgilerin hücre durumunda tutulup tutulmayacağını belirler.

- Çıkış kapısı ise hücrede saklanan bilginin dışarıya ne ölçüde aktarılacağını yönetir.

Bu yapı sayesinde LSTM, geçmiş bilgileri kaybetmeden uzun diziler üzerinde etkili öğrenme gerçekleştirebilir.



Şekil 3.8 LSTM biriminin yapısı: giriş kapısı, unutma kapısı, çıkış kapısı ve hücre durumu etkileşimi

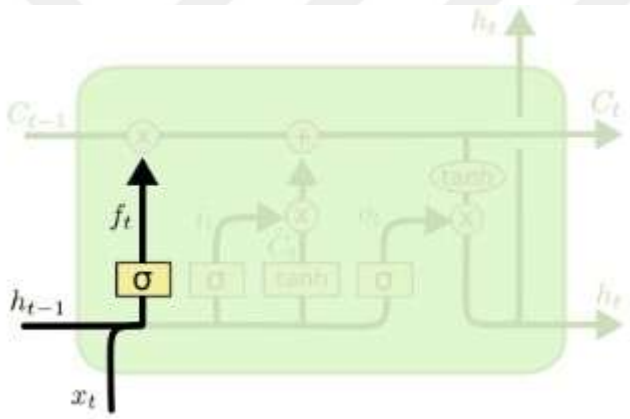
Şekil 3.8'de, bir LSTM hücresinin temel yapısı ve bilgi akışı görselleştirilmiştir (Xia vd. 2020). LSTM hücresindeki bilgi işleme sürecinin ilk adımı, hücre belleğinde yer alan bilgilerin ne kadarının korunacağına ve ne kadarının silineceğine karar verilmesidir. Bu işlev, genellikle unutma kapısı (forget gate) olarak adlandırılan ve sigmoid aktivasyon fonksiyonu kullanan bir sinir ağı katmanı tarafından gerçekleştirilir.

Bu katman yani unutma kapısı, bir önceki zaman adımındaki gizli durum h_{t-1} ve mevcut giriş x_t bilgilerini dikkate alarak, 0 ile 1 arasında bir değer üretir. Bu değer, önceki hücre durumu C_{t-1} deki bilginin ne ölçüde korunacağını veya silineceğini belirler. Eğer çıktı

1’e yakınsa bilgi korunur, 0’a yakınsa bilgi silinir.

Bu mekanizma, LSTM’nin uzun süreli bağımlılıkları etkili bir şekilde öğrenmesini sağlar. Böylece model, geçmişe yönelik çıkarım ve tahminlerde yüksek doğruluk elde edibilir.. Örneğin; bir hasta hakkında zamana bağlı olarak elde edilen laboratuvar sonuçları, medikal görüntüler veya klinik bulguların uzun vadeli ilişkileri bu mimari sayesinde analiz edilebilir.

Bu yapı sayesinde LSTM modelleri, tıbbi zaman serilerinde veya karmaşık çok-adımlı dizilerde, geçmiş bilgiler ışığında geleceği tahmin etmede geleneksel RNN mimarilerinden daha başarılı sonuçlar verir (Xia vd. 2020).



Şekil 3.9 LSTM hücresinde unutma kapısının (ft) işlevsel hesaplama süreci

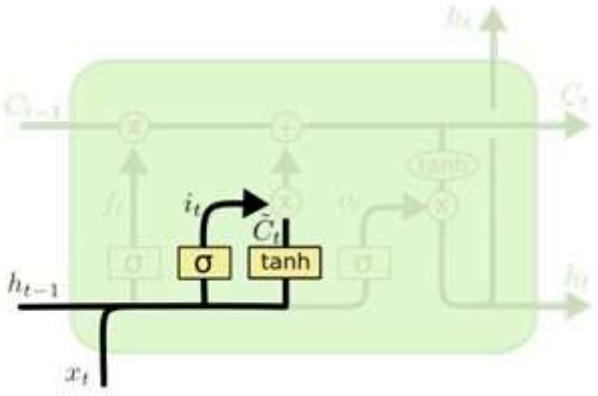
Şekil 3.9’da, bir LSTM hücresinde yer alan ve literatürde genellikle unutma kapısı (forget gate) olarak adlandırılan bileşenin işlevsel konumu ve hesaplama süreci görselleştirilmiştir (Olah 2015). Bu kapı, hücre belleğinde geçmişe dair hangi bilgilerin korunacağına ve hangi bilgilerin silineceğine karar vermektedir. Şeklin sol alt köşesinden hücreye doğru gelen iki ok, bir önceki zaman adımındaki gizli durum h_{t-1} ile mevcut zaman adımındaki giriş vektörü x_t bilgilerini temsil etmektedir. Bu iki bilgi, unutma kapısına iletilerek birlikte değerlendirilir. Sonrasında bu bileşenler, şeklin alt kısmında sarı renkle gösterilen sigmoid aktivasyon fonksiyonu (σ) ile işlenir. Bu işlem sonucunda 0 ile 1 aralığında bir değer üretilir:

- Eğer değer 1'e yakınsa, önceki hücre durumundaki bilgi tamamen korunur.
- Eğer değer 0'a yakınsa, bilgi tamamen silinir.

Bu çıktılar, LSTM hücresinin önceki hücre durumu olan C_{t-1} ile çarpılarak yeni hücre durumu C_t için geçerli olacak bilgi taşıma oranını belirler. Yani unutma kapısı, LSTM'nin geçmişten gelen bilgiyi ne ölçüde unutmadan geleceğe taşıyacağını denetleyen önemli bir bileşendir. Unutma kapısını formülü aşağıdaki gibidir:

$$f_t = \sigma\ (\ W_f.[h_{t-1}, x_t] + b_f\ ) \tag{3.11}$$

Unutma kapısının çıktısı olan f_t, h_{t-1} ve x_t ile birleştirilip, bu birleşim üzerine W_f ağırlık matrisiyle çarpılır. Sonrasında bias değeri b_f nin eklenmesiyle elde ettiğimiz sonucun sigmoid aktivasyon fonksiyonundan geçirilmesiyle hesaplanır. Bu hesaplamanın sonucuna göre her bir hücre bileşeni için 0 ile 1 aralığında bir değer üretilerek hangi bilgilerin korunacağına yönelik karar mekanizması oluşturulur.



Şekil 3.10 LSTM hücresinde bellek durumunun güncellenme süreci

Şekil 3.10'da, LSTM hücresinde yer alan giriş kapısının bellek durumu üzerindeki etkisi görselleştirilmiştir (Olah 2015). Giriş kapısı, dışarıdan gelen bilginin ne kadarının hücre durumuna ekleneceğini kontrol eden bir filtre işlevi görür. Bu yapı, önceki zaman adımına ait gizli durum h_{t-1} ile mevcut zamandaki veri x_t olmak üzere iki temel bileşeni dikkate almaktadır. Bu iki veri bir araya getirilerek yalnızca giriş kapısına özgü olarak tanımlanan ağırlık matrisi W_i ve bias terimi b_i üzerinden işlenir. Sonrasında sigmoid aktivasyon

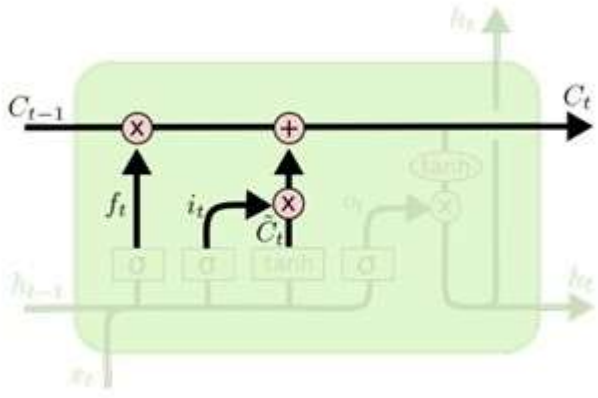
fonksiyonu uygulanarak i_t adı verilen [0,1] aralığında giriş kapısının çıktısı oluşturulur. Oluşturulan bu çıktı hücreye dışarıdan ne kadar bilginin alınacağına karar verir. İlgili hesaplama formülü aşağıdaki gibidir.

$$i_t = \sigma\ (\ W_i.[h_{t-1},x_t] + b_i\ ) \tag{3.12}$$

Önceki gizli durum h_{t-1} ile mevcut giriş bilgisi x_t, birleştirilerek doğrusal dönüşüme tabii tutulur ve ardından tanh aktivasyon fonksiyonundan geçirilerek aday hücre durumu $\tilde{C}_t$ elde edilir. Aday hücre durumu, hücre belleğine aktarılacak yeni bilgiyi ifade eder. Bu noktada i_t ve $\tilde{C}_t$ birlikte değerlendirildiğinde alınacak bilgisinin miktarıyla içeriği eş zamanlı olarak belirlenmiş olur.

$$\tilde{C}_t = \tanh\ (\ W_C.[h_{t-1},x_t] + b_C\ ) \tag{3.13}$$

Bu denklemde W_C aday hücre durumu için tanımlanan ağırlık matrisini, b_C ise buna karşılık gelen bias terimini ifade eder. Yukarıdaki şekilde sarı renkle gösterilen katmanlar, bu süreci açıkça göstermekle beraber ok yönleri bilgi akışını ve bileşenler arasındaki etkileşimi yansıtır. Sonuç olarak, i_t giriş kapısının hücreye alınacak bilginin miktarını, $\tilde{C}_t$ ise, aktarılacak bilginin içeriğini belirler. Bu iki bileşenin etkileşimi, LSTM hücresinin uzun vadeli bağımlılıkları etkili bir şekilde öğrenebilmesini sağlar.



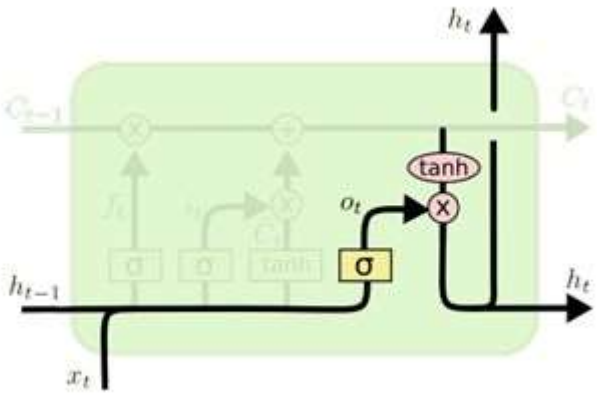
Şekil 3.11 LSTM hücre durumunun güncellenmesi süreci

Şekil 3.11'de yeni hücre durumu (C_t) oluşturulurken izlenen adımlar gösterilmektedir (Olah 2015). Bir önceki zaman adımına ait hücre durumu (C_{t-1}), unutma kapısından (f_t) elde edilen değerle çarpılarak geçmiş bilgi içeriğinden hangilerinin korunarak aktarılacağı belirlenir. Diğer taraftan, giriş kapısı çıktısı i_t ile aday hücre durumu $\tilde{C}_t$ çarpılarak yeni alınacak olan bilginin miktarı ve içeriği belirlenir. Akabinde iki ayrı bilgi akışı bu noktada bir araya gelir bunlar geçmişten gelip korunan bilgiler ve mevcut girişten üretilen yeni bilgilerdir. Bu iki bileşen toplanarak güncellenmiş hücre durumu C_t oluşturulur. Şekilde görüldüğü üzere çarpı (X) ve artı (+) simgeleri sırayla çarpma ve toplama işlemlerini, oklar ise bilginin akış yönünü ifade etmektedir. Bu işlem aşağıdaki gibi formülize edilmiştir.

$$C_t = f_t \,.\, C_{t-1} + i_t \,.\, \tilde{C}_t \tag{3.14}$$

$$C_t = f_t \,.\, C_{t-1} + i_t \,.\, \tanh\,(\, W_C.[h_{t-1},x_t] + b_C\,) \tag{3.15}$$

olarak ifade edilir. Bu süreç, LSTM yapısının uzun vadeli bilgiyi muhafaza edip, yeni öğrenilen bilgileri de sistematik bir şekilde entegre etmesine olanak sağlar.



Şekil 3.12 LSTM hücresinde çıkış vektörünün hesaplaması

Şekil 3.12'de LSTM hücresinin çıktısı olan h_t vektörünün üretim adımları gösterilmektedir (Olah 2015). Bu işlem, güncellenmiş hücre durumu C_t ile çıkış kapısı çıktısı o_t arasında kurulan iki aşamalı etkileşime dayalı gerçekleştirilir. İlk adımda, hücre durumu C_t, tanh fonksiyonu aracılığıyla [-1,1] aralığına dönüştürülür. Bu sayede hücre belleğinde yer alan bilginin yoğunluğunu ayarlayarak çıktı değerlerinin dengeli aralıkta

temsil edilmesini sağlar. Akabinde dönüştürülmüş değer, çıkış kapısından elde edilen o_t değeriyle çarpılır. Burada o_t, önceki zaman adımına ait gizli durum h_{t-1} ve mevcut zamandaki giriş vektörü x_t bazı alınarak sigmoid fonksiyonu ile hesaplanan değer olup, çıktı bilgisinin hangi ölçüde dışarı aktarılacağını belirler. Bu iki bileşenin çarpımı sonucu elde edilen h_t, LSTM hücresinin anlık zaman adımındaki çıktısını temsil eder. Şekilde yer alan tanh bloğu, hücre durumun dönüştürülmesini, sigmoid bloğu ise çıkış kapısı hesaplamasını ifade etmektedir. Bilgi akışının yönünü gösteren oklar ise işlemler arasındaki sıralı bağlantıyı gösterir. Bu mekanizma sayesinde LSTM, hem geçmiş bilgiye dayalı belleği koruyup aynı zamanda bu bilginin uygun bir kısmını modelin sonraki katmanlarına aktarıp öğrenme sürecine katkı sağlayabilir. LSTM hücresinin çıkış vektörünün hesaplama formülü aşağıdaki gibidir.

$$o_t = \sigma\,(W_0\,[h_{t-1}\,,\,x_t] + b_0) \tag{3.16}$$

$$h_t = o_t * \tanh\,(C_t) \tag{3.17}$$

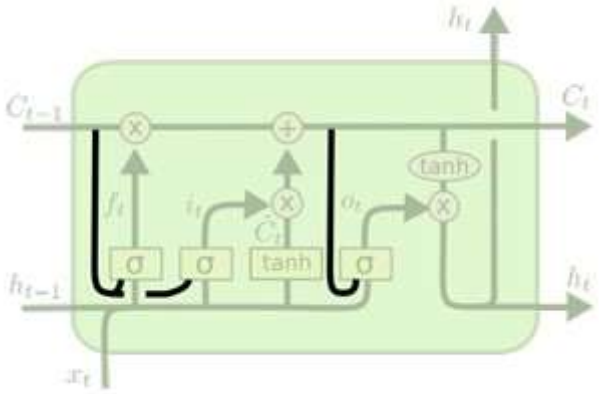
3.4.1 LSTM Varyantları

LSTM ağları, zamana bağlı değişen veri yapılarının modellenmesinde oldukça başarılı sonuçlar veren ve geniş uygulama alanına sahip yapay sinir ağı mimarileridir. Ancak zaman içerisinde karşılaşılan farklı sorunlara yanıt verebilmek ve bu zorluklara karşı daha kapsamlı çözümler geliştirebilmek amacıyla, LSTM'nin çeşitli varyantları ortaya çıkartılmıştır.

Bu varyantlar; hesaplama verimliliğini artırmak, bazı görevlerde daha yüksek başarı elde etmek veya mimarideki bazı karmaşık yapıların sadeleştirilmesini sağlamak amacıyla geliştirilmiştir. Bazı modeller parametre sayısını azaltarak daha hafif çalışmayı hedeflerken, bazıları ise mimariye yeni bileşenler ekleyerek öğrenme kabiliyetini artırmayı amaçlamıştır.

Örneğin, Rahman ve Siddiqi (2019) tarafından yapılan bir çalışmada, geleneksel LSTM yapısının önceki zaman adımına ait hücre bilgisini yalnızca kapılar üzerinden aktarması yerine, doğrudan erişim sağlayan yeni bir yapı önerilmiştir. Bu yapı, peephole connection

(gözleme deliği bağlantısı) olarak adlandırılan mekanizma sayesinde giriş, unutma ve çıkış kapılarının hücre içeriğine doğrudan erişmesine olanak tanır. Böylece, çıkış kapısı kapalı bile olsa önceki hücre belleği üzerinde değerlendirme yapılabilir ve daha güçlü zaman bağımlılıkları öğrenilebilir hale gelir.



Şekil 3.13 Peephole bağlantılı LSTM yapısının mimarisi

$$f_t = \sigma\,(\,W_f\,.\,[C_{t-1}, h_{t-1}, x_t] + b_f\,) \tag{3.18}$$

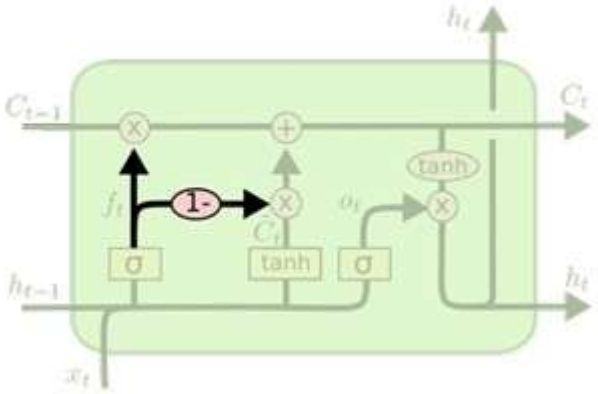
$$i_t = \sigma\,(\,W_i\,.\,[C_{t-1}, h_{t-1}, x_t] + b_i\,) \tag{3.19}$$

$$o_t = \sigma\,(\,W_o\,.\,[C_t, h_{t-1}, x_t] + b_o\,) \tag{3.20}$$

Şekil 3.13'te peephole yapısında, her kapının girdileri yalnızca mevcut zamandaki veri x_t ve bir önceki zaman adımına ait gizli durum h_{t-1} ile sınırlı kalmayıp aynı zamanda önceki zaman adımına ait hücre durumu c_{t-1} de hesaplamalara dahil edilir (Olah 2015). Bu girdilere bias terimleri de eklenerek her kapının karar mekanizması daha kararlı bir hale getirilir. Özellikle kapıların, hücre belleğini pasif bir veri deposu olarak görmeyip, aktif bir bilgi kaynağı olarak değerlendirmesi bu modelin temel farkını ortaya koymaktadır. Peephole LSTM, yapısal genişletmeler sayesinde geleneksel LSTM mimarisine kıyasla sıralı veriler üzerindeki ilişkileri daha etkin biçimde öğrenmekle kalmayıp çıktılarda daha isabetli sonuçlar üretebilmektedir. Ayrıca zamanla değişen örüntülerin yakalanmasında bu model birçok avantajlar sunmaktadır (Rahman ve Siddiqi 2019).

Şekil 3.14'te geleneksel LSTM hücresinin sadeleştirilmiş bir hali olarak düşünebiliriz (Olah 2015). Bu yapıda, standart bir LSTM'de ayrı ayrı tanımlanan giriş ve unutma

kapıları, tek bir kapı mekanizmasında birleştirilerek bilgi akışı daha sade bir şekilde yönetilmektedir.



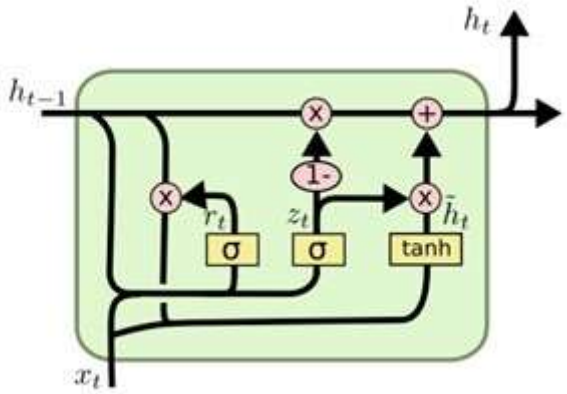
Şekil 3.14 Birleşik giriş ve unutma kapılı LSTM yapısının mimarisi

Bu sayede modelin hesaplama iş yükü azaltılırken aynı zamanda bilgi güncelleme süreci de etkin bir şekilde uygulanmaktadır.

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t \tag{3.21}$$

Yapı, geçmişte öğrenilen bilginin ne kadarının tutulacağına, yeni bilginin ise hangi ölçüde ekleneceğine yönelik karar süreci, öğrenilen parametrelere göre hesaplanıp sigmoid (σ) fonksiyonuyla [0,1] aralığına denk gelecek şekilde uygulanır. Burada elde edilen değer, önceki zaman adımına ait hücre durumu c_{t-1} ile çarpılarak korunacak olan bilgi belirlenirken, aynı zamanda 1 eksiği alınarak yeni bilgiye ne kadar yer açılacağı belirlenmektedir. Burada çift yönlü etki sayesinde hem eski bilginin ne kadarının unutulacağını hem de yeni bilginin hangi oranda hücreye dahil edileceği aynı anda denetlenir. Tanh aktivasyon fonksiyonu, mevcut zamandaki veri x_t ve önceki zaman adımına ait gizli durum h_{t-1} ile etkileşime girerek aday hücre durumu üretilip daha önce 1-f olarak elde edilen değerle çarpılarak hücreye eklenir. Böylece güncellenmiş hücre durumu C_t, hem önceki bilgileri saklayıp, yeni bilgileri de ekleyerek dengeli bir şekilde güncellenmiş olur. Mimarinin sağ kısmında, çıkış kapısı geleneksel LSTM yapısı ile aynı şekilde çalışmaktadır. Sigmoid fonksiyonu ile üretilen o_t değeri, tanh aktivasyon çıktısıyla çarpılarak yeni gizli durum h_t elde edilir. Bu süreç, modelin dış katmanlara

aktaracağı bilgiyi belirler. Bu yapıyı, GRU mimarisi ile LSTM arasında bir geçiş modeli olarak değerlendirebiliriz. GRU yapısında giriş ve unutma kapıları birleştirilmiştir ancak GRU, LSTM'den farklı olarak hücre durumu yerine gizli durumu kullanır.



Şekil 3.15 Gated recurrent unit (GRU) yapısal mimarisi

Şekil 3.15'te LSTM'nin bir diğer varyantı ise Gated Recurrent Unit GRU modelidir (Olah 2015). Gated Recurrent Unit, Cho ve arkadaşları tarafından, farklı zaman ölçeklerinde ortaya çıkan bağımlılıkları esnek bir şekilde öğrenebilen tekrarlayan sinir ağı mimarisi olarak geliştirilmiştir. LSTM yapısıyla benzer şekilde, GRU da bilgi akışını kontrol eden kapı mekanizmaları içerir. Ancak LSTM'den farklı olarak, GRU'nun yapısında ayrı bir bellek hücresi bulunmamaktadır (Chung vd. 2014). GRU modeli, güncelleme kapısı z_t ve sıfırlama kapısı r_t olmak üzere iki temel kapı üzerinden çalışır. Bu kapılar, gizli durumun h_t nasıl güncelleneceğini ve geçmiş bilginin hangi oranda kullanılacağını belirler. Güncelleme kapısının z_t formülü aşağıdaki gibidir:

$$z_t = \sigma\,(W_z\,.\,[h_{t-1}\,,\,x_t]) \tag{3.22}$$

Güncelleme kapısı, önceki gizli durum h_{t-1} ile mevcut zamandaki veriyi x_t beraber değerlendirerek hangisinin ne oranda ne kadar kullanılacağını belirler. Sigmoid aktivasyonu (σ) kullanıldığı için çıktı 0 ve 1 aralığında olacaktır. Sıfırlama kapısı r_t, geçmiş bilginin hangi kısmının kullanım dışı bırakılacağını belirler. Eğer modelin, geçmiş bilgilerden ziyade daha çok yeni verilere yönelmesi isteniyorsa, sıfırlama kapısı düşük bir değer üretir. Bu sayede, geçmiş bilgiyi önemsemeyerek yeni gelen veriye göre karar verecektir. Sıfırlama kapısının r_t formülü aşağıdaki gibidir:

$$r_t = \sigma\,(W_r\,.\,[h_{t-1}\,,\,x_t]) \tag{3.23}$$

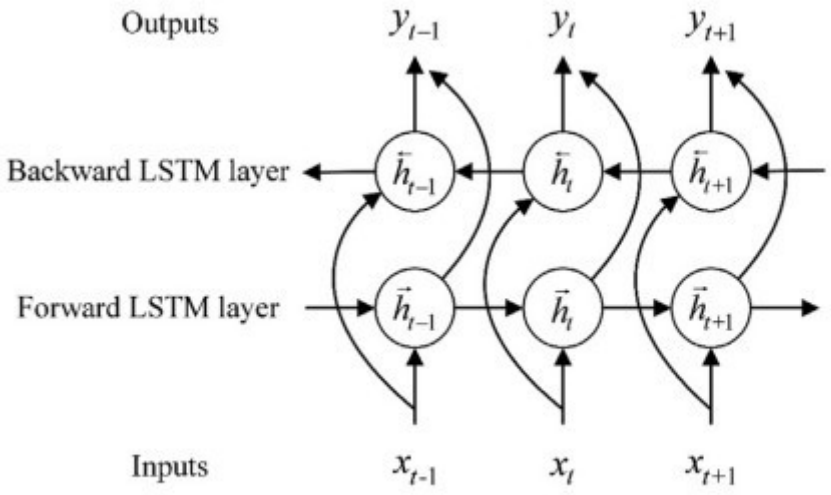
Akabinde sıfırlama kapısının etkilediği önceki gizli durum ile mevcut zamandaki veriyi birleştirdikten sonra tanh aktivasyon fonksiyonuyla aday gizli durum $\widetilde{h_t}$ oluşturulur. Bu aday bilgi, kullanılmaya hazırdır fakat tek başına yeterli değildir. Son olarak güncelleme kapısı devreye girer ve bu bilgiyle eskisi birleştirilerek son hali oluşturulur. Aday gizli durumun $\widetilde{h_t}$ formülü aşağıdaki gibidir:

$$\widetilde{h_t} = \tanh\,(\mathrm{W}.\,[r_t * h_{t-1}\,,\,x_t]) \tag{3.24}$$

Son adımda ise güncelleme kapısı aracılığıyla önceki gizli durum h_{t-1} ile yeni aday bilgi $\widetilde{h_t}$ arasında bir denge kurulur. Böylece model, geçmişteki bilgiyi tamamen silmeden önce gerekli kısmını korur ve buna yeni gelen bilgiyi uygun ölçüde ekleyerek güncellenmiş gizli durumu oluşturur. Formülü aşağıdaki gibidir:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h_t} \tag{3.25}$$

LSTM'nin bir diğer varyantı da çift yönlü BiLSTM modelidir. BiLSTM modeli, ilk olarak Schuster ve Paliwal'in çalışmalarıyla geliştirilmiştir (Schuster ve Paliwal 1997).



Şekil 3.16 Çift yönlü uzun kısa süreli bellek (BiLSTM) ağının yapısı

Şekil 3.16'da BiLSTM yapısı üç katmandan oluşmaktadır. Bunlar giriş katmanı, LSTM katmanları ve çıkış katmanıdır (Xia vd. 2020). Giriş katmanında x_{t-1}, x_t ve x_{t+1} olmak üzere giriş vektörleri yer alır. Bu vektörler, aynı anda hem ileri yönlü hem de geri yönlü olarak LSTM katmanlarına aktarılır. Böylece model, her adımda yalnızca geçmişteki bilgiyi değil, gelecekteki bilgileri de dikkate alarak veriyi işler. Gizli katmanlar ise iki ayrı bileşenden oluşur. Bunlar ileri yönlü LSTM ve geri yönlü LSTM'dir. İleri yönlü LSTM katmanı, veriyi zaman akışına uygun olarak t-1'den t+1'e doğru işler. Bu işlemin sonucunda her bir zaman adımı için $\overrightarrow{h_t}$ olarak ifade edilen gizli durum vektörü üretilir. Buna karşılık, geri yönlü LSTM katmanı ters yönde t+1'den t-1'e doğru işler. Bu işlemin sonucunda yine benzer şekilde $\overleftarrow{h_t}$ olarak ifade edilen gizli durum vektörü üretilir. Sonuç olarak her bir zaman adımı t için, biri ileri yönden diğeri de geri yönden gelen iki ayrı gizli durum vektörü elde edilmiş olur. Son aşama olan çıkış katmanında ise ileri yönlü ve geri yönlü LSTM katmanlarından elde edilen iki gizli durum vektörü birleştirilir ve her zaman adımı için kesin çıktı olan y_t hesaplanır. Bu işlem, genellikle iki vektörün yan yana eklenmesiyle yapılır: $[\overrightarrow{h_t};\overleftarrow{h_t}]$. Ayrıca zaman dizisindeki tüm adımlar için tekrarlanır. Böylece model, her bir veriyi hem geçmiş hem de gelecek bilgileri dikkate alarak değerlendirecektir. Bu sayede veriyle ilgili anlam bütünlüğü korunarak modelin yaptığı tahminler daha doğru bir hale gelecektir. BiLSTM'nin matematiksel denklemleri aşağıdaki gibidir:

$$\overrightarrow{h_t} = \text{LSTM}\ (x_t\ ,\ \overrightarrow{h_{t-1}}) \tag{3.26}$$

Yukarıdaki denklem, ileri yönlü LSTM hücresini tanımlamaktadır. Zaman adımı t'deki ileri yönlü gizli durum ($\overrightarrow{h_t}$), o andaki giriş verisi x_t ile bir önceki ileri gizli durum ($\overrightarrow{h_{t-1}}$) kullanılarak hesaplanır. Burada veri soldan sağa, yani geçmişten bugüne doğru ilerler.

$$\overleftarrow{h_t} = \text{LSTM}\ (x_t\ ,\ \overleftarrow{h_{t+1}}) \tag{3.27}$$

Yukarıdaki denklem, geri yönlü LSTM hücresini tanımlamaktadır. Aynı giriş verisi x_t, bu kez gelecekteki gizli durum olan ($\overleftarrow{h_{t+1}}$) ile işlenir. Burada veri sağdan sola, yani gelecekten geçmişe doğru ilerler.

$$y_t = W_{\overrightarrow{hy}} \overrightarrow{h_t} + W_{\overleftarrow{hy}} + \overleftarrow{h_t} + b_y \quad (3.28)$$

Yukarıdaki denklem, y_t çıktısının nasıl üretildiğini gösteriyor. İleri yönlü ve geri yönlü gizli durum vektörleri ayrı ayrı ağırlık matrisleriyle çarpılır. Elde edilen değerler toplanıp bias değeri olan b_y eklenir. Bu işlemlerin sonucunda y_t değeri hesaplanmış olur.

3.5 ARIMA

ARIMA modeli, 1976 yılında Box ve Jenkins tarafından geliştirilmiştir (Box ve Jenkins 1976). ARIMA yöntemi, tek değişkenli zaman serilerinin tahmini ve kontrolü amacıyla kullanılan bir modelleme tekniğidir (Yeşilyuva 2013). ARIMA (Otoregresif Entegre Hareketli Ortalama) modeli, zaman serisi verilerini modellemek için yaygın olarak kullanılan bir yöntemdir. ARIMA modelleri, zaman serisi verilerindeki trendleri, mevsimselliği ve diğer yapısal bileşenleri yakalamayı amaçlar. Model, geçmiş verilere dayanarak gelecekteki değerleri tahmin etmeyi amaçlar ve bu yönüyle finans, ekonomi ve mühendislik gibi birçok uygulama alanında kullanılır (Box vd. 2015). ARIMA modeli üç temel bileşeni bir araya getirir: Otoregresif Süreç (AR), Hareketli Ortalama (MA) ve Fark Alma (I).

- Otoregresif (AR) Bileşeni: Gelecekteki değerleri geçmiş verilerin doğrusal kombinasyonlarını kullanarak tahmin eder. AR bileşeni, parametre p ile temsil edilir ve modelde geçmiş p dönemin değerlerini içerir. Bu yapı, belirli bir zaman periyodunun kendi geçmiş değerleriyle olan örüntüsünü modellemeye dayanır. (Arunkumar vd. 2021).
- Farklama (I) Bileşeni: Zaman serisini durağan hale getirmek için kullanılır. I bileşeni d parametresi ile temsil edilip, d parametresi serinin kaç kez farklanacağını belirler. Yani durağan olmayan serileri işleyebilmek için farklama işlemiyle seriyi durağan hale getirme işlevini yerine getirir.
- Hareketli Ortalama (MA) Bileşeni: Gelecekteki değerleri geçmiş hata terimlerinin doğrusal kombinasyonlarını kullanarak tahmin eder. MA bileşeni, q parametresi ile temsil edilir ve geçmiş q dönemin hata terimlerini içerir. MA, belirli bir zaman

adımında yapılan tahmin hatasına bakarak, sonraki zaman adımındaki değerleri tahmin etmeye çalışır (Arunkumar vd. 2021).

ARIMA modelleri, zaman serisi verilerini analiz etmek ve tahmin etmek için güçlü araçlardır. Ancak modelin doğru uygulanabilmesi için verilerin durağan olması gereklidir. Durağan olmayan verilerde, veri serisi fark alınarak durağan hale getirilebilir (Hyndman ve Athanasopoulos 2018). Bu çalışmada, ARIMA modeli yalnızca tek değişkenli verilere uygulanmıştır. Meme kanseri veri seti üzerinde uygulanan ARIMA modeli, her hastanın teşhis verilerini zaman içinde kullanarak tahminler yapmayı amaçlamaktadır. Bu model, biyomedikal verilerin zaman serisi analizinde özellikle etkili olabilir (Shumway ve Stoffer 2017). ARIMA modeli, finansal piyasa analizi Hamilton (1994), ekonomik göstergelerin tahmin edilmesi Makridakis vd. (1998) ve çeşitli mühendislik uygulamaları gibi geniş bir uygulama yelpazesinde kullanılmaktadır. Bu modelin avantajlarından biri, veri setindeki trendleri ve mevsimselliği yakalama yeteneğidir. Ancak modelin doğruluğu ve performansı seçilen p, d ve q parametrelerine bağlıdır. Bu nedenle, model parametrelerinin dikkatlice belirlenmesi ve optimize edilmesi önemlidir (Box vd. 2015). Aşağıda, p dereceli AR modeli ve q dereceli MA modeli matematiksel olarak gösterilmiştir:

$$y_t = \mathrm{C} + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t \tag{3.29}$$

$$y_t = \mathrm{C} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{3.30}$$

ARIMA modelleri, AR modeli, fark alma ve MA modeli birleştirilerek oluşturulur. Fark alma (I), tahmin üretmek amacıyla seriyi fark alarak durağanlaştırma işlemidir (Arunkumar vd. 2021). ARIMA modelinin matematiksel biçimi aşağıdaki gibidir:

$$y_t = \mathrm{C} + \varphi_1 y_{t-1} + \varphi_p y_{t-p} + \ldots + \varphi_n y_{t-n} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3.31}$$

Bu yapı sayesinde, serinin geçmiş değerlerine dayalı olarak hem düzey hem de hata terimleriyle ilişki kurularak geleceğe yönelik tahminler yapılabilir. ARIMA, bu bileşenleri sistematik bir çerçevede bir araya getirdiği için diğer tahmin modellerinden

ayrılır. ARIMA (p,d,q) ifadesindeki her bir parametre aşağıdaki bileşenleri temsil eder:

- p, otoregresyon katsayılarının sayısını,
- d, kaç kez farklama işleminin uygulandığını,
- q ise kaç gecikmeli hata teriminin kullanıldığını belirtir.

Örnek olarak ARIMA(1,1,1) modeli şu anlama gelir: bir önceki değere otoregresyon uygulanır (p=1), veri bir kez farklanarak durağanlaştırılır (d=1) ve bir önceki tahmin hatası modele dahil edilir (q=1). Buna karşılık ARIMA(1,0,1) modelinde, seriye farklama uygulanmadan direkt olarak otoregresif ve hareketli ortalama bileşenleri kullanılır. ARIMA modellemesinde dikkat edilmesi gereken ilk adım, zaman serisinin durağan olup olmadığını kontrol etmektir. Çünkü ARIMA, sadece durağan serilerde sağlıklı sonuçlar vermektedir.

Bir serinin durağan olma durumu, serinin ortalama ve varyansının zaman içerisinde değişmemesi ve veriler arasındaki ilişkinin sadece gözlemler arası zamana bağlı kalması durumudur. Bu özelliği anlamak için öncelikle serinin grafiği incelenir. Akabinde verilerin zamanla sabit bir ortalama etrafında dalgalanıp dalgalanmadığına bakılır.

Eğer grafik, verilerin ortalamadan sapıp yeniden mevcut ortalama seviyesine geliyorsa genellikle durağanlık işareti olarak yorumlanır. Grafiksel incelemeler subjektif olabileceği için daha kesin sonuçlar ancak istatistiksel testlerle mümkün olabilir. Modelin uygun parametrelerini belirleyebilmek için öncelikle ACF (Autocorrelation Function) ve PACF (Partial Autocorrelation Function) grafiklerinden yararlanılır. ACF, serinin geçmiş değerleriyle olan genel ilişki düzeyini gösterirken, PACF ise her bir gecikmenin seriye olan etkisini diğer gecikmelerin etkisinden bağımsız olarak gösterir. Bu grafiklerin analiziyle AR(p) ve MA(q) parametrelerine dair ilk tahminler yapılır. Genellikle p değeri PACF, q değeri ise ACF grafiği yardımıyla belirlenir. Sonrasında model kurulup parametreler tahmin edilir ve sonuçların doğruluğu MSE (Ortalama Kare Hatası), MAE (Ortalama Mutlak Hata) ve RMSE (Kök Ortalama Kare Hatası) gibi hata ölçümleriyle değerlendirilir. Bu değerlerin düşük olması, modelin daha doğru tahminler yaptığını gösterir.

3.6 Performans Değerlendirmesi

Modelin uygunluğu, genel tahmin MSE, MAE, RMSE gibi çeşitli parametreler kullanılarak değerlendirilmiştir. MSE (Ortalama Kare Hatası), MAE (Ortalama Mutlak Hata) ve RMSE (Kök Ortalama Kare Hatası) gibi istatistiksel hata ölçümleri, makine öğrenimi ve veri analizinde modellerin performansını değerlendirmek için yaygın olarak kullanılır. Ortalama kare hatası (MSE), tahmin edilen değerler ile gerçek değerler arasındaki farkların karelerinin ortalamasıdır ve şu formül ile hesaplanır:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{3.32}$$

MAE, modelin yaptığı tahminlerle gerçek değerler arasındaki farkların mutlak değerinin ortalamasıdır. Pozitif ya da negatif hataların yönü dikkate alınmadan büyüklüğü hesaplanır ve formülü aşağıdaki gibidir:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{3.33}$$

RMSE, MSE değerinin karekökü alınarak işlem yapılır. Bu sayede hata birimi, hedef değişkenle aynı hale gelmektedir. RMSE, aşağıdaki formül ile hesaplanır:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)} \tag{3.34}$$

Accuracy doğruluk anlamına gelmektedir. Doğruluk, modelin doğru olarak tahmin etmiş olduğu tüm örneklerin (hem pozitif hem de negatif sınıflar) toplam örnek sayısına oranıdır. Aşağıdaki formül ile hesaplanır:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.35}$$

Yukarıdaki formülde TP değeri True Positive kısaltmasından gelmekte olup gerçek pozitif anlamına gelmektedir. TN değeri, True Negative kısaltmasından gelmekte olup gerçek negatif anlamına gelmektedir. FP değeri, False Positive kısaltmasından gelmekte olup yanlış pozitif anlamına gelmektedir. FN değeri ise False Negative kısaltmasından

gelmekte olup yanlış negatif anlamına gelmektedir.

Sensetive yani duyarlılık, gerçek pozitif örnekler arasında modelin doğru tanıdığı örneklerin oranıdır. Aşağıdaki formül ile hesaplanır:

$$Sensitivity = \frac{TP}{TP+FN} \tag{3.36}$$

4. BULGULAR

Çizelge 4.1'de, meme kanseri teşhisine yönelik olarak geliştirilen KAN, ARIMA ve LSTM modellerinin performansları, üç temel hata metriği olan MSE, RMSE ve MAE açısından karşılaştırılmıştır.

MSE değerleri incelendiğinde, KAN modeli 0.0309 ile en düşük değere sahip olup, tahminlerinin gerçek değerlere en yakın model olduğunu göstermektedir. ARIMA modeli 0.2553 değeriyle orta düzeyde bir performans sergilerken, LSTM modeli 13.3826 gibi oldukça yüksek bir MSE değeriyle en büyük hata payına sahip model olarak öne çıkmaktadır.

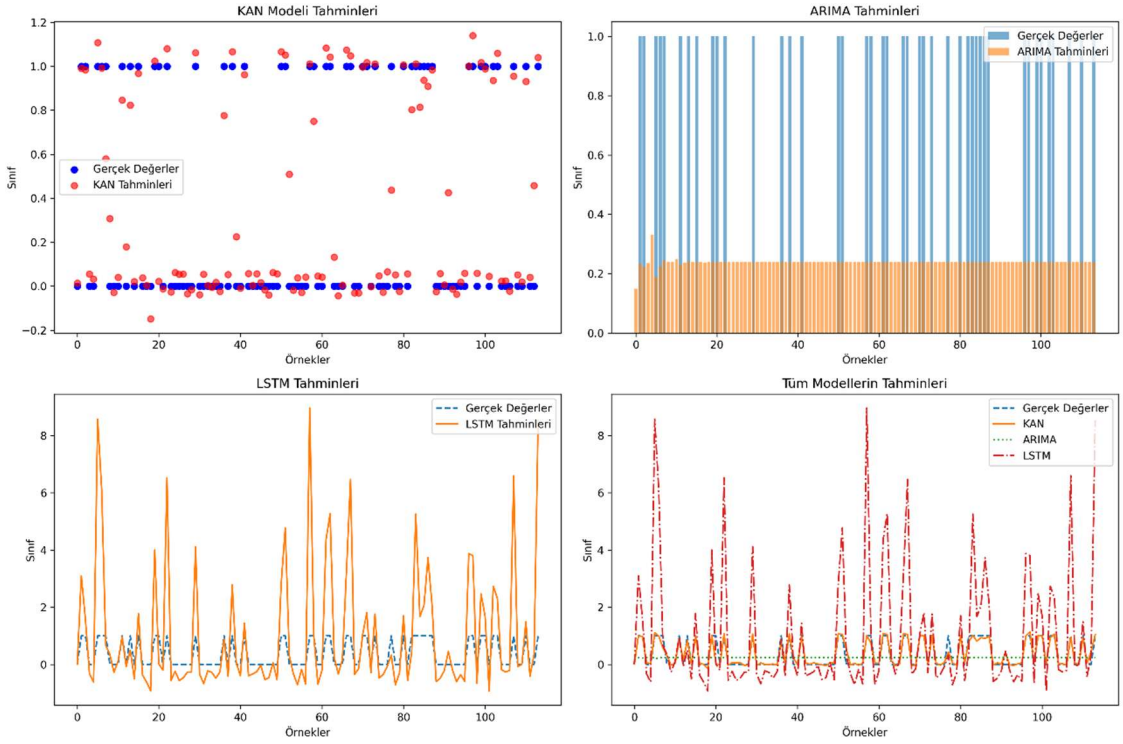
RMSE değeri, tahmin hatalarının büyüklüğünü doğrudan ifade etmesi bakımından önem taşır. Bu bağlamda, KAN modeli 0.1758 değeri ile en düşük RMSE'ye sahip olup, tahminlerinin genellikle düşük hata içerdiğini göstermektedir. ARIMA modeli 0.5053 değeri ile daha yüksek hata büyüklüğüne işaret ederken, LSTM modeli 3.6582 ile en yüksek RMSE değerini almıştır. Bu durum, LSTM modelinin tahminlerinde daha büyük sapmalar içerdiğini ortaya koymaktadır.

MAE açısından da benzer bir tablo gözlemlenmektedir. KAN modeli, 0.1108 değeriyle en düşük mutlak hata ortalamasına sahiptir ve bu durum, modelin genel olarak en doğru tahminleri ürettiğini göstermektedir. ARIMA modeli 0.4368 ile orta düzeyde hata üretirken, LSTM modeli 2.7679 ile en yüksek MAE değerini alarak en hatalı model olarak öne çıkmaktadır.

Çizelge 4.1 KAN, ARIMA ve LSTM Modellerinin Hata Metrikleri Karşılaştırması (Veri seti 1)

Metrikler	KAN	ARIMA	LSTM
MSE	0.0309	0.2553	13.3826
RMSE	0.1758	0.5053	3.6582
MAE	0.1108	0.4368	2.7679

MSE: Ortalama Kare Hatası,MSE: Kök Ortalama Kare Hatası,MAE: Ortalama Mutlak Hata



Şekil 4.1 ARIMA, LSTM ve KAN modellerinin meme kanseri tanısında tahmin performansı (Veri seti 1)

Şekil 4.1'de, KAN, ARIMA ve LSTM modellerinin meme kanseri teşhisine yönelik tahmin çıktıları karşılaştırmalı olarak görselleştirilmiştir. Bu grafiksel analiz, her bir modelin sınıflandırma kararlılığı, doğruluk seviyesi ve tahmin hataları açısından değerlendirilmesini mümkün kılmaktadır.

ARIMA modeli, özellikle sabit olmayan ve sıklıkla değişkenlik gösteren çıktılar üretmiştir. Gerçek sınıflar ile uyumsuzluğu belirgin şekilde görülen bu model, yanlış pozitif ve yanlış negatif oranları açısından zayıf bir performans sergilemiştir. Sınırlı doğrusal yapısı nedeniyle ARIMA, karmaşık yapısal ilişkileri modelleme konusunda yetersiz kalmış; bu durum duyarlılık (sensitivity) ve özgüllük (specificity) gibi metriklerde düşük başarıya neden olmuştur.

LSTM modeli, zaman serisi modellemesinde avantaj sağlamasına rağmen bazı örneklerde yüksek sapmalar ve sıfırın çok üzerinde tahmin değerleri üretmiştir. Bu durum, modelin sınıf sınırlarını aşan öngörülerde bulunarak doğruluk ve özgüllük açısından tutarsızlıklar

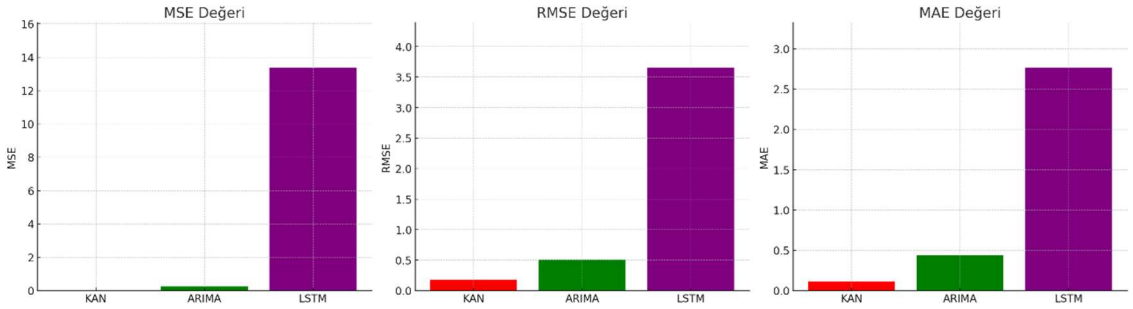
yarattığını göstermektedir. Her ne kadar LSTM, geçici desenleri daha iyi yakalama eğiliminde olsa da, doğrulama setindeki dengesiz öğrenme çıktıları modelin güvenilirliğini kısıtlamaktadır.

KAN modeli ise tüm modeller içinde en stabil ve gerçeğe en yakın sonuçları üretmiştir. Gerçek etiketlerle olan yüksek örtüşme, yanlış sınıflandırma oranlarının düşüklüğü ve tahmin eğrilerinin sınıf sınırlarında kalması KAN modelinin hem doğruluk (accuracy) hem de duyarlılık ve özgüllük gibi temel metriklerde üstün performans sergilediğini ortaya koymaktadır. Görsel olarak da KAN modelinin tahmin noktalarının gerçeğe sıkıca hizalanması, bu modelin meme kanseri teşhisinde güvenilir bir alternatif olduğunu desteklemektedir.

Şekil 4.2'de, KAN, ARIMA ve LSTM modellerine ait tahmin performansları; MSE, RMSE ve MAE metrikleri üzerinden karşılaştırmalı olarak görselleştirilmiştir. Grafiklerden elde edilen bulgulara göre, KAN modeli, tüm hata metriklerinde MSE, RMSE ve MAE en düşük değerlere sahiptir. Bu durum, KAN modelinin tahminlerinde diğer modellere göre çok daha düşük sapma ile çalıştığını ve meme kanseri teşhisinde yüksek doğruluk sağladığını göstermektedir. KAN modeli, hata oranlarını minimize etmesi açısından literatürde öne çıkan modellerle karşılaştırıldığında istikrarlı ve güvenilir bir alternatif olarak öne çıkmaktadır.

ARIMA modeli, KAN modeline kıyasla daha yüksek hata oranlarına sahip olmakla birlikte, LSTM modeline göre daha başarılı bir performans sergilemiştir. Özellikle RMSE ve MAE değerlerinde KAN'a yakın seyreden ARIMA, zaman serisi yapısına dayalı verilerde kabul edilebilir doğruluk seviyeleri sağlamaktadır.

LSTM modeli ise, her üç metrikte de açık ara en yüksek hata değerlerini üretmiştir. LSTM'nin hem MSE hem RMSE hem de MAE değerlerinin yüksek olması, bu modelin incelenen veri seti bağlamında istikrarsız ve düşük doğruluklu tahminler yaptığını göstermektedir. Bu bulgu, LSTM modelinin özellikle bu çalışma kapsamında ele alınan veri yapısı üzerinde yetersiz genelleme kapasitesine sahip olduğunu ortaya koymaktadır.



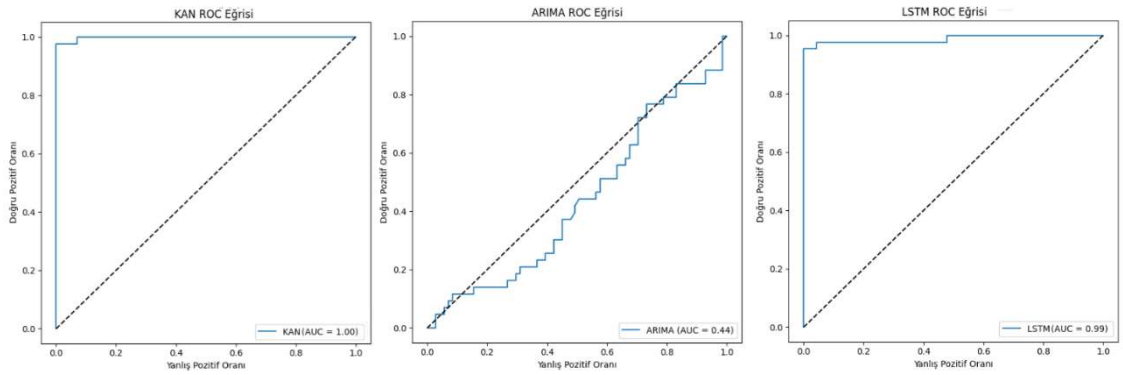
Şekil 4.2 Meme kanseri tanısına yönelik KAN, ARIMA ve LSTM modellerinin hata metrikleri (Veri seti 1)

Şekil 4.3'te, KAN, ARIMA ve LSTM modellerinin ROC (Receiver Operating Characteristic) eğrileri yer almakta olup, her bir modelin sınıflandırma başarımını grafiksel olarak göstermektedir.

KAN Modeli ROC eğrisi neredeyse sol üst köşeye yapışacak şekilde ilerlemektedir ve eğri altındaki alan AUC = 1.00 olarak hesaplanmıştır. Bu durum, KAN modelinin sınıflandırma görevinde mükemmele yakın bir performans sergilediğini göstermektedir. Model, pozitif sınıfları negatiflerden oldukça doğru bir şekilde ayırt edebilmekte ve yüksek bir duyarlılık ve özgüllük dengesine sahiptir. Bu tür bir ROC eğrisi, overfitting riski olmadan ideal sınıflandırıcıya oldukça yaklaştığını işaret eder.

ARIMA Modeli ise ROC eğrisinde diyagonal çizgiye oldukça yakın seyretmektedir ve AUC değeri yalnızca 0.44 olarak raporlanmıştır. Bu değer, rastgele tahmin eden bir sınıflayıcıdan bile daha kötü bir performansı göstermektedir. ROC eğrisinin şekli, modelin pozitif sınıfları ayırt etme yetisinin zayıf olduğunu ve çoğu durumda hem yanlış pozitif hem de yanlış negatif sınıflandırmalara yol açtığını ortaya koymaktadır.

LSTM Modeli için çizilen ROC eğrisi, KAN modeline benzer şekilde üst sol köşeye çok yakın olup AUC değeri 0.99'dur. Bu, LSTM modelinin sınıflandırma başarımının oldukça yüksek olduğunu ve doğru pozitifleri yüksek başarıyla yakaladığını göstermektedir. Ancak bazı lokal eşiklerde hatalı sınıflandırmalar gözlemlenebileceği, eğrinin küçük dalgalanmalarından anlaşılmaktadır. Yine de AUC değeri, LSTM modelinin güçlü ve genellenebilir bir sınıflayıcı olduğunu doğrulamaktadır.



Şekil 4.3 Meme kanseri tanısına yönelik KAN, LSTM ve ARIMA modellerinin ROC eğrileri ve AUC değerleri (Veri seti 1)

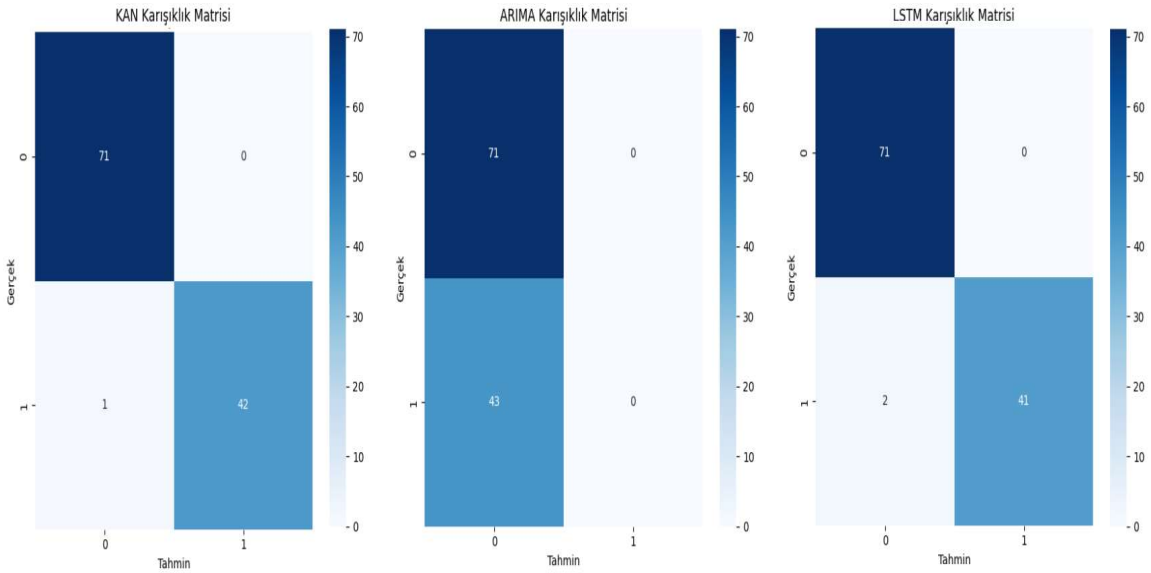
Şekil 4.4'te, KAN, ARIMA ve LSTM modellerine ait karışıklık matrisleri karşılaştırmalı olarak sunulmakta olup, meme kanseri teşhisinde her bir modelin sınıflandırma başarımı detaylı şekilde analiz edilmiştir. Bu matrisler, modellerin gerçek pozitif, gerçek negatif, yanlış pozitif ve yanlış negatif tahmin sayılarını açıkça ortaya koyarak, klinik karar destek sistemleri açısından ne derece güvenilir olduklarını göstermektedir.

KAN modeli, 71 doğru negatif (TN) ve 42 doğru pozitif (TP) tahmin ile en başarılı sınıflandırmayı gerçekleştirmiştir. Modelin sadece 1 yanlış negatif (FN) üretmiş olması, yani bir malign vakayı benign olarak hatalı sınıflandırması, genel teşhis doğruluğu açısından ihmal edilebilir düzeydedir. Özellikle hiçbir yanlış pozitif tahminin olmaması, modelin özgüllük (specificity) açısından mükemmel bir performans sergilediğini ortaya koymaktadır. Bu bağlamda KAN modeli, hem duyarlılık hem de doğruluk açısından en üstün model olarak öne çıkmaktadır.

ARIMA modeli, her ne kadar 71 benign vakayı doğru şekilde tahmin etmiş olsa da, malign vakaların hiçbirini doğru sınıflandıramamış; tamamını benign olarak işaretlemiştir. Bu durum, modelin duyarlılığının sıfır olduğunu göstermektedir. ARIMA'nın tüm pozitif sınıflarda başarısız olması, tıbbi teşhis gibi kritik bir alanda kullanımını son derece riskli hale getirmektedir. Model yalnızca negatif sınıfları ayırt edebilmekte, bu da ciddi yanlış negatif sonuçlara yol açarak hastalığın gözden kaçırılmasına neden olabilir.

LSTM modeli ise 71 doğru negatif ve 41 doğru pozitif tahminle, KAN modeline oldukça

yakın bir performans sergilemiştir. Modelin 2 yanlış negatif ve 0 yanlış pozitif üretmesi, duyarlılığının KAN modeline göre biraz daha düşük olduğunu, ancak özgüllük bakımından benzer seviyede olduğunu göstermektedir. LSTM, zamansal desenleri öğrenme kabiliyeti sayesinde oldukça başarılı sonuçlar vermiştir.



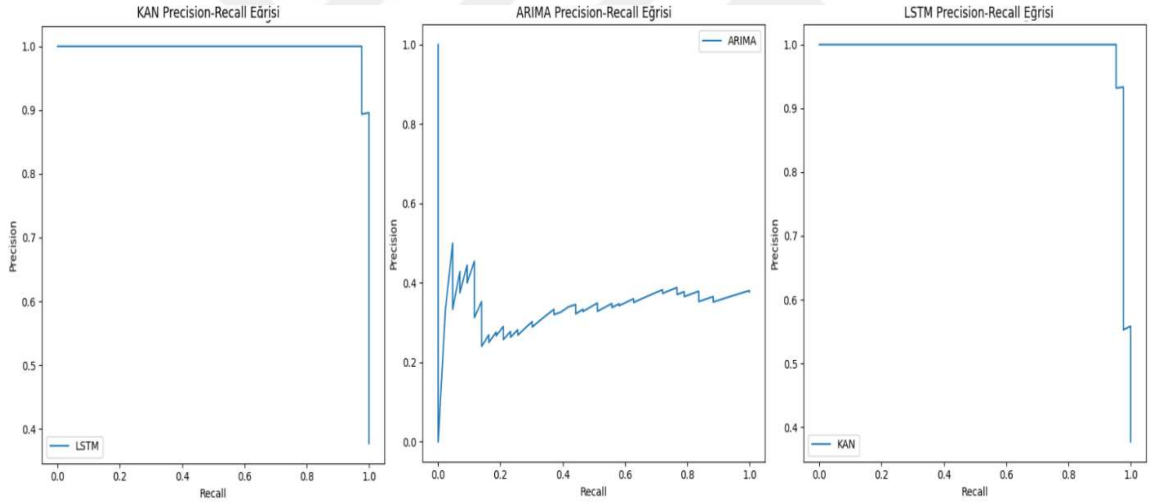
Şekil 4.4 KAN, ARIMA ve LSTM modellerine ait karışıklık matrisleri (Veri seti 1)

Şekil 4.5'te sunulan Precision-Recall (Kesinlik-Duyarlılık) eğrileri, KAN, ARIMA ve LSTM modellerinin meme kanseri teşhisi üzerindeki pozitif sınıfları tanıma başarısını detaylı olarak görselleştirmektedir. Bu eğriler özellikle dengesiz veri setlerinde, ROC eğrilerine göre daha hassas bir değerlendirme sunar; çünkü yalnızca pozitif sınıfın tahmin performansına odaklanır.

KAN modeline ait Precision-Recall eğrisi, neredeyse kusursuz bir performans sergileyerek %100'e yakın kesinlik (precision) ve duyarlılık (recall) ile uzun süre sabit kalmakta, yalnızca uç noktalarda hafif bir düşüş göstermektedir. Bu durum, KAN modelinin pozitif vakaları büyük ölçüde doğru ve eksiksiz şekilde sınıflandırdığını göstermektedir. Grafik, modelin yalnızca çok az sayıda yanlış pozitif ve yanlış negatif üretmiş olduğunu ortaya koyarak, tıbbi uygulamalarda klinik güvenilirliğini desteklemektedir.

ARIMA modelinin eğrisi ise belirgin şekilde dağınık ve zayıftır. Düşük precision seviyeleri, artan recall ile birlikte daha da azalmış ve modelin sınıflandırma kararlılığı bozulmuştur. Bu durum, ARIMA modelinin pozitif sınıfları ayırt etme kapasitesinin son derece düşük olduğunu göstermekte olup, sağlık alanında güvenle kullanılamayacak kadar yetersiz bir model olduğunu ortaya koymaktadır.

LSTM modelinin eğrisi, KAN modeline benzer biçimde yüksek bir kesinlik ve duyarlılık oranı sunmaktadır. Eğri boyunca büyük ölçüde 1.0 seviyesine yakın kalan precision ve recall değerleri, modelin pozitif sınıfları başarıyla tespit ettiğini göstermektedir. Ancak eğrinin sonunda gözlenen küçük bir düşüş, LSTM modelinin bazı uç durumlarda hata yapabileceğini ima etmektedir. Yine de, genel performans açısından LSTM oldukça başarılı bir sınıflayıcı olarak değerlendirilebilir.



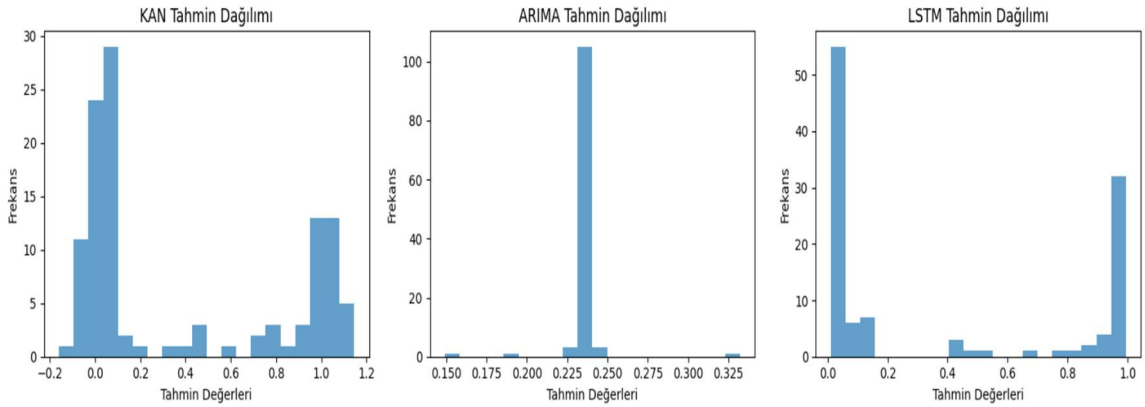
Şekil 4.5 KAN, ARIMA ve LSTM modellerine ait precision recall eğrileri (Veri seti 1)

Şekil 4.6'da, KAN, ARIMA ve LSTM modellerinin meme kanseri teşhisi üzerine ürettikleri tahmin değerlerinin dağılımları histogramlar aracılığıyla sunulmaktadır. Bu dağılımlar, her bir modelin sınıflandırma kararlarını ne kadar net verdiğini, sınıf ayrımında ne derece başarılı olduğunu ve belirsizlik durumlarını nasıl yönettiğini ortaya koymaktadır.

KAN modeline ait tahmin dağılım grafiği incelendiğinde, tahmin değerlerinin büyük oranda 0 (benign) ve 1 (malign) sınıflarına yakın iki farklı kümede toplandığı

görülmektedir. Bu durum, KAN modelinin sınıflar arasında oldukça belirgin bir ayrım yapabildiğini ve belirsizlik yaşamadan güçlü sınıflandırmalar gerçekleştirdiğini göstermektedir. Modele ait tahminlerin yalnızca küçük bir kısmı ara değerlere dağılmış olup, bu durum modelin kararsızlık oranının oldukça düşük olduğunu işaret etmektedir. Bu yapı, modelin ROC eğrisi ve karışıklık matrisi üzerinden elde edilen yüksek başarı oranlarıyla da örtüşmektedir.

ARIMA modeline ait tahmin dağılımı ise modelin sınıflandırma performansındaki zayıflığı açıkça göstermektedir. Tahmin değerlerinin çok büyük bir kısmının 0.225 civarında yoğunlaştığı gözlemlenmektedir. Bu homojen dağılım, modelin pozitif ve negatif örnekleri ayırt etme becerisinin olmadığını, dolayısıyla neredeyse tüm örnekler için benzer ve anlamlı olmayan tahminler ürettiğini ortaya koymaktadır. Bu durum, karışıklık matrisinde görülen yüksek yanlış sınıflandırma oranları ve ROC eğrisindeki düşük AUC değeri ile de doğrudan ilişkilidir. LSTM modelinin tahmin dağılımı ise daha dengeli bir yapıya sahiptir. Model, tıpkı KAN gibi tahminlerini büyük ölçüde 0 ve 1 etiketlerine yakın değerlerde üretmiştir. Ancak bununla birlikte, LSTM modelinde orta düzeyde bir yayılım gözlemlenmekte, bu da modelin bazı örneklerde kararsız kaldığını göstermektedir. Ara değerlerdeki örnekler, LSTM'nin zaman zaman sınıflar arasında kesin ayrım yapmada güçlük çektiğini ve sınıf sınırlarında daha fazla belirsizlik yaşadığını ortaya koymaktadır. Bu yapı, LSTM'nin genel olarak yüksek performans sergilemesine rağmen, KAN modeline kıyasla daha fazla hata yapmasına neden olmuştur.



Şekil 4.6 KAN, ARIMA ve LSTM modellerinin meme kanseri teşhisi üzerine ürettikleri tahmin değerlerinin dağılımları (Veri seti 1)

Çizelge 4.2'de, meme kanseri teşhisine yönelik olarak kullanılan ikinci veri setiyle eğitilen KAN, ARIMA ve LSTM modellerinin MSE değerleri incelendiğinde, KAN modeli 0.1267 ile en düşük değere sahip olup, tahminlerinin gerçek değerlere en yakın ve en düşük sapmayla gerçekleştiğini göstermektedir. Bu sonuç, KAN modelinin tahmin hatalarının hem sistematik hem de istatistiksel açıdan oldukça düşük olduğunu ortaya koymaktadır. LSTM modeli 0.1418 değeriyle ikinci sırada yer alırken, ARIMA modeli 0.3255 ile en yüksek MSE değerini üretmiş ve bu bağlamda en fazla hata yapan model olmuştur. RMSE metrikleri incelendiğinde de benzer bir sıralama söz konusudur. KAN modeli 0.3559 değeriyle en düşük kök ortalama hata oranına sahiptir ve bu durum, tahminlerinin büyüklük olarak da hedefe en yakın çıktığını göstermektedir. LSTM modeli 0.3766 ile KAN'a yakın bir performans sergilerken, ARIMA modelinin 0.5706'lık değeri, hataların daha büyük sapmalar içerdiğini ve modelin genel tahmin doğruluğunun sınırlı olduğunu göstermektedir.

MAE (açısından da en düşük değer 0.2419 ile yine KAN modeline aittir. Bu metrik, modelin yaptığı bireysel tahminlerin ne ölçüde hedef sınıfa yakın olduğunu gösterdiğinden, KAN modelinin genel olarak en doğru tahminleri ürettiği söylenebilir. LSTM modeli 0.2811 ile ikinci sırada yer almakta; ARIMA modeli ise 0.5159 değeriyle açık farkla en fazla mutlak hata üreten model olarak dikkat çekmektedir.

Çizelge 4.2 KAN, ARIMA ve LSTM Modellerinin Hata Metrikleri Karşılaştırması (Veri seti 2)

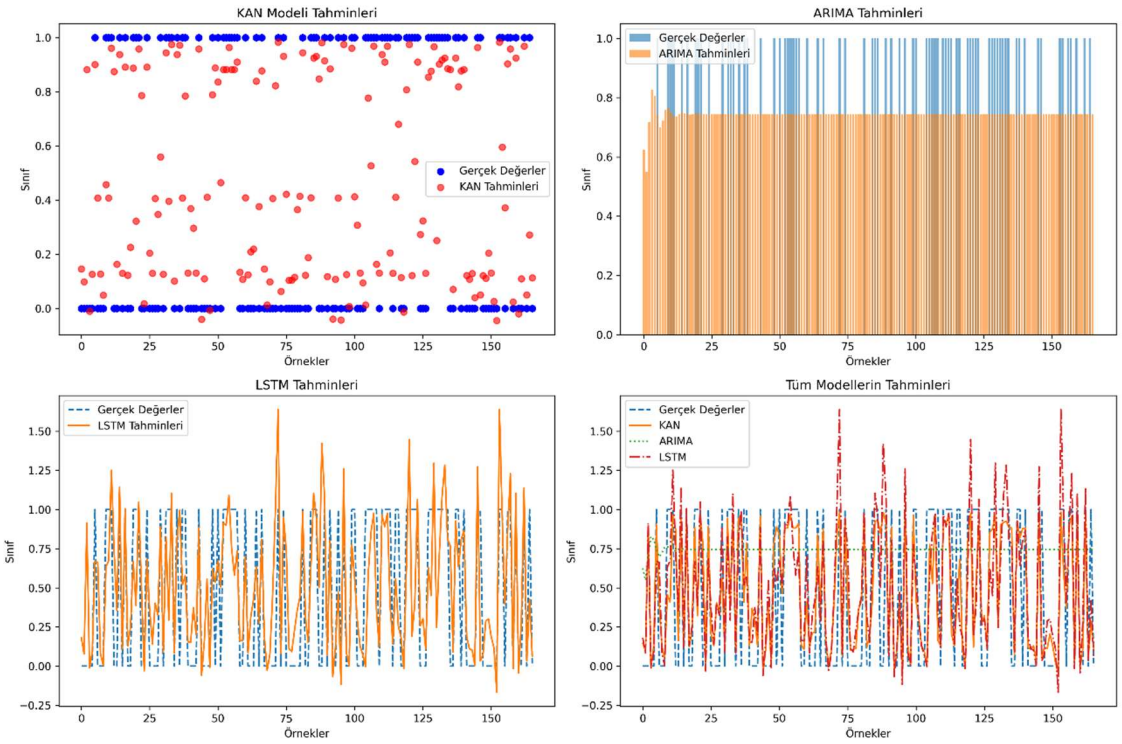
Metrikler	**KAN**	**ARIMA**	**LSTM**
MSE	0.1267	0.3255	0.1418
RMSE	0.3559	0.5706	0.3766
MAE	0.2419	0.5159	0.2811

MSE: Ortalama Kare Hatası,MSE: Kök Ortalama Kare Hatası,MAE: Ortalama Mutlak Hata

Şekil 4.7'de, KAN, ARIMA ve LSTM modellerinin ikinci veri seti üzerinde meme kanseri teşhisine yönelik tahmin çıktıları karşılaştırmalı olarak sunulmuştur. ARIMA modeli, çıktılarında ciddi bir sabitlik eğilimi göstermekte ve bu durum modelin büyük çoğunlukla benzer tahminler üretmesine neden olmaktadır. Gerçek değerlerle arasında ciddi bir tutarsızlık gözlemlenmekte olup, bu durum yanlış negatif ve yanlış pozitif

oranlarının yüksek olmasıyla sonuçlanmıştır. Görselde de açıkça görüldüğü üzere, ARIMA tahminleri gerçek sınıfları yeterince ayırt edememekte ve sınırlı yapısı gereği karmaşık ilişkileri modellemede başarısız kalmaktadır.

LSTM modeli, zaman serisi yapılarının modellenmesinde teorik avantajlar sunsa da, bu çalışmada oldukça dalgalı ve sınıf dışına taşan tahmin değerleri üretmiştir. Bazı örneklerde sıfırın oldukça üzerinde ya da altında çıkan tahminler, modelin sınıf sınırlarına uyum gösteremediğini ve genelleme performansının zayıf kaldığını göstermektedir. Bu da LSTM'nin doğruluk ve özgüllük metriklerinde istikrarsız sonuçlar üretmesine neden olmuştur. KAN modeli, hem grafiksel olarak hem de tahmin değerlerinin gerçek sınıflarla uyumu bakımından en başarılı sonuçları göstermektedir. Tahmin noktalarının doğru sınıf sınırlarında yoğunlaşması, yanlış sınıflandırma oranlarının minimum düzeyde kalmasına katkı sağlamıştır. KAN, özellikle sınıf içi ayrımlar konusunda yüksek doğruluk sergileyerek duyarlılık (sensitivity) ve özgüllük (specificity) gibi metriklerde en iyi performansı göstermiştir. Şekil üzerinde de KAN tahminlerinin, gerçek sınıf etiketleriyle neredeyse birebir örtüştüğü dikkat çekmektedir.

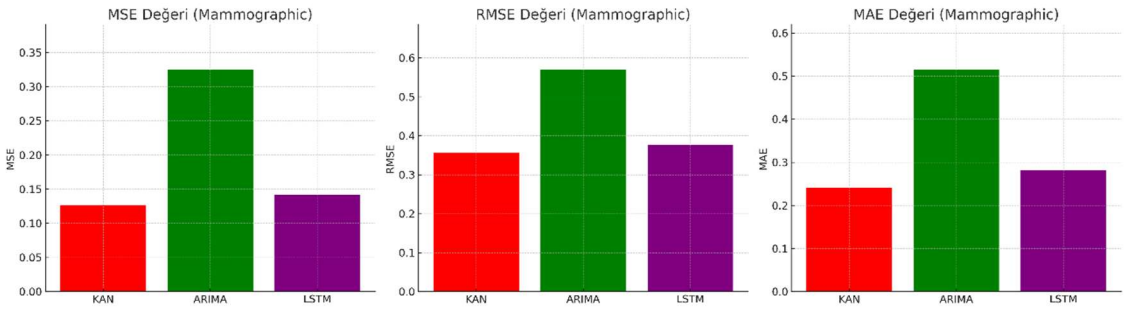


Şekil 4.7 ARIMA, LSTM ve KAN modellerinin meme kanseri tanısında tahmin performansı (Veri seti 2)

Şekil 4.8’de, Mammographic veri seti kullanılarak eğitilen KAN, ARIMA ve LSTM modellerde KAN modeli, tüm hata metriklerinde en düşük değerlere sahiptir. MSE, RMSE ve MAE açısından en az sapma gösteren model olması, KAN’ın tahminlerinin gerçek değerlere en yakın çıktılar ürettiğini kanıtlamaktadır. Bu bulgu, KAN modelinin meme kanseri teşhisinde yüksek doğruluk, düşük varyans ve istikrarlı sonuçlar verdiğini ortaya koymaktadır. Grafiklerde de görüldüğü gibi, bu model, tahmin hatalarını minimumda tutarak güvenilirliğini pekiştirmektedir.

ARIMA modeli, LSTM modeline kıyasla daha başarılı bir performans sergilemiş, ancak KAN’a göre daha yüksek hata oranları üretmiştir. Özellikle RMSE ve MAE değerlerinde orta düzeyde bir hata seviyesi sunan ARIMA, zaman serisi karakteri taşıyan veriler için kabul edilebilir performans ortaya koymuştur. Ancak sınırlı model kapasitesi nedeniyle, ARIMA'nın doğrulukta KAN kadar başarılı olmadığı da açıktır.

LSTM modeli, hem MSE, hem RMSE hem de MAE açısından en yüksek hata değerlerine sahip model olmuştur. Bu durum, LSTM’nin Mammographic veri seti bağlamında yeterli genelleme yapamadığını, dolayısıyla sınıflandırma doğruluğunun düştüğünü göstermektedir. Derin öğrenme tabanlı yapısına rağmen, bu modelin mevcut veri yapısı üzerinde tutarlı ve düşük hatalı tahminler üretmede zayıf kaldığı anlaşılmaktadır.



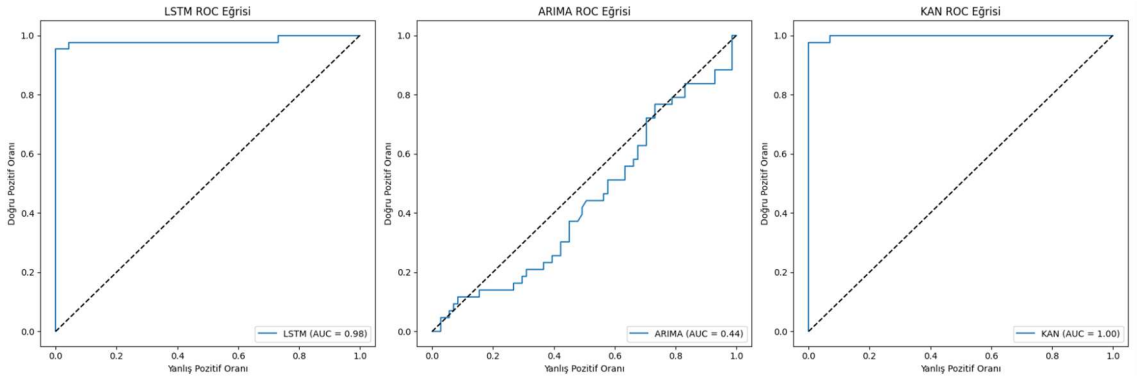
Şekil 4.8 Meme kanseri tanısına yönelik KAN, ARIMA ve LSTM modellerinin hata metrikleri (Veri seti 2)

Şekil 4.9'da, meme kanseri teşhisine yönelik geliştirilen KAN, ARIMA ve LSTM modellerinin ROC eğrileri sunulmuştur. KAN modeli, ROC eğrisinde neredeyse ideal sınıflayıcıya karşılık gelen sol üst köşeye yapışan bir yapıya sahiptir ve AUC = 1.00

değeri ile maksimum performans göstermiştir. Bu durum, KAN modelinin hem yüksek duyarlılık (sensitivity) hem de yüksek özgüllük (specificity) sağlayarak pozitif ve negatif sınıfları çok başarılı biçimde ayırt ettiğini göstermektedir. Elde edilen bu eğri ve AUC değeri, KAN modelinin bu veri seti üzerinde mükemmel genelleme kapasitesine sahip olduğunu ve sınıflandırma performansının en üst seviyede olduğunu ortaya koymaktadır.

ARIMA modeli, ROC eğrisi boyunca diyagonale çok yakın bir seyir izlemekte olup AUC = 0.44 değeriyle şans düzeyinin altında bir başarı sergilemiştir. Bu durum, ARIMA modelinin pozitif sınıfları ayırt etme becerisinin çok zayıf olduğunu ve rastgele bir tahminleyiciden daha kötü performans gösterdiğini açıkça ortaya koymaktadır. Modelin yüksek oranda yanlış pozitif ve yanlış negatif kararlar verdiği bu eğriden anlaşılmaktadır.

LSTM modeli ise ROC eğrisinde üst sol köşeye oldukça yakın bir seyir izlemiş ve AUC = 0.98 değeriyle yüksek bir sınıflandırma performansı göstermiştir. Eğrinin düzgün formu, modelin pozitif sınıfı büyük doğrulukla tahmin ettiğini gösterirken, bazı eşik değerlerinde lokal hatalar yapabileceğini de işaret etmektedir. Yine de AUC değeri, LSTM'nin bu problemde etkili bir sınıflayıcı olduğunu doğrulamaktadır.



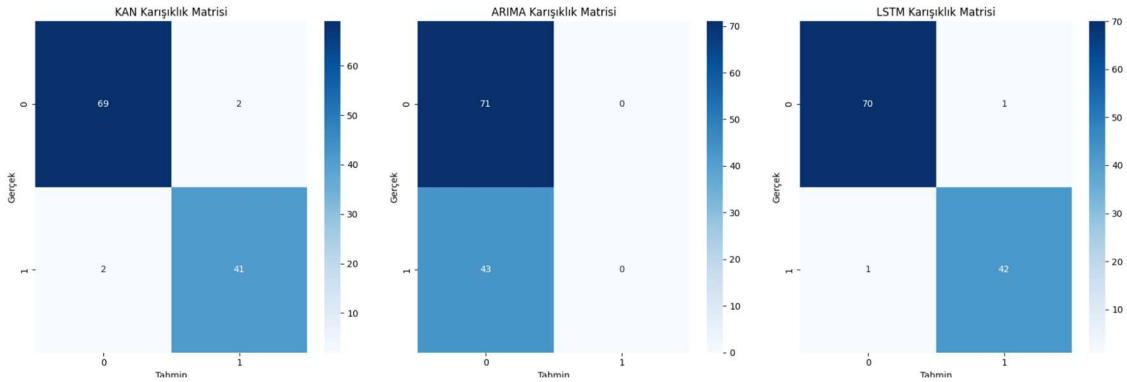
Şekil 4.9 Meme kanseri tanısına yönelik KAN, LSTM ve ARIMA modellerinin ROC eğrileri ve AUC değerleri (Veri seti 2)

Şekil 4.10'da, KAN modeli, 69 doğru negatif ve 41 doğru pozitif tahmin ile yüksek doğruluk oranı sergilemiştir. Model yalnızca 2 yanlış negatif ve 2 yanlış pozitif tahmin yapmıştır. Bu, modelin duyarlılık (sensitivity) ≈ %95.3, özgüllük (specificity) ≈ %97.2, ve doğruluk (accuracy) ≈ %97.4 düzeyinde çalıştığını göstermektedir. KAN modeli hem benign (iyi huylu) hem de malign (kötü huylu) sınıfları dengeli biçimde ayırt edebilmiş;

yanlış sınıflandırma oranlarını minimumda tutmuştur. Klinik uygulamalar açısından bu model, yanlış negatiflerin ve yanlış pozitiflerin düşük tutulması sayesinde güvenilir bir teşhis aracı olarak değerlendirilebilir.

ARIMA modeli, 71 benign vakayı doğru tahmin etse de, malign vakaların tamamını hatalı (FN = 43) şekilde benign olarak tahmin etmiştir. Bu durum, modelin duyarlılığının %0 olduğunu ve pozitif sınıfı ayırt edemediğini göstermektedir. ARIMA yalnızca negatif sınıflarda başarılı olurken, malign vakaların tamamını gözden kaçırmıştır. Bu da ARIMA'nın klinik kullanım açısından son derece riskli bir model olduğunu ortaya koymaktadır. Özellikle hastalık teşhisinde yanlış negatif oranı bu kadar yüksek olan bir modelin güvenilirliği sorgulanmaktadır.

LSTM modeli, 70 doğru negatif ve 42 doğru pozitif tahminle oldukça dengeli bir performans ortaya koymuştur. Yalnızca 1 yanlış negatif ve 1 yanlış pozitif tahmin üretmiştir. Bu sonuçlar doğrultusunda, LSTM modelinin duyarlılığı (sensitivity) ≈ %97.7, özgüllüğü (specificity) ≈ %98.6 ve genel doğruluğu (accuracy) ≈ %98.3 olarak hesaplanabilir. LSTM, KAN modeline yakın bir başarı sergileyerek özellikle zamansal örüntüleri içeren verilerde etkili bir sınıflayıcı olarak değerlendirilmiştir.



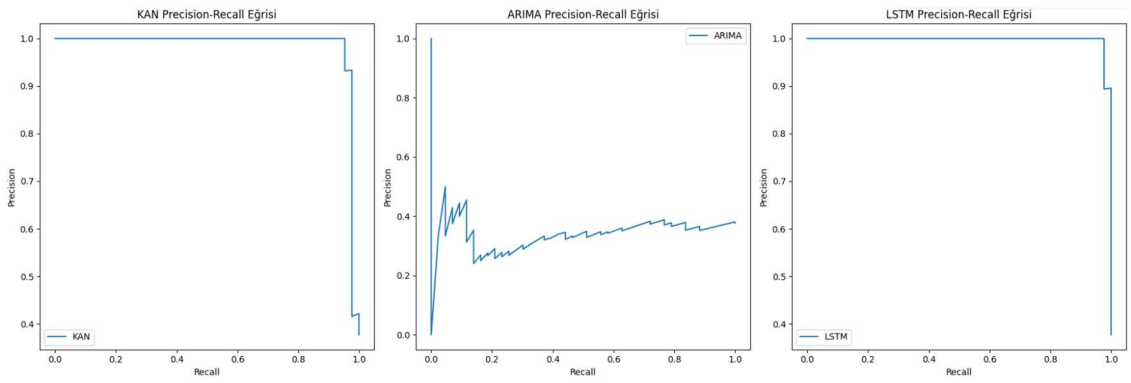
Şekil 4.10 KAN, ARIMA ve LSTM modellerine ait karışıklık matrisleri (Veri seti 2)

Şekil 4.11'de, KAN modeli, precision-recall eğrisinde neredeyse ideal bir profil sergilemiştir. Eğri, uzun bir süre boyunca kesinlik (precision) ve duyarlılık (recall) değerlerini 1.0 düzeyinde korumuş ve yalnızca uç eşik noktalarında küçük düşüşler

göstermiştir. Bu durum, modelin yanlış pozitif ve yanlış negatif sayısını son derece düşük tuttuğunu ve özellikle pozitif (malign) sınıfları güvenilir biçimde tespit edebildiğini göstermektedir. Bu bulgu, KAN modelinin sağlık alanındaki tanısal uygulamalar için klinik olarak oldukça güvenilir bir araç olduğunu ortaya koymaktadır.

ARIMA modeli, precision-recall eğrisi açısından en zayıf performansı gösteren model olmuştur. Eğrinin düşük precision seviyelerinde seyrettiği ve artan recall ile daha da keskin düşüşler yaşadığı gözlenmiştir. Bu dağılım, modelin pozitif sınıfları belirleme konusunda kararsız, tutarsız ve düşük güvenilirlikte tahminler yaptığını göstermektedir. Söz konusu durum, özellikle yanlış negatiflerin hayati önem taşıdığı tıbbi tanılama senaryolarında ARIMA'nın güvenli bir model olmadığını göstermektedir.

LSTM modeli, KAN modeli ile benzer şekilde oldukça başarılı bir precision-recall eğrisi üretmiştir. Model, büyük ölçüde 1.0'a yakın kesinlik ve duyarlılık değerleri sunmuş, yalnızca eğrinin son bölümünde çok az sapma göstermiştir. Bu, LSTM'nin pozitif sınıfları yüksek başarı ile tahmin ettiğini ve sınıflandırma hatalarının çok az olduğunu ifade etmektedir. LSTM modeli, özellikle zaman bağımlı örüntülerin güçlü olduğu veri setlerinde etkili bir sınıflayıcı olarak değerlendirilebilir.

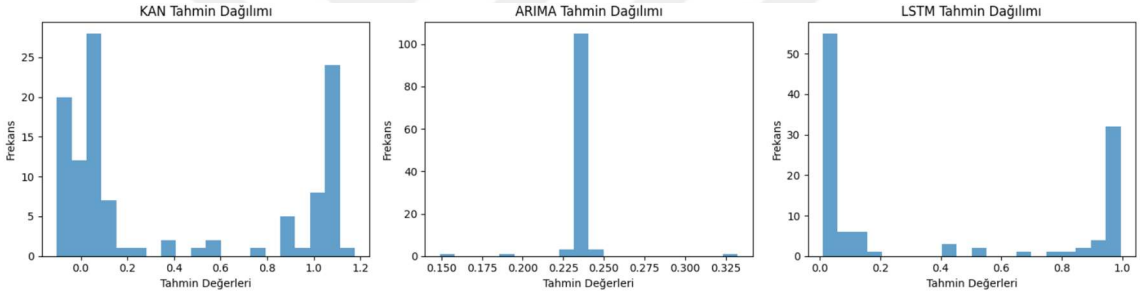


Şekil 4.11 KAN, ARIMA ve LSTM modellerine ait precision recall eğrileri (Veri seti 2)

Şekil 4.12'de, KAN modeli, tahmin değerlerini belirgin biçimde 0 ve 1 çevresinde yoğunlaştırmıştır. Grafik, modelin büyük çoğunlukla ya "benign" (0) ya da "malign" (1) sınıflarına yakın değerler ürettiğini ve sınıflar arasında net bir ayrım yapabildiğini ortaya koymaktadır. Ara tahminlerin sayıca azlığı, KAN modelinin sınıf sınırlarında kararsız kalmadığını ve yüksek sınıflandırma güvenliği sunduğunu göstermektedir.

ARIMA modeli ise tüm modeller arasında en zayıf tahmin dağılımını üretmiştir. Tahmin değerlerinin neredeyse tamamı tek bir noktada, yaklaşık 0.225 civarında toplanmıştır. Bu durum, ARIMA'nın veri setindeki farklı sınıfları ayırt etme konusunda başarısız olduğunu göstermektedir. Model, hem benign hem de malign örnekler için benzer tahminler üretmiş ve böylece belirleyici olmayan, anlamsız çıktılar vermiştir.

LSTM modeli, tahminlerini KAN modeline benzer şekilde uç değerlere yakın üretmiş; yani 0 (benign) ve 1 (malign) civarında belirgin iki yoğunluk alanı oluşturmuştur. Ancak, ara tahmin değerlerinin LSTM'de biraz daha geniş bir dağılıma sahip olduğu görülmektedir. Bu durum, modelin bazı örneklerde sınıf sınırında kaldığını ve kesin bir karar vermekte zorlandığını göstermektedir.



Şekil 4.12 KAN, ARIMA ve LSTM modellerinin meme kanseri teşhisi üzerine ürettikleri tahmin değerlerinin dağılımları (Veri seti 2)

5. TARTIŞMA

Bu çalışmanın bulguları, Wisconsin Meme Kanseri ve Breast Cancer Coimbra veri setleri kullanılarak meme kanseri teşhisine uygulanan üç farklı makine öğrenimi modeli KAN, ARIMA ve LSTM hakkında önemli içgörüler sunmaktadır. Elde edilen sonuçlar, Kolmogorov-Arnold Ağı (modelinin, MSE, RMSE ve MAE gibi tüm hata metriklerinde en düşük değerleri elde ettiğini ve genel olarak en yüksek sınıflandırma performansını gösterdiğini ortaya koymuştur. Bu bulgu, KAN modelinin nöral ağların esnekliğini çekirdek temsilleriyle birleştirme yeteneği sayesinde, karmaşık ve doğrusal olmayan biyomedikal ilişkileri başarılı bir şekilde modelleyebildiğini göstermektedir. Nasser ve Yusof'un (2023) sistematik derlemesinde de hibrit derin öğrenme yöntemlerinin teşhis doğruluğu açısından tekil modellere göre daha yüksek başarı sağladığı vurgulanmıştır.

ARIMA modeli, zaman serisi analizlerinde yaygın olarak kullanılan klasik bir yöntem olmasına rağmen, bu çalışmada sınırlı bir başarı elde etmiştir. LSTM modeline göre daha düşük hata oranları göstermiştir; ancak KAN modeline kıyasla daha düşük doğruluk ve duyarlılık sergilemiştir. ARIMA'nın bu performansı, modelin yalnızca doğrusal desenleri yakalayabilmesi ve biyomedikal verilerde sıklıkla rastlanan karmaşık ilişkileri yansıtamaması ile açıklanabilir. Bununla birlikte, Tarawneh vd. (2022), basit karar algoritmalarının belirli koşullarda oldukça başarılı olabildiğini ve özellikle sınıflandırma kurallarının net olduğu veri setlerinde avantaj sağladığını belirtmektedir.

LSTM modeli, literatürde sıralı verilerle çalışma ve uzun vadeli bağımlılıkları modelleme konusundaki başarısıyla tanınmaktadır (Dutta vd. 2020). Ancak bu çalışmada, beklenenin aksine, en yüksek hata oranlarına sahip olmuştur. Bu durum, modelin hiperparametre ayarlarının yetersiz olması, overfitting gibi yapısal sorunlar ya da veri setinin küçük boyutu ve homojenliği gibi nedenlerle açıklanabilir. Dutta vd. (2020) çalışmasında da benzer şekilde, LSTM temelli modellerin performansının veri seti büyüklüğü ve eğitme stratejilerine büyük ölçüde bağlı olduğu rapor edilmiştir.

Ayrıca, destek vektör makineleri gibi alternatif yöntemlerin, özellikle manyetik rezonans görüntüleme verileriyle yapılan sınıflandırmalarda yüksek başarı sağladığı Vidić vd. (2018) çalışmasında gösterilmiştir. Bu bulgu, klasik yöntemlerin karmaşık tıbbi verilerde

uygun biçimde yapılandırıldığında rekabetçi performans sunabileceğini ve sadece derin öğrenme modellerine odaklanmanın tek çözüm olmadığını ortaya koymaktadır. Çalışmanın sonuçları aynı zamanda, model başarımının yalnızca algoritmanın karmaşıklığıyla değil, veri setinin özellikleriyle olan uyumuyla da doğrudan ilişkili olduğunu göstermektedir. Örneğin, Ciurescu vd. (2025), 2D mamografi verilerinde yapay zekâ uygulamalarıyla elde edilen sonuçların, model mimarisi kadar verinin niteliğiyle de belirlendiğini ortaya koymuşlardır. Bu, model seçiminde sadece teknik parametrelere değil, aynı zamanda alan spesifik uyarlamalara da dikkat edilmesi gerektiğini vurgulamaktadır.

Sonuç olarak, bu çalışma KAN modelinin yüksek performansını vurgularken, klasik yöntemlerin (ARIMA gibi) belirli sınırlamalara sahip olduğunu ve karmaşık modellerin (LSTM gibi) dikkatli bir yapılandırma ve iyileştirme gerektirdiğini ortaya koymaktadır. Sağlık alanındaki uygulamalarda hibrit modellerin daha fazla araştırılması hem teşhis doğruluğunu artırmak hem de klinik karar destek sistemlerinin güvenilirliğini geliştirmek adına önemli bir adımdır. Gelecekteki araştırmalar, daha büyük, çeşitli ve gerçek yaşamdan alınan veri setlerinde bu modellerin yeniden test edilmesine, hiperparametre optimizasyonuna ve yeni özellik mühendisliği tekniklerinin entegre edilmesine odaklanmalıdır.

6. SONUÇ

Bu çalışma, meme kanseri teşhisinde farklı makine öğrenimi modellerinin performanslarını karşılaştırmalı olarak değerlendirmiş ve Kolmogorov-Arnold Ağı (KAN) modelinin hem doğruluk hem de hata metrikleri bakımından diğer modellerden üstün olduğunu ortaya koymuştur. Özellikle MSE, RMSE ve MAE gibi temel ölçütlerde KAN modelinin sağladığı düşük hata oranları, bu yöntemin biyomedikal verilerdeki karmaşık ve doğrusal olmayan ilişkileri etkili biçimde modelleyebildiğini göstermiştir. Bu bağlamda KAN modeli, ARIMA ve LSTM gibi yaygın modelleri geride bırakarak, klinik teşhis süreçlerinde kullanılabilecek güçlü ve güvenilir bir alternatif olarak öne çıkmaktadır.

Araştırma bulguları, yalnızca model performanslarını karşılaştırmakla kalmayıp, makine öğrenimi algoritmalarının uygulandığı bağlama ve veri türüne duyarlılığının ne denli önemli olduğunu da vurgulamaktadır. Sağlık alanında, özellikle meme kanseri gibi kritik hastalıkların erken teşhisi için geliştirilen modellerde, doğruluk oranı kadar güvenilirlik ve yorumlanabilirlik de temel öneme sahiptir. Bu açıdan KAN'ın sunduğu açıklanabilir yapı, kara kutu yaklaşımına kıyasla klinik karar süreçlerinde daha fazla güven sunmaktadır.

Ancak bu çalışma bazı sınırlamalara da sahiptir. Öncelikle, analiz yalnızca Wisconsin Meme Kanseri ve Breast Cancer Coimbra veri setleri üzerinden gerçekleştirilmiş ve tek bir hastalık türü ile sınırlı kalmıştır. Bu nedenle elde edilen bulguların genellenebilirliği sınırlıdır ve farklı veri setleri ile farklı tıbbi durumlarda yeniden test edilmesi gerekmektedir. Ayrıca, LSTM ve ARIMA modelleri için sınırlı sayıda hiperparametre kombinasyonu denenmiş olup, daha derinlemesine optimizasyon veya topluluk modelleri (ensemble learning) ile bu yöntemlerin performansı geliştirilebilir.

Gelecekteki araştırmaların, bu bulguları farklı popülasyonlardan alınmış, daha büyük ve çeşitli veri setlerinde doğrulaması önem arz etmektedir. KAN modelinin diğer ileri düzey derin öğrenme mimarileriyle (örneğin Transformer tabanlı sistemlerle) entegre edilmesi, daha güçlü ve ölçeklenebilir teşhis sistemlerinin geliştirilmesine katkı sağlayabilir. Ayrıca, sadece hücresel özelliklere dayalı verilerle sınırlı kalmaksızın; genetik bilgi,

hastalık geçmişi, yaşam tarzı verileri ve görüntüleme sonuçları gibi çok modlu verilerin modele dahil edilmesi, modelin hem performansını hem de klinik uygulanabilirliğini daha da artıracaktır.

Sonuç olarak bu çalışma, makine öğrenimi modellerinin dikkatli seçimi ve özelleştirilmesinin tıbbi teşhis sistemlerinin doğruluğu, güvenilirliği ve klinik değeri üzerinde belirleyici olduğunu göstermektedir. KAN modeli bu açıdan önemli bir potansiyel sunmakta olup, ileri araştırmalarla desteklenmesi durumunda sağlık alanındaki yapay zekâ temelli teşhis sistemlerinde öncü rol oynayabilecek niteliktedir.

7. KAYNAKLAR

Abdoli G, 2020, Comparing Prediction Accuracy of LSTM and ARIMA, Universidade Federal da Paraíba, 9.

American Cancer Society,, 2020, Breast Cancer Facts & Figures 2019-2020, American Cancer Society, Inc., 76s, Atlanta.

Anderson J A, 1984, A Theory for the Scotch Tape Model, Mathematical Biosciences, 72, 17–29.

Arnold V I, 1957, On Functions of Three Variables, Doklady Akademii Nauk SSSR, 114, 679–681.

Arunkumar S, Kumaravel A, Sivasankar A, 2021, AR Model-Based Time Series Forecasting: A Case Study, International Journal of Advanced Research in Computer Science, 12, 45–52.

Bao W, Yue J, Rao Y, 2017, A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and LSTM, PloS One, 12, e0180944.

Berry D A, Cronin K A, Plevritis S K, Fryback D G, Clarke L, Zelen M, Feuer E J vd., 2005, Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer, New England Journal of Medicine, 353, 1784–1792.

Box G E P, Jenkins G M, Reinsel G C, Ljung G M, 2015, Time Series Analysis: Forecasting and Control, John Wiley & Sons, 712s, New Jersey.

Chen B, Zhu X, Principe J C, 2012, Kernel Adaptive Filtering: A Comprehensive Introduction, Wiley-IEEE Press, 332s, New Jersey.

Chimmula V K R, Zhang L, 2020, Time Series Forecasting of COVID-19 Transmission in Canada Using LSTM Networks, Chaos, Solitons & Fractals, 135, 109864.

Ciurescu S, Cerbu S, Dima C N, Borozan F, Pârvănescu R, Ilaș D G, Sas I vd., 2025, AI in 2D Mammography: Improving Breast Cancer Screening Accuracy, Medicina, 61, 809.

Doi K, 2007, Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential, Computerized Medical Imaging and Graphics, 31, 198–211.

Dutta S, Mandal J K, Kim T H, Bandyopadhyay S K, 2020, Breast Cancer Prediction Using Stacked GRU-LSTM-BRNN, Applied Computer Systems, 25, 163–171.

Esteva A, Kuprel B, Novoa R A, Ko J, Swetter S M, Blau H M, Thrun S, 2017,

Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks, Nature, 542, 115–118.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Bray F vd., 2015, Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012, International Journal of Cancer, 136, E359–E386.

García F, Guijarro F, Oliver J, Tamošiūnienė R, 2023, Foreign Exchange Forecasting Models: ARIMA and LSTM Comparison, Engineering Proceedings, 39, Makale ID: 81.

Gencer K, Basçiftçi F, 2021, Time Series Forecast Modeling of Vulnerabilities in Android OS, Sustainable Computing, 30, 100515.

Gers F A, Schmidhuber J, Cummins F, 2000, Learning to Forget: Continual Prediction with LSTM, Neural Computation, 12, 2451–2471.

Goodfellow I, Bengio Y, Courville A, 2016, Deep Learning, MIT Press, 775s, Cambridge.

Greff K, Srivastava R K, Koutník J, Steunebrink B R, Schmidhuber J, 2016, LSTM: A Search Space Odyssey, IEEE Transactions on Neural Networks and Learning Systems, 28, 2222–2232.

Hamilton J D, 1994, Time Series Analysis, Princeton University Press, 799s, Princeton.

Hochreiter S, Schmidhuber J, 1997, Long Short-Term Memory, Neural Computation, 9, 1735–1780.

Hyndman R J, Athanasopoulos G, 2018, Forecasting: Principles and Practice, OTexts, 382s, Melbourne.

Kandemir-Cavas C, Cavas L, 2019, A Comparative Study of ARIMA and LSTM in Forecasting ICU Bed Occupancy Rates, Procedia Computer Science, 158, 157–162.

Kırbaş İ, Sözen A, Tuncer A D, Kazancıoğlu F Ş, 2020, Comparative Analysis and Forecasting of COVID-19 Cases in Various European Countries with ARIMA, NARNN and LSTM Approaches, Chaos, Solitons & Fractals, 138, Makale ID: 110015.

Kolmogorov A N, 1961, On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables, American Mathematical Society Translations: Series 2, 17, 369-373.

Liu Y, Yao X, Li Y, Wu X, 2020, Kernel Adaptive Filtering: A Comprehensive Survey, IEEE Transactions on Neural Networks and Learning Systems, 31, 1430–1449.

Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljačić M, Tegmark M vd., 2024, Blood: Kolmogorov-Arnold Networks, arXiv preprint arXiv:2404.19756.

Makridakis S, Wheelwright S C, Hyndman R J, 1998, Forecasting: Methods and Applications, John Wiley & Sons, 642s, New York.

Martis R J, Acharya U R, Min L C, Suri J S, 2013, Automated Diagnosis of Atrial Fibrillation Using Bayesian Paradigm, Knowledge-Based Systems, 54, 269–275.

Mougiakakou S G, Prountzou A, Vasiloglou M F, 2010, A Real Time Simulation Model of Glucose-Insulin Metabolism, IEEE Transactions on Biomedical Engineering, 57, 2466–2470.

Nasser M, Yusof U K, 2023, Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction, Diagnostics, 13, Makale ID: 161.

Pandey B, Shukla A, Khamparia A, 2024, An Optimized Hybrid ARIMA-LSTM Model for Time Series Forecasting of Agricultural Production in India, Microbial Data Intelligence, Springer, 107–119, Singapore.

Patricio M, Pereira J, Crisóstomo J, Matafome P, Gomes M, Seiça R, Caramelo F, 2018, Using Resistin, Glucose, Age and BMI to Predict the Presence of Breast Cancer, BMC Cancer, 18, Makale ID: 29.

Schuster M, Paliwal K K, 1997, Bidirectional Recurrent Neural Networks, IEEE Transactions on Signal Processing, 45, 2673–2681.

Shahid F, Zameer A, Muneeb M, 2020, Predictions for COVID-19 with Deep Learning Models of LSTM, GRU and Bi-LSTM, Chaos, Solitons & Fractals, 140, 110212.

Shumway R H, Stoffer D S, 2017, Time Series Analysis and Its Applications: With R Examples, Springer, 562s, New York.

Siami-Namini S, Siami Namin A, 2018, Forecasting Economics and Financial Time Series: ARIMA vs. LSTM, arXiv:1803.06386.

Siegel R L, Miller K D, Wagle N S, Jemal A, 2023, Cancer Statistics, 2023, CA: A Cancer Journal for Clinicians, 73, 17–48.

Skaane P, Bandos A I, Gullien R, Eben E B, Ekseth U, 2008, Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-Based Screening Program, Radiology, 248, 401–410.

Smith R A, Duffy S W, Gabe R, Tabár L, Yen A M F, 2013, The Randomized Trials of Breast Cancer Screening: What Have We Learned?, Radiologic Clinics, 45, 843–857.

Street W N, Wolberg W H, Mangasarian O L, 1993, Nuclear Feature Extraction for Breast Tumor Diagnosis, IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology, 861–870.

Sundermeyer M, Schlüter R, Ney H, 2012, LSTM Neural Networks for Language Modeling, Thirteenth Annual Conference of the International Speech Communication Association, 194–197.

Tarawneh O, Otair M, Husni M, Abuaddous H Y, Tarawneh M, Almomani M A, 2022, Breast Cancer Classification Using Decision Tree Algorithms, International Journal of Advanced Computer Science and Applications, 13, Makale ID belirtilmemiş.

Taslim D G, Murwantara I M, 2022, Comparative Study of ARIMA and LSTM, ICITACEE, IEEE, 231–235.

Topol E J, 2019, High-Performance Medicine: The Convergence of Human and Artificial Intelligence, Nature Medicine, 25, 44–56.

Van Vaerenbergh S, Santamaría I, De Vito E, 2012, A Sliding-Window Kernel RLS Algorithm and Its Application to Nonlinear Channel Identification, International Journal of Adaptive Control and Signal Processing, 26, 306–319.

Verywell Health, 2023, Stages of Breast Cancer: Understanding the Classification System, Verywell Health, https://www.verywellhealth.com/stages-of-breast-cancer-8362458, Erişim Tarihi: 23.05.2025.

Vidić I, Egnell L, Jerome N P, Teruel J R, Sjøbakk T E, Østlie A, Goa P E vd., 2018, Support Vector Machine for Breast Cancer Classification Using Diffusion-Weighted MRI Histogram Features: Preliminary Study, Journal of Magnetic Resonance Imaging, 47, 1205–1216.

Wolberg W H, Street W N, Mangasarian O L, 1994, Machine Learning Techniques to Diagnose Breast Cancer from Fine-Needle Aspirates, Cancer Letters, 77, 163–171.

Xia Y, Xu Z, Hu J, Xing Y, 2020, Bidirectional LSTM for Named Entity Recognition in Clinical Texts, BMC Bioinformatics, 21, 519.

Yeşilyuva A, 2013, Zaman Serisi Analizi ve ARIMA Modeli, Afyon Kocatepe Üniversitesi, Yüksek Lisans Tezi, 85s, Afyonkarahisar.

Zhao Z, Zheng P, Xu S, Wu X, 2019, Object Detection with Deep Learning: A Review, IEEE Transactions on Neural Networks and Learning Systems, 30, 3212–3232.