



Makale / Research Paper

Derin Öğrenme ve Görüntü İşleme Yöntemlerini Kullanarak Yüz ve Göz Hareketleri İle Bilgisayar Kontrolü

Muhammet Fatih ÇAPŞEK^{1a}, Abdulkadir KARACI^{2b}

¹Kastamonu Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü.
Kastamonu/TÜRKİYE

²Samsun Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, Samsun/TÜRKİYE
abdulkadir.karaci@samsun.edu.tr

Received/Geliş: 15.06.2022

Accepted/Kabul: 07.09.2022

Öz: Bu çalışmada bu bireyler için kafa ve göz hareketleri ile fareyi kontrol edebilecekleri yapay zeka destekli bir sistem geliştirilmiştir. Bu sistemde, kamera üzerinden elde edilen görüntülerden Haar Cascade, Dlib ve Open CV kütüphaneleri vasıtasıyla gerçek zamanlı olarak yüz hareketleri ve gözler tespit edilmektedir. Haar Cascade yüz bölgesini tespit etmek için kullanılırken, Dlib kütüphanesi tespit edilen bu yüz görüntüsünden sağ ve sol göz bölgesi görüntülerini elde etmek için kullanılmaktadır. Elde edilen bu göz bölgeleri görüntüleri 2874 göz verisiyle eğitilen CNN modeline giriş olarak verilmekte ve gözün kapalı ya da açık olduğu tespit edilmektedir. CNN modeli 1500 açık ve 1374 kapalı göz durumunu temsil eden halka açık göz görüntü veri seti üzerinde eğitilmiştir. Sol gözün kapatılıp açılma durumu farenin sol tıklamasını, sağ gözün kapatılıp açılma durumu ise farenin sağ tıklamasını sağlamaktadır. Bunun yanı sıra, fare hareketini modellemek için Haar Cascade ile tespit edilen yüzün konumu kullanılmaktadır. Test sonuçlarına göre sistemin gözleri ve gözlerin açık kapalı durumunu doğru bir şekilde tespit ettiği, her iki göz içinde göz kırpma olayını CNN ile doğru bir şekilde sınıflandırdığı gözlenmiştir. Ancak fare hareketinin modellenmesinde yavaşlık ya da yüz hareketine tam uyum sağlamama durumlarının yaşandığı tespit edilmiştir.

Anahtar Kelimeler: Haar cascade, cnn, dlib, görüntü işleme, fare kontrolü

Computer Control with Face and Eye Movements Using Deep Learning and Image Processing Methods

Abstract: In this study, an artificial intelligence-supported system was developed for these individuals to control the mouse with head and eye movements. In this system, facial movements and eyes are detected in real-time from the images obtained from the camera through the Haar Cascade, Dlib, and Open CV libraries. While Haar Cascade is used to detect the face region, the Dlib library is used to obtain right and left eye region images from this detected face image. These eye region images obtained are given as input to the CNN model trained with 2874 eye data and it is determined whether the eye is closed or open. The CNN model was trained on a public eye image dataset representing 1500 open and 1374 closed eye states. Closing and opening of the left eye cause the left click of the mouse and closing and opening of the right eye cause the right click of the mouse. In addition, the position of the face detected by Haar Cascade is used to model mouse movement. According to the test results, it has been observed that the system correctly detects the eyes and the open-closed state of these eyes, and correctly classifies the blinking event in both eyes with CNN. However, it has been determined that there are cases of slowness or incomplete adaptation to facial movement in the modeling of mouse movement.

Keywords: Haar cascade, cnn, dlib, image processing, mouse control

Bu makaleye atıf yapmak için

Çapşek, M.F., Karaci, A., "Derin Öğrenme ve Görüntü İşleme Yöntemlerini Kullanarak Yüz ve Göz Hareketleri İle Bilgisayar Kontrolü" El-Cezerî Fen ve Mühendislik Dergisi 2022, 9(4); 1170-1177.

How to cite this article

Çapşek, M.F., Karaci, A., "Computer Control with Face and Eye Movements Using Deep Learning and Image Processing Methods" El-Cezerî Journal of Science and Engineering, 2022, 9(4); 1170-1177.

ORCID ID: ^a 0000-0002-2462-1608; ^b 0000-0002-2430-1372

1. Giriş

Bilgisayar kullanabilmek bütün insanlar için bir ihtiyaçtır. Sunduğu çözümler sonucu yoğun bir yaygınlaşma oranına sahip olan bilgisayar kullanımı, klavye, fare ve dokunmatik yüzey gibi bilgisayar giriş aygıtları aracılığı ile gerçekleşmektedir. Bazı engellerden veya zorunluluklardan dolayı bilgisayar kullanmakta zorluklar yaşayan bireylere yardımcı olabilmek için farklı çalışmalar yapılmaktadır. Ancak hareket engeline sahip olan kişiler bilgisayar giriş aygıtları ile fiziksel temas sağlayamamaktadır ve bu sebeple bilgisayar kullanamamaktadır. Bu durumdaki insanlar için ergonomik bir kullanım sağlaması açısından bu sorunun ortadan kaldırılması önemlidir [1]. Engelli bireyler eğitim hizmetlerinden yeterince yararlanamamaktadır. Bunun sonucu olarak, ekonomik hayata katılım seviyeleri oldukça düşmektedir. İşte bu bireylerin gerek eğitim seviyelerinin yükseltilmesi gerekse istihdama katılabilmesini sağlamadaki en önemli araçlardan biri bilişim teknolojileri ve hizmetleridir [2].

Literatürde bu alanda yapılmış birkaç çalışma göze çarpmaktadır. Baştuğ ve ark. [3], engelliler için sanal fare kontrolünü kamera kullanarak göz takibi ile gerçekleştirmiştir. Göz hareketlerinin takibi diğer yöntemlere göre daha zordur. Çünkü engelleyici birçok sebep bulunmaktadır. Bunlar; görüntüdeki belli bir alanı kaplayan kirpikler, göz kapakları, iris kusurları, kozmetik lensler ve kafa hareketleridir [3]. Sadece göz hareketleriyle çalışan bir sistem tutarlı olmayabilir. Bu nedenle yüz ile hareketi sağlamak ve göz kırparak tıklama eylemlerini gerçekleştirmek daha verimli olacaktır. Atay ve ark. [4], göz takip yöntemini kullanarak bir iletişim ve kontrol sistemi geliştirmişlerdir. Geliştirdikleri sistemde göz hareketleri ve göz kırpmasını kullanarak ışık, klima, kapı, yatak kontrolü, yemek, su, ilaç isteme, tuvalet ihtiyacı bildirme ve teşekkür etme gibi komutları gerçekleştirmişlerdir. Uygulama tasarımında Python programlama dili üzerinde Dlib ve Opencv kütüphanelerini kullanmışlardır. Literatürdeki ilgili çalışmalara bakıldığında sadece göz hareketleri kullanılarak fare kontrolü sağlayan sistemler geliştirildiği görülmektedir. Bu çalışmada ise web kamerası kullanılarak hem göz hem de baş hareketleri ile fare kontrolü gerçekleştirilmiştir.

Bu çalışmada, bilgisayar kullanımı için gerekli olan ve fiziksel temas gerektiren fare kullanımına ihtiyaç duyulmadan baş ve göz hareketleri ile bilgisayar kontrolünün gerçekleştirilmesi amaçlanmaktadır. Bu amaçla, Derin Öğrenme ve görüntü işleme temelli bir uygulama geliştirilmiştir. Böylece fiziksel engelli kişilere bilgisayar kullanımı konusunda fayda sağlanacaktır. Çalışmanın ikinci bölümünde sistemi geliştirmek için kullanılan yöntemler ayrıntılı bir şekilde açıklanmaktadır. Üçüncü bölümünde sistemin gerçek zamanlı kullanılmasından elde edilen deneysel sonuçlar ve CNN modelinin performans değerleri sunulmaktadır. Son olarak dördüncü bölümünde elde edilen sonuçlar özetlenmektedir.

2. Materyal ve Metotlar

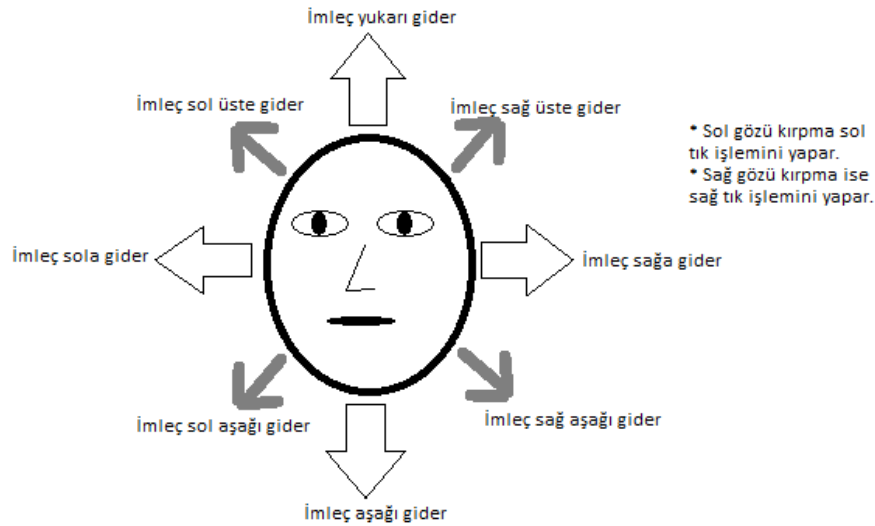
Yüz hareketleri ve göz kırpma durumları ile fare kontrolünü sağlamak amacıyla geliştirilen sistemin mimarisi Şekil 1'de gösterilmektedir. Bu şekilde de görüldüğü üzere kamera aktif edildikten sonra görüntüden kareler elde edilmektedir. Sonrasında yüz algılama işlemi gerçekleştirilerek kalibrasyon işlemi gerçekleştirilmektedir. Bu işlemde, kullanıcının yüz bölgesini ekranın ortasına konumlandırması gerekmektedir. Böylece ekran çözünürlüğüne göre fare imleci de ekranın tam orta noktasına konumlandırılmaktadır. Kalibrasyon işleminden sonra gözlerin sınırları tespit edilmektedir. Tespit edilen bu sınırlar içindeki gözler görüntü olarak kırılmaktadır. Kırılan bu görüntüler eğitilmiş CNN modeline giriş olarak verilerek sol, sağ ya da iki göz kırpma durumu tespit edilmektedir. Sonrasında bu kırpma işlemlerine göre fareye tek tıklama, sağ tıklama ve çift tıklama görevleri atanmaktadır.



Şekil 1. Geliştirilen sistemin mimarisi

Yüz tanımlama sistemi biri yüzü tespit etme diğeri ise yüzü takip etme olmak üzere iki temel basamaktan oluşmaktadır. Çalışma kapsamında yüzü tespit etmek ve yüzün konumunu belirlemek için OpenCV kütüphanesindeki Haar Cascade yöntemi kullanılmıştır. Yüzün konumuna göre fare imlecini hareket ettirmek için ise Pyautogui kütüphanesinden faydalanılmıştır. Göz bölgesinin tespit edilmesi için Dlib Kütüphanesi kullanılırken ilgi bölgesinin kırılması için OpenCV kütüphanesi kullanılmıştır. Bunun yanı sıra, sağ göz kırpma, sol göz kırma ve iki göz kırpma durumunun tespit edilmesi ise CNN derin öğrenme modeli ile gerçekleştirilmiştir. Bu işlemlere göre aşağıdaki işlemler yerine getirilmektedir. Bu işlemler ayrıca Şekil 2’de de gösterilmektedir.

- Yüzün konumuna göre imleç hareket ettirilir.
- Sağ göz kırıldığında farenin sağ tuşa tıklama işlevi yerine getirilir.
- Sol göz kırıldığında farenin sol tuşuna tıklama işlevi gerçekleştirilir.
- İki göz kırıldığında ise farenin çift tıklama işlevi gerçekleştirilir.



Şekil 2. Yüz ve göz durumuna göre gerçekleştirilen işlemler

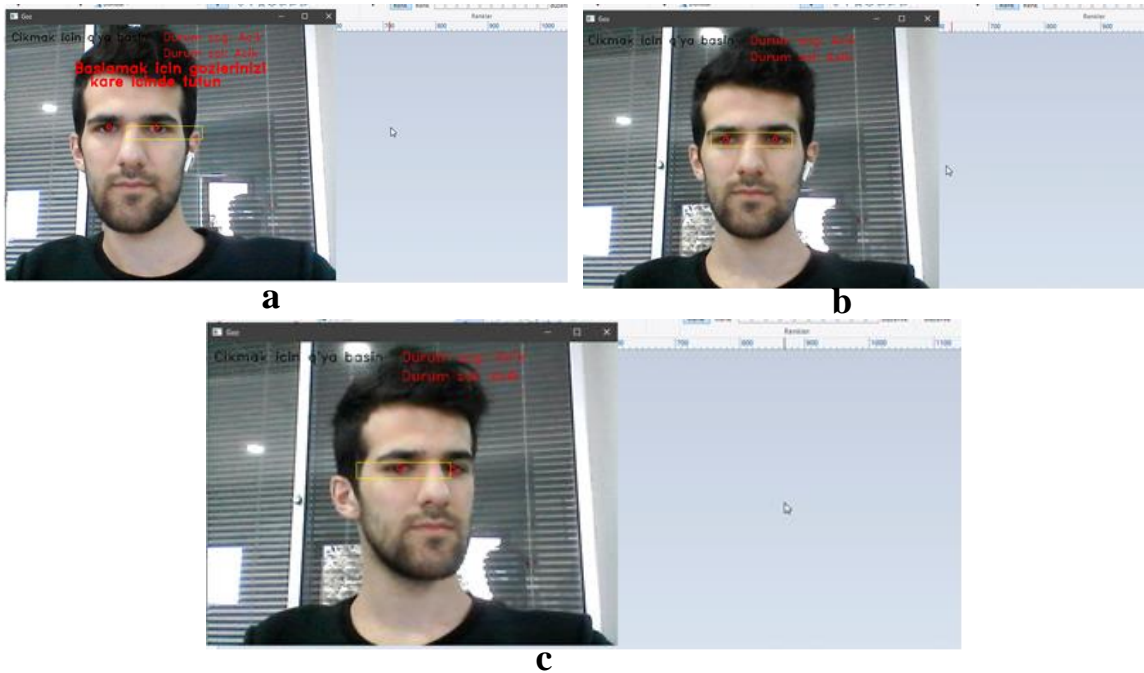
2.1. Yüz Konumunun ve Göz Bölgesinin Algılanarak Fare Kontrolünün Gerçekleştirilmesi

OpenCV kütüphanesi birçok yöntem sunarak yüz, göz, ağız ya da diğer objeler gibi nesnelere tanımasına ve takip edilmesine olanak tanır. Bu yöntemlerden biri Haar Cascade’dir. Haar Cascade,

Paul Viola ve Michael Jones tarafından 2001 yılında yayınlanan makine öğrenmesi nesne algılama algoritmasıdır. Geliştirilen sistemde yüz bölgesinin ve konumunun tespit edilmesi için Haar Cascade yöntemi kullanılmıştır.

Yüz bölgesi ve konum tespiti yapıldıktan sonra bu konuma göre farenin kontrol edilmesi önem arz etmektedir. Bunun için “PyAutoGUI” kütüphanesi kullanılmıştır. PyAutoGUI, fare ve klavyeyi kontrol etmeyi sağlayan, böylece otomatik görevler yapan botlar yazılmasına yardımcı bir Python kütüphanesidir. Fare kontrolü yapabilmek için ekrandaki koordinat sisteminin anlaşılması gerekir. Bu koordinat sistemi x (yatay) ve y (dikey) düzlemlerden oluşmaktadır. Sol üst (x, y) koordinatı (0,0) başlangıç noktasıdır, x sağa doğru artış gösterirken y aşağı doğru artış gösterir. Ekran boyut sınırları her bilgisayarda farklılık gösterebilir. Kullanıcının yüz hareketlerinden elde edilen veriler kullanılarak fare imleci ilgili konuma taşınır.

Fare imleci hareketinin ekran boyutuna göre daha sağlıklı bir şekilde gerçekleştirilmesi için mutlaka bir kalibrasyon işlemi de gerçekleştirilmelidir. Böylece yüz ve fare hareketi daha uyumlu olmaktadır. Kalibrasyon işleminde, daha önce de belirtildiği üzere hem yüzün hem de buna bağlı olarak farenin ekranın orta noktasına konumlandırılması sağlanmaktadır. Şekil 3, kalibrasyon öncesi, kalibrasyon anı ve kalibrasyon sonrası örnek görüntüleri sunmaktadır. Burada da görüleceği üzere kalibrasyon sonrasında fare imleci ekranın ortasına konumlanmaktadır. Ayrıca kalibrasyona başlamak için göz bölgesi ekranın tam ortasına yerleştirilen kare içinde tutulmalıdır. Böylece yüz bölgesinin ekranın ortasına konumlandırılması sağlanmaktadır. Buna göre de fare imleci kalibrasyonu yapılmaktadır.



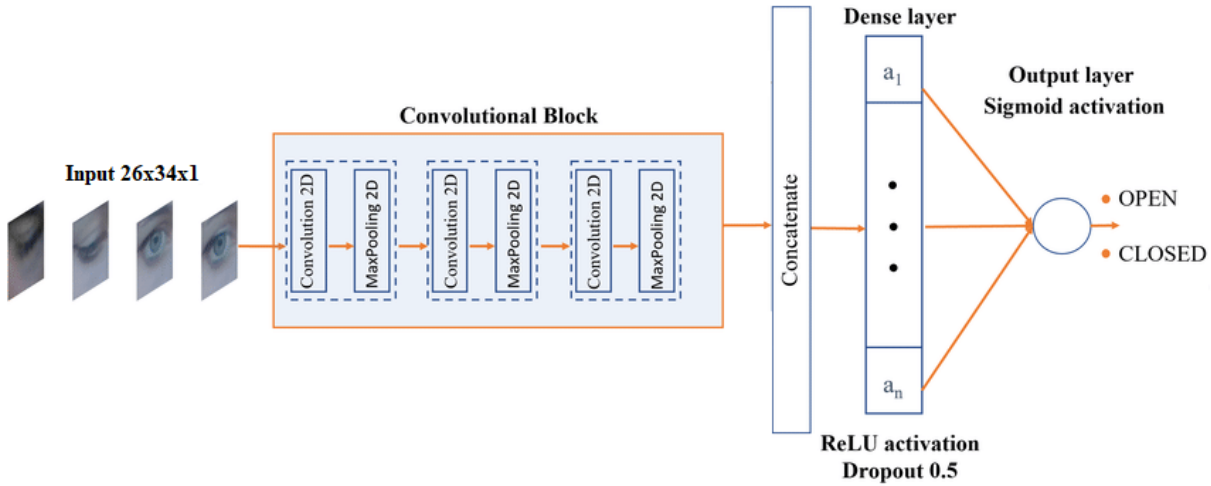
Şekil 3. Kalibrasyon işlemi: (a) kalibrasyon öncesi, (b) kalibrasyon, (c) kalibrasyon sonrası

Çalışmadaki diğer bir önemli özellik göz kırpma hareketleri ile fare tıklamasının modellenmesidir. Bunu yapabilmek için öncelikle gözlerin görüntü olarak tespit edilerek yüz bölgesinden kırılması gerekir. Bu işlem “Dlib” kütüphanesi ile gerçekleştirilmiştir. Bu kütüphane, gözler için 6 çift koordinat vermektedir. Böylece gözün sınırlarını belirleyen bir çember oluşturulmaktadır. Dlib, makine öğrenmesi (machine learning) algoritmalarını içeren C++ kütüphanesidir. Yüz tanıma (face recognition), yüz tespiti (face detection), yüz modelleme (facial landmarks) gibi uygulamalar Dlib kütüphanesi ile kolayca geliştirilebilir.

2.3. CNN Modeli ve Göz Kırpma Tespiti

Geliştirilen sistemde fare tıklama işlemlerini modellemek için göz kırpmaya hareketleri kullanılmaktadır. Hangi gözlerin kırıldığını tespit etmek için mimarisi çalışma kapsamında oluşturulan CNN modeli kullanılmıştır. CNN modelinin oluşturulmasında ve eğitilmesinde Keras ve Tensorflow kütüphaneleri kullanılmıştır. CNN modelinin mimarisi Şekil 4'te gösterilmektedir.

Model, yeniden etkinleştirme özelliğine sahip üç evrimsel katmandan ve bu katmanların ardından gelen üç maksimum havuzlama katmanından oluşmaktadır. Bu katmanların ardından bir flatten katmanı ve iki dense katmanı gelmektedir. Dense katmanlarda Relu aktivasyon fonksiyonu kullanılmaktadır. Modelin çıkış katmanında ise sigmoid aktivasyon fonksiyonuna sahip iki nöron bulunmaktadır. Bu nöronlar gözün açık ya da kapalı durumunu temsil etmektedir. Yani CNN modeli ikili sınıflandırma yapmaktadır.



Şekil 4. CNN modelinin mimarisi

2.3.1. CNN Modelinin Eğitilmesi

CNN modelinin eğitilmesinde simple-blink-detector (<https://github.com/iparaskev/simple-blink-detector/blob/master/train/dataset.csv>) veri seti kullanılmıştır. Veri setinde, toplam 2874 açık-kapalı göz verisi bulunmaktadır. Bu veriler RGB renk kodu olarak tutulmaktadır ve açık-kapalı etiket bilgisine sahiptir. Bu RGB kodları öncelikle resim haline getirilmiş ve sonrasında 26x34 piksele yeniden boyutlandırılmıştır. Bu ön işlemlerden sonra model eğitimi gerçekleştirilmiştir. Model eğitimi sırasında dışarda tutma (hold-out) doğrulama tekniği kullanılmıştır. Veri setinin %80'i eğitim, %20'si ise test için ayrılmıştır. Model, Adam optimizasyon algoritması, ikili çapraz entropi kayıp fonksiyonu ve 32 batch boyutu ile 50 epok eğitilmiştir.

2.3.2. Performans Metrikleri

CNN modelinin sınıflandırma sonuçlarının değerlendirilmesinde doğruluk, duyarlılık, kesinlik ve F1-puanı metrikleri kullanılmıştır. Bu metrikler sırasıyla Eşitlik 1-4'teki gibi hesaplanır.

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

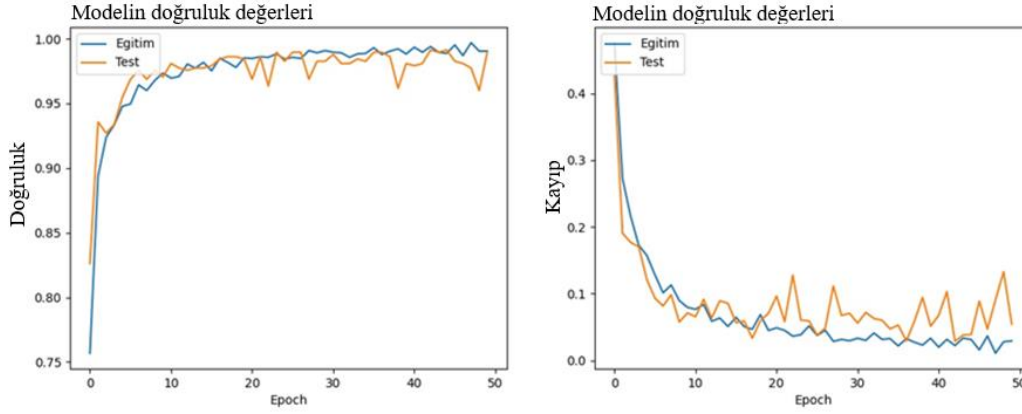
$$\text{Duyarlılık} = TP / (TP + FN) \quad (2)$$

$$\text{Kesinlik} = TP / (TP + FP) \quad (3)$$

$$F1 - \text{puanı} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall}) \quad (4)$$

3. Deneysel Sonuçlar

Bu bölümde geliştirilen sistemin örnek senaryolar üzerinden elde edilen deneysel sonuçları ve CNN modelinin eğitim ve test performans sonuçları sunulmaktadır. Şekil 5, CNN modelinin test ve eğitim verileri için doğruluk ve kayıp eğrilerini göstermektedir. Bu eğrilerde de görüleceği üzere hem eğitim hem de test verileri için doğruluk değeri 1'e yaklaşırken kayıp değeri 0'a yaklaşmaktadır. Bu da modelin iyi eğitildiğinin göstergesidir.



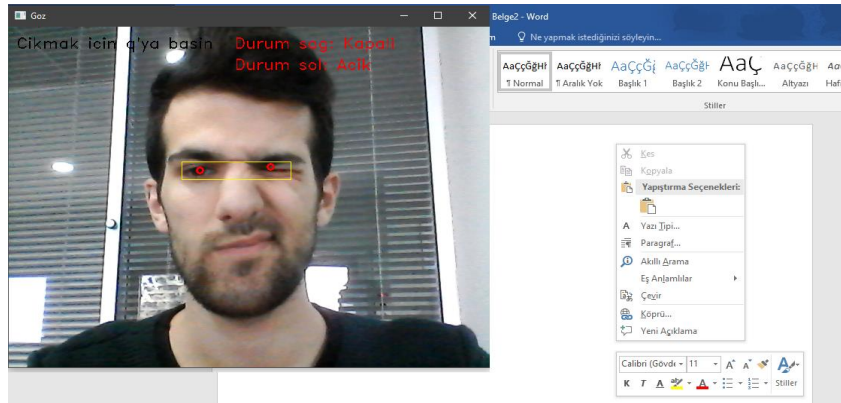
Şekil 5. Eğitim ve test verileri için accuracy ve loss grafikleri

Çalışma kapsamında eğitilen CNN modelinin test verileri için ortalama doğruluk, duyarlılık ve F1-puanı değerleri Tablo 1'de sunulmaktadır. Bu tabloda da görüleceği üzere CNN modeli göz kırpmaya durumunu %97 doğrulukla sınıflandırmıştır. Bunun yanı sıra gözün kapalı olma durumunu ise %98 oranında doğru tespit etmiştir. Bu oranların yeterli olduğu gerçek-zamanlı testlerle de doğrulanmıştır.

Tablo 1. CNN modelinin test verileri için sınıflandırma performansı değerleri

Doğruluk	F1-puanı	Kesinlik	Duyarlılık
%97	%99	%99	%98

Gerçekleştirilen deneysel çalışmada CNN modelinin gerçek zamanlı performansı da test edilmiştir. Bunun için sağ ve sol göz gerçek zamanlı olarak kırılarak CNN modeline giriş olarak verilmiştir. Modelin yeterli bir düzeyde sınıflandırma yaptığı gözlenmiştir. Modelin göz kırpmaya durumunu gerçek zamanlı olarak tespiti ile ilgili ekran görüntüsü Şekil 6'da gösterilmektedir. Bu şekilde de görüleceği üzere kullanıcı sağ gözünü kırpmakta, CNN bunu algılamakta ve buna bağlı olarak farenin sağ tuşuna tıklama işlemi gerçekleştirilmektedir.



Şekil 6. Sağ göz kırpmaya ile farenin sağ tuş tıklama işleminin gerçekleştirilmesi

Ayrıca sistemin performansını gerçek zamanlı olarak test etmek için 30 cm, 40 cm ve 50 cm uzaklıklardan çeşitli görevler yerine getirilerek bu görevlerin gerçekleştirilme süreleri kaydedilmiştir. İlk görev yeni dosya oluşturma ile ilgilidir. Bu görevde farenin sol ve sağ tuşlarına tıklama işlemi sağ ve sol göz kırpma hareketleri ile modellenmektedir. Diğer görev dosya açma görevidir. Bu görevde ise çift göz kırpma ile farenin çift tıklama görevi modellenmektedir. Dosya silme işlemi ise yeni dosya oluşturma görevine benzer bir görevdir. Tablo 2, bu görevleri yerine getirilmesi için gereken süreleri göstermektedir. Kameraya olan mesafe arttıkça sistemin algılaması azalmakta ve süre buna bağlı olarak az da olsa uzamaktadır. Ayrıca sistem performansında ortam aydınlatması ve kamera kalitesi de etkili olmaktadır.

Tablo 2. Görevleri gerçekleştirmek için geçirilen süreler

Uzaklık	Görevler		
	Yeni dosya oluşturma	Dosya açma	Dosya silme
30 cm	17sn	7sn	12sn
40 cm	18sn	7sn	11sn
50 cm	20sn	9sn	14sn

5. Sonuçlar

Bu çalışmada, yüz hareketleri ve göz kırpma durumları kullanılarak fare kullanımı sağlayacak derin öğrenme destekli gerçek zamanlı bir sistem geliştirilmiştir. Deneysel çalışma sonuçlarına göre sistemin yüz hareketleri ile fare konumunu başarılı bir şekilde kontrol edebildiği görülmüştür. Bunun yanı sıra sistemin göz kırpma durumlarını hızlı bir şekilde algılayarak farenin sol, sağ ve çift tıklama işlevini modelleyebildiği ortaya konulmuştur. Ayrıca sistem farklı kişiler tarafından da test edilmiş ve benzer sonuçlar elde edilmiştir. Bu sistem engelliler için eğitim hayatını kolaylaştırmada önemli bir rol oynayabilir. Fiziksel engelli bireylerin çalışma hayatını kolaylaştırabilecek bir sistem olabilir ve engellilerin günlük hayatta bilgisayar kullanımını kolaylaştırabilir. Yüksek çözünürlüklü ve düşük ortam aydınlatmasında bile net görüntüler elde edebilen bir kamera kullanılması sistemin doğru sonuçlar vermesinde etkili olacaktır. Aynı zamanda ortam aydınlatmasının düzeyi de sistem performansı üzerinde rol oynamaktadır.

Gelecek çalışma olarak, sanal bir klavye üzerinden yine yüz ve göz hareketleri ile yazı yazma özelliği de eklenecektir. Ayrıca sistemin fiziksel engelli bireyler tarafından test edilmesi sağlanarak elde edilen sonuçlar sunulacaktır.

Teşekkür

Bu çalışma Tubitak 2209-A, Üniversite Öğrencileri Araştırma Projeleri Destekleme Programı tarafından desteklenmiştir. Destekleri için TÜBİTAK'a teşekkür ederiz.

Yazar(lar)ın Katkıları

MFC ve AK fikri ortaya koydu. MFC sistemi geliştirdi. AK sistemin geliştirilmesinde ve yöntemlerde fikir ve yönlendirmeleri ile destek sağlamıştır. MFC makaleyi yazdı. AK makalenin tümünü gözden geçirerek ana yapıyı ve içeriği düzenledi. Her iki yazar da makalenin son halini okudu ve onayladı.

Çıkar Çatışması

Yazarlar, çıkar çatışması olmadığını beyan eder.

Not: Bu çalışma ICAIAME 2022 konferansında sunulmuş olup, özet metin olarak basılmıştır.

Kaynaklar

- [1]. Dönmez, M., Çağıltay, K., Fiziksel engelliler için göz hareketlerini izleme yöntemi ile bilgisayar kullanma sistemi, Eğitimde FATİH Projesi Eğitim Teknolojileri Zirvesi, Ankara, 2016. <https://hdl.handle.net/11511/75795>
- [2]. Kalaç, M. Ö., Teke, İ., Kılınç, M.. Türkiye'de engellilere yönelik bilişim teknolojileri ve hizmetleri üzerine genel değerlendirme, BİLTEVT Uluslararası Engelsiz Bilişim Kongresi, Manisa, 1-14, 2018.
- [3]. Baştuğ, R. S., Yeşilkaya, B., Unay, M. ve Akan, A., Virtual mouse control by webcam for the disabled, Medical Technologies National Congress (TIPTEKNO), Gazi Magosa 1-4, 2018. doi: 10.1109/TIPTEKNO.2018.8597147.
- [4]. Atay, B., Korkusuz, N.A., Fiziksel engelliler için göz takip yöntemi kullanılarak iletişim ve kontrol sisteminin geliştirilmesi, Uluslararası Engelsiz Bilişim 2021 Kongresi, Manisa, 2021.