

# Türk İşaret Dili Alfabeti Algılama Uygulaması

02210224064

Amr Nawaf Walıdı, İnönü Üniversitesi

## Öz

Derin öğrenme ve bilgisayarlı görme teknolojilerinin ilerlemesiyle birlikte jest ve işaret dili tanıma sistemlerinde önemli ilerlemeler kaydedilmiştir. Bu çalışmada, özel bir görüntü veri kümesi üzerinde eğitilmiş bir CNN modeli kullanılarak bir Türk İşaret Dili (TSL) alfabe algılama sistemi geliştirilmiştir. Veri kümesi, alfabe sınıfı başına 300'den fazla elle yakalanmış ve önceden işlenmiş el işareti görüntüsünden oluşmuş ve değişkenliği artırmak için genişletilmiştir. Alfabe tahmini için bir Flutter uygulamasından gerçek zamanlı görüntü girişi almak üzere Flask tabanlı bir arka uç hizmeti uygulanmıştır. Çalışmanın amacı, çeşitli koşullar altında TSL el işaretlerini algılamada CNN modelinin etkinliğini değerlendirmektir. Performansı değerlendirmek ve sınıflandırma zorluklarını vurgulamak için bir karışıklık matrisi kullanılmıştır. Model, eğitim ve belirli doğrulama aşamalarında yüksek doğruluk elde etse de, ara sıra sınıf karışıklığı ve aydınlatma veya el pozisyonundaki değişiklik gibi sınırlamalar devam etmektedir ve bu da daha fazla iyileştirmeye açık alanları göstermektedir.

## 1. Giriş

Doğal, karmaşık ve eksiksiz diller olan işaret dillerinde anlamı iletmek için manuel eklemler ve manuel olmayan işaretleyiciler kullanılır. Diğer ulusal işaret dilleri gibi Türk İşaret Dili (TSL), Türkiye'nin sağır ve işitme güçlüğü çeken nüfusu için temel bir iletişim aracıdır. TSL, zengin yapısı, benzersiz sözdizimi ve grameri nedeniyle konuşulan Türkçeden farklıdır.

Derin öğrenmedeki gelişmeler sayesinde, bilgisayarlı görme alanı makinelerin görsel verileri şaşırtıcı bir doğrulukla yorumlamasını mümkün kılmıştır. LeCun ve diğerleri (1998) tarafından tanıtılmasından ve ardından AlexNet (Krizhevsky ve diğerleri, 2012) tarafından popülerleştirilmesinden bu yana, Evrişimli Sinir Ağları (CNN'ler), el hareketi tanıma gibi görüntü sınıflandırma görevlerinde olağanüstü verimlilik göstermiştir. Bu ağlar, özelliklerin mekansal hiyerarşilerini öğrenmek için evrişim katmanlarını kullandıkları için işaret dili alfabelerini karakterize eden ayrıntılı el şekillerini ve yönelimlerini belirlemek için uygundur.

Birkaç çalışma, işaret dili tanımda CNN'lerin kullanımını incelemiş olsa da (özellikle Amerikan İşaret Dili (ASL) için), Türk İşaret Dili hakkında çok az araştırma ve yaygın olarak

erişilebilen veri kümeleri bulunmaktadır. Pigou ve diğerleri (2015), sürekli jest tanımda derin sinir ağlarının gücünü göstermiş ve Starner ve diğerleri (1998), giyilebilir ve masaüstü platformları kullanarak gerçek zamanlı ASL tanıma konusunda en erken çabalardan birini sunmuştur. Bu önceki çalışmalara göre, bu mevcut proje, CNN tabanlı derin öğrenmeye dayalı statik bir TSL alfabe işareti algılama sistemi oluşturmaya odaklanmaktadır.

Gerçek zamanlı el algılama, herhangi bir işaret tanıma sisteminin en önemli bileşenidir. MediaPipe tabanlı CVZone, sağlam el algılama ve izleme yetenekleri sağlar. El bölgelerini gerçek zamanlı olarak tanıma yeteneği, özel veri kümelerinin kolayca oluşturulmasını kolaylaştırır ve böylece derin öğrenme modelleri oluşturma sürecini hızlandırır (Hassan, CVZone).

Bu çalışmanın amacı, kişiselleştirilmiş bir veri kümesinin oluşturulması, genelleme için artırmanın uygulanması ve bir CNN modelinin eğitilmesi yoluyla sağlam bir TSL alfabe tanıma sistemi kurmaktır. Daha sonra sistem, uygulanabilirliğini göstermek için bir mobil uygulamada uygulama için doğruluk testine tabi tutulur ve test edilir. Bu çalışma yalnızca TSL araştırmasındaki boşluğu doldurmakla kalmaz, aynı zamanda daha erişilebilir iletişim araçları için bir fırsat yaratır.

## 2. Materyal ve Yöntem

### 2.1 Veri Seti

Türkçe alfabe için el hareketlerinin fotoğraflanmasıyla kişisel bir veri seti oluşturuldu. Görüntü toplama süreci, arka plan gürültüsünü önlemek için kontrollü ışık ortamında bir web kamerası kullanılarak gerçekleştirildi. CVZone'un el algılama modülü, her kareden elin ilgi alanını kırpma ve çıkarma için kullanıldı ve daha sonra görsel değişkenliği azaltmak için boş bir arka plan üzerinde sabit bir boyuta (örneğin 300x300 piksel) normalleştirildi. Her sınıfın yeterli temsilini sağlamak için tüm el işaretleri yaklaşık 300 kez kaydedildi. İşaretler, Q, W ve X gibi yerel karakterler hariç olmak üzere yalnızca Türkçe alfabe için harflerden oluşuyordu. Toplam 29 sınıf hazırlandı ve veriler 85-15 oranında eğitim ve test setlerine ayrıldı.

### 2.2 Veri Arttırma

Modelin farklı el pozisyonlarına, yönelimlerine ve aydınlatmaya genelleme yapma yeteneğini geliştirmek için çeşitli artırma teknikleri kullanıldı. Bunlar şunları içeriyordu:

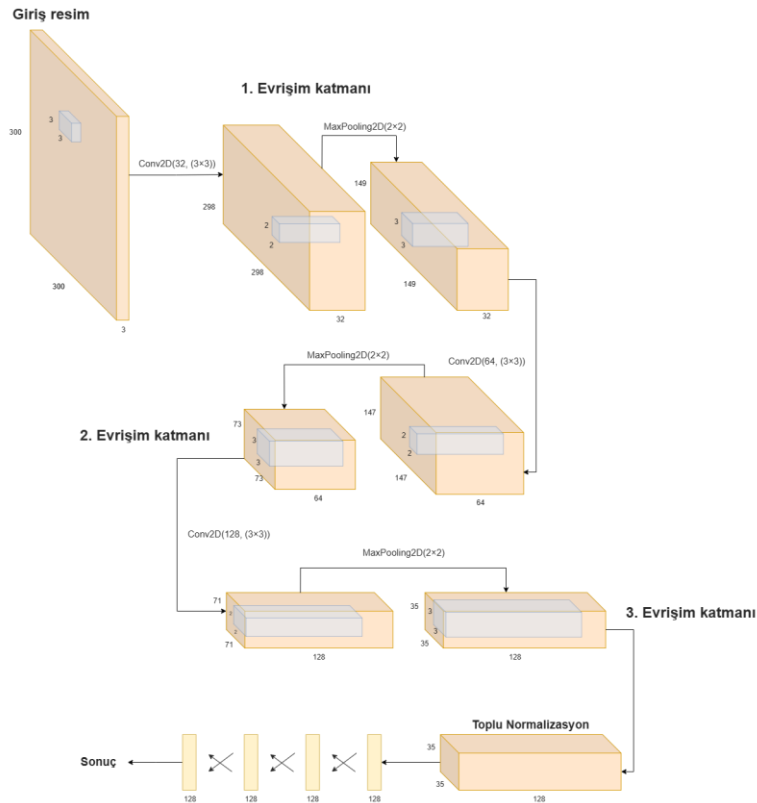
- **Rotation:** Doğal el eğimini simüle etmek için görüntülerin rastgele küçük döndürülmesi

- **Translation:** El pozisyonu değişikliklerini simüle etmek için yatay ve dikey kaydırmalar.
- **Zoom:** Mesafe tutarsızlığını simüle etmek için rastgele yakınlaştırma/uzaklaştırma.
- **Flip:** Ayna hareketlerinin çeşitliliğini simüle etmek için yatay çevirmeler.
- **Shearinng:** Hareket sunumu sırasında küçük hareketleri simüle etmek için bozulma.

Bu değişiklikler veri setindeki çeşitliliği artırarak model aşırı uyumunu önledi ve gerçek dünyadaki sağlamlığı iyileştirdi

### 2.3 Model Mimarisi

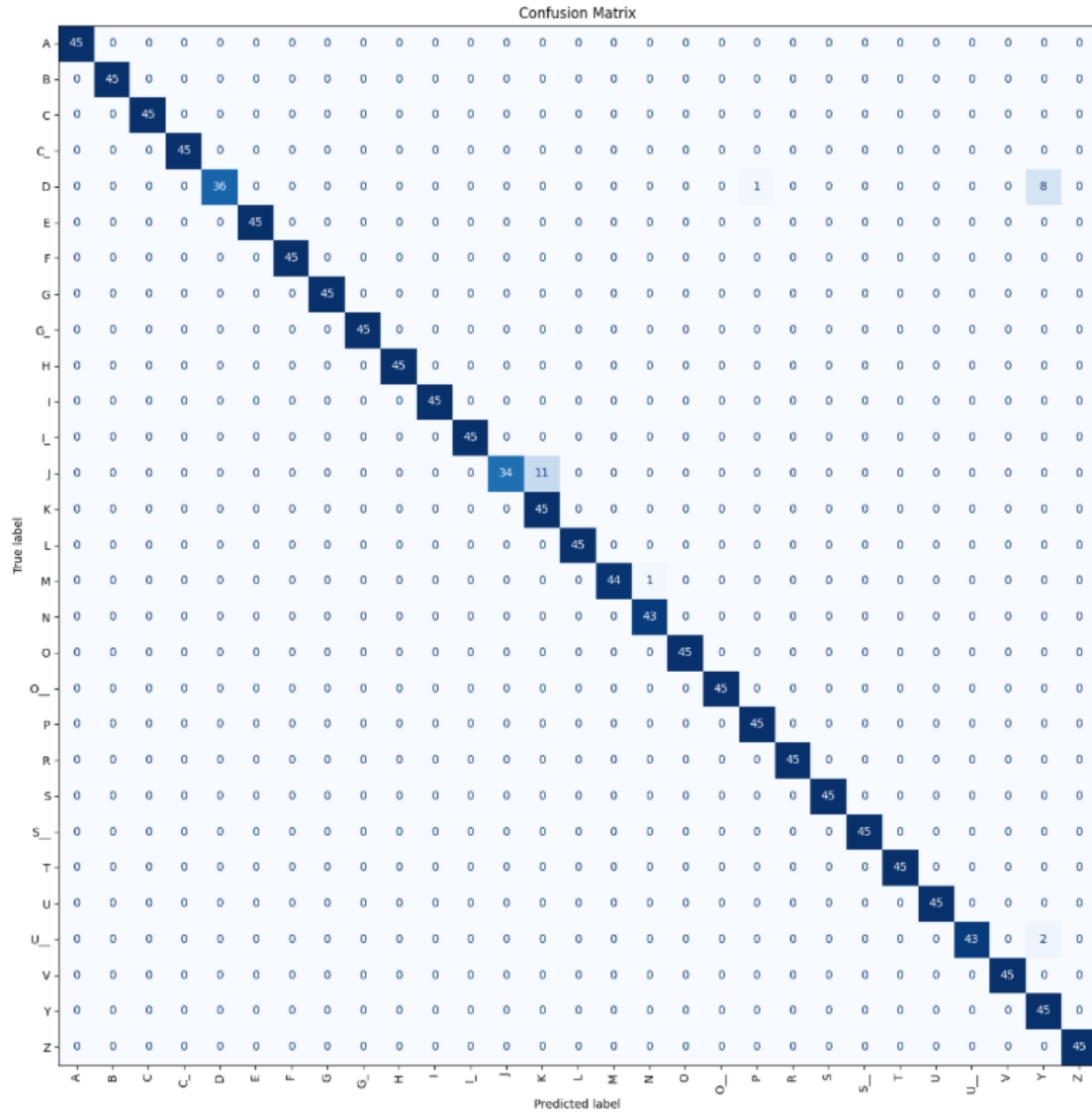
Düşük seviyeden yüksek seviyeye özellikler elde etmek için bir dizi evrişimsel katmanla bir Evrişimsel Sinir Ağı kullanıldı, ardından aşağı örnekleme için maksimum havuzlama katmanları kullanıldı. Her bir evrişimden sonra yakınsamayı hızlandırmak ve kararlılığı iyileştirmek için toplu normalizasyon tanıtıldı. Eğitim sırasında nöronların bir kısmını rastgele devre dışı bırakarak aşırı uyumu önlemek için bırakma katmanları tanıtıldı. Son katmanlar, özellik boyutunun azaltılması için küresel ortalama havuzlama katmanı ve özellik yorumlaması için ReLU ile etkinleştirilen yoğun (tam bağlı) katmanları içeriyordu. Softmax etkinleştirilmiş çıktı katmanı, 29 TSL harfinin sınıf olasılıklarını sundu.



Şekil 1: CNN modelinin görsel bir temsili

## 2.4 Model Eğitimi ve Değerlendirmesi

Model, kategorik çapraz entropi kaybı ve parametre başına öğrenme oranı uyarlaması olan Adam optimize edicisi ile eğitildi. Aşırı uyumu önlemek ve en iyi performans gösteren modeli korumak için erken durdurma ve model kontrol noktası uygulandı. Test, verilen test setinde bir karışıklık matrisinde gerçekleştirildi ve sınıf bazında içgörüler sağladı. Görsel olarak belirsiz harfler (örneğin, 'J' ve 'K') yanlış sınıflandırma kalıpları için kapsamlı bir şekilde incelendi.



Şekil 2: Test setinde gerçekleştirilen karışıklık matrisi

## 2.5 Sistem Dağıtımı

Modelin pragmatik kullanılabilirliğini ölçmek için, gerçek bir algılama sistemine entegre edildi. Modeli barındırmak ve Flutter tabanlı bir mobil uygulamadan görüntü girişlerini kabul etmek için bir Flask sunucusu kullanıldı. Uygulama, telefon kamerası aracılığıyla el hareketi görüntüleri yakaladı ve tahmin edilmek üzere arka uca gönderdi. Sunucu, görüntüleri modelle uyumlu hale getirmek amacıyla tahminden önce görüntüleri önceden işler. Beklenen harf, kullanıcı arayüzünde görüntülenir. Bu istemci-sunucu gösterimi, modelin gerçek mobil ortamlardaki potansiyelini gösterdi.

## 3. Bulgular ve Tartışma

CNN modeli, son epochlerde %98'in üzerinde bir orana ulaşarak iyi bir eğitim doğruluğuna sahipti. Doğrulama doğruluğu yaklaşık %96 ile en yüksek seviyedeydi ve bu, modelin statik TSL alfabe işaretlerini tanımada iyi olduğunu gösteriyordu. Erken epochlerde doğrulama doğruluğundaki dalgalanmalar aşırı uyum eğilimi gösterdi, ancak bu, bırakma ve veri artırma kullanılarak bir dereceye kadar azaltıldı.

Karmaşıklık matrisi, çoğu işaretin doğru bir şekilde tahmin edilmesine rağmen, minimal görsel farka sahip belirli işaretlerin yanlış sınıflandırmalara yol açtığını gösterdi. Örneğin, 'J' ve 'k' harfleri için el şekilleri, veri kümesinde görsel olarak ayırt edilemiyordu ve doğru bir şekilde tahmin edilmesi zordu. Bu, daha ayırt edici jest temsiline veya gelişmiş özellik çıkarma yöntemlerine olan ihtiyacı vurgular.

Bir mobil uygulamaya entegrasyon, modelin gerçek zamanlı yanıtını ve kullanılabilirliğini doğruladı. Ancak, değişen aydınlatma ve dağınık arka planlar gibi çevresel faktörlerin performans üzerinde orta düzeyde bir etkisi oldu. Arka plan segmentasyonu ve dinamik hareket tanıma özelliklerinin birleştirilmesi sistem güvenilirliğini artırmak için kullanılabilir.

## 4. Sonuç

Bu proje, Türk İşaret Dili alfabelerini tespit etmek için CNN tabanlı bir sistemin başarılı bir şekilde uygulanmasını göstermektedir. Özelleştirilmiş bir veri setinin oluşturulması ve veri artırmanın kullanılmasıyla, doğru bir model geliştirildi ve bir istemci-sunucu mimarisi aracılığıyla mobil bir ortamda dağıtıldı. Model oldukça doğruydı ve işletim için kullanılabilirdi, ancak benzer işaretler arasında ayırım yapma ve değişen durumlarda tutarlı performans konusunda sorunlar devam etmektedir.

## Kaynak

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.
- Starner, T., Weaver, J., & Pentland, A. (1998). Real-time American Sign Language recognition using desk and wearable computer-based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12), 1371–1375.
- Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2015). Sign language recognition using convolutional neural networks. In *European Conference on Computer Vision (ECCV) Workshops*.
- Hassan, M. (n.d.). CVZone. Retrieved from <https://github.com/cvzone/cvzone>