

Atf İçin: Pacal, I. ve Alaftekin, M. (2023). Türk İşaret Dilinin Sınıflandırılması için Derin Öğrenme Yaklaşımları. *İğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(2), 760-777.

To Cite: Pacal, I. & Alaftekin, M. (2023). Deep Learning Approaches for Classification of Turkish Sign Language. *Journal of the Institute of Science and Technology*, 13(2), 760-777.

Türk İşaret Dilinin Sınıflandırılması için Derin Öğrenme Yaklaşımları

Ishak PACAL¹, Melek ALAFTEKİN^{2*}

Öne Çıkanlar:

- Türkçe İşaret Dili Tanıma
- CNN-tabanlı Sınıflandırma

Anahtar Kelimeler:

- Türk İşaret Dili
- CNN
- ResNet
- Sınıflandırma

ÖZET:

İşaret dili, sağır ve dilsiz bireylerin duygularını, düşüncelerini ve sosyal kimliklerini çevrelerine aktarabilmek için kullandıkları sözsüz bir iletişim aracıdır. İşaret dili, sağır ve dilsiz bireyler ile toplumun geri kalan bireyleri arasındaki iletişimde kilit bir role sahiptir. Normal insanlar arasında işaret dilinin çok yaygın bilinmemesi ve insan-bilgisayar etkileşiminin artmasıyla birlikte birçok işaret dili tanıma sistemleri geliştirilmiştir. Bu çalışmada, Türk işaret dili için literatürde ilk kez rakam temelli yeni bir veri seti önerilmiştir. Türk işaret dilinin otonom bir şekilde sınıflandırılması ve bu konuda sıkıntı yaşayan insanların iletişimini bilgisayar temelli yapabilmesi için en güncel derin öğrenme yaklaşımları önerilen veri setine uygulanmıştır. Çalışmada özellikle CNN tabanlı VGG, ResNet, MobileNet, DenseNet ve EfficientNet gibi en güncel ve popüler mimariler kullanılmıştır. Deneysel çalışmalarda ResNet152 modeli, %98.76 doğruluk, %98.85 kesinlik, %98.81 duyarlılık ve %98.80 F1-skoru ile diğer modellere göre daha iyi performans gösterdiği gözlemlenmiştir. Ayrıca, deneysel çalışmalarda kullanılan diğer modellerin hepsi %90'ın üzerinde bir başarı oranı göstererek önerilen veri setinin etkililiğini desteklemektedir. Bu, CNN modellerinin Türk işaret dilini tanımayı başarılı bir şekilde tespit yapabildiğini göstermektedir.

CNN-Based Approaches for Automatic Recognition of Turkish Sign Language

Highlights:

- Turkish Sign Language Recognition
- CNN-based Classification

Keywords:

- Turkish Sign Language
- CNN
- ResNet
- Classification

ABSTRACT:

Sign language is a nonverbal communication tool used by deaf and dumb individuals to convey their feelings, thoughts and social identities to their environment. Sign language has a key role in communication between deaf and dumb individuals and the rest of the society. Many sign language recognition systems have been developed with the increase in human-computer interaction and the fact that sign language is not widely known among normal people. In this study, a new number-based data set for Turkish sign language is proposed for the first time in the literature. The most up-to-date deep learning approaches have been applied to the proposed data set in order to classify Turkish sign language autonomously and to enable computer-based communication of people who have difficulties in this regard. The most up-to-date and popular architectures such as CNN-based VGG, ResNet, MobileNet, DenseNet and EfficientNet were used in the study. In experimental studies, it has been observed that the ResNet152 model performs better than other models with 98.76% accuracy, 98.85% precision, 98.81% sensitivity and 98.80% F1-score. Additionally, the other models used in experimental studies all show a success rate above 90%, supporting the effectiveness of the proposed data set. This shows that CNN models can successfully detect Turkish sign language.

¹ Ishak PACAL ([Orcid ID: 0000-0001-6670-2169](https://orcid.org/0000-0001-6670-2169)), İğdır Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İğdır, Türkiye

² Melek ALAFTEKİN ([Orcid ID: 0000-0001-7440-1913](https://orcid.org/0000-0001-7440-1913)), İğdır Üniversitesi, Mühendislik Fakültesi, Mekatronik Mühendisliği Bölümü, İğdır, Türkiye

*Sorumlu Yazar/Corresponding Author: Melek ALAFTEKİN, e-mail: melekalftekin@hotmail.com

Bu çalışma Melek ALAFTEKİN'in Yüksek Lisans tezinden üretilmiştir.

GİRİŞ

Dünya sağlık örgütünün 2021 yılı verilerine göre, dünya nüfusunun %5'inden fazlası (432 milyon yetişkin ve 34 milyon çocuk) işitme kaybı yaşamaktadır. 2050 yılına kadar işitme kaybı yaşayan insanların sayısının 700 milyonu bulacağı tahmin edilmektedir (WHO, 2021). İşaret dili, sağır ve dilsiz bireylerin duygularını, düşüncelerini ve sosyal kimliklerini çevrelerine aktarabilmek için kullandıkları sözsüz bir iletişim aracıdır. İşaret dili, sesli iletişimin aksine parmaklar, el, kol, vücut duruşu, jest ve mimik gibi vücudun farklı bölgelerinin kullanıldığı görsel tabanlı bir iletişim sağlar. İşaret dili, konuşma diline benzer şekilde kendine özgü bir kelime dağarcığı, anlamlı hareketleri ve gramer yapısı bulunmaktadır. Normal insanlar arasında işaret dili yaygın değildir ve işaret dilini yorumlayabilen çok az insan vardır. Bu durum, sağır ve dilsiz insanların iletişimde ciddi zorluklar yaşamasına sebep olmaktadır. Bu yüzden, işaret dilinin tercüme edilmesi iletişimde önemli bir rol oynamaktadır. Günümüzde bu iletişim engelini kaldırmak, sağır ve dilsiz insanların etkileşimlerini ve sosyal yaşam deneyimlerini arttırmak adına pek çok işaret dili tanıma sistemi geliştirilmiştir. Sağır ve dilsiz insan nüfusunun her geçen gün artması da işaret dili tanıma sistemlerinin önemini arttırmıştır.

İşaret dili tanıma sistemleri, sağır ve dilsiz bireyler ile toplumun geri kalanı arasında ortak bir iletişim tabanı oluşturmak amacıyla, işaretleri takip eden, tanımlayan ve bunları anlamlı ifadelere dönüştüren sistemlerdir. İşaret dili tanıma sistemleri iki yaklaşıma dayanmaktadır: Sensor ve görme tabanlı. Sensöre dayalı çalışmalarda, veri eldivenleri (Shukor ve ark., 2015), microsoft kinect (Ren ve ark., 2013), sıçrama hareket kontrolörü (Naglot ve Kulkarni, 2016), derinlik kamerası (Almeida ve ark., 2014) gibi yapılar kullanılarak tanıma işlemi gerçekleştirilir. Bu tarz sistemler kullanıcıya harici bir sistem kullanmayı zorunlu kıldığı için ve maliyetli olmasından dolayı çok tercih edilmemektedir. Görsel tanımaya dayalı tanıma sistemleri ise el hareketleri, jest ve mimik görüntüleri ile ilgilenir. Bu yüzden görme tabanlı sistemler bilgisayarla görme alanındaki trend konulardan biri haline getirmiştir. Son yıllarda, makine öğreniminin bir alt kümesi olan derin öğrenme, sağlam ve gerçek zamanlı bir işlem için görme tabanlı işaret dili tanımda popüler hale gelmiştir (Rastgoo ve ark., 2021). Derin öğrenme mimarisi, büyük veri kümelerini işleyebilmesi, yüksek genelleme kabiliyetine sahip olması popüler bir araştırma konusu haline getirmiştir (Lecun ve ark., 2015). Derin öğrenme, görüntü sınıflandırma, nesne algılama, görüntü elde etme, anlamsal bölütleme, ve insan pozu tahmini gibi bilgisayarla görme alanında iyi bir performans sergilemiştir (Guo ve ark., 2016). Bu da görme tabanlı olan işaret dili çalışmalarında büyük bir önem ve güvenilirlik kazanmasını sağlamıştır. Popüler bir derin öğrenme mimarisi olan Evrişimli Sinir Ağı (Convolutional Neural Network-CNN), büyük görsel veri kümeleri üzerinde işlem yaparak daha doğru bir sınıflandırma ve algılama yapması görüntü işleme alanında araştırmacılar için çok tercih edilen bir mimari haline getirmiştir (Yu ve ark., 2017). CNN'ler bilgisayarla görü olmak üzere konuşma işleme (Nam ve Lee, 2021), yüz tanıma (Sajjanhar ve ark., 2018), tıbbi görüntü analizi (Pacal ve ark., 2020; Alıcı-Karaca ve ark., 2022, Karaman ve ark., 2023), moleküler düzeyde analiz (Ozkok ve Celik, 2023) ve çeşitli hastalıkların patolojik incelenmesi (Karagoz ve ark., 2023) gibi birçok alanda yaygın olarak kullanılmaktadır. CNN mimarisi, hiyerarşik bir yapıya sahiptir. Düşük özelliklerden yüksek özellikleri otomatik olarak çıkararak daha hızlı ve verimli bir sınıflandırma yapma yeteneğine sahiptir. CNN mimarilerinin, özellikle büyük veri kümeleri üzerinde etkin bir şekilde çalışılmasına ve birçok parametreye izin vermesi, video veya statik büyük veri kümelerinde görüntü içeriğini analiz etme, tanıma ve algılama işlemlerinde kullanımını arttırmıştır (Saqib ve ark., 2021, Pacal ve ark., 2021). İşaret dili tanıma sistemlerindeki temel amaç, ilgili el işaretinin doğru bir sınıflandırma ile gerçek zamanlı tanınmasını sağlamaktır. CNN mimarileri, tanıma ve sınıflandırma işlemlerini hızlı ve doğru bir şekilde yapabilme yeteneğinden dolayı işaret dili tanıma sistemleri ile ilgili yapılan birçok çalışmada

araştırmacılar tarafından tercih edilmektedir. Ayrıca, CNN'nin arka plan karmaşası sorununu da aza indirmesi sınıflandırma üzerindeki performansına olumlu diğer bir etkisidir.

Son yıllarda, bilgisayar teknolojilerindeki çalışmalara paralel olarak insan-bilgisayar etkileşiminin ve akıllı tanıma sistemleri gibi görsel tanımaya dayalı sistemlerin artmasıyla birlikte görsel tanıma popüler bir araştırma konusu haline getirmiştir (Al-Hammadi ve ark., 2020). Bu teknolojik ilerlemelerle birlikte, günümüzde, sağlık, eğitim, iletişim hizmeti, bilgi teknolojisi gibi birçok alanda insan-bilgisayar etkileşimi karşımıza çıkmaktadır. Sağır ve dilsiz bireylerin, herhangi bir tercümana ihtiyaç duymadan çevresi ile sağlıklı ve hızlı bir iletişim kurması ve günlük yaşama hızlı ayak uydurmalarını sağlamak için işaret dili tanıma sistemlerinin kullanımını önemli bir ölçüde arttırmıştır. Bilgisayar, telefon ve robotik araçlar gibi insan-bilgisayar etkileşimini arttıran sistemlerde görsel tanımanın popülerlik kazanması işaret dili tanıma sistemlerinin de görsel tanıma alanında popüler konular arasında yerini almasını sağlamıştır. İşaret dili tanıma sistemleri, daha sağlam ve gerçek zamanlı bir tanıma yapmak için daha fazla görsel hesaplama gücüne ihtiyaç duyar. Bu yüzden, CNN mimarilerinin görsel hesaplama gücünün fazla olması, ham verilerden doğru bilgileri elde etmesi ve sağlam bir sınıflandırma yapması, işaret dili tanıma sistemlerinde trend bir çalışma konusu haline getirmiştir. Bu çalışma ile birlikte Türk işaret diline ait rakamlardan oluşan bir veri seti üzerine güncel CNN mimarilerini uygulayarak tanıma sistemi üzerindeki etkileri incelemeye çalışılmıştır.

Özetle, bu makalenin başlıca katkıları şunlardır:

- Türk işaret dilindeki rakamlardan oluşan özgün bir veri seti oluşturularak literatüre kazandırılmıştır.
- Türk işaret dili tanıma için, statik el hareketlerini sınıflandırma görevini düşük hata oranlarıyla çözebilen ve benzer rakamlara ait el işaretleri arasındaki küçük farkı ayırt edebilen güncel CNN mimarileri ile yeni bir uçtan uca eğitilebilmesini sunmaktadır. (2 ile 7 ve 3 ile 8 benzerdir.)
- Basit veri artırma tekniklerinin ağırlıklı olarak doğru verimli bir artırması incelenmiştir.
- Görüntü içindeki ilgili el işaretinin sınıflandırılması için veri seti üzerinde herhangi bir ön işlem yapılmadan uçtan uca bir eğitim gerçekleştirilmiştir.

Literatür Taraması

İşaret dilinin insanlar arasında yaygın olarak bilinmemesi ve bir tercümana ihtiyaç duyulması dünyada var olan diğer diller için de önemli bir sorundur. Literatüre baktığımızda, son yıllarda, bu sorunu azaltmayı amaçlayan pek çok işaret dili tanıma sistemleri ile ilgili çalışmalar mevcuttur. Özellikle son yıllarda, ilerleyen teknolojiyle birlikte insan-bilgisayar etkileşiminin iyice artması işaret dili tanıma sistemlerini literatür de aktif araştırma konusu haline getirmiştir. Arka plan karmaşıklığı, karmaşık el ve mimik hareketleri, gürültü ve genelleme kabiliyeti gibi işaret tanıma sistemlerinin karşı karşıya olduğu pek çok zorluk vardır. Bu nedenle birçok araştırmacı bu sorunların üstesinden gelmek için farklı çözümler önermiştir. İşaret dili, el hareketlerini ve yüz ifadelerini içeren görsel bir iletişim olması bilgisayarlı görme algoritmaları için uygun bir test-yazılım yatağı haline getirmiştir (Wu ve Huang, 1999).

İşaret dili tanıma, bilgisayarla görme alanında önemli ve zorlu bir çalışma alanı olarak kabul edilir. Bunun başlıca nedenleri arasında işaret dillerinin karmaşık görsel ve jest-mimik yapısının olması ve otomatik tanıma ile ilgili az sayıda veri tabanı ve çalışmanın olmasıdır (Rezende ve ark., 2021). Son yıllarda, CNN mimarileri, görsel verilerdeki özellikleri öğrenme kapasitesi, görüntü analitik görevlerinde yazarlar tarafından oldukça etkili bir şekilde kullanılmaktadır. Özellikle işaret dili tanıma gibi görsel tabanlı tanıma görevlerinde CNN tabanlı yaklaşımlar etkin çalışmalar için yeni fırsatlar sunmaktadır. Hint işaret dilinde, Rao ve ark. (2018) özçekim işaret dili hareketlerini sınıflandırmak için

bir CNN mimarisi önerdiler. Yaptıkları çalışmada mobil kullanım için herhangi bir uygulamaya gerek duymadan özçekim modunda işaretleri yakalayan bir tanıma sistemi geliştirdiler. 5 farklı gönüllü tarafından çeşitli arka planlar kullanarak 5 farklı görüş açısında 200 işaretten oluşan bir veri seti kullandılar. 2 sn süren videolardan oluşan özgün veri setini, 4 evrişim katmanı ve her evrişim katmanında farklı boyutlarda filtreler sahip bir CNN mimarisi ile sınıflandırdılar. Önerdikleri CNN mimarisi ile daha az eğitim süresi ve doğrulama kaybı elde ederek sınıflandırmada %92.88 tanıma oranına ulaştılar. Benzer bir şekilde, Khari ve ark. (2019), CNN mimarilerinden biri olan ince ayarlı bir VGG19 (Visual Geometry Group-19) modeli kullanarak bir RGB ve RGB-D statik hareket tanıma yöntemi önerdiler. Amerikan işaret dili veri setinde, yazarlar önerdikleri model ile %94.8 tanıma oranı elde ettiler. Myagila ve Kivalo (2021), Tanzanya'da konuşma engelliler için derin öğrenme tekniklerinden Destek Vektör Makinesi (Support Vector Machine- SVM) ve CNN'in performanslarını değerlendiren bir çalışma sundular. Veri setindeki görüntülerin boyutlarını küçültmek ve işlem hızını azaltmak için temel bileşen analizi kullandılar. Birleştirilmiş 5x2cv F testini kullanarak her iki modelin performanslarındaki farkı karşılaştırmaya çalıştılar. Yaptıkları çalışmanın neticesinde, CNN'nin doğruluk, hatırlama ve kesinlik gibi tüm parametrelerde %96 oranında bir başarımlı gösterdiği ve SVM modelinden %1 daha yüksek doğruluk ve hassasiyet değerine sahip olduğunu buldular. Suri ve Gupta (2019), yapılan el hareketlerini kaydeden özel tasarlanmış bir IMU (Inertial Measurement Unit) cihazı ve CNN mimarisi kullandıkları bir tanıma sistemi önerdiler. IMU cihazı, Hint işaret dilinde tanımlı işaret dili cümlelerini genel cümle ya da soru cümlesi olarak farklı iki grupta kaydeder. Her iki cümle tipini sınıflandırmak için tek boyutlu iki CNN mimarisi kullandılar. Önerdikleri sistemin etkinliğini ölçmek için iki farklı cümle tipinden oluşan veri setini birleştirerek geleneksel bir CNN mimarisinde eğiterek bir karşılaştırma yaptılar. Yaptıkları çalışmalar sonucunda, önerilen CNN mimarisinde genel cümleler için %94.20 ve soru cümleleri için %95.00 iken geleneksel CNN mimarisi ile yapılan çalışmada ise %93.50'lik bir sınıflandırma doğruluğu elde ettiler. CNN tabanlı başka bir çalışmada, Ma ve ark. (2022) tarafından önerildi. Önerilen yöntemde, CNN mimarisinin performansını iyileştirmek için bir TSM (Two-Stream Mixed) içeren bir yapı kullandılar. Amerikan işaret dili (AİL) veri seti üzerinde yapılan çalışma ön işleme, TSM bloğu ve CNN olmak üzere üç aşamadan oluşmaktadır. Sistemde, art arda yapılan iki dinamik el hareketi görüntüsü, ön işlem kısmında yeniden boyutlandırma, dönüştürme ve büyütme işlemlerinden geçer. Ön işlem aşamasından gelen girdiler, TSM bloğunda toplama ve birleştirme işlemleri yapılarak füzyon özellik haritalarını çıkartılır. Çıkartılan özellikler sınıflandırma işlemi için CNN mimarisine aktarılır. Sistemin performansını değerlendirmek için TSM-LeNet, TSM-AlexNet, TSM-ResNet18 ve TSM-ResNet50 olmak üzere 4 tane CNN tabanlı model ve geleneksel CNN mimarisi ile karşılaştırdılar. Deneysel sonuçlar, TSM bloğuna sahip CNN modellerinin sınıflandırma için daha iyi bir performans gösterdiğini ve bu modeller içinde ise TSM-ResNet50 modelinin MNIST ve AİL veri kümeleri üzerinde %97,57 ile en iyi doğruluğa sahip olduğunu gözlemlediler.

İşaret dili tanıma sistemlerinde yaşanan en büyük problemlerden biri arka plan karmaşıklığıdır. Araştırmacılar, bu sorunu çözmek için tek düze bir arka plana sahip veri kümesi ya da el işaretinin arka plandan ayrılmasını sağlayan bir dizi ön işlem teknikleri önerdiler. Böyle bir sorunla baş etmek için Gangrade ve Bharti (2020), Hint işaret dili için microsoft kinect sensörüne ve CNN tabanlı bir sistem önerdiler. Ten rengi, arka plan ve el ve yüzün örtüşmesinde oluşabilecek sorunları çözmek için sensör yardımıyla elin derinlik görüntüsünü aldılar. Bu işlem ile birlikte arka plan karmaşıklığı, ışık ve aydınlatma gibi ortam şartlarında ilgili el görüntüsünü ayırmayı başardılar. Algılanan el işaretinin gerçek zamanlı algılamasını sağlamak için CNN'i kullanarak %99.3 doğrulama elde ettiler. Alawwad ve ark. (2021), Arap işaret dilinde Faster R-CNN tabanlı bir tanıma sistemi önerdiler. Veri seti, standart telefon kameraları yardımıyla çekilmiş, farklı arka planlara sahip el hareketlerini içeren 15.360 görüntüden

oluşmaktadır. Görüntünün özelliklerini çıkarmak ve haritalamak ve görüntü içindeki ilgili elin konumunu öğrenmek için Faster R-CNN'i yeniden tasarladılar. Önerdikleri çalışmada farklı arka planlar kullanarak %93'lük bir doğrulama elde ettiler. Tasmere ve Ahmed (2020), Bengal işaret dilinde harfleri tanımayaya yönelik bir işaret dili tanıma sistemi önerdiler. İlgili el işaretlerini HSV (Hue Saturation Value) ve YCbCr renk uzaylarını kullanarak arka plandan ayırdılar. 37 harften oluşan Bengal işaret dili için oluşturulan veri seti, 6 farklı gönüllüden alınan toplam 3.219 görüntü içermektedir. Elde ettikleri yeni veri setini 37 sınıfa ayırarak derin bir CNN mimarisi ile eğittiler. Yapılan deneysel çalışmaların sonucunda %99.22'lik doğruluk oranı elde ettiler. Özcan ve Baştürk (2021), hastanelerin acil servisinde çok sık kullanılan kelimelerin sınıflandırılması için transfer öğrenme ile GoogLeNet tabanlı bir CNN modeli önerdiler. 13.186 görüntü içeren ERUSLR adındaki veri seti ile eğittiler. Modelin performansını arttırmak için grid arama (Grid Search-GS), rastgele arama (Random Search-RS) ve genetik algoritma (Genetic Algorithm-GA) gibi yöntemler ile modelin eğitim parametreleri optimize ettiler. GA destekli modelinin daha iyi sınıflandırma yaptığını ve GoogLeNet tabanlı CNN modellerinin daha düşük bir başarımla elde ettiklerini gördüler. İşaret dili tanıma sistemlerinde araştırmacıların karşılaştıkları bir diğer zorluklardan biri de art arda yapılan dinamik el hareketlerinin doğru bir şekilde sınıflandırılmamasıdır. Böyle bir sorunu çözmek için Luqman ve El-Sayed (2022), işaret dilindeki dinamik el hareketlerini tanımak için CNN mimarisini LSTM ile birleştirdikleri bir çalışma önerdiler. Önerdikleri sistemi, verimlilik, ölçeklenebilirlik ve hesaplama süresi olmak üzere üç kademede incelediler. Dinamik el hareketlerinin uzamsal bilgileri için CNN ve zamansal bilgileri içinde LSTM'den yararlanarak sistemi beslediler. Video karelerindeki ardışık yapılan hareketler arasındaki mutlak zaman farkı hesaplanarak tanıma sistemi beslediler. Önerilen modelleri değerlendirmek için iki işaret diline ait üç veri seti kullandılar. Yapılan deneysel çalışmalar neticesinde, CNN-LSTM modellerinin dinamik hareketlerde daha iyi uzamsal-zamansal özellikler yakaladığını ve %99'un üzerinde tanıma doğruluğu ile diğer tekniklerden daha iyi performans gösterdiğini gözlemlediler. Suliman ve ark. (2021), özellik çıkarmak için CNN mimarisi olan AlexNet'i ve sınıflandırma için uzun kısa süreli hafıza (Long short-term memory-LSTM) sistemini önerdiler. LSTM ile videoların sıralı yapısı hafızada tutmak için kullandılar. Model, işaretleri yapanlara bağlı olarak %95.9'luk elde ederken farklı kişiler için %43.62'lik bir genel tanıma doğruluğu elde ettiler.

Literatürde, yazarlar CNN tabanlı çalışmalarda çoğunlukla doğru bir sınıflandırma ve gerçek zamanlı bir algılama üzerinde durmuşlardır. Bhushan ve ark. (2022), Naive Bayes, k-en yakın komşu (K-nearest neighbor-KNN), rastgele orman, XGBoost, destek vektör sınıflandırıcı (Support Vector Classifier-SVC), lojistik regresyon, stokastik gradyan içeren sınıflandırıcı (Stochastic Gradient Descent Classifier-SGDC) ve CNN gibi makine öğreniminde popüler olan sınıflandırma tekniklerini işaret dili MNIST veri kümesi üzerinde uygulayarak karşılaştırmalı bir çalışma sundular. En iyi sınıflandırma sonuçlarını CNN algoritmasında %91.41 ve rastgele orman algoritmasında ise %84.43 doğruluk oranı elde ettiler. Yazarlar, CNN'nin diğer algoritmalara kıyasla sınıflandırmada daha iyi bir performansa sahip olduğunu kanıtlamış oldular. Benzer bir sınıflandırma çalışması Marais ve ark. (2022) tarafından önerildi. Sınıflandırmada daha doğru sonuçlar elde etmek için LSA64 veri setindeki el işaretlerine farklı özellik çıkarma ve segmentasyon algoritmaları uyguladılar. Farklı ön işlemden geçirilen görüntüler, ResNet ve Pruned VGG ağı gibi popüler CNN mimarileri ile eğittiler. Görüntülerdeki ilgili el işaretinin koordinatını belirlemek için MediaPipe Holistic ve sınıflandırmak için 1D-CNN kullandılar. Yaptıkları çalışma sonucunda ham verileri içeren veri setinde Pruned VGG ile yapılan sınıflandırmada %95.50 doğruluk ve MediaPipe Holistic ile koordinatları belirlenen işlenmiş veri setinde 1D-CNN kullanılarak yapılan sınıflandırma da %94.91'lik bir doğruluk elde ettiler. İşaret dili algılama sistemlerindeki diğer ana zorluklardan biri gerçek zamanlı tanıma olmasıdır. Aiouez ve ark. (2022), Arap işaret dili alfabeti

için gerçek zamanlı bir algıma sistemi önerdiler. Sistemin gerçek zamanlı olması ve daha doğru bir sınıflandırma elde etmek için derin öğrenme tabanlı nesne algılama modeli olan YOLOv5'i önerdiler. Farklı arka plan, farklı boyutlarda eller, aydınlatma koşulları ve aksesuarlı/aksesuarsız yaklaşık 15.000 görüntü içeren 28 Arapça işaretten oluşan veri setini, YOLOv5'in farklı versiyonları ile eğittiler. Sistem performansını değerlendirmek için Faster R-CNN ile karşılaştırdılar. Yapılan çalışmalar neticesinde en iyi sonucu uyarlanmış YOLOv5'te elde ettiler. Gerçek zamanlı diğer bir çalışma ise Wang ve ark. (2022) tarafından önerildi. İnsanların işaret dilini duymasına yardımcı olmak için işaret dilini seslere çevirmek üzere DeepSLR adlı yeni bir gerçek zamanlı bir sistem önerdiler. Kullanıcılardan işaretleri almak için bir IMU sensörü ve çok kanallı sEMG sensörleri ile gömülü iki kol bandı kullandılar. Sensörler yardımıyla yakalanan hareketler herhangi bir segmentasyon işlemi yapılmadan dikkat tabanlı bir kodlayıcı-kod çözücü ve çok kanallı bir CNN modeli önerdiler. Akıllı telefonlarda uygulanabilen DeepSLR, 4 kelimelik bir cümleyi 1.1 saniyede algılama yeteneği ve %10.8'lik ortalama kelime hatası ile gerçek zamanlı bir tanıma sundu. Nuñez-Prieto ve ark. (2019), standart bir akıllı telefon kamerası kullanılarak alınan el işareti verileri için FPGA (Field Programmable Gate Array) tabanlı bir CNN mimarisi ile gerçek zamanlı bir tanıma sistemi önerdiler. CNN tabanlı ZynqNet (Gschwend, 2020) mimarisinin kullanıldığı çalışmada veri artırma ve öğrenme aktarımı teknikleri kullanılarak sınıflandırma doğruluğunu %80,1'e kadar arttırdılar. Sistemin bellek gereksinimlerini ve veri işleme modellerini geliştirmek için çeşitli donanımsal teknikler uygulayarak CNN mimarisinin 23,5 FPS performans sunan bir FPGA'ya gömülebileceğini gösterdiler. Yaptıkları deneysel çalışmalar neticesinde yazarlar, 3D CNN mimarilerinin, donanım hızlandırıcı sistemlerde video kareleri içindeki uzamsal ve zamansal özelliklerin çıkarılması için kullanılabilmesi ve dinamik el hareketi tanıma sistemlerinde kullanım olasılığının önünü açtılar.

Özetlenen çalışmalarda görüldüğü üzere araştırmacılar tarafından çeşitli işaret dilleri için farklı algılama yöntemleri önerilmiştir. Bunlarla birlikte, Türk işaret dili için literatüre baktığımızda sınırlı sayıda çalışmanın mevcut olduğu görülmektedir. Bu çalışma, Türk işaret dilindeki rakamları sınıflandırmak için derin öğrenme yöntemlerinin kullanımına odaklanmaktadır. Veri setindeki işaretlerin sınıflandırılması için en popüler ve güncel derin öğrenme yöntemlerinin performansı incelenmiştir. Derin öğrenme modellerinin performansını artırmak için transfer öğrenme ve veri artırma teknikleri de kullanılmıştır. Bu çalışmanın amacı, CNN tabanlı mimarilerin kullanımıyla işaret dili hareketlerinin doğru bir şekilde sınıflandırılması ve otomatik olarak tanınmasıdır. Bu sayede sağır-dilsiz insanların normal hayata adapte olmaları ve tercümana olan ihtiyaçlarının azaltılması hedeflenmektedir.

MATERYAL VE METOT

Derin Öğrenme

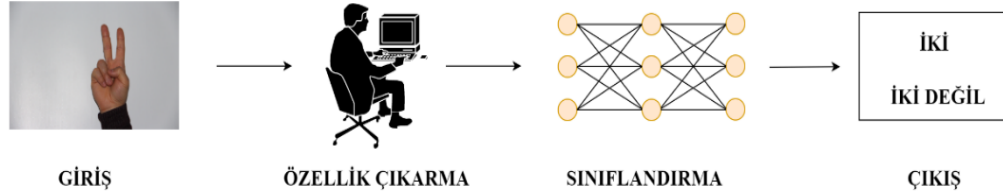
Bilgisayar teknolojilerinin hızlı gelişimiyle birlikte bilginin hızla yayılması ve beraberinde veri çokluğunu ortaya çıkarttı. Veri çokluğu içerisinde istenilen bilgilerin hızlı ve doğru bir şekilde elde edilmesi amaçlanmaktadır. Bunun için veri madenciliği, veri analitik, makine öğrenmesi gibi öğrenme teknikleri önerilmiştir. Son yıllarda, yapay zeka ve makine öğreniminin alt alanlarından biri olan derin öğrenme, birçok bilimsel disiplinde kullanılarak en yaygın hesaplama yaklaşımı haline geldi ve bu alanda çok iyi sonuçlar elde etti (Alzubaidi ve ark., 2021). Geleneksel makine öğrenimi teknikleri, ham veriler üzerinden özellik çıkarmak ile sınırlıdır (Halbouni ve ark., 2022). Derin öğrenme ise yapılandırılmamış veya etiketlenmemiş ham verilerden aşamalı olarak özellik çıkararak bu verilerden doğru bir sınıflandırma gerçekleştiren makine öğreniminden geliştirilmiş bir yaklaşımdır.

Araştırmacıların derin öğrenmeye olan ilgilerinin artmasıyla birlikte anlamsal ayrıştırma (Bordes ve ark., 2012), görüntü tanıma (He ve ark., 2016), oyun oynama (Justesen ve ark., 2020), tıbbi görüntü

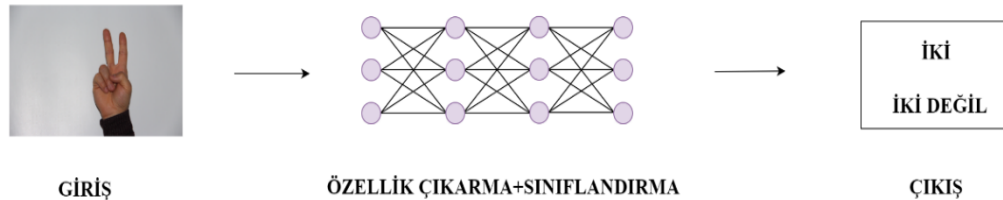
analizi (Pacal ve ark., 2022) ve siber güvenlik (Burukanlı ve ark., 2019) gibi çeşitli alanlarda yapay zeka yaygın olarak kullanılmaktadır. Derin öğrenme, bir görüntüde pikseller ile işlem yaparak piksel başına düşen yoğunluk vektör değeri, bir dizi kenar veya belirli bir şekle sahip bölgeleri daha soyut bir hale getirerek öğrenme görevini daha basit bir hale getirir (Ongsulee, 2017; Pacal, 2022). Günlük hayatımızda, çok sık kullandığımız Siri ve Google Asistan gibi sanal asistanlar, Netflix ve Youtube gibi kişiselleştirilmiş platformlarda da derin öğrenme algoritmaları kullanılmıştır. Derin öğrenmenin devasa büyüklükteki veri kümelerinin eğitimini kolaylaştırması ve teknolojinin ilerlemesiyle birlikte geliştirilen grafik işleme birimlerinin (GPU'lar) muazzam bir paralel hesaplama gücü sağlamasıyla derin öğrenmede işlem hızı artmıştır. Bu da gerçek dünya uygulamalarında derin öğrenmeyi popüler hale getirmiştir.

Derin öğrenme, makine öğrenimi metodolojilerinden yararlanır fakat öğrenme mekanizmaları birbirinden farklıdır. Derin öğrenme ve geleneksel makine öğrenimi arasındaki temel fark Şekil 1 'de gösterilmiştir. Geleneksel makine öğrenmesi sürecinde, işlenmiş veri seti denetimli öğrenme algoritmaları kullanılarak eğitim yapılır. Makine öğreniminde özellik çıkarma adımı dışardan manuel olarak yapılır. Derin öğrenme ise etiketlenmemiş ham verileri kullanarak eğitim sırasında çok katmanlı sinir ağı tarafından özellik çıkarma adımı otomatik olarak gerçekleştirilir (Karaman ve ark, 2023). Eğitim sırasında ham verileri kullanılabilmesi derin öğrenme algoritmalarının benzersiz özelliklerinden biridir (Chaudhuri ve ark., 2011). Çok katmanlı sinir ağlarında hiyerarşik yapılar kullanarak gerekli bilgiler doğrudan ham veriler üzerinden sağlaması daha doğru bir sonuca götürür.

GELENEKSEL MAKİNE ÖĞRENİMİ



DERİN ÖĞRENME



Şekil 1. Geleneksel Makine Öğrenimi ve Derin Öğrenme

Derin öğrenme, çok sayıda özellik katmanına sahip olan etkili bir makine öğrenimi algoritması olarak geliştirilmiştir. Derin öğrenme algoritmaları, büyük veriler üzerinde kendi kendine öğrenebilmesi ve ön-eğitim (pre-training) olmaksızın özellik çıkarma işlemini zahmetsiz bir şekilde yapması hızlı bir eğitim sürecine ve daha doğru sonuçlara ulaşılmasını sağlamıştır (Fan ve ark., 2019). Bu sebeple, derin öğrenme algoritmaları bilgisayarla görme alanında araştırmacıların kapsamlı çalışılmasına izin vererek nesne algılama alanında ilham verici bir performans sergilemiştir. Derin öğrenme, yapay sinir ağı,

evrişimli sinir ağı, tekrarlayan sinir ağı gibi yaklaşımları mevcuttur. Bu yaklaşımlardan, insan beyin sinir yapısı ilham alınarak oluşturulan evrişimli sinir ağları, nesnelere sınıflandırılmasında daha az hata payı bırakarak mevcut nesne algılama algoritmalarından daha iyi olduğunu kanıtladığı bir nesne algılama yarışması olan ImageNet sınıflandırmasında dikkatleri üzerine çekmeyi başararak literatürdeki yerini almıştır (Krizhevsky ve ark., 2012). Derin öğrenme modelleri çeşitli optimizasyon teknikleri ile hem model bazında hem de hiper-parametre düzeyinde optimize edilerek başarımları daha da arttırılabilir (Karaman ve ark., 2022).

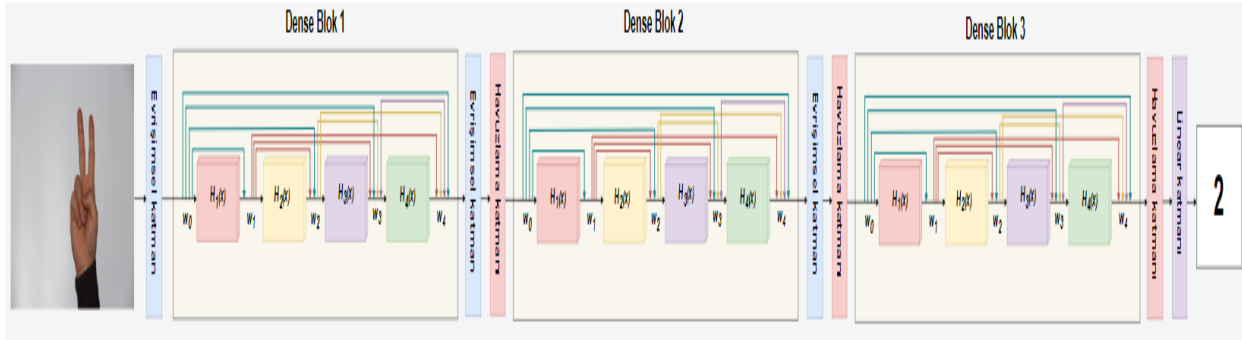
CNN Yaklaşımları

CNN mimarisi, derin öğrenme alanındaki en önemli temsili ağlardan biridir (Li ve ark., 2021). CNN mimarisi, çok sayıda girdi üzerinde evrişim operatörleri kullanarak otomatik olarak işlem yapabilme yeteneğine sahiptir. Bu nedenle, son yıllarda, bilgisayarla görme alanında dikkate değer bir performans göstermiştir (Yu ve ark., 2017). CNN mimarisi katmanlı bir yapıya ve hiyerarşik öğrenme özelliğine sahiptir. Her katmandaki özellik, kendinden bir önceki katmanın yerel alanının evrişimli çekirdek ağırlıklarının aktarılmasıyla oluşturulur. Bu, CNN'de eğitim parametrelerini büyük ölçüde azaltır. Böylelikle, CNN, görüntüden özellikleri eksiksiz bir şekilde çıkararak nesne algılama için uçtan uca bir öğrenme modeli sunar (Zhiqiang ve Jun, 2017).

Evrişimli sinir ağlarının genel mimarisi, bir giriş katmanı, bir çıkış katmanı ve sırasıyla bir evrişim katmanı, aktivasyon katmanı, havuzlama katmanı, tam bağlı katman (Fully Connected-FC) ve kayıp katmanı içeren gizli katmandan oluşur. Evrişimli sinir ağının ilk katmanı olan evrişim katmanıdır. Bu katman, giriş verileri üzerine evrişim adı verilen doğrusal bir işlem uygulayarak bir sonraki katmanın girdisini belirler. Bu katmanda bulunan çeşitli filtrelerin yardımıyla giriş verilerinin en temel özellikleri çıkartılır. Aktivasyon katmanı, sinir ağının aktifleşmesini sağlayarak ağın çıktısını belirleyen katmandır. Bu katman, matematiksel bir dönüşüm uygulayarak girdinin bir sonraki katman için etkinleştirilip etkinleştirilmeyeceğine karar verir. Bu yüzden aktivasyon katmanı sinir ağları için önemli bir parametredir. Havuzlama katmanı, genellikle iki ardışık evrişim katmanı arasına dahil edilir. Maksimum, minimum ve ortalama gibi çeşitli havuzlama teknikleri kullanılarak girdinin derinliği değiştirilmeden alt örnekleme yapar ve girdinin uzamsal boyutunu küçültür. Tam bağlı katman (Fully Connected Layer-FC), kendinden önceki tüm katmanlar ile bağlantılıdır. Evrişim katmanları tarafından çıkarılan ve havuz katmanları tarafından alt örnekleme özellik haritaları bu katmanına girdi olarak bağlanır. Bu katmanın çıktısı ile nihai sonuç elde edilir. Ayrıca, Tam bağlantılı katmandan sonra gelen ve eğitim sırasında, tahmin ve gerçek değeri karşılaştırarak, aralarındaki farkı olabildiğince en aza indiren bir kayıp katmanı vardır.

1989 yılından günümüze kadar CNN mimarisinin çeşitli modelleri piyasaya sürüldü. 2012 yılında Krizhevsky ve ark. tarafından AlexNet (Krizhevsky ve ark., 2012) mimarisi nesne algılamada kullanarak araştırmacıların ilgisini büyük ölçüde CNN üzerine çekmiştir. Bu atılım ile yeni birçok CNN tabanlı mimarilerin hızla geliştirilmesiyle literatürdeki yerini sağlamlaştırmıştır. Bu çalışmada, son 10 yılda oldukça yaygın olan bazı CNN modellerini açıklayacağız. Buradaki temel amacımız, Türk işaret dili veri setini kullanarak verdiğimiz CNN modellerin performanslarını karşılaştırmaktır. VGG mimarisi, Simonyan ve Zisserman tarafından geliştirilen popüler bir CNN mimarisidir. Genel olarak VGG, görüntülerin sınıflandırmasında ve yerelleştirmede önemli sonuçlar elde etmiştir. 2014-ILSVRC yarışmasında artan derinliği, homojen topolojisi ve sadeliği ile ön plana çıkmıştır. Çok fazla parametreye sahip VGG-16 ve VGG-19 versiyonları sırasıyla 16 ve 19 katmandan oluşmaktadır (Simonyan and Zisserman, 2014). Bu çalışmada kullanılan bir diğer popüler CNN mimarisi ResNet (He ve ark., 2016), 152 katmana sahip ve diğer modellere göre son derece derin bir yapıdır. CNN'lerde artık öğrenme

kavramını tanıtarak çok derin ağların eğitilebilmesini sağlayarak etkin bir öğrenme sağladı. Katman sayısına bağlı olarak ResNet-34, ResNet-50 ve ResNet-152 gibi türleri geliştirilmiştir. Bir diğer ağ mimarisi, mobil uygulamaları için tasarlanmış MobileNet'tir (Howard ve ark., 2017). MobileNet, derinliği ayrılabilir evrişimden faydalanarak parametre sayısı, işlem hacmi ve model karmaşıklığı değiştirilebilir bir mimariye sahiptir. Aerodinamik yapıya sahip olması, donanımsal olarak küçük yapı mobil ve gömülü cihazlarda kullanılmasını kolaylaştırır. MobileNet, çok daha küçük yapılarda hafif derin sinir ağ olması ve az parametre sahip olması diğer birçok popüler CNN mimarilerine göre yüksek bir performansa ulaşmıştır. Bu çalışmada kullanılan bir diğer ağ DenseNet'tir (Huang ve ark., 2017). Şekil 2'de gösterilen DenseNet mimarisi, ResNet mimarisinden yararlanılan ağ yapısında, bir katmanda üretilen bilgi bir sonraki katmana ekleyerek bir sonraki katmanın girdisini oluşturmaktadır. Bu sayede, DenseNet sinir ağının tüm katmanları ileri besleme yapılarak hem ağ parametre sayısı azaltılmış hem de ağ içinde özellik yayılması ile bilginin yeniden kullanılması sağlanmıştır. DenseNet'in diğer bir avantajı ise kaybolan gradyan sorununu hafifleterek daha doğru ve daha verimli olmasıdır. EfficientNet mimarisi, 2019 yılında, Google Research Brain ekibi tarafından daha verimli ve etkin bir sinir ağı tasarımı oluşturulabilmek için önerilmiştir (Tan ve Le, 2019). EfficientNet, bir bileşik katsayı uygulanarak derinlik, genişlik ve çözünürlük boyutlarını eşit şekilde ölçeklenebilen evrişimli sinir ağı mimarisi ve ölçeklendirme tekniğidir. Yazarlar, ağ genişliğini, derinliğini ve çözünürlüğünü rastgele ölçeklendirmek yerine bir dizi önceden ayarlanmış ölçeklendirme parametreleri ile ölçeklenmesi gerektiğini savunmuşlardır. EfficientNet mimarisi, hesaplama gücünü verimli bir şekilde kullanarak yüksek doğruluk elde etmeyi amaçlayan ve ölçeklenebilir bir CNN mimarisi olarak tanımlanabilir. Bir sinir ağının ölçeklenebilmesi, ağın büyüklüğü arttıkça parametre sayısının artması ve küçüldükçe parametre sayısının azalması anlamına gelir. Ayrıca, EfficientNet mimarisi, ImageNet veri seti üzerinde eğitilerek diğer öğrenme modelleri ile karşılaştırıldığında, daha az sayıda parametre ile en yüksek doğruluğa ulaşabilmesiyle öne çıkmaktadır.

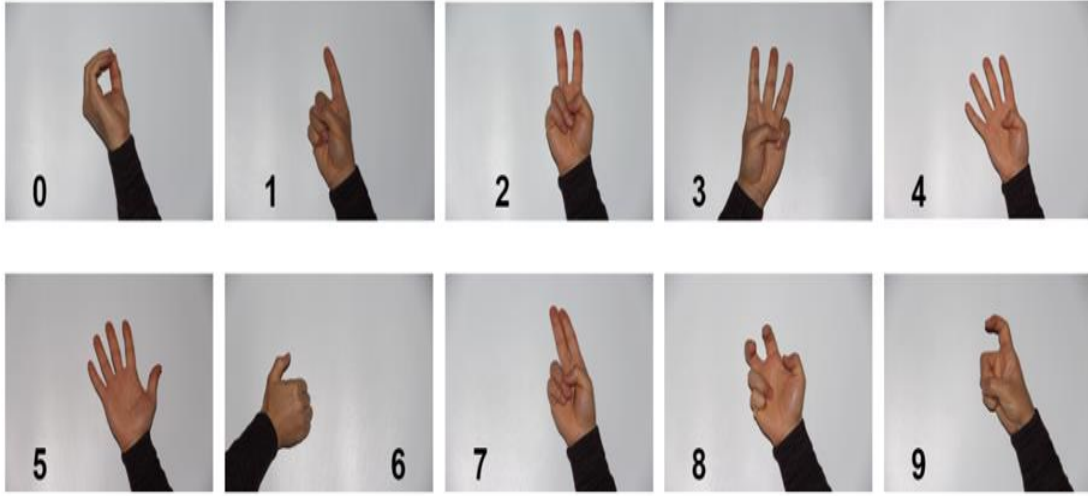


Şekil 2. DenseNet Mimarisi

Türk İşaret Dili için Özgün Veri seti

CNN mimarilerinin performanslarını değerlendirmek için Türk işaret diline özgü 0'dan 9'a kadar olan rakamları temsil eden el hareketlerini içeren görüntülerinden oluşan bir veri seti kullanılmıştır. İşaret dili tanıma sistemlerinde kullanılan işaret dili veri setleri kendine özgü hareketler içeren görüntülerden oluşmaktadır. Türk işaret dili veri seti oluşturulurken bir uzmana danışılarak özgün bir veri seti oluşturulmuştur. Şekil 3'te Türk işaret dilindeki tüm rakam hareketlerinin RGB görüntüleri gösterilmektedir. Ayrıca, oluşturduğumuz veri seti literatürde Türk işaret dilindeki rakamları içermesinden dolayı da özgünlüğünü korumaktadır. Veri setindeki tüm görüntüler, CANON 700d fotoğraf makinesi ile 6 farklı gönüllüden çekilmiştir. Her bir rakam için tüm opsiyonlar düşünülerek el pozisyon ve açısında sürekli değişiklikler yapıldı. Böylece, deneysel veri setinin daha çeşitli olması planlandı. Toplam 1.500 görüntü içeren veri seti elde ettik. Giriş görüntülerinin daha kaliteli olması ve

eğitim sırasında arka plan karmaşıklığı önlemek için tek düze bir renkten (beyaz) oluşan bir arka plan tercih edilmiştir.

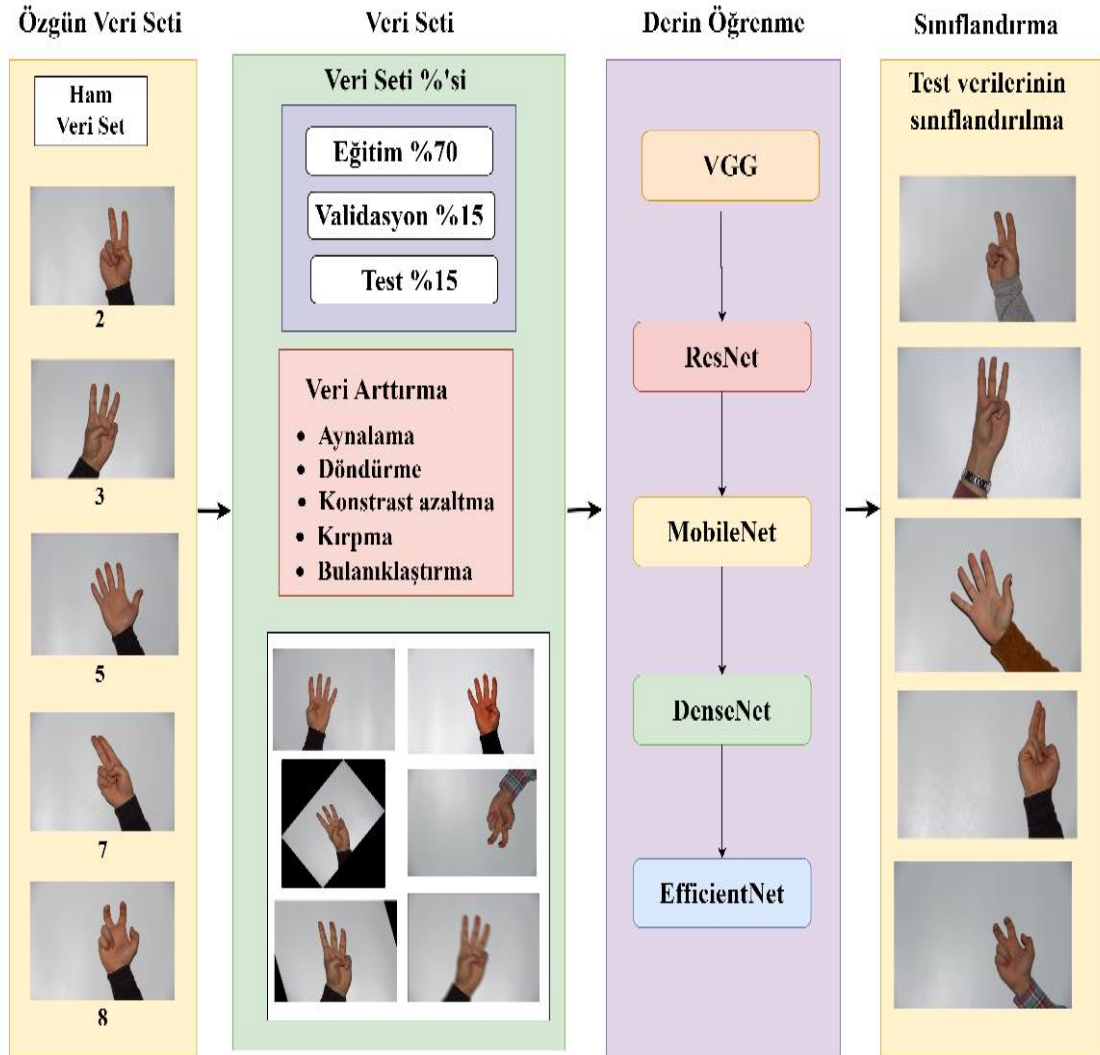


Şekil 3. Türk İşaret Dilinde Rakamlar

Önerilen Yöntem

Deneyisel sonuçları daha objektif bir şekilde değerlendirmek için her bir CNN oluşturduğumuz Türk işaret dili veri seti ile eğitilmiştir. Hem popüler hem de güncel CNN mimarileri olan VGG, ResNet, MobileNet, DenseNet ve EfficientNet mimarileri kullanarak daha geniş bir çalışma yürütülmüştür. Çalışmada, aynı çözünürlüğe sahip 5184x3456 orijinal boyutlarında görüntüler kullanılmıştır. Türk işaret dili veri seti üzerinde derin öğrenme mimarilerin performansı karşılaştırılırken izlenen yol Şekil 4'te gösterilmektedir. Bu mimarilerde daha kararlı bir çalışma sunmak için tüm mimarilerde SGD (Stochastic Gradient Descent) optimizasyon algoritması, 0.001 öğrenme oranı (adımı) ve 0.9 momentum olarak kullanılmaktadır. Derin öğrenme mimarileri son derece veriye aç sinir ağlarıdır (Li ve ark., 2018). Derin öğrenmeden iyi bir performans modeli oluşturmak ve aşırı öğrenme problemini önlemek için veri setinin büyüklüğü ve çeşitliliği önemli bir faktördür. Bu sorunu düzgün bir şekilde ele almak için önerilen iki yaygın yöntem vardır: Transfer öğrenme ve veri artırma teknikleri. Transfer öğrenme, önceden eğitilmiş bir sinir ağından kazanılan bilginin yeni bir görevde kullanılmasını sağlayan bir makine öğrenme tekniğidir (Pan ve Yang, 2010). Geleneksel makine öğreniminde eğitim ve test verilerinin, girdi özellik uzayı ve veri dağıtım özelliklerinin eşleştirilmesi zor ve maliyetli olabilir (Weiss ve ark., 2016). Bu nedenle, Transfer öğrenme, kazanılan bilgilerle yüksek performanslı bir öğrenme ihtiyacını karşılar. Transfer öğrenme, genellikle küçük ve sınırlı veri setlerinde kullanılarak iyi bir öğrenme için gerekli olan veri ihtiyacını karşılar. Türk işaret dili rakamlarından oluşan veri setinin sınırlı olması ve modellerin daha etkin bir öğrenme süreci için bu yöntemle yer verilmiştir. Bu çalışmada, her bir modelin ImageNet ağırlıklarını kullanarak transfer öğrenme gerçekleştirilmiştir. Diğer bir yöntem veri artırma tekniğidir. Veri artırma, eğitim veri setlerinin boyutunu ve kalitesini değiştiren bir dizi teknik içerir. Eğitim sırasında veri çeşitliliği arttırmak için kontrast ayarı, döndürme, aynalama, parlaklık gibi veri büyütme teknikleri de kullanıldı. Türk işaret dili veri seti tek başına eğitim için sınırlı ve küçük olduğu için modellerin test verileri üzerinde genelleme kabiliyetini arttırmak için farklı veri artırma teknikleri kullanılmıştır. Bu iki yöntemle hem veri boyutu hem de veri çeşitliliği artırılmıştır. Modellerin eğitilmesi ve test edilmesi için 1.500 görüntüden oluşan veri seti belirli oranlarda ayrılmıştır. Önerdiğimiz veri setinde her bir sınıfa ait görüntülerin %70'i eğitim, %15'i doğrulama ve %15'i ise

modelin genelleme kabiliyetini ölçmek için test verisi olarak ayrılmıştır. Bir modelin başarımı, modelin daha önce görmediği test verileri ile ölçülmektedir.



Şekil 4. Sistem akış diyagramı

BULGULAR VE TARTIŞMA

Uygulama Detayları

Bu çalışmadaki, tüm deneyler aşağıdaki özelliklere sahip bir bilgisayarda gerçekleştirilmiştir. Derin öğrenme platformları için en uygun işletim sistemi olan Linux'un en güncel versiyonu olan Ubuntu 22.04 kullanılmıştır. Donanım açısından bu bilgisayar, Intel® Core™ i7-12700K İşlemci (25M Önbellek, 5.00 GHz'e kadar) işlemci, 64 GB DDR5 (5200Mhz) RAM ve NVIDIA RTX 3090 grafik kartından oluşur. NVIDIA RTX 3090 grafik kartı, 10496 CUDA çekirdeği, 328 tensör çekirdeği içerir ve 24 GB GDDR6X bellek ile 384 bit bellek arabirimi kullanır. Programlama dili olarak Python, çerçeve olarak PyTorch ve NVIDIA CUDA Toolkit 11.7 kullanılmıştır.

Değerlendirme Metrikleri

Derin öğrenme algoritmalarının performanslarını genelleme yeteneklerini değerlendirmek için kullanılan birçok değerlendirme metriği mevcuttur. Sınıflandırma ve regresyona dayalı algoritmalar için farklı metrik değerleri kullanılır. Bu çalışmada sınıflandırma algoritmalarında en yaygın kullanılan

Doğruluk, Kesinlik, Duyarlılık ve F1-skoru metrik değerleri kullanılmıştır. Bu metriklere ait formüller aşağıda Çizelge 1’de verilmiştir.

Çizelge 1. Sınıflandırma metrikleri

Metrik	Açıklama	Formül
Doğruluk	Doğruluk, doğru tahmin sayılarının toplam tahmin sayısına oranıdır. (DP: Doğru Pozitif, DN: Doğru Negatif, YP: Yanlış Pozitif, YN: Yanlış Negatif)	$\frac{DP + DN}{DP + DN + YP + YN}$
Kesinlik	Doğru tahmin edilen pozitif sonuçların, Doğru pozitif (DP) değerlerinin, tahmin edilen tüm pozitif sonuçlara (DP+YP) oranı olarak hesaplanır.	$\frac{DP}{DP + YP}$
Duyarlılık	Doğru tahmin edilen pozitif sonuçların (DP) oranını ölçmek için kullanılır. Aynı zamanda modelin hassasiyetini ölçen bir metriktir.	$\frac{DP}{DP + YN}$
F1-skoru	Kesinlik ve hatırlama değerlerinin harmonik ağırlığını hesaplar.	$2x \frac{Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık}$

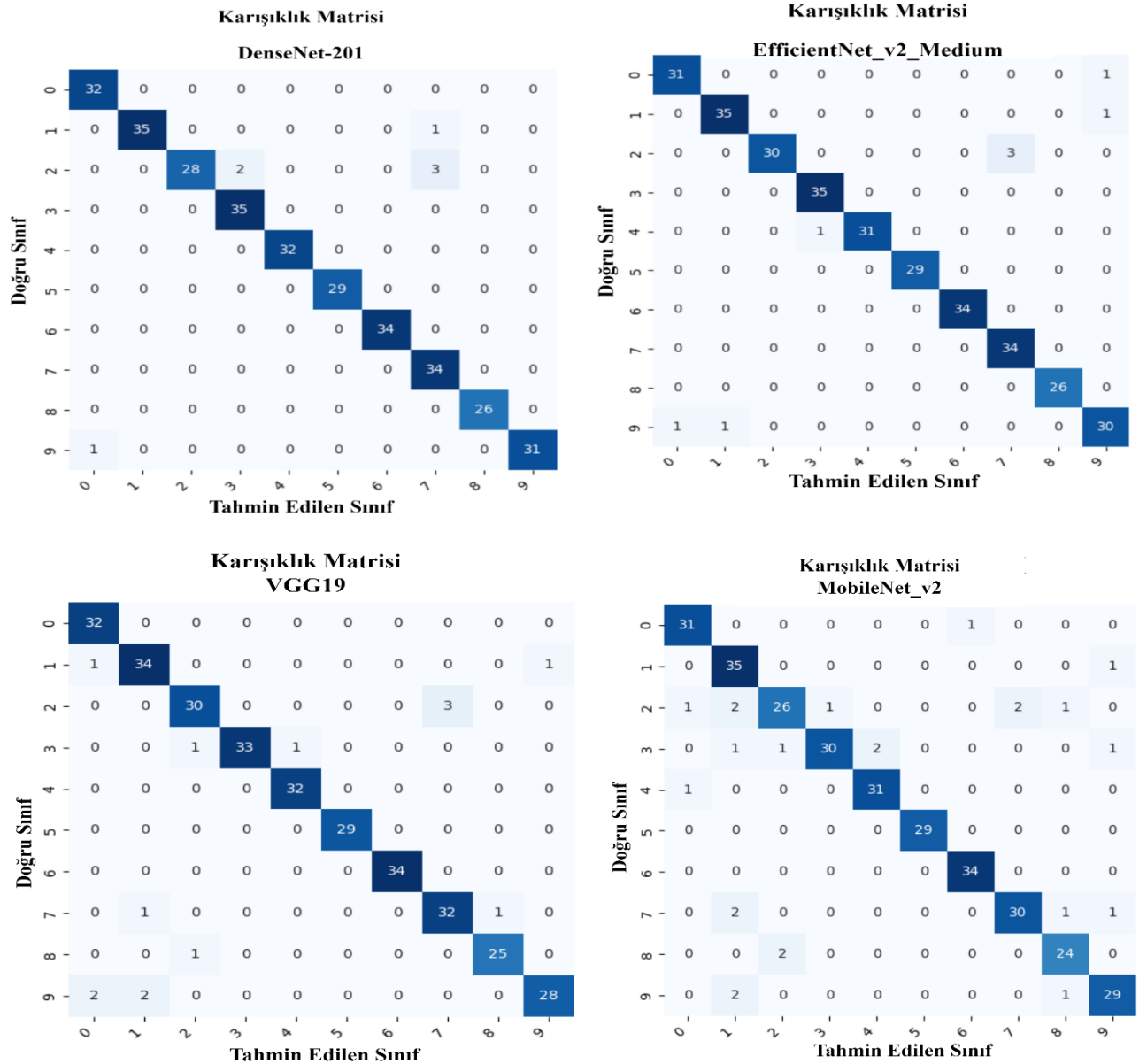
Deneysel Sonuçlar

Literatürde, farklı konuşma dillerine ait işaret dili tanıma, jest-mimik tanıma ve hareket tanıma gibi birçok çalışma mevcuttur. Sunulan modellerin performansını gösteren en önemli etkenlerden biri de kullandıkları veri setleridir. Yapılan çalışmaları incelediğimizde her çalışma kendi özel veri setini kullanmaktadır. Bu da modellerin farklı işaret diller üzerindeki objektifliğini etkiler. Yaptığımız çalışmada, güncel ve popüler derin öğrenme modellerinin Türk işaret dili üzerindeki performanslarını incelemek için deneysel bir çalışma sunulmuştur. Her bir model için doğruluk, kesinlik, duyarlılık, F1-skor değerleri ve kullanılan modellerin parametre sayıları Çizelge 2’de verilmiştir. Çizelge 2’de en yüksek metrik değerleri kalın yazı tipi ile vurgulanmıştır.

Çizelge 2. Derin öğrenme modeli için deneysel sonuçlar

Model	Doğruluk	Kesinlik	Duyarlılık	F1-skoru	Parametre sayısı
Resnet18	0.9412	0.9416	0.9424	0.9414	11.180.616
Resnet34	0.9505	0.9527	0.9513	0.9516	21.288.776
Resnet50	0.9319	0.9326	0.9329	0.9322	23.524.424
Resnet101	0.9505	0.9534	0.9503	0.9519	42.508.356
Resnet152	0.9876	0.9885	0.9881	0.9880	58.152.004
EfficientNet_v2_s	0.9412	0.9438	0.9422	0.9425	20.182.612
EfficientNet_v2_m	0.9754	0.9771	0.9756	0.9762	20.182.612
EfficientNet_v2_l	0.9751	0.9742	0.9749	0.9746	52.863.480
VGG13	0.9226	0.9259	0.9248	0.9242	128.967.236
VGG16	0.9351	0.9351	0.9349	0.9351	134.276.932
VGG19	0.9567	0.9582	0.9574	0.9574	128.967.236
MobileNet_v2	0.9257	0.9281	0.9266	0.9269	2.261.827
DenseNet121	0.9628	0.9655	0.9632	0.9641	6.957.956
DenseNet161	0.9778	0.9791	0.9789	0.9784	26.480.836
DenseNet169	0.9779	0.9803	0.9785	0.9791	12.491.140
DenseNet201	0.9783	0.9811	0.9789	0.9797	18.100.612

Çizelge 2’deki modellerin deneysel sonuçlarına baktığımızda Türk işaret dili üzerinde her bir model etkili sonuçlar vermiştir. Modellerin performanslarını daha sağlam bir şekilde değerlendirmek için kesinlik ve duyarlılık metriklerinin harmonik ortalaması olan F1-skorunu kullanılmıştır. Çizelge 2’yi genel olarak değerlendirdiğimizde modellerin hepsi %90’dan fazla değerler olarak Türk işaret dilinde başarılı sonuçlar verdiler. Modellerin performanslarını değerlendirdiğimizde tüm metrik değerlerinde en başarılı model ResNet152’ye aittir. Her bir modeli kendi içinde değerlendirdiğimizde ResNet mimarilerinden ResNet152 en yüksek F1-skoru olarak %98.8 değerine sahipken en düşük değer %94,1 ResNet50 modeline aittir. EfficientNet mimarilerinde, EfficientNet_v2_m %97.6 değeri ile en yüksek değeri alırken EfficientNet_v2_s %94.2 ile en düşük değere sahiptir. VGG ve DenseNet modellerinde ise modeller derinleştikçe performanslarının arttığı görülmektedir. En yüksek değerler, VGG19 için %95.7 ve Densenet201 içinse %97.9 F1-skor değerlerine sahiplerdir. Öte yandan, MobileNet_v2 mimarisi %92.6’lık F1-skoru ile diğer modellere kıyasla en düşük performansı göstermiştir.



Şekil 5. Bazı modellerin karışıklık matrisi

Şekil 5'te bazı modellerin performanslarının görselleştiren karışıklık matrisleri verilmiştir. Türk işaret dilinde 2 ve 7 rakamlarına ait el işareti görsel olarak birbirine yakındır. Özellikle veri seti oluşturulurken bu rakamların el ve parmak pozisyonları hemen hemen aynı olmakla beraber parmak açıklıkları arasında farklılık vardır. Modeller çoğunlukla 2 rakamını değerlendirirken hata vermişlerdir. DenseNet201, EfficientNet_v2_m ve VGG19 modelleri 3'er tane 7 rakamını 2 olarak görürken MobileNet_v2 ise 2 tanesini 7 rakamı olarak görmektedir. Genel olarak, karışıklık matrisine baktığımızda MobileNet_v2 modeli en düşük sınıflandırma performansını göstermiştir.

SONUÇ

Sağır ve dilsiz insanların bir tercümana ihtiyaç duymadan çevreleriyle sağlıklı ve doğru bir iletişim kurmaları için işaret dili çeviri sistemleri önemli bir yere sahiptir. Bununla birlikte, işaret dili kullanıcılarının herhangi harici bir donanıma ihtiyaç duymadan akıllı telefon, tablet gibi teknolojik aletleri kullanarak günlük yaşama uyum sağlamaları da önemli bir diğer etkidir. Bu da işaret dili tanıma sistemlerinin bilgisayarla görü alanında popüler hale gelmesini sağlamıştır. Bu çalışmada, Türk işaret dilindeki rakamlara ait oluşturduğumuz özgün veri setini kullanarak güncel CNN mimarilerinin detaylı bir karşılaştırmasını yaptık. Deneylerde tek düze bir arka plan kullanılarak oluşturulan veri seti herhangi bir ön işlem yapılmadan ham haliyle kullanılmıştır. Modellerin sınıflandırma performansını artırmak için basit veri artırma teknikleri de kullanılmıştır. Yapılan bu deneysel çalışmalar, Türk işaret dili rakamlarına ait görüntülerinin sınıflandırılması için güncel derin öğrenme tekniklerinin oldukça başarılı olduğunu ortaya koymuştur. Derin mimariler ve hafif mimari olarak adlandırılan az katmana sahip mimariler %90'ın üzerinde başarı göstermiştir. Bu çalışmayla birlikte, işaret dili tanıma sistemlerinde derin öğrenme mimarilerinin sınıflandırma kapasiteleri göz önüne alınarak gelecekteki işaret dili çalışmaları için önemli bir referans olmasını umuyoruz. İşaret dili, görsel işaretler kullanılarak iletişim kurulan bir dil yapısına sahip olduğu için karmaşık bir yapıya sahiptir. İşaret dili, el ve jest-mimik hareketlerinin doğru bir şekilde sınıflandırması ve bunlardan anlamlı ifadelerin çıkarılması önemli bir işlemdir. Bu yüzden, işaret dili tanıma sistemlerinde sınıflandırma önemli bir yere sahiptir. Ayrıca, yaptığımız bu çalışmayla literatüre ilk defa Türk işaret dili için rakamlardan oluşan bir özgün bir veri kümesi kazandırılarak Türk işaret dili çalışmalarında yazarlara katkı sağlanmasını umuyoruz.

Gelecekte yapılacak çalışmalar için harf, kelime veya dinamik hareketlerden oluşan bir yeni ve özgün bir veri seti geliştirilebilir. Bunun yanı sıra, sınıflandırma doğruluğunu artırmak için, farklı ışıklandırma, aydınlatma, arka plan ve el-parmak aksesuarlarına sahip görüntüler veri setine eklenerek CNN için veri çeşitliliği sağlanabilir.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

KAYNAKLAR

- Aiouez, S., Hamitouche, A., Belmadoui, M. S., (Belattar, K., & Souami, F. (2022). Real-time Arabic Sign Language Recognition based on YOLOv5. *PROCEEDINGS OF THE 2ND INTERNATIONAL CONFERENCE ON IMAGE PROCESSING AND VISION ENGINEERING*, (s. 17-25). doi:10.5220/0010979300003209
- Alawwad, R. A., Bchir, O., & Ismail, M. M. (2021). Arabic Sign Language Recognition using Faster. *International Journal of Advanced Computer Science and Applications*, 12(3), 692-700.

- Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., & Mekhtiche, M. A. (2020). Hand Gesture Recognition for Sign Language Using 3DCNN. *IEEE Access*, 8, 79491 - 79509.
- Alici-Karaca, D., Akay, B., Yay, A., Suna, P., Nalbantoglu, O. U., Karaboga, D., . . . Baran, M. (2022). A new lightweight convolutional neural network for radiation-induced liver disease classification. *Biomedical Signal Processing and Control*, 73. doi:10.1016/j.bspc.2021.103463
- Almeida, S. G., Guimarães, F. G., & Ramirez, J. A. (2014). Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors. *Expert Systems with Applications: An International Journal*, 14(6), 7259–7271.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Ayad Al-Dujaili, Y. D., Al-Shamma, O., Santamaría, J., . . . Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8(1), 1-74.
- Bhushan, S., Alshehri, M., Keshta, I., Chakraverti, A. K., Rajpurohit, J., & Abugabah, A. (2022). An Experimental Analysis of Various Machine Learning Algorithms for Hand Gesture Recognition. *Electronics*, 11(6). doi:10.3390/electronics11060968
- Bordes, A., Glorot, X., Weston, J., & Bengio, Y. (2012). Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing. *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (s. 127-135). PMLR.
- Burukanlı, M., Budak, Ü., & Çıbuk, M. (2019). Saldırı Tespit Sistemlerinde Makine Öğrenme Metotlarının Kullanımı. *Uluslararası Bilim ve Mühendislik Sempozyumu*, 20(22), 1052-1057.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.
- Deafness and hearing loss. (2021, Nisan 1). World Health Organization(WHO): <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> adresinden alındı
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. doi:10.1109/CVPR.2009.5206848
- Fan, J., Ma, C., & Zhong, Y. (2019). *A selective overview of deep learning*. arXiv:1904.05526 . adresinden alındı
- Gangrade, J. B. (2020). Vision-based hand gesture recognition for Indian sign language using convolution neural network. *IETE Journal of Research*, 1-10.
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & S.Lew, M. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27-48.
- Gschwend, D. (2020). Zynqnet: An fpga-accelerated embedded convolutional neural network. arXiv preprint arXiv:2005.06892.
- Halbouni, A., Gunawan, T. S., Habaebi, M. H., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). Machine Learning and Deep Learning Approaches for CyberSecurity: A Review. *IEEE Access* (10), 19572 - 19585. doi:10.1109/ACCESS.2022.3151248
- 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 (CVPR)*, (s. 770-778).
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. <https://arxiv.org/abs/1704.04861> adresinden alındı

- Huang, G., Liu, Z., Maaten, L. v., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (s. 4700-4708).
- Justesen, N., Bontrager, P., Togelius, J., & Risi, S. (2020). Deep Learning for Video Game Playing. *IEEE Transactions on Games*, 12(1), 1 - 20.
- Karaman, A., Karaboga, D., Pacal, I., Akay, B., Basturk, A., Nalbantoglu, U., Sahin, O. (2022). Hyper-parameter optimization of deep learning architectures using artificial bee colony (ABC) algorithm for high performance real-time automatic colorectal cancer (CRC) polyp detection. *Applied Intelligence*. <https://doi.org/10.1007/s10489-022-04299-1>
- Karaman, A., Pacal, I., Basturk, A., Akay, B., Nalbantoglu, U., Coskun, S., Sahin, O., & Karaboga, D. (2023). Robust real-time polyp detection system design based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC). *Expert Systems with Applications*, 221. <https://doi.org/10.1016/j.eswa.2023.119741>
- Karagoz, M. A., Akay, B., Basturk, A., Karaboga, D., & Nalbantoglu, O. U. (2023). An unsupervised transfer learning model based on convolutional auto encoder for non-alcoholic steatohepatitis activity scoring and fibrosis staging of liver histopathological images. *Neural Computing and Applications*, 1-15.
- Khari, M., Garg, A., Crespo, R. G., & Verdú, E. (2019). Gesture Recognition of RGB and RGB-D static Images using Convolutional Neural Networks. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(7), 22-27.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems* 25.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Li, Y., Ding, L., & Gao, X. (2018). On the Decision Boundary of Deep Neural Networks. <https://arxiv.org/abs/1808.05385> adresinden alındı
- Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999 - 7019.
- LUQMAN, H., & ELALFY, E. (2022). Utilizing motion and spatial features for sign language gesture recognition using cascaded CNN and LSTM models. *Turkish Journal of Electrical Engineering and Computer Sciences*, 30(7), 2508-2525.
- Ma, Y., Xu, T., & Kim, K. (2022). Two-Stream Mixed Convolutional Neural Network for American Sign Language Recognition. *Sensors*, 22(16), 5959.
- Marais, M., Brown, D., Connan, J., & Boby, A. (2022). An Evaluation of Hand-Based Algorithms for Sign Language Recognition. *2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*. IEEE. doi:10.1109/icABCD54961.2022.9856310
- Myagila, K., & Kilavo, H. (2021). A Comparative Study on Performance of SVM and CNN in Tanzania Sign Language Translation Using Image Recognition. *Applied Artificial Intelligence*, 1-16. doi:10.1080/08839514.2021.2005297
- Naglot, D., & Kulkarni, M. (2016). Real time sign language recognition using the leap motion controller. *International conference on inventive computation technologies (ICICT)*. 3, s. 1-5. IEEE.
- Nam, Y., & Lee, C. (2021). Cascaded convolutional neural network architecture for speech emotion recognition in noisy conditions. *Sensors*, 21(13), 4399.

- Núñez-Prieto, R., Gómez, P. C., & Liu, L. (2019, October). A real-time gesture recognition system with fpga accelerated zynqnet classification. In *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)* (pp. 1-6). IEEE.
- Ongsulee, P. (2017). Artificial intelligence, machine learning and deep learning. *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)* (s. 1-6). IEEE.
- Ozkok, F. O., & Celik, M. (2023). Classification of High Resolution Melting Curves Using Recurrence Quantification Analysis and Data Mining Algorithms. In *Smart Applications with Advanced Machine Learning and Human-Centred Problem Design* (pp. 641-650). Cham: Springer International Publishing.
- Özcan, T., & Baştürk, A. (2020). ERUSLR: A new Turkish sign language dataset and its recognition using hyperparameter. *ournal of the Faculty of Engineering and Architecture of Gazi University*, 36(1), 527-542.
- PACAL, İ. (2022). Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images. *Journal of the Institute of Science and Technology*, 1917–1927.
<https://doi.org/10.21597/jist.1183679>
- Pacal, I., & Karaboga, D. (2021). A robust real-time deep learning based automatic polyp detection system. *Computers in Biology and Medicine*, 134.
<https://doi.org/10.1016/J.COMPBIOMED.2021.104519>
- Pacal, I., Karaboga, D., Basturk, A., Akay, B., & Nalbantoglu, U. (2020). A comprehensive review of deep learning in colon cancer. *Computers in Biology and Medicine*, 126.
<https://doi.org/10.1016/J.COMPBIOMED.2020.104003>
- Pacal, I., Karaman, A., Karaboga, D., Akay, B., Basturk, A., Nalbantoglu, U., & Coskun, S. (2022). An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets. *Computers in Biology and Medicine*, 141.
<https://doi.org/10.1016/J.COMPBIOMED.2021.105031>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
- Rao, G. A., Syamala, K., Kishore, P. V., & Sastry, A. S. (2018). Deep convolutional neural networks for sign language recognition. *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*. doi:10.1109/SPACES.2018.8316344
- Rastgoo, R., Kiania, K., & Escalerab, S. (2021). Sign Language Recognition: A Deep Survey. *Expert Systems with Applications*, 164, 113794. doi:10.1016/j.eswa.2020.113794
- Ren, Z., Yuan, J., Meng, J., & Zhang, Z. (2013). Robust Part-Based Hand Gesture Recognition Using Kinect. *IEEE Transactions on Multimedia*, 15(5), 1110 – 1120.
- Rezende, T. M., Almeida, S. G. M., & Guimarães, F. G. (2021). Development and validation of a Brazilian sign language database for human gesture recognition. *Neural Computing and Applications*, 33(16), 10449-10467.
- Sajjanhar, A., Wu, Z., & Wen, Q. (2018). Deep learning models for facial expression recognition. *2018 digital image computing: Techniques and applications (dicta)* (s. 1-6). IEEE.
- Saqib, S., Ditta, A., Khan, M., Kazmi, S. A., & Alquhayz, H. (2021). Intelligent Dynamic Gesture Recognition Using CNN Empowered by Edit Distance. *Computers, Materials and Continua*, 66(2), 2061-2076.
- Shukor, A. Z., Miskon, M. F., Jamaluddin, M. H., binAli@Ibrahim, F., FareedAsyraf, M., & binBahar, M. B. (2015). A new data glove approach for Malaysian sign language detection. *Procedia Computer Science*, 76, 60-67.

- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. <https://arxiv.org/abs/1409.1556> adresinden alındı
- Suliman, W., Deriche, M., Luqman, H., & Mohandes, M. (2021). Arabic Sign Language Recognition Using Deep Machine Learning. (s. 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)). IEEE. doi:10.1109/ISAECT53699.2021.9668405
- Suri, K., & Gupta, R. (2019). Convolutional neural network array for sign language recognition using wearable IMUs. In 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 483-488). IEEE.
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International conference on machine learning* (s. 6105-6114). PMLR.
- Tasmere, D., & Ahmed, B. (2020). Hand Gesture Recognition for Bangla Sign Language Using Deep Convolution Neural Network. 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI). IEEE. doi:10.1109/STI50764.2020.9350484
- Wang, Z., Zhao, T., Ma, J., Chen, H., Liu, K., Shao, H., . . . Ren, J. (2022). Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System. *IEEE Transactions on Mobile Computing*, 21(7), 2398 - 2410.
- Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, 3(1), 1-40.
- Wu, Y., & Huang, T. S. (1999). Vision-Based Gesture Recognition: A Review. In *International gesture workshop* (s. 103-115). Berlin Heidelberg: Springer.
- Yu, S., Jia, S., & Xu, C. (2017). Convolutional neural networks for hyperspectral image classification. *Neurocomputing*, 219, 88-98.
- Zhiqiang, W., & Jun, L. (2017). A review of object detection based on convolutional neural network. 2017 36th Chinese Control Conference (CCC) (s. 11104-11109). IEEE.