

Böbrek Hastalıklarının Tanısında Derin Öğrenme Modellerinin Performans Analizi

Serkan ÖZDEMİR^{1*}, Burakhan ÇUBUKÇU²

¹ Bilecik Şeyh Edebali Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği ABD, Bilecik, Türkiye

² Bilecik Şeyh Edebali Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Bilecik, Türkiye

Alındı/Received: 05/02/2025; Kabul/Accepted: 02/06/2025; Yayın/Published: 30/06/2025

*Sorumlu yazar e-posta: serkanhan75@hotmail.com

Öz

Kronik böbrek hastalığı, dünya nüfusunun %10-15'ini etkileyen ve tedavi edilmediğinde ciddi sağlık sorunlarına yol açabilen önemli bir küresel sağlık problemidir. Erken teşhis edilmediğinde böbrek yetmezliği ve kardiyovasküler komplikasyonlara neden olabilir. Bu çalışmada, KBH teşhis ve sınıflandırmasında DL modellerinin etkinliği, CT-KIDNEY veri seti kullanılarak değerlendirilmiştir. InceptionV3, ResNet50, CNN, CNN+LSTM ve VGG16 mimarileri doğruluk, kesinlik, hatırlama ve F1-skor metrikleri ile karşılaştırılmıştır. CNN+LSTM modelinin %91,0 doğruluk ile en yüksek sınıflandırma başarısını gösterdiğini ortaya koymuştur. VGG16 (%90,6), InceptionV3 (%89,6) ve ResNet50 (%88,9) gibi transfer öğrenme tabanlı modeller de %88-%91 aralığında rekabetçi ve yüksek doğruluk oranları elde etmiştir. Buna karşın, standart CNN modeli %27,4 gibi oldukça düşük bir doğruluk oranı sergileyerek, bu spesifik görev için yetersiz kalmıştır. Öğrenme eğrileri transfer öğrenme modellerinin ve CNN+LSTM'in başarılı öğrenme süreçleri sergilediğini, ancak özellikle CNN ve CNN+LSTM'de aşırı öğrenme eğilimleri (CNN-LSTM'de ayrıca test doğruluğunda yüksek varyans) olduğunu göstermiştir. ResNet50 en dengeli öğrenme profilini sunmuştur. Gelecekteki çalışmalarda, daha büyük ve çeşitli veri setleri kullanmayı, IoT tabanlı gerçek zamanlı tanı sistemlerine entegrasyonunu sağlamayı ve farklı tıbbi görüntüleme modalitelerinde uygulanabilirliğini incelemeyi hedefleyebilir.

Anahtar Kelimeler: Kronik böbrek hastalığı, derin öğrenme, CT-KIDNEY, sınıflandırma, evrimsel sinir ağları, öğrenme aktarımı.

Performance Analysis of Deep Learning Models in the Diagnosis of Kidney Diseases

Abstract

Chronic kidney disease is a major global health problem affecting 10-15% of the world's population and can lead to serious health problems if left untreated. If not diagnosed early, it can lead to kidney failure and cardiovascular complications. In this study, the effectiveness of DL models in diagnosing and classifying CKD was evaluated using the CT-KIDNEY dataset. InceptionV3, ResNet50, CNN, CNN+LSTM and VGG16 architectures were compared using accuracy, precision, recall and F1 score metrics. The CNN+LSTM model showed the highest classification success with an accuracy of 91.0%. Transfer learning based models such as VGG16 (90.6%), InceptionV3 (89.6%) and ResNet50 (88.9%) also achieved competitive and high accuracy rates in the range of 88%-91%. In contrast, the standard CNN model was inadequate for this particular task, with a very low accuracy of 27.4%. The learning curves showed that the transfer learning models and CNN+LSTM exhibited successful learning processes, but there were tendencies to overlearn, especially in CNN and CNN+LSTM (CNN-LSTM also showed a high variance in test accuracy). ResNet50 showed the most balanced learning profile. Future work could aim at using larger and more diverse datasets, integrating them into IoT-based real-time diagnostic systems, and investigating their applicability in different medical imaging modalities.

Key Words: Chronic kidney disease, deep learning, CT-KIDNEY, classification, convolutional neural networks, transfer learning.

Atıf / Töcite: Demirel Yılmaz B, Eren B, Sağır D, Başardı Gökçe A, Mercan S (2024). Histological Evaluation of the Protective Role of β -glucan Against Cisplatin-Induced Hepatotoxicity. Karamanoğlu Mehmetbey University Journal of Engineering and Natural Sciences, 6(1): 46-51.

1. GİRİŞ

Kronik böbrek hastalığı (KBH), böbreklerin kanı yeterince filtreleyememesi sonucu ortaya çıkan bir durumdur. Böbrekler, fazla suyu ve zararlı maddeleri kandan süzerek idrar yoluyla vücuttan uzaklaştırır. Ancak, KBH durumunda bu işlev bozulur ve atıklar vücutta birikerek bel ağrısı, ishal, burun kanaması, ateş, deri döküntüleri, kusma ve karın ağrısı gibi belirtilere neden olur. Hastalık zamanla ilerleyerek vücudun diğer sistemlerini etkiler ve çeşitli sağlık sorunlarına yol açar. İleri evrelere ulaştığında ise ölümcül olabilecek ciddi sonuçlara neden olabilir. Dünya genelinde yaklaşık 850 milyon kişinin farklı nedenlerden dolayı böbrek hastalıklarına sahip olduğu tahmin edilmektedir (Lal ve ark. 2024). 2019 Dünya Böbrek Günü raporuna göre, her yıl en az 2,4 milyon kişi böbrek hastalıkları nedeniyle hayatını kaybetmekte ve bu hastalıklar, dünya çapında en hızlı artan ölüm nedenlerinden biri olarak 6. sırada yer almaktadır. KBH, özellikle düşük gelirli ülkelerde teşhis, önleme ve tedavi imkanlarının sınırlı olması nedeniyle giderek büyüyen bir halk sağlığı sorunu haline gelmiştir (Goel 2022).

Son yıllarda tıbbi görüntüleme ve biyomedikal veri analizinde derin öğrenme algoritmalarının kullanımı, böbrek hastalıklarının tanı ve tedavi süreçlerini iyileştirme potansiyeli sunmaktadır. Derin öğrenme, yapay sinir ağları kullanarak karmaşık veri yapılarından anlamlı özellikler çıkarma yeteneğine sahip güçlü bir makine öğrenimi yöntemidir. Özellikle, manyetik rezonans görüntüleme (MRI), bilgisayarlı tomografi (BT) ve ultrason gibi tıbbi görüntüleme modalitelerinde böbrek yapılarının ve hastalığa özgü anomalilerin analizinde önemli bir rol oynamaktadır. KBH, böbrek fonksiyonunun kademeli olarak azalması ve bu durumun böbrek hasarına veya yetmezliğine yol açabilmesidir. Hastalık ilerledikçe teşhis koymak daha zor hale gelir. Doktor görüşmeleri ve veri kullanılarak KBH'nin çeşitli aşamalarını değerlendirmek, erken teşhis ve müdahaleyi kolaylaştırabilir. Bu amaçla, araştırmacılar öğrenme sürecinden ilham alarak KBH'yi kategorize etmek için bir optimizasyon tekniği önermektedir. Yapay zekâ, dünyadaki pek çok şeyi mümkün kılma potansiyeline sahip olup, bazı doktorlar, hastaların vücutlarını tarayarak hastalıklarını analiz edebilecek teknolojilere heyecanla yaklaşmaktadır (Venkatrao ve Kareemulla 2023).

Bu zorlukların üstesinden gelmek için derin öğrenme (DL) modelleri umut verici bir çözüm olarak ortaya çıkmıştır. Derin öğrenme tekniklerinin KBH tespitinde uygulanması, tıp alanında ve hasta sonuçlarında büyük bir potansiyele sahiptir. Derin öğrenme algoritmaları, tanı hassasiyetini ve hızını artırarak tıp alanında daha hızlı ve doğru teşhisler yapılmasına yardımcı olabilir. Bu tekniklerin kullanımıyla, tıp profesyonelleri teşhis hatalarını azaltabilir, tedavi planlamasını iyileştirebilir ve nihayetinde KBH bağlamında hasta bakımı ve sonuçlarını geliştirebilir. Derin öğrenmenin teşhis süreçlerine entegrasyonu, sağlık uygulamalarını büyük ölçüde iyileştirme potansiyeline sahiptir. Bu teknoloji, tıp uzmanlarına hastaların prognozunu ve genel sağlığını

olumlu yönde etkileyebilecek değerli bir araç sunabilir (Zhang ve ark. 2021).

Ayrıca, çeşitli ve geniş veri setlerinden yararlanarak bu modeller, tıbbi görüntülerde insan gözünden kaçabilecek karmaşık desenleri algılayabilir. Bu yetenek, KBH işaret eden ince anormallikleri, özellikle de tedavinin en etkili olduğu erken aşamalarda tespit etmek için çok önemlidir. KBH tanı ve tedavisi için doğruluğu yüksek sınıflandırılması yenilikçi yaklaşımlar gerektirir. Özellikle konvolüsyonel sinir ağları (CNN) gibi derin öğrenme teknikleri, tıbbi görüntü analizinde devrim yaratmıştır. Bu algoritmalar, tıbbi görüntülerden karmaşık desenleri ve özellikleri çıkarmada üstün bir performans sergileyerek glomerüler anormalliklerin kesin şekilde tanımlanmasını sağlamaktadır (Elhoseny ve ark. 2019).

Bu çalışmanın amacı, KBH, derin öğrenme algoritmaları kullanılarak verimli bir şekilde tahmin ve sınıflandırma yoluyla teşhis edilip edilemeyeceğini analiz etmektir. Bu hedefe ulaşmak için çalışmada beş farklı derin öğrenme algoritması ResNet50, VGG, InceptionV3, CNN ve CNN+LSTM uygulanmıştır. Bu algoritmalar, KBH hastalıklarını sınıflandırmada doğruluk ve hata oranları açısından kapsamlı bir şekilde karşılaştırılmıştır. Bu bağlamda, çalışmanın literatüre temel katkıları şu şekilde özetlenebilir:

- ResNet50, VGG, InceptionV3, CNN ve CNN+LSTM modellerinin KBH teşhisindeki performansları sistematik olarak analiz edilmiştir.
- InceptionV3 modelinin %0,9840 doğruluk oranıyla veri kümesinde üstün performans sergilediği gösterilmiş ve KBH teşhisinde en uygun model olarak önerilmiştir.
- Çalışma, yalnızca doğruluk metriklerini değil, aynı zamanda kesinlik, geri çağırma ve F1-Skor gibi ek değerlendirme kriterlerini kullanarak kapsamlı bir analiz sunmuştur.

Makalenin geri kalanı şu şekilde yapılandırılmıştır: Bölüm 2, makine öğrenimi (ML) ve derin öğrenme (DL) algoritmalarını kullanan araştırmacılar tarafından gerçekleştirilen ilgili çalışmaları ve mevcut literatürdeki eksiklikleri incelemektedir. Bölüm 3, çalışmanın materyal ve metodunun detaylı açıklamalarıyla birlikte ele alınmasını sağlamaktadır. Bölüm 4, elde edilen bulguları, bölüm 5'te ayrıntılı bir tartışma ile birlikte çalışmanın sınırlamaları sunulmaktadır. Son olarak bölüm 6, araştırmanın sonuçlarını özetlemektedir.

2. İLGİLİ ÇALIŞMALAR

Birçok çalışma, derin öğrenme modelleri kullanarak KBH'yi tespit ve analiz etmeye çalışmıştır. Mevcut bazı çalışmaların özeti aşağıdaki şekilde sunulmuştur:

KBH taranması, yüksek gelirli ülkelerde bile zor bir süreçtir. Sabanayagam ve ark. (2020) çalışmalarında, retina görüntülerinden kronik böbrek hastalığını tespit etmek için bir yapay zekâ derin öğrenme algoritması (DLA) geliştirilmiştir. Singapur ve Çin'deki üç kesitsel çalışmadan elde edilen verilerle, DLA, risk faktörleri (yaş,

cinsiyet, etnik köken, diyabet ve hipertansiyon) ve bu ikisinin birleşimi olan hibrit bir model kullanılarak üç farklı model eğitilmiştir. Modeller, alıcı işletim karakteristik eğrisinin altındaki alan (AUC) ile değerlendirilmiştir. DLA hem geliştirme hem de doğrulama aşamalarında başarılı performans göstermiş ve kronik böbrek hastalığı tanısına katkı sağlayabilecek bir araç olarak önerilmiştir. Böbrek hastalığının tespiti için gelişmiş makine öğrenimi algoritmaları geliştirilmiştir.

Kumar ve ark. (2023) fuzzy derin sinir ağları kombinleyen yeni bir derin öğrenme modeli sunmakta olup, böbrek hastalığının tanı ve tahmininde kullanılmaktadır. Sonuçlar, önerilen modelin 0,9923 doğruluk oranı ile mevcut yöntemlerden daha başarılı olduğunu göstermektedir. Önerilen yaklaşım, mevcut bilgi madenciliği sınıflandırmalarına kıyasla sınıflandırma doğruluğu, hassasiyet, F1-Skor ve duyarlılık metriklerinde iyileştirme sağlamaktadır. Diyabet ve yüksek tansiyon, KBH için ana nedenlerdir. KBH'nin erken tespiti, doktorlar için zorlayıcıdır.

Singh ve ark. (2022) araştırmalarında, KBH'nin erken tanısı için yeni bir derin öğrenme modeli önerilmiştir. Model, Recursive Feature Elimination (RFE) ile seçilen önemli özellikler (Hemoglobin, Serum Kreatinin, Hipertansiyon vb.) kullanarak, sınıflandırma amacıyla makine öğrenimi modellerine verilmiştir. Sonuçlar, önerilen derin öğrenme modelinin %100 doğrulukla diğer dört sınıflandırıcıyı (SVM, KNN, Lojistik Regresyon, Naif Bayes) geride bıraktığını göstermektedir. Bu yaklaşım, KBH tespiti için nefrologlar için yararlı bir yaklaşımdır.

Zhang ve ark. (2024) çalışmalarında, böbrek taşı teşhisinde erken tanı için derin öğrenme ve metaheuristik optimizasyon tekniklerini birleştiren bir metodoloji önerilmektedir. InceptionV4 modelinin performansı, Combined Dwarf Mongoose Optimization (CDMO) algoritması ile optimize edilmiştir. Bu metodoloji, CT Böbrek Veri setine uygulanarak mevcut yöntemlerle karşılaştırılmış ve 0,9814 doğruluk, 0,9774 hassasiyet, 0,9633 özgüllük ve 0,9251 F1-skor ulaşılmıştır. Sonuçlar, önerilen yaklaşımın böbrek taşı teşhisinde yüksek doğruluk ve etkinlik sunduğunu göstermektedir.

Kumar ve ark. (2023) koronal CT görüntülerinde böbrek taşlarını otomatik tespit etmek için derin öğrenme ve bilgisayar tabanlı bir teşhis sistemi geliştirmişlerdir. Önerilen yöntem, giriş görüntülerinden soyut ve derin özellikler çıkararak 0,9856 doğruluk elde etmiştir ve CNN, ResNet, AlexNet gibi geleneksel yaklaşımları geçmiştir. 1799 koronal CT taraması kullanılarak yapılan testlerde, model 10 katlı çapraz doğrulama (CV) ile doğrulanmıştır. Otomatik model, BT görüntülerini kullanarak böbrek taşlarını 0,9856 doğrulukla tespit etmiştir.

Tahir ve Abdulrahman (2023) böbrek taşlarını tespit etmek için Discrete First Chebyshev Wavelets Transform (DFCWT) tekniği ve CNN kombinasyonu önermişlerdir. DFCWT, tıbbi görüntülerdeki safsızlıkları ortadan kaldırarak görüntü kalitesini artırmıştır. DFCWTCNN algoritması, derin öğrenme ve DFCWT kullanarak böbrek taşlarını başarıyla tespit etti ve bu yaklaşımın diğer tıbbi

durumlarda da uygulanabileceğini göstermiştir. Algoritma, 0,9884 doğruluk elde etmiştir.

Sundaramoorthy ve Jayachandru (2023) böbrek taşları, kistler ve tümörlerin tanısında derin öğrenme kullanarak otomatik sınıflandırma yapan bir model geliştirmişlerdir. Model, VGG-16 ve AlexNet'ten ilham alarak %79 doğruluk elde etmiştir. CT tarama görüntülerinin doğruluk için önemli olduğu vurgulandı, ancak ultrason ve MR görüntüleri de dikkate alınmıştır. Görüntü işleme ve gürültü kaldırma teknikleri ile doğruluk artırılmıştır. Ancak modelin mevcut yöntemlerle karşılaştırması yapılmadı ve CT tarama görüntülerine bağımlılık, modelin sınırlı ortamlarda uygulanabilirliğini kısıtlayabilir.

Sapra ve ark. (2023) ultrason görüntülerinden böbrek yapılarının çıkarılmasını amaçlayan bir sistem geliştirmişlerdir. Yöntem, görüntü kalitesini artırmak için Gabor filtresi ve histogram eşitleme kullanılmıştır. Hücre segmentasyonu ve bölge temelli segmentasyon karşılaştırılarak en iyi sonuç bölge temelli segmentasyonla elde edilmiştir. Sonrasında CNN-ELM modeliyle böbrek alanı üzerinde eğitim yapıldı ve 0,9850 doğruluk elde edilmiştir. Ancak, çalışma farklı veri setleri üzerinde test edilmemiştir.

Rani ve ark. (2023) CT görüntülerinden böbrek taşlarını tespit etmek için bir yöntem sunmuşlardır. Bu yöntem, ön işleme, segmentasyon, özellik çıkarımı ve sınıflandırma aşamalarını içermektedir. 3×3 medyan filtre ve DWT ile tuz ve karabiber gürültüsü temizlendikten sonra taşlar watershed segmentasyonu ile izole edilmiştir. Özellik çıkarımı GLCM kullanılarak yapıldı ve sınıflandırma PNN ile gerçekleştirilmiştir. Yöntem, 0,868 doğruluk oranı ile başarılı olmuştur. Ancak, çalışma farklı veri setlerinde doğrulanmamıştır. Yüzey güçlendirilmiş Raman spektroskopisi (SERS), hızlı ve non-invaziv bir teknik olarak hastalık teşhisinde umut vaat etmektedir.

Yang ve ark. (2023) tavllanmış AgNPs/PSB SERS alt tabakası geliştirilmiş ve AlexNet, ResNet, TCN, MCNN gibi derin öğrenme algoritmalarıyla birleştirilmiştir. MCNN, 0,9470 ve 0,9200 doğruluk oranlarıyla en iyi sonucu vermiştir. Bu yöntem, bağışıklık ve kronik hastalıkların erken teşhisinde önemli bir potansiyel taşımaktadır. KBH, erken tanı ve kesin tahmin gerektiren önemli bir küresel sağlık sorunudur.

Jeyalakshmi ve ark. (2024) genetik, klinik ve görüntüleme verilerini değerlendiren yenilikçi bir DL sistemi sunmuşlardır. Önerilen model, karmaşık desenleri otomatik olarak çıkararak biyobelirteç tespitinde duyarlılık (0,945), özgüllük (0,922), doğruluk (0,908) ve AUC-ROC (0,94) ile mevcut yöntemleri geride bırakmıştır. Çalışma, CKD'nin erken teşhisi ve bireyselleştirilmiş tedavi için derin öğrenme ile devrim niteliğinde bir adım sunmaktadır.

KBH, dünya nüfusunun %10-15'ini etkileyen ve ciddi sağlık sorunlarına yol açan önemli bir hastalıktır. CKD'nin erken ve doğru tespiti, komplikasyonların azaltılması için kritik öneme sahiptir.

Debal ve Sitote (2022) çalışmalarında CKD evrelerinin ikili ve çoklu sınıflandırması için Random Forest (RF), Destek Vektör Makinesi (SVM) ve Karar Ağacı (DT)

kullanılmıştır. Özellik seçimi için özyinelemeli eleme ve çapraz doğrulama uygulanmış, modeller on katlı çapraz doğrulama ile değerlendirilmiştir. RF, SVM ve DT'den daha iyi performans göstermiştir.

Akter ve ark. (2022), çalışmalarında KBH'nin tahmin ve sınıflandırılması için yedi derin öğrenme algoritması (ANN, LSTM, GRU, Bidirectional LSTM, Bidirectional GRU, MLP, Simple RNN) kullanılmıştır. Algoritmalar, pre-proses edilmiş CKD veri setlerinden özellikler çıkarılarak uygulanmıştır. Sonuçlar, ANN, Simple RNN ve MLP algoritmalarının sırasıyla %99, %96 ve %97 doğruluk sağladığını ve modelin geleneksel yöntemlere göre üstün performans gösterdiğini ortaya koymuştur. Çalışma, KBH tahmini için derin öğrenme algoritmalarının entegrasyonunu ve Internet of Medical Things (IoMT) kullanımını öneriyor.

Alsubibany ve ark. (2021) IoT ortamında KBH tanısı için derin öğrenmeye dayalı klinik karar destek sistemleri (EDL-CDSS) geliştirilmişlerdir. EDL-CDSS, IoT cihazlarıyla toplanan tıbbi verileri kullanarak CKD'nin evrelerini tespit ve sınıflandırmak için derin inanç ağı (DBN), kernel ekstrem öğrenme makinesi (KELM) ve CNN-GRU modellerini birleştirir. Aykırı değer tespiti için Adaptive Synthetic (ADASYN) ve hiperparametre ayarı için kuazi-opozisyonel kelebek optimizasyon algoritması (QOBOA) kullanılmıştır. Simülasyonlar, EDL-CDSS'nin mevcut yöntemlere göre üstün performans gösterdiğini ortaya koymuştur.

Saif ve ark. (2023) çalışmalarında, KBH'nin 6 veya 12 ay öncesinde tahmin edilmesi için üç model önerilmiştir: CNN, LSTM ve derin ansambl model. Ansambl model, CNN, LSTM ve LSTM-BLSTM modellerini çoğunluk oylama ile birleştirir. İki farklı veri seti üzerinde yapılan deneylerde, derin ansambl modelinin 6 aylık ve 12 aylık tahminlerde sırasıyla 0,993 ve 0,992 doğrulukla en iyi sonuçları verdiği bulunmuştur.

Ma ve ark. (2020) Heterojen Modifiye Yapay Sinir Ağı (HMANN), Kronik Böbrek Yetmezliğinin erken tespiti, segmentasyonu ve teşhisi için IoMT platformunda önermişlerdir. Ayrıca, önerilen HMANN, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı (MLP) ile Backpropagation (BP) algoritmasıyla sınıflandırılmaktadır. Önerilen algoritma, ultrason görüntüsü üzerinde çalışmakta olup, görüntüdeki böbrek ilgi alanı segmentasyonu gerçekleştirilir. Böbrek segmentasyonunda, önerilen HMANN yöntemi yüksek doğruluk elde etmekte ve konturun belirlenme süresini önemli ölçüde azaltmaktadır. KBH sınıflandırmasında, temel özellikleri belirlemek için İstatistiksel Olarak Eşdeğer İmza (SES) yöntemi kullanılmıştır. SES, birden fazla prediktif özellik alt kümesi tanımlayabilir.

Yogesh ve ark. (2024) LASSO ile birleşik özellik seçimi ve LSTM ile GRU ağlarını içeren topluluk derin öğrenme modeli önermişlerdir. Modelin performansı, doğruluk, keskinlik, geri çağırma ve F1-Skoru ile değerlendirilmiş ve hibrit yöntem, bireysel sınıflandırıcılara göre %2 daha iyi sonuçlar vermiştir. Ayrıca bazı özelliklerin sınıflandırma üzerinde minimal etkisi olduğu bulunmuştur.

Çalışmalar, KBH ve böbrek taşlarının erken tespiti ve tanısı için yapay zekâ ve DL yöntemlerinin etkili olduğunu göstermektedir. Ancak mevcut literatür incelendiğinde bazı önemli boşluklar dikkat çekmektedir.

Literatürdeki çalışmaların birçoğu, sınırlı veri setleri üzerinde değerlendirilmiş ve farklı veri setleri veya popülasyonlarda genel performans yeterince test edilmemiştir. Benzer şekilde bizim çalışmamız da CT-KIDNEY veri seti üzerinde odaklanmış olup, çalışmamızın birincil amacı bu spesifik veri seti üzerinde kapsamlı bir model karşılaştırması gerçekleştirerek en uygun derin öğrenme yaklaşımını belirlemektir.

Mevcut literatürde kullanılan modellerin klinik ortamlarda uygulanabilirliği ve gerçek zamanlı tanı süreçlerine entegrasyonu hakkında sınırlı bilgi bulunmaktadır. Bu önemli boşluk, çalışmamızın gelecek araştırma yönerimiz olarak belirlenmiş olup, sonuç bölümünde de belirttiğimiz gibi ileride IoT tabanlı gerçek zamanlı tanı sistemleri ile entegrasyon araştırmaları planlanmaktadır.

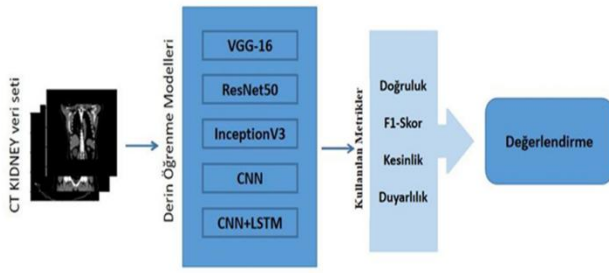
Bazı modeller yüksek doğruluk oranları sunsa da çoğu çalışma diğer yöntemlerle doğrudan karşılaştırma yapmamış veya kullanılan özellik seçimi tekniklerinin biyolojik anlamlılığını detaylandırmamıştır. Çalışmamız bu noktada literatüre katkı sağlamakta olup, beş farklı derin öğrenme modelini (InceptionV3, ResNet50, CNN, CNN+LSTM ve VGG16) aynı veri seti üzerinde sistematik olarak karşılaştırmıştır.

Özellikle, KBH'nin farklı evrelerinin ayrımı ve taş teşhisinde coğrafi, genetik ve klinik değişkenlerin etkisinin daha derinlemesine araştırılması gerekmektedir. Ayrıca, yöntemlerin IoT tabanlı sağlık izleme platformlarına entegrasyonu gibi yenilikçi yaklaşımlarda teknik zorluklar ve maliyet unsurları da ele alınmamıştır.

Gelecekteki araştırmalarda, bu çalışmada elde ettiğimiz bulguları temel alarak, daha geniş ve çeşitlendirilmiş veri setlerinde kapsamlı testlerin yapılması, model genelleştirilebilirliğinin artırılması, biyolojik olarak anlamlı özellik seçiminin teşvik edilmesi ve gerçek zamanlı tanı sistemlerinin geliştirilmesi hedeflenmektedir. Bu adımlar, KBH ve böbrek taşı teşhisi alanında ilerleme sağlamak için kritik gerekliliklerdir ve araştırma programımızın gelecek fazlarında ele alınacaktır.

3. MATERYAL VE METOD

Bu çalışmada, KBH analizi CT KIDNEY veri seti üzerinde çeşitli derin öğrenme yöntemleri kullanılarak gerçekleştirilmiştir. Çalışmada kullanılan veri seti, açık kaynaklı bir veri seti olan böbrek taşı, tümör, normal ve kist bulgularına ait bir veri setidir. Bu veri seti üzerinde Inception-V3, ResNet50, CNN, CNN + LSTM ve VGG16 modelleri kullanılmıştır. Testler, NVIDIA GeForce GTX 860M, Intel(R) HD Graphics 4600 ve 64GB Sistem RAM 'ine sahip, Google Colab ortamında gerçekleştirilmiştir. Çalışmanın genel akışı Şekil 1'de sunulmaktadır. Bu bölümün devamında veri seti ve kullanılan yöntemler ile ilgili detaylar verilmektedir.



Şekil 1. Çalışmanın genel akış şeması

3.1. CT KIDNEY Veri Seti

Veri seti, Bangladeş'in Dakka şehrindeki bir hastanenin PACS (Resim Arşivleme ve İletişim Sistemi) ve çalışma istasyonlarından toplanmıştır. Bu veri setinde yer alan hastalar, böbrek tümörü, kist, normal bulgular veya taş tanısı almış bireylerdir. Araştırmaya katılan tüm kişiler, veri toplama öncesinde bilgilendirilmiş onam alarak gönüllü olmuşlardır. Veri toplama ve deneyler, Dakka Central International Medical College and Hospital (DCIMCH) yetkilileri tarafından onaylanmış ve geçerli düzenlemelere uygun şekilde yapılmıştır. Veri toplama sürecinde hem kontrastlı hem de kontrastsız çalışmalarla yapılan koronal ve aksiyal kesitler seçilmiştir. Dicom çalışmaları dikkatlice incelenip, her bir teşhis için ilgili bölgeyi içeren bir Dicom görüntü grubu oluşturulmuştur. Ardından, hastaların kişisel bilgileri ve meta verileri görüntülerden çıkarılmış ve Dicom görüntüleri kayıpsız JPEG/JPG formatına dönüştürülmüştür. Veri etiketleme, Philips IntelliSpace Portal 9,0 aracılığıyla yapılmış ve dönüştürme işlemi için Sante Dicom Editor aracı kullanılmıştır. Etiketleme işlemi sonrasında, görüntüler doktorlar ve tıbbi teknologlar tarafından yeniden kontrol edilmiştir. Oluşturulan veri seti, toplamda 12,446 benzersiz veri içermekte olup, bu veriler arasında 3,709 kist, 5,077 normal, 1,377 taş ve 2,283 tümör bulgusu yer almaktadır. Bu veri seti, Kaggle platformuna yüklenmiş ve diğer araştırmacıların kullanabilmesi amacıyla kamuya açık hale getirilmiştir. Veri setinden alınan örnekler, radyologların tanı koyarken kullandığı bulgu alanlarını kırmızı işaretlerle göstermektedir (Islam ve ark. 2022). Bu sayıdaki resmi işlemekte bilgisayarlar zorlandığı için her bir veriden 288 tane alınarak bu çalışma yürütüldü. Bu şekilde az veriseti ile bu modellerin çalışma performanslarının karşılaştırma imkanı elde edildi.

3.2. Veri Hazırlama ve Ön İşleme

Görsel verilerin yeniden boyutlandırılması, orijinal görüntülerin belirli bir boyuta yeniden örneklenmesi süreci olup, işleme ve analiz verimliliğini artırmak için kritik bir ön işleme adıdır. Bu çalışmada, tüm karın ve ürogram CT taraması görüntüleri, 224×224 piksel boyutuna yeniden boyutlandırılmıştır.

3.3. Derin Öğrenme Modelleri

Bu çalışmada, çeşitli derin öğrenme modelleri kullanılmıştır. Kullanılan modeller arasında VGG16, ResNet50, Inceptionv3, CNN ve CNN+LSTM yer almaktadır. Bu bölümün alt başlıklarında kullanılan modellerin özelliklerine yer verilmiştir.

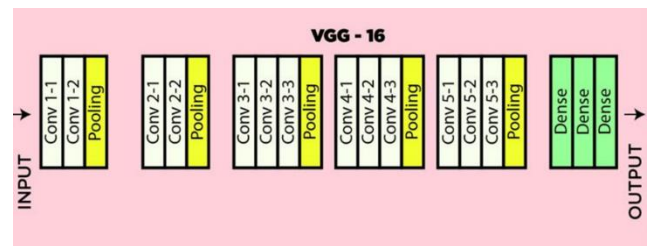
Tablo 1. Modellerin parametreleri

Hiper-parametre	VGG16	InceptionV3	CNN	CNN+LSTM	ResNet50
Dropout oranı	0.5	0.5	0.5, 0.3 (son katman)	0.3, 0.4 (son katman)	0.5
Batch boyutu	32	32	32	32	32
Epoch sayısı	10	10	30 (erken durdurma ile)	50 (erken durdurma ile)	10
Öğrenme oranı	Adam varsayılan	1,00E-04	1,00E-02	1,00E-03	1,00E-04
Optimizasyon algoritması	Adam	Adam	Adam	Adam	Adam
Aktivasyon fonksiyonu (gizli katmanlar)	ReLU	ReLU	ReLU	ReLU, tanh (LSTM)	ReLU
Filtre sayıları (evrimsizli katmanlar)	-	-	32,64,128	32,64,128	-
Çekirdek boyutu	-	-	3x3	3x3	-
LSTM birimleri	-	-	-	64	-
Tam bağlantılı katman boyutu	512	512	128	256	512

Her bir modelin eğitimi sırasında kullanılan hiper-parametreler Tablo 1'de ayrıntılı olarak sunulmaktadır. Yapılan deneysel çalışmalar neticesinde bu parametre kombinasyonlarının, daha etkili performans sonuçları sağladığı gözlemlenmiştir.

3.3.1. VGG-16

Bu ağ mimarisi, AlexNet prensiplerine benzer şekilde üretilmiş olup, giriş olarak 224×224×3 boyutunda RGB formatındaki görüntüleri alır. Konvolüsyon katmanında 3×3 boyutunda filtreler kullanılır. 2014 yılında tanıtılan VGG-16 mimarisi, toplamda 16 katmandan oluşur; bunlardan 13'ü konvolüsyonel ve 3'ü tam bağlantılıdır. Mimarinin son katmanında softmax sınıflandırıcısı kullanılır. VGG-16 mimarisinden farklı olarak, VGG-19 mimarisine 3 ek konvolüsyon katmanı eklenmiş ve toplamda 19 katmanlı olarak tasarlanmıştır, bunlardan 16'sı konvolüsyon katmanlarıdır (Simonyan ve Zisserman 2014). Şekil 2, VGG16 mimarisinin katmanlarını sunmaktadır (Ul Hassan 2018).

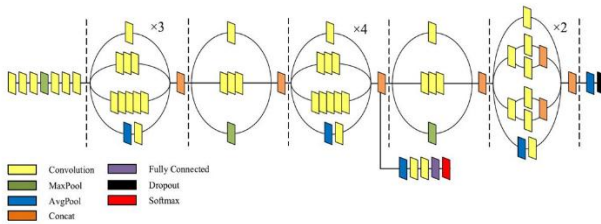


Şekil 2. VGG16 mimarisi

Bu çalışmada, transfer öğrenme yöntemleri kullanılarak VGG16 modeli, böbrek hastalığı sınıflandırması için özelleştirilmiştir. İlk olarak, ImageNet üzerinde eğitilmiş ağırlıkları içeren önceden eğitilmiş bir VGG16 modeli temel olarak kullanılmıştır. Özellik haritalarını sıkıştırmak ve daha fazla bilgi yoğunluğu elde etmek için pooling='max' ayarı tercih edilmiştir. Girdi boyutu, modelle uyumluluk sağlamak için 224×224 piksel olarak belirlenmiştir. Önceden eğitilmiş model, bir Sequential modelin içine entegre edilmiştir. Daha sonra, modelin üzerine çeşitli katmanlar eklenmiştir. Bu katmanlar, özelliklerin düzleştirilmesi için bir Flatten katmanı, sınıflandırma yeteneğini artırmak için 512 nöronlu ve ReLU aktivasyon fonksiyonuna sahip bir Dense katmanı, eğitim sırasında ağırlıkların stabilize edilmesi amacıyla bir Batch Normalization katmanı ve aşırı öğrenmeyi önlemek için bir Dropout katmanından oluşmaktadır. Çıkış katmanı ise, 4 sınıf için olasılık tahminleri sağlayan softmax aktivasyon fonksiyonuna sahip bir Dense katman olarak tasarlanmıştır. Modelin derlenmesinde, optimizasyon algoritması olarak Adam, kayıp fonksiyonu olarak ise sparse_categorical_crossentropy kullanılmıştır. Modelin eğitim süreci, eğitim ve doğrulama veri setlerinde 10 epoch boyunca gerçekleştirilmiştir. Eğitim sırasında doğruluk ve kayıp değerleri kaydedilmiş ve modelin performansı detaylı bir şekilde incelenmiştir. Eğitim sonunda, model mimarisi özetlenmiş ve doğrulama veri setindeki performansı analiz edilmiştir.

3.3.2. Inception-V3

Inception-V3 çok katmanlı model olarak, üç tür Inception modülüne sahip olup, Inception A, Inception B ve Inception C olarak bilinir. Bu modüller, veri özelliklerini ayırt etme konusunda üstün başarı göstermekte olup, bitki sınıflandırma çalışmalarında umut verici sonuçlar elde edilmiştir (Xia ark. 2017). Google tarafından geliştirilen TensorFlow kütüphanesi, transfer öğrenme yöntemlerinin kullanılmasında yüksek esneklik sağlamakta olup, Kumar ve ark. (2023) tarafından da belirtilmiştir ve bu durum, derin öğrenmeye dayalı görüntü tanıma ve sınıflandırma görevlerinde araştırma alanlarını açmaktadır (Ali ark. 2021). Şekil 3'te gösterilen Inception-V3 modeli, konvolüsyonel, maksimum havuzlama, birleştirme, dropout, tam bağlantılı ve ortalama havuzlama katmanlarından oluşan çok yönlü bir mimariye sahiptir (Lee ark. 2017). Bu mimari, özellik çıkarımının etkinliğini artırmakta ve görüntü özelliklerinin yorumlanmasına yardımcı olmaktadır. Şekil 3, inception-V3 mimarisini sunmaktadır.



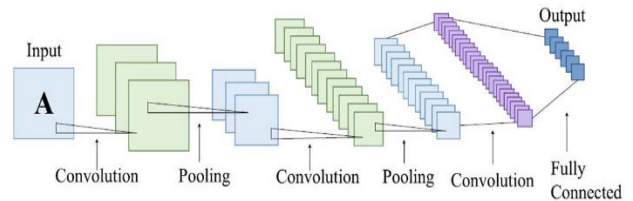
Şekil 3. Inception-V3 mimarisi

Bu çalışmada, CT böbrek hastalığı sınıflandırması için önceden eğitilmiş InceptionV3 modeli kullanılmış ve transfer öğrenme yöntemleri uygulanmıştır. Model,

ImageNet üzerinde eğitilmiş ağırlıkları içerecek şekilde yapılandırılmış ve temel özellik çıkarımını korumak amacıyla tüm katmanları dondurulmuştur. Daha sonra, modelin üzerine ek katmanlar eklenerek özelleştirilmiştir. Bu katmanlar arasında, özellik haritalarını düzeltmek için bir Flatten katmanı, 512 nöronlu ve ReLU aktivasyon fonksiyonuna sahip tam bağlı bir katman, Batch Normalizasyon ve Dropout katmanları ile softmax aktivasyonlu çıkış katmanı bulunmaktadır. Eğitim sürecinde, Adam optimizasyonu algoritması kullanılmış ve kayıp fonksiyonu olarak sparse_categorical_crossentropy tercih edilmiştir. Modelin performansı, eğitim ve doğrulama setlerindeki kayıp ve doğruluk değerleri görselleştirilerek analiz edilmiştir. Doğrulama veri kümesi üzerindeki sınıflandırma başarısını değerlendirmek amacıyla bir karışıklık matrisi oluşturulmuş ve bu matriste modelin her bir sınıf için doğru ve yanlış tahminleri incelenmiştir.

3.3.3. Konvolüsyonel Sinir Ağları

Konvolüsyonel Sinir Ağları (CNN), derin öğrenme alanında en yaygın ve etkili ağ mimarilerinden biridir. Özellikle görüntü işleme ve sınıflandırma görevlerinde yüksek performans sergileyen CNN, görsel verileri analiz etmek için özel olarak tasarlanmış katmanlardan oluşur. CNN, görüntüdeki yerel özellikleri tanıyabilen konvolüsyonel katmanlar, bu özellikleri daha genel ve soyut temsillere dönüştüren havuzlama katmanları ve sınıflandırma işlemini gerçekleştiren tam bağlantılı katmanlar kullanır. Bu yapı, görüntülerin doğrudan işlenmesini sağlar, böylece karmaşık ön işleme gereksinimlerini ortadan kaldırır. Ayrıca, ağırlık paylaşımı ve yerel alıcı alan gibi özellikler sayesinde, ağır eğitim parametrelerini önemli ölçüde azaltır ve genel performansını artırır. CNN, tıbbi görüntüleme, yüz tanıma, nesne tespiti gibi birçok alanda başarıyla kullanılmaktadır (Wu 2017). Şekil 4 CNN mimarisini sunmaktadır.



Şekil 4. CNN mimarisi

Bu çalışmanın metodolojisi, CT böbrek hastalığı görüntülerinin sınıflandırılması için CNN kullanılarak yapılandırılmış bir yaklaşım içermektedir. Bu CNN modeli, ardışık konvolüsyon ve havuzlama katmanlarıyla derin özellikler çıkartacak şekilde tasarlanmıştır.

Model, ilk olarak 32 filtreli bir 'Conv2D' katmanı ve 'ReLU' aktivasyon fonksiyonu ile başlar. Aşırı öğrenmeyi önlemek için her konvolüsyon katmanından sonra 'Dropout' uygulanır ve özellik haritasını küçültmek için 'MaxPooling2D' katmanı eklenir. Bu yapı, sırasıyla 64 ve 128 filtreli ek konvolüsyon katmanlarıyla tekrarlanır. Tam bağlantılı katmanlar, özellikleri düzleştiren bir 'Flatten' katmanı, ardından 128 nöronlu bir yoğun katman, 'BatchNormalization' ve bir çıkış katmanı içerir. Çıkış

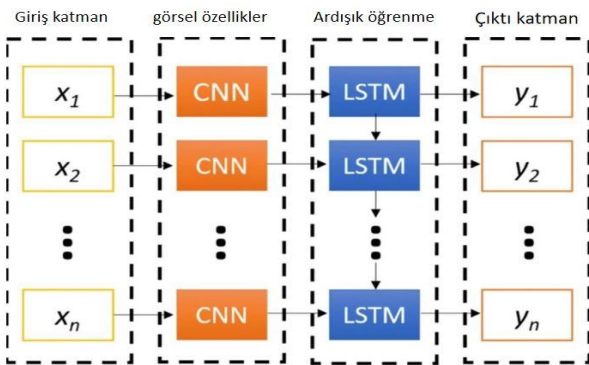
katmanı, 4 sınıflı temsil eden 'softmax' aktivasyon fonksiyonuyla sonlandırılır.

Model, Adam optimizasyon algoritması ve 'sparse_categorical_crossentropy' kayıp fonksiyonu kullanılarak derlenmiştir. Eğitim sırasında doğrulama doğruluğunda iyileşme olmadığında erken durdurma sağlamak için 'EarlyStopping', en iyi model ağırlıklarını kaydetmek için ise 'ModelCheckpoint' kullanılmıştır. Eğitim sonuçları görselleştirilmiş ve hem doğruluk hem de kayıp değerleri epoch bazında çizdirilmiştir. Doğrulama veri seti üzerinde modelin performansını ölçmek için tahmin edilen etiketlerle gerçek etiketler arasında bir karışıklık matrisi hesaplanmış ve bu matris, sonraki analizlerde kullanılmak üzere bir dosyaya kaydedilmiştir.

3.3.4. CNN+LSTM

Bu çalışmada, CNN ve LSTM birleşimiyle bir model tasarlanmıştır. Modelin ilk kısmı, üç konvolüsyonel katmandan oluşan bir CNN yapısı kullanır. İlk katmanda, 32 filtreli bir 3×3 boyutunda konvolüsyon işlemi yapılır ve ardından Dropout katmanı ile aşırı uyum önlenir. Bu katman, MaxPooling katmanı ile devam eder. İkinci ve üçüncü konvolüsyon katmanlarında sırasıyla 64 ve 128 filtreler kullanılır ve her iki katman için de Dropout ve MaxPooling işlemleri uygulanır. Sonrasında, son konvolüsyon katmanındaki çıktılar, bir LSTM katmanına beslenmek üzere yeniden şekillendirilir. Bu LSTM katmanı, sıralı verilere dair özellikleri öğrenir ve modelin sıralı ilişkileri anlamasına yardımcı olur. LSTM katmanının ardından, bir gizli katman (Dense katmanı) eklenir, burada 256 nöron bulunur ve bu katmanda BatchNormalization ile öğrenme süreci hızlandırılır. Aşırı uyumun engellenmesi için bir Dropout katmanı da eklenmiştir.

Modelin son katmanında, 4 sınıflı sınıflandırmak için softmax aktivasyon fonksiyonu ile bir çıkış katmanı eklenmiştir. Model, Adam optimizer'ı ve 1e-3 öğrenme oranı ile derlenmiştir. Kayıp fonksiyonu olarak sparse categorical crossentropy kullanılmıştır ve doğruluk metriği izlenmiştir. Bu yapı, derin öğrenme yöntemlerinin güçlü birleşimiyle, görsel verilerdeki sıralı ve uzamsal ilişkileri etkili bir şekilde öğrenmek için tasarlanmıştır (Khorram ve ark. 2021). Şekil 5, CNN+LSTM mimarisini sunmaktadır.



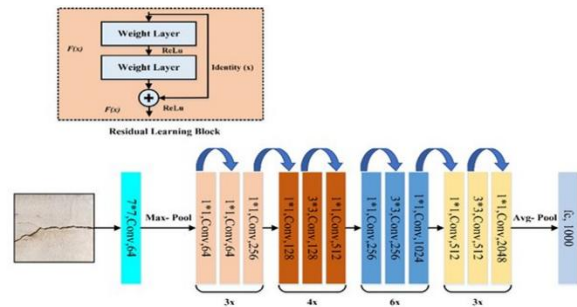
Şekil 5. CNN+LSTM mimarisini

Bu çalışmada, zaman-mekansal özellikleri aynı anda öğrenebilen bir CNN+LSTM modeli, böbrek hastalığı

sınıflandırması için tasarlanmıştır. Model, uzamsal özelliklerin çıkarılması için bir CNN ve zaman içindeki ilişkileri öğrenmek için LSTM katmanlarını birleştirmiştir. İlk olarak, CNN kısmında üç konvolüsyon katmanı kullanılmıştır. Bu katmanlar, 32, 64 ve 128 filtreye sahiptir ve her biri 3×3 boyutunda çekirdekler kullanarak özellik çıkarımı gerçekleştirmiştir. Her konvolüsyon katmanının ardından aşırı öğrenmeyi önlemek için Dropout ve veri boyutlarını küçültmek için MaxPooling katmanları eklenmiştir. Son konvolüsyon katmanının çıkışı, LSTM için sıralı bir giriş haline getirilebilmesi amacıyla yeniden şekillendirilmiştir. LSTM katmanı, mekansal ilişkileri dikkate alarak zaman içindeki özellikleri öğrenmek için kullanılmıştır. Bu katman, 64 LSTM birimiyle yapılandırılmıştır. Tam bağlı katmanlar arasında, ReLU aktivasyon fonksiyonuna sahip 256 nöronlu bir gizli katman, modelin genelleştirme kapasitesini artırmak için Batch Normalization ve aşırı öğrenmeyi önlemek için Dropout katmanları yer almıştır. Çıkış katmanı, 4 sınıfa ilişkin olasılık tahminleri yapmak için softmax aktivasyon fonksiyonuna sahiptir. Model, Adam optimizasyon algoritması ve learning_rate=1e-3 ile derlenmiştir, kayıp fonksiyonu olarak sparse_categorical_crossentropy kullanılmıştır. Eğitim süreci, 50 epoch boyunca gerçekleştirilmiş ve erken durdurma kriterleri ile doğrulama doğruluğunu iyileştiren en iyi modeli kaydetmek için geri çağırma fonksiyonları eklenmiştir. Eğitim sonrası, modelin doğrulama veri kümesi üzerindeki performansı bir karışıklık matrisi (confusion matrix) kullanılarak değerlendirilmiştir.

3.3.5. ResNet50

He ve ark. (2016), 2015 yılında ImageNet Büyük Ölçekli Görsel Tanıma Yarışması'nda (ILSVRC 2015) birincilik elde eden rezidüel sinir ağı (ResNet50) modelini önermiştir. ResNet50, katmanlar arasında rezidüel bağlantılar sunarak kayıpları azaltır, bilgi kazanımını korur ve eğitim sürecindeki performansı artırır. Bir katmandaki rezidüel bağlantı, o katmanın çıktısının, girişinin konvolüsyon işlemine tabi tutulmuş hali ile girişin toplamı olduğu anlamına gelir. ResNet50 modelinin mimarisine ait bir blok diyagramı Şekil 6'da gösterilmektedir.



Şekil 6. ResNet-50 mimarisini

Bu çalışmada, ResNet50 transfer öğrenme modeli özelleştirilerek kullanılmıştır. Model, ImageNet veri kümesi üzerinde önceden eğitilmiş ağırlıklarla başlatılmıştır ve sınıflandırma katmanı çıkarılmıştır.

Önceden eğitilmiş ResNet50 modeli, giriş boyutu (224×224×3) ve maksimum havuzlama işlemiyle adapte

edilmiştir. Modelin transfer öğrenme aşamasında, önceden eğitilmiş ağırlıkların değiştirilmemesi için ağırlıklar dondurulmuştur. ResNet50 modelinin ardından sırasıyla yassılaştırma (flatten), tam bağlantılı (dense) ve sınıflandırma katmanları eklenmiştir. Gizli katman olarak, 512 nörona sahip ReLU aktivasyon fonksiyonu kullanılan bir katman eklenmiş, ardından Batch Normalization ile

öğrenme süreci optimize edilmiştir. Aşırı uyumu önlemek için Dropout (0,5) uygulanmıştır. Son katman olarak, softmax aktivasyon fonksiyonuna sahip bir çıkış katmanı eklenmiş ve 4 sınıflı sınıflandırma işlemi gerçekleştirilmiştir. Bu yapı, önceden öğrenilmiş özellikleri yeniden kullanarak sınıflandırma görevinde performansı artırmayı amaçlamaktadır.

Tablo 2. Kullanılan modellerin özeti

Model	Parametreler	Katman	Özellikler
Inceptionv3	22,855,972	53	İçerideki çevreler, paralel evrişimler ve boyut azaltma
ResNet50	24,640,900	55	Kalıntı bağlantılarıyla derin ağlarda gradyan kaybını önleme
CNN+LSTM	-	-	CNN görsel özellikleri, LSTM ise zamansal dinamikleri işler
CNN	-	-	Evrişim katmanlarıyla uzamsal ilişkileri öğrenme
VGG16	27,602,772	23	Küçük filtreler ve maksimum havuzlama

Tablo 2 kullanılan modellerin özetini sunmuştur. CNN+LSTM modelde toplamda 15 katman bulunmaktadır. Model, sırasıyla 32 filtre ile evrişim işlemi (Conv2D), 32 kanalda dropout uygulanması (Dropout), boyut küçültme işlemi (MaxPooling2D), 64 filtre ile evrişim işlemi (Conv2D), 64 kanalda dropout uygulanması (Dropout), boyut küçültme işlemi (MaxPooling2D), 128 filtre ile evrişim işlemi (Conv2D), 128 kanalda dropout uygulanması (Dropout), boyut küçültme işlemi (MaxPooling2D), çıktı tek boyutlu hale getirilmesi (Flatten), 128 nöronlu tam bağlantılı katman (Dense), 128 kanalda dropout uygulanması (Dropout) ve son olarak 4 sınıfa çıkış veren son katman (Dense) içermektedir. CNN modelinde toplam 13 katman bulunmaktadır. Model, sırasıyla 32 filtre ile evrişim işlemi (Conv2D), 32 kanalda dropout uygulanması

(Dropout), boyut küçültme işlemi (MaxPooling2D), 64 filtre ile evrişim işlemi (Conv2D), 64 kanalda dropout uygulanması (Dropout), boyut küçültme işlemi (MaxPooling2D), 128 filtre ile evrişim işlemi (Conv2D), 128 kanalda dropout uygulanması (Dropout), boyut küçültme işlemi (MaxPooling2D), çıktı tek boyutlu hale

getirilmesi (Flatten), 128 nöronlu tam bağlantılı katman (Dense), 128 kanalda dropout uygulanması (Dropout) ve son olarak 4 sınıfa çıkış veren katman (Dense) içermektedir.

3.4. Metrikler

Kullanılan derin öğrenme yöntemlerini karşılaştırmak için literatürde yaygın olarak kullanılan doğruluk, kesinlik, geri çağırma ve F1-Skor metrikleri kullanılmaktadır (Senthil ve Vidyaathulasiraman 2019; Aydın ve Şamlı 2020; Doğru ve Altuntaş 2023). Bu değerlerin bir karışıklık matrisindeki görünümüne bir örnek Şekil 7'de gösterilmektedir (Sulistyowati ark. 2023).

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

Şekil 7. Karmaşıklık matrisi

Doğruluk, Kesinlik, Duyarlılık ve F1-Skor 1-4 denklemlerinde gösterildiği gibi hesaplanır.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

4. BULGULAR

Bu bölümde, çalışmada kullanılan veri setine dayalı olarak derin öğrenme yöntemlerinin VGG16, ResNet50, Inceptionv3, CNN ve CNN+LSTM karşılaştırması sunulmaktadır.

Her model 10 epoch için eğitilmiştir. Ancak CNN ve CNN+LSTM modelleri için erken durdurma (early stopping) mekanizması kullanılmış ve belirli bir epoch'tan sonra eğitim durdurulmuştur. Bu modellerde, sırasıyla 30 ve 50 epoch belirtilmiş olsa da erken durdurma nedeniyle daha erken sonlanmışlardır. Erken durdurma, modelin eğitim verilerine aşırı uyum sağlamasını önlemek için

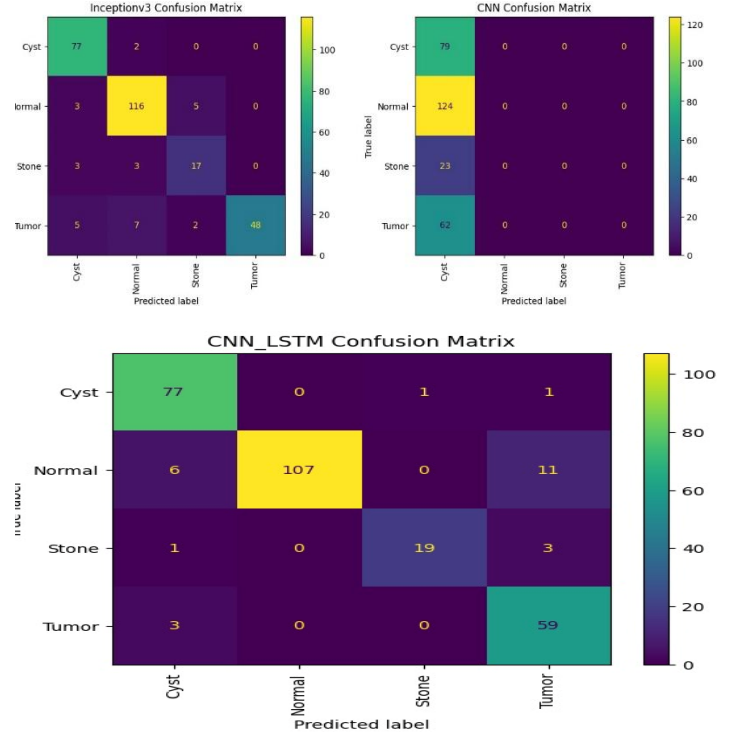
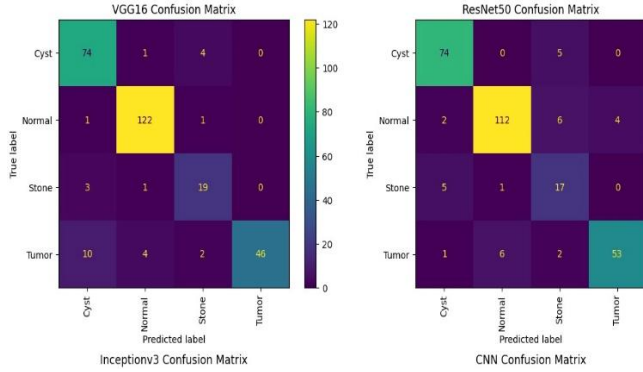
$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - \text{Skor} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

eğitimi erken durdurur, bu sayede daha iyi genelleme yeteneği sağlar. Veri seti, %80 eğitim ve %20 doğrulama (validation) olarak ayrılmıştır. Yani, verinin %80'i modeli eğitmek için, %20'si ise modelin performansını değerlendirmek için kullanılmıştır. Tüm modellerde batch boyutu 32 olarak ayarlanmıştır. Şekil 8, farklı derin öğrenme modellerinin (VGG, ResNet50, InceptionV3, CNN ve CNN_LSTM) dört sınıflı bir sınıflandırma problemindeki performansını göstermektedir. Bu matrisler, her bir modelin performansını dört ayrı sınıf için göstermektedir: Cyst (Kist), Normal, Stone (Taş) ve

Tumor (Tümör). Bu matrisler, modellerin her bir sınıfı diğer sınıflardan ayırt etme kapasitesini açıkça göstermektedir.



Şekil 8. Modellerin karmaşıklık matrisleri

Çalışmada kullanılan derin öğrenme modellerinin CT-KIDNEY veri seti üzerindeki sınıflandırma performansları, Şekil 8'de sunulan ayrıntılı karmaşıklık matrisleri ve Tablo 2'de özetlenen nicel metrikler aracılığıyla kapsamlı bir şekilde değerlendirilmiştir. Karmaşıklık matrisleri (Şekil 8), her bir modelin Kist, Normal, Taş ve Tümör sınıfları arasındaki doğru ve yanlış sınıflandırma sayılarını görselleştirerek, modellerin hangi sınıfları ayırt etmekte zorlandığını veya hangi sınıfları birbirine karıştırdığını ayrıntılı olarak ortaya koymaktadır. Tablo 2 ise bu performansları özetleyen Doğruluk, F1-Skor, Kesinlik ve Duyarlılık metriklerinin makro ortalama değerlerini sunmaktadır.

Elde edilen sonuçlar incelendiğinde (Tablo 2), test edilen modeller arasında en yüksek performansı CNN+LSTM hibrit modelinin sergilediği açıkça görülmektedir. Bu model, %91,0 gibi dikkat çekici bir genel doğruluk oranına ulaşmıştır. Daha da önemlisi, %90,1'lik F1-Skor değeri, modelin Kesinlik (%90,8) ve Duyarlılık (%90,4) metrikleri arasında güçlü bir denge kurduğunu göstermektedir. Bu yüksek değerler, CNN+LSTM modelinin hem ilgili böbrek durumunu doğru bir şekilde tespit etme (yüksek duyarlılık) hem de sağlıklı veya farklı bir duruma sahip böbreği yanlışlıkla hasta olarak etiketlememe (yüksek kesinlik) konusunda diğer modellere kıyasla üstün bir yetenek sergilediğini işaret etmektedir.

VGG16 modeli, %90,6'lık doğruluk oranı ile CNN+LSTM modeline oldukça yakın bir genel performans göstermiştir. Bu, VGG16'nın da KBH sınıflandırması için oldukça etkili bir mimari olduğunu teyit etmektedir. Ancak, VGG16'nın F1-Skor (%86,7), Kesinlik (%87,4) ve Duyarlılık (%87,5) değerleri, CNN+LSTM modelinin metriklerine göre bir miktar daha düşüktür. Bu durum, VGG16'nın genel doğruluğu yüksek

olsa da, yanlış pozitif ve yanlış negatif tahminler arasındaki denge veya belirli sınıflardaki performansı açısından CNN+LSTM kadar optimize olmayabileceğini düşündürmektedir.

InceptionV3 ve **ResNet50** modelleri de sırasıyla %89,6 ve %88,9 doğruluk oranları ile KBH sınıflandırmasında başarılı ve kabul edilebilir performanslar sergilemişlerdir. Bu iki modelin metrikleri birbirine oldukça yakındır (InceptionV3: F1 %86,0, Kesinlik %87,2, Duyarlılık %85,6; ResNet50: F1 %86,8, Kesinlik %87,0, Duyarlılık %86,9). Bu sonuçlar, her iki derin mimarinin de CT görüntülerinden anlamlı özellikler çıkararak sınıflandırma yapabildiğini, ancak en iyi performansı gösteren CNN+LSTM ve VGG16'nın bir miktar gerisinde kaldıklarını göstermektedir.

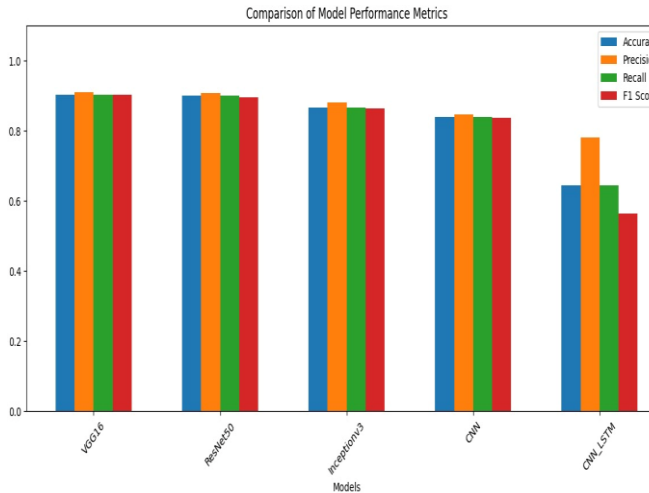
Değerlendirilen modeller arasında en dikkat çekici ve farklı sonuç **standart CNN** modeline aittir. Bu model, %27,4 gibi oldukça düşük bir doğruluk oranı elde etmiştir. Diğer metrikler (F1-Skor %10,8, Kesinlik %6,9, Duyarlılık %25,0) de bu modelin sınıflandırma görevinde başarısız olduğunu açıkça ortaya koymaktadır. Özellikle %6,9'lık Kesinlik değeri, modelin yaptığı pozitif tahminlerin çok büyük bir kısmının yanlış olduğunu göstermektedir. %25,0'lık Duyarlılık değeri ise (dört sınıf olduğu düşünüldüğünde), modelin muhtemelen tüm örnekleri veya büyük çoğunluğunu tek bir sınıfa (örneğin, veri setindeki en yaygın sınıf) atama eğiliminde olduğunu düşündürmektedir, bu da diğer sınıfları neredeyse hiç tanımadığı anlamına gelir. Bu durum, Şekil 8'deki CNN karmaşıklık matrisinde de muhtemelen tek bir sütunda yoğunlaşma olarak gözlemlenecektir. Sonuç olarak, temel bir CNN mimarisi, bu spesifik CT görüntüsü sınıflandırma görevinin karmaşıklığı için yetersiz kalmıştır.

Genel olarak bulgular, daha derin (VGG16, ResNet50, InceptionV3) veya hibrit (CNN+LSTM) mimarilerin,

basit bir CNN'e kıyasla CT görüntülerinden böbrek patolojilerini ayırt etmede belirgin şekilde daha etkili olduğunu güçlü bir şekilde desteklemektedir. Özellikle **Tablo 2:** CT-KIDNEY veri setinde derin öğrenme modellerinin sonuçları

Model	Doğruluk	F1-Skor	Kesinlik	Duyarlılık
CNN+LSTM	0,910	0,901	0,908	0,904
VGG16	0,906	0,867	0,874	0,875
Inceptionv3	0,896	0,860	0,872	0,856
ResNet50	0,889	0,868	0,870	0,869
CNN	0,274	0,108	0,069	0,250

Şekil 9, çalışmada değerlendirilen beş farklı derin öğrenme modelinin (CNN+LSTM, VGG16, InceptionV3, ResNet50 ve CNN) performansını dört temel metrik (Doğruluk, F1-Skor, Kesinlik ve Duyarlılık) açısından karşılaştırmalı olarak sunmaktadır. Grafik, her model için bu dört metriğin skor değerlerini (0 ile 1,0 arasında) farklı renklerdeki çubuklarla görselleştirerek, modellerin göreceli başarılarını ve her bir metriğe göre güçlü/zayıf yönlerini bir bakışta anlamayı kolaylaştırmaktadır.



Şekil 9. CT-KIDNEY veri setinde derin öğrenme modellerinin sonuçlarının grafiksel gösterimi

Grafik incelendiğinde, CNN+LSTM modelinin tüm metriklerde en yüksek veya en yükseğe çok yakın değerlere sahip olduğu açıkça görülmektedir. Doğruluk (0,910), F1-Skor (0,901), Kesinlik (0,908) ve Duyarlılık (0,904) çubuklarının tümü 0,9 seviyesinin üzerinde yer almakta ve diğer modellere kıyasla tutarlı bir üstünlük sergilemektedir. Bu durum, CNN+LSTM modelinin hem genel sınıflandırma başarısı (yüksek doğruluk) hem de kesinlik ile duyarlılık arasındaki denge (yüksek F1-Skor) açısından en iyi performansı gösterdiğini görsel olarak teyit etmektedir.

VGG16, InceptionV3 ve ResNet50 modelleri de oldukça başarılı sonuçlar sergilemektedir. Bu üç modelin doğruluk çubukları sırasıyla 0,906, 0,896 ve 0,889 seviyelerinde olup, hepsi %88'in üzerinde performans göstermiştir. Diğer metrikler (F1-Skor, Kesinlik, Duyarlılık) açısından da bu üç model genellikle 0,85 ile 0,88 arasında, birbirlerine yakın ve yüksek değerler almıştır. Bu, her üç modelin de KBH sınıflandırması için

CNN ve LSTM'in birleştirildiği hibrit yaklaşım, bu veri seti ve görev için en umut verici sonuçları vermiştir.

yetkin olduğunu, ancak CNN+LSTM'in ulaştığı zirve performansının biraz altında kaldıklarını göstermektedir.

Grafikteki en çarpıcı bulgu ise CNN modelinin performansınıdır. Diğer dört modelin aksine, CNN modelinin tüm metrikleri temsil eden çubukları belirgin bir şekilde kısadır. Doğruluk çubuğu sadece 0,274 seviyesine ulaşırken, F1-Skor (0,108) ve özellikle Kesinlik (0,069) değerleri son derece düşüktür. Duyarlılık değeri (0,250) ise, dört sınıflı bir problemde modelin muhtemelen rastgele tahminden biraz daha iyi veya tüm tahminleri tek bir sınıfa yönlendirme eğiliminde olduğunu düşündürmektedir. Bu görsel kontrast, standart CNN mimarisinin bu özel görev için diğer daha karmaşık veya hibrit modellere kıyasla açıkça yetersiz kaldığını vurgulamaktadır.

Özetle, Şekil 9, Tablo 2'deki sayısal verileri etkili bir şekilde görselleştirerek, CNN+LSTM modelinin üstünlüğünü, VGG16, InceptionV3 ve ResNet50'nin güçlü performansını ve temel CNN modelinin zayıflığını net bir biçimde ortaya koymaktadır.

Şekil 10, her bir modelin eğitim süreci boyunca (epoch'lar ilerledikçe) nasıl öğrendiğini ve öğrendiklerini görmediği yeni verilere (doğrulama veya test verisi) ne kadar iyi genellebildiğini gösterir.

Modellerin öğrenme dinamikleri ve genelleme yetenekleri, eğitim ve doğrulama/test doğruluklarının epoch'lar boyunca değişimini gösteren öğrenme eğrileri incelenerek analiz edilmiştir.

VGG16 modeli, eğitim verisi üzerinde hızlı bir öğrenme eğrisi sergilemiştir; eğitim doğruluğu 10 epoch sonunda yaklaşık %98 seviyesine ulaşmıştır. Doğrulama (validation) doğruluğu da başlangıçta belirgin bir artış göstermiş, yaklaşık 6. epoch civarında %90 seviyesinde bir platoya ulaşmış ve sonraki epoch'larda bu seviyede kararlılık kazanmıştır. Eğitim ve doğrulama eğrileri arasındaki gözlemlenen fark, modelin bir miktar aşırı öğrenme (overfitting) eğilimi gösterdiğine işaret etmektedir. Bununla birlikte, doğrulama doğruluğunun ulaştığı yüksek ve stabil plato, modelin önemli ölçüde genelleme yeteneği kazandığını ve optimum performansına 6 ila 10 epoch aralığında yaklaştığını göstermektedir.

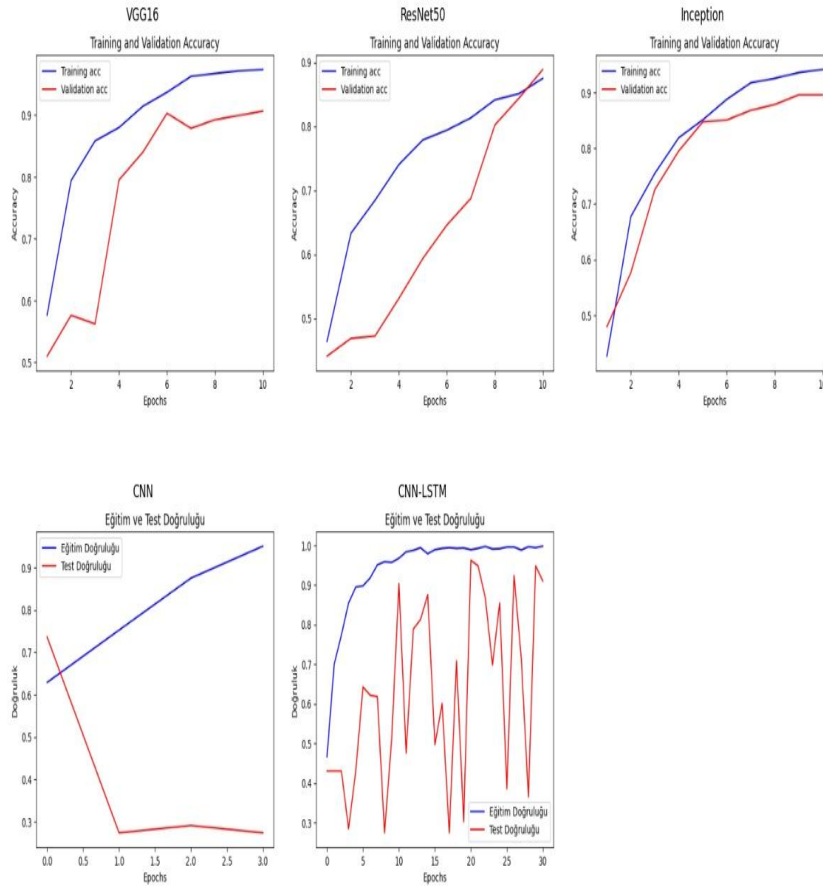
ResNet50 mimarisi, daha dengeli bir öğrenme süreci ortaya koymuştur. Eğitim doğruluğu, 10 epoch sonunda yaklaşık %88'e ulaşarak istikrarlı bir artış göstermiştir. Dikkat çekici bir şekilde, doğrulama doğruluğu eğrisi, eğitim eğrisini yakından takip etmiş ve özellikle son epoch'larda belirgin bir artışla 10. epoch'ta eğitim doğruluğuna oldukça yakın bir değere (~%89) ulaşmıştır. İki eğri arasındaki bu minimal fark, ResNet50'nin değerlendirilen eğitim süresi boyunca etkili bir genelleme yaptığını ve aşırı öğrenmenin oldukça sınırlı kaldığını göstermektedir. Her iki eğrinin de hala yükselme eğiliminde olması, eğitim süresinin uzatılmasıyla potansiyel olarak daha ileri performans iyileştirmelerinin mümkün olabileceğini düşündürmektedir.

InceptionV3 modelinin öğrenme dinamikleri, VGG16'ya benzer şekilde, eğitim verisi üzerinde hızlı bir yakınsama göstermiş ve doğruluk yaklaşık %94 seviyesine ulaşmıştır. Doğrulama doğruluğu da bu eğilimi yakından takip ederek, 10 epoch sonunda %90 civarında yüksek bir platoda stabilize olmuştur. Bu durum, modelin hem öğrenme hem de genelleme açısından yetkin olduğunu göstermektedir. Eğitim ve doğrulama doğruluğu arasındaki farkın VGG16'ya kıyasla potansiyel olarak biraz daha dar olması, aşırı öğrenmenin bir miktar daha iyi kontrol edildiğine işaret edebilir. Doğrulama eğrisinin platoya ulaşması, modelin yaklaşık 10 epoch civarında optimal performansına yaklaştığını göstermektedir.

Standart CNN modelinin öğrenme eğrisi, ciddi bir genelleme sorununu ortaya koymuştur. Model, eğitim verisini oldukça hızlı bir şekilde öğrenerek 3 epoch içinde %95'in üzerinde bir doğruluğa ulaşmıştır. Ancak, test doğruluğu başlangıçtaki (~%73) değerinden sonra ilk epoch'un ardından dramatik bir düşüş yaşayarak %30'un altına inmiş ve bu düşük seviyede kalmıştır. Eğitim ve test doğrulukları arasındaki bu aşırı ayrışma, modelin eğitim

verisini ezberlerken yeni verilere genelleme yapamadığı şiddetli bir aşırı öğrenme (overfitting) durumunu açıkça göstermektedir. Bu bulgu, daha önce rapor edilen düşük genel sınıflandırma metrikleriyle (%27,4 doğruluk) tutarlıdır ve modelin karmaşıklığının veri seti için uygun olmadığını veya kullanılan regülerizasyon mekanizmalarının yetersiz kaldığını düşündürmektedir.

CNN-LSTM hibrit modeli ise farklı bir öğrenme profili sergilemiştir. Eğitim doğruluğu çok hızlı bir şekilde artarak yaklaşık 15. epoch sonrasında %100'e yakın bir seviyeye ulaşmış ve bu seviyeyi korumuştur, bu da eğitim verisinin tamamen ezberlendiğini göstermektedir. Test doğruluğu ise başlangıçta önemli ölçüde artsa ve %90-95 aralığında tepe noktalarına ulaşsa da, eğitim süreci boyunca belirgin dalgalanmalar (yüksek varyans) göstermiştir. Test doğruluğundaki bu kararsızlık, eğitim doğruluğunun doyumluğa ulaşmasıyla birlikte, modelin hem aşırı öğrenme eğiliminde olduğunu hem de eğitiminin stabil olmadığını göstermektedir. Modelin genelleme performansı epoch'lar arasında önemli ölçüde değişkenlik göstermiştir. Fakat en iyi sonucu üretmeyi başarmıştır.



Şekil 10. Derin öğrenme modellerinde eğitim doğrulukları

5. TARTIŞMA

Kronik böbrek hastalığının (KBH) erken teşhisi, hastalığın ilerlemesini yavaşlatmak ve ciddi komplikasyonları önlemek adına büyük önem taşımaktadır. Bilgisayarlı tomografi (BT) görüntüleri bu teşhis sürecinde değerli bilgiler sunarken, derin öğrenme (DL) modelleri bu

görüntülerin otomatik analizinde umut verici bir potansiyel sergilemektedir. Bu çalışma, KBH teşhisi ve sınıflandırması için InceptionV3, ResNet50, temel bir Evrişimli Sinir Ağı (CNN), CNN+LSTM ve VGG16 gibi farklı DL mimarilerinin performansını, literatürde de

kullanılan CT-KIDNEY veri seti üzerinde, özellikle **288 örnek gibi kısıtlı bir veri** kullanarak değerlendirmiştir.

Çalışmanın bulguları, test edilen modeller arasında %91,0 doğruluk ile en yüksek performansı CNN+LSTM modelinin sergilediğini göstermektedir. VGG16 (%90,6), InceptionV3 (%89,6) ve ResNet50 (%88,9) gibi transfer öğrenme modelleri de %88-%91 aralığında rekabetçi sonuçlar vermiştir. Bu sonuçlar, özellikle veri kısıtlamaları altında bile DL modellerinin KBH sınıflandırmasında %90 civarında bir doğruluğa ulaşabildiğini göstermesi açısından önemlidir. Ancak, temel CNN modelinin %27,4 gibi oldukça düşük bir doğruluk oranı sergilemesi, bu spesifik modelin implementasyonu veya eğitimi ile ilgili olası sorunlara işaret etmektedir.

Bu çalışmanın sonuçlarını, aynı CT-KIDNEY veri setini kullanan literatürdeki diğer çalışmalarla karşılaştırmak, elde edilen bulguların değerini ve bağlamını daha iyi anlamamızı sağlar. Örneğin, Mehedi ve ark. (2022) VGG16 ile %99,48 ve InceptionV3 ile %97,38 gibi olağanüstü doğruluk oranları elde etmişlerdir. Benzer şekilde, Bingol ve ark. (2023), yine aynı veri seti üzerinde kendi geliştirdikleri hibrit bir modelle %99,37 doğruluğa ulaşmışlardır. Zhang ve ark. (2024) ise, özel bir optimizasyon algoritması (CDMO) ile optimize ettikleri InceptionV4 modeliyle %98,14 doğruluk raporlamışlardır. Bu çalışmalar, uygun koşullar ve yeterli veri ile CT-KIDNEY veri setinde %98-%99 doğruluk seviyelerine ulaşılabilirliğini göstermektedir.

Literatürdeki bu yüksek başarı oranları ile karşılaştırıldığında, bu çalışmada elde edilen %91'lik en yüksek doğruluk ilk bakışta düşük görünebilir. Ancak, bu çalışmanın **önemli bir gücü, bu sonuçları sadece 288 veri örneği gibi oldukça sınırlı bir veri seti ile elde etmiş olmasıdır**. Tıbbi görüntüleme alanında genellikle büyük ve etiketlenmiş veri setlerine ulaşmanın zorluğu göz önüne alındığında, **kısıtlı veri ile %90'ın üzerinde bir doğruluğa ulaşmak, kullanılan yöntemlerin veri verimliliğini (data efficiency) ve pratik uygulanabilirliğini göstermesi açısından değerlidir**. Bu durum, çalışmanın daha gerçekçi klinik senaryolara yakın bir bağlamda değerlendirilmesini sağlar.

Diğer bir dikkat çekici nokta, bu çalışmada en iyi performansı gösteren CNN+LSTM modelidir. İncelenen literatürdeki diğer çalışmalarda (Mehedi ve ark. 2022; Bingol ve ark. 2023; Zhang ve ark. 2024) bu spesifik hibrit mimari doğrudan ele alınmamıştır. CNN+LSTM modelinin, bu çalışmanın kendi deney koşulları ve veri kısıtlamaları altında standart VGG16, InceptionV3 ve ResNet50 modellerinden daha iyi performans göstermesi, **bu yaklaşımın potansiyel bir katkısı olarak değerlendirilebilir**. LSTM katmanlarının, BT görüntülerindeki kesitler arası ardışık bilgiyi veya uzamsal bağımlılıkları yakalayarak, özellikle kısıtlı veri durumunda sınıflandırma performansına olumlu katkı

sağlama potansiyeli bulunmaktadır. Bu bulgu, daha ileri araştırmalar için ilginç bir başlangıç noktası sunmaktadır.

Model bazında karşılaştırmalara bakıldığında; VGG16 (%90,6) ve InceptionV3 (%89,6) için elde edilen sonuçlar, Mehedi ve arkadaşları tarafından raporlanan %99,48 ve %97,38'lik değerlerin altında kalsa da, Bingol ve arkadaşları tarafından raporlanan ince ayarlı InceptionV3 (%87,75) sonucuyla karşılaştırılabilir düzeydedir. ResNet50 için elde edilen %88,9'luk doğruluk ise, Bingol ve arkadaşlarının [2] %86,06'lık sonucundan biraz daha iyidir. Bu durumlar, kısıtlı veri göz önüne alındığında, transfer öğrenme modellerinin bu çalışmada makul bir performans sergilediğini desteklemektedir.

6. SONUÇ

Bu çalışmada, CT-KIDNEY veri seti üzerinde farklı derin öğrenme modellerinin performansları karşılaştırılmış ve KBH erken teşhisi ve sınıflandırmasında derin öğrenmenin etkili bir yöntem olduğu gösterilmiştir. Çalışmada, KBH sınıflandırması için DL modellerinin performansını oldukça kısıtlı bir veri seti (288 örnek) üzerinde değerlendirerek önemli bilgiler sunmaktadır. %91'e varan doğruluk oranları, veri azlığına rağmen elde edilen kayda değer bir başarı elde edilmiştir. Özellikle CNN+LSTM modelinin bu kısıtlı veri setinde diğer standart modellere göre daha iyi performans göstermesi, çalışmanın özgün bir yönü olarak öne çıkmaktadır. Literatürdeki daha yüksek doğruluk oranları, genellikle daha büyük veri setleri ve/veya gelişmiş optimizasyon ve ön işleme teknikleri ile elde edilmiştir. Gelecekteki çalışmalar, bu çalışmada elde edilen umut verici sonuçları temel alarak, literatürdeki gelişmiş teknikleri (örn. segmentasyon, veri artırma, ince ayar) kısıtlı veri senaryolarına adapte etmeyi ve özellikle CNN+LSTM mimarisinin potansiyelini daha derinlemesine araştırmayı hedefleyebilir.

Gelecekteki çalışmalar, veri çeşitliliğini artırarak modelin genelleme kabiliyetini daha da geliştirebilir ve IoT tabanlı gerçek zamanlı tanı sistemleri ile entegrasyonunu araştırabilir. Bu yenilikçi yaklaşımlar, KBH gibi küresel sağlık sorunlarının çözümüne önemli katkılar sağlayacaktır. Gelecekteki çalışmalar, daha geniş ve çeşitli veri setleri üzerinde modelin genelleme kapasitesini artırmayı, IoT tabanlı gerçek zamanlı tanı sistemlerine entegrasyonunu sağlamayı ve farklı tıbbi görüntüleme modalitelerinde uygulanabilirliğini araştırmayı hedefleyebilir.

Çıkar Çatışması Beyanı

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı Beyanı

Yazar-1: Fikir, Orijinal Taslak Oluşturma, Yazım, Metot Oluşturma, Deneysel Çalışma, Verilerin Düzenlenmesi, Sonuç ve Tartışma

Yazar-2: Fikir, Yazım, Revize, Metot Oluşturma, Sonuç ve Tartışma

KAYNAKLAR

- Ahmed F, Abbas S, Athar A, Shahzad T, Khan Wa, Alharbi M, ... Ahmet A (2024). Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence. *Scientific Reports*, 14(1): 6173.
- Akter R, Golam M, Doan VS, Lee Jm, Kim DS (2022). Iomt-net: Blockchain-integrated unauthorized uav localization using lightweight convolution neural network for internet of military things. *IEEE Internet of Things Journal*, 10(8): 6634-6651.
- Ali L, Alnajjar F, Jassmi HA, Gocho M, Khan W, Serhani MA (2021). Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors*, 21(5):1688.
- Alsubhibany SA, Abdel-Khalek S, Algarni A, Fayomi A, Gupta D, Kumar V, Mansour RF (2021). Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment. *Computational Intelligence and Neuroscience*, 2021(1):4931450.
- Aydın ZBG, Şamlı R (2020). A comparison of software defect prediction metrics using data mining algorithms. *Journal of Innovative Science and Engineering (JISE)*, 4(1): 11-21.
- Bingöl H, Yıldırım M, Yıldırım K, Alatas B (2023). Automatic classification of kidney CT images with relief based novel hybrid deep model. *PeerJ Computer Science*, 9:e1717. doi: 10.7717/peerj-cs.1717
- Debal DA, Sitote TM (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, 9(1):109.
- Doğru Ş, Altuntaş V (2023). Prediction of Cancer in DNA Sequences Using Unsupervised Learning Methods. *Journal of Innovative Science and Engineering*, 7(1): 40-47.
- Elhoseny M, Shankar K, Uthayakumar J (2019). Intelligent diagnostic prediction and classification system for chronic kidney disease. *Scientific reports*, 9(1):9583.
- Goel A (2022). *Deployed deep learning kidney segmentation for polycystic kidney disease MRI*. *Radiol Artif Intell*, 4 (2): e210205.
- He K, Zhang X, Ren S, Sun J (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.
- Islam MN, Hasan M, Hossain MK, Alam MGR, Uddin MZ, Soylu A (2022). Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. *Scientific Reports*, 12(1):1-14.
- Jeyalakshmi G, Lloyd FV, Subbulakshmi K, Vinduevi G (2024). Application of deep learning in identifying novel biomarkers for chronic kidney disease progression. In *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, 353-358. IEEE.
- Khorram A, Khalooei M, Rezghi M (2021). End-to-end CNN+ LSTM deep learning approach for bearing fault diagnosis. *Applied Intelligence*, 51(2): 736-751.
- Kumar K, Pradeepa M, Mahdal M, Verma S, RajaRao MVLN, Ramesh JVN (2023). A deep learning approach for kidney disease recognition and prediction through image processing. *Applied Sciences*, 13(6): 3621.
- Kumar S, Ratan R, Desai JV (2022). Cotton disease detection using tensorflow machine learning technique. *Advances in Multimedia*, 2022(1): 1812025.
- Lal S, Chanchal AK, Kini J, Upadhyay GK (2024). FPGA implementation of deep learning architecture for kidney cancer detection from histopathological images. *Multimedia Tools and Applications*, 1-19.
- Lee SH, Chan CS, Mayo SJ, Remagnino P (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, 71: 1-13.
- Ma F, Sun T, Liu L, Jing H (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*, 111: 17-26.
- Mehedi MHK, Haque E, Radin SY, Rahman MAU, Reza MT, Alam MGR (2022). Kidney tumor segmentation and classification using deep neural network on CT images. 2022 International Conference on Innovations in Science, Engineering and Technology (ICISSET), 406-411. doi:10.1109/ICISSET54810.2022.10034638.
- Patro KK, Allam JP, Neelapu BC, Tadeusiewicz R, Acharya UR, Hammad M, ... Pławiak P (2023). Application of Kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal CT images. *Information Sciences*, 640:119005.

- Rani S, Malu G, Sherly E (2023). Kidney stone detection from CT images using probabilistic neural network (PNN) and watershed algorithm. In 2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS), 1-6. IEEE.
- Sabanayagam C, Xu D, Ting DS, Nusinovici S, Banu R, Hamzah H, ... Wong TY (2020). A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *The Lancet Digital Health*, 2(6):e295-e302.
- Saif D, Sarhan AM, Elshennawy NM (2023). Deep-kidney: an effective deep learning framework for chronic kidney disease prediction. *Health Information Science and Systems*, 12(1):3.
- Sapra P, Mary SSC, Chauhan A, Parte SA, Nishant N (2023). Hybrid convolutional neural network and extreme learning machine for kidney stone detection. In 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), 936-942. IEEE.
- Senthil K, Vidyaathulasiraman (2021). Ovarian cancer diagnosis using pretrained mask CNN-based segmentation with VGG-19 architecture. *Bio-Algorithms and Med-Systems*, (0):000010151520210098.
- Simonyan K, Zisserman A (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Singh V, Asari VK, Rajasekaran R (2022). A deep neural network for early detection and prediction of chronic kidney disease. *Diagnostics*, 12(1):116.
- Sulistyowati T, Purwanto P, Alzami F, Pramunendar RA (2023). VGG16 deep learning architecture using imbalance data methods for the detection of apple leaf diseases. *Moneter: Jurnal Keuangan dan Perbankan*, 11(1):41-53.
- Sundaramoorthy S, Jayachandru K (2023). Designing of enhanced deep neural network model for analysis and identification of kidney stone, cyst, and tumour. *SN Computer Science*, 4(5):466.
- Tahir FS, Abdulrahman AA (2023). Kidney stones detection based on deep learning and discrete wavelet transform. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(3):1829.
- Ul Hassan M (2018). Vgg16-convolutional network for classification and detection. Neurohive. Available at: <https://neurohive.io/en/popular-networks/vgg16/> [Accessed 10.4.2019].
- Venkatrao K, Kareemulla S (2023). HDLNET: a hybrid deep learning network model with intelligent IoT for detection and classification of chronic kidney disease. *IEEE Access*.
- Wu J (2017). Introduction to convolutional neural networks. National Key Lab for Novel Software Technology, Nanjing University, China, 5(23):495.
- Xia X, Xu C, Nan B (2017). Inception-v3 for flower classification. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC), 783-787, IEEE.
- Yang J, Chen X, Luo C, Li Z, Chen C, Han S, ... Chen C (2023). Application of serum SERS technology combined with deep learning algorithm in the rapid diagnosis of immune diseases and chronic kidney disease. *Scientific Reports*, 13(1):15719.
- Yogesh N, Shrinivasacharya P, Naik N (2024). Novel statistically equivalent signature-based hybrid feature selection and ensemble deep learning LSTM and GRU for chronic kidney disease classification. *PeerJ Computer Science*, 10:e2467.
- Zhang L, Zhang J, Gao W, Bai F, Li N, Sheykhahmad FR (2024). A novel approach for automated diagnosis of kidney stones from CT images using optimized InceptionV4 based on combined dwarf mongoose optimizer. *Biomedical Signal Processing and Control*, 94:106356.
- Zhang X, Agborbesong E, Li X (2021). The role of mitochondria in acute kidney injury and chronic kidney disease and its therapeutic potential. *International Journal of Molecular Sciences*, 22(20):11253.

