

TÜRKİYE CUMHURİYETİ
YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



DERİN ÖĞRENME İLE SES SPEKTROGRAMI
GÖRÜNTÜLERİ ÜZERİNDEN YORGUNLUK TESPİTİ

17011044 – Uğur KESKİN

BİLGİSAYAR PROJESİ

Danışman
Dr.Öğr.Üyesi Hamza Osman İLHAN

Haziran, 2020

TEŞEKKÜR

Proje sürecinde sağladıkları katkılarıyla ve düşünceleri ile bana destek veren proje danışmanı hocam Dr .Öğr. Üyesi Hamza Osman İLHAN'a ve Arş. Gör. Hacer BAYIROĞLU hocama teşekkürlerimi sunarım.

Uğur KESKİN

İÇİNDEKİLER

SİMGE LİSTESİ	v
KISALTMA LİSTESİ	vi
ŞEKİL LİSTESİ	vii
TABLO LİSTESİ	ix
ÖZET	x
ABSTRACT	xi
1 Giriş	1
2 Ön İnceleme	2
2.1 Mevcut Çalışmalar	2
3 Fizibilite	4
3.1 Teknik Fizibilite	4
3.1.1 Yazılım Fizibilitesi	4
3.1.2 Donanım Fizibilitesi	4
3.2 Yasal Fizibilite	5
3.3 Ekonomik Fizibilite	5
3.4 Zaman Fizibilitesi	6
4 Sistem Analizi	7
4.1 Verilerin Tanımı	7
4.1.1 Verilerin Toplanması ve Düzenlenmesi	7
4.1.2 Verilerin Düzenlenmesi	8
4.2 Transfer Learning ve CNN Modelleri	9
4.3 Değişkenlerin ve Parametrelerin Optimizasyonu	10
4.4 Aktivasyon ve Hata Fonksiyonları	10
5 Sistem Tasarımı	12
5.1 CNN Modellerinin Eğitilmesi	12

5.1.1	Verilerin Düzenlenmesi	12
5.1.2	Görüntü Biçimlendirme ve Çeşitlendirme	12
5.1.3	Verilerin CNN Modellerinin Eğitiminde kullanılması	13
5.1.4	CNN Modellerinin Seçilmesi	13
5.2	CNN Yapısının Optimize Edilmesi	14
5.2.1	Öğrenme Katsayısı	14
5.2.2	Batch Boyutu	14
5.2.3	CNN Katmanları	15
5.3	Modelin Kaydedilmesi ve Yüklenmesi	16
6	Uygulama	17
7	DeneySEL Sonuçlar	21
8	Sonuç	27
	Referanslar	28
	Özgeçmiş	29

SİMGE LİSTESİ

KISALTMA LİSTESİ

5-NN	5-Nearest Neighbour
2D	2 Dimensional
ANN	Artificial Neural Network
BN	Bayesian Network
CNN	Convolutional Neural Network
DL	Deep Learning
EC	Ensemble Classifier
KSS	Karolinska Sleepiness Scale
LDA	Linear Discriminant Analysis
MLP	Multilayer Perceptron
PC	Parzen Classifier
SGD	Stochastic gradient descent
SOL	Sleep Onset Latency
SVM_l	Support Vector Machine Linear Kernel
SVM_r	Support Vector Machine Radial Kernel
VCM	Voice Correlation Metric
WHO	World Health Organization

ŞEKİL LİSTESİ

Şekil 3.1	Gant Diyagramı	6
Şekil 4.1	Kullanım Senaryosu Diyagramı	7
Şekil 4.2	Dosya Düzeni	8
Şekil 4.3	Transfer Learning ile Sınıflandırma	9
Şekil 4.4	VGG16 ve Resnet CNN Modelleri	9
Şekil 4.5	Softmax Fonksiyonu	10
Şekil 4.6	Categorical Cross Entropy Hata Fonksiyonu	10
Şekil 5.1	Belirli Oranlarda Yatay Olarak Kaydırılmış Ses Spektrogramı Görüntüsü	13
Şekil 5.2	Overfitting ve Underfitting	14
Şekil 5.3	Verilerin Elde Edilmesi ve Eğitilmesi Aşamaları	15
Şekil 6.1	Arayüz Görüntüsü	17
Şekil 6.2	Dosya Diyalogu Menüsü	18
Şekil 6.3	Sınıflandırma Sonuçlarının Gösterilmesi	18
Şekil 6.4	Test Edilecek Klasörün Seçilmesi	19
Şekil 6.5	Klasör Sınıflandırma Sonuçlarının Gösterilmesi	19
Şekil 6.6	Klasör Sınıflandırma Sonuçlarının Gösterilmesi	20
Şekil 7.1	1 veya 10 olarak puanlandırılmış harf veriseti üzerindeki test sonuçları	23
Şekil 7.2	1 veya 10 olarak puanlandırılmış kelime veriseti üzerindeki test sonuçları	23
Şekil 7.3	1 veya 10 olarak puanlandırılmış harf ve kelime birleşimi veriseti üzerindeki test sonuçları	23
Şekil 7.4	1, 2, 9 veya 10 olarak puanlandırılmış harf veriseti üzerindeki test sonuçları	24
Şekil 7.5	1, 2, 9 veya 10 olarak puanlandırılmış kelime veriseti üzerindeki test sonuçları	24
Şekil 7.6	1, 2, 9 veya 10 olarak puanlandırılmış harf ve kelime birleşimi veriseti üzerindeki test sonuçları	24
Şekil 7.7	1, 2, 3, 8, 9 veya 10 olarak puanlandırılmış harf veriseti üzerindeki test sonuçları	25

Şekil 7.8	1, 2, 3, 8, 9 veya 10 olarak puanlandırılmış kelime veriseti üzerindeki test sonuçları	25
Şekil 7.9	1, 2, 3, 8, 9 veya 10 olarak puanlandırılmış harf ve kelime birleşimi veriseti üzerindeki test sonuçları	25
Şekil 7.10	Sırasıyla 1, 2, 9 veya 10 olarak sınıflandırılmış Harf, Kelime ve Birleştirilmiş Veriseti Üzerindeki Sonuçlar	26

TABLO LİSTESİ

Tablo 3.1	Gereken Minimum Özellikler	5
Tablo 3.2	Google Colab Sistem Özellikleri	5
Tablo 3.3	Maliyetler	5
Tablo 7.1	Spektrogram tipi, Ses tipi, Yorgunluk seviyesi ve sınıfa göre veri sayıları	21
Tablo 7.2	En Yüksek Doğruluk Oranını Sağlayan Parametreler	22
Tablo 7.3	En başarılı modellerin Şekil 7.10 den türetilen performans ölçütleri	26

DERİN ÖĞRENME İLE SES SPEKTROGRAMI GÖRÜNTÜLERİ ÜZERİNDEN YORGUNLUK TESPİTİ

Uğur KESKİN

Bilgisayar Mühendisliği Bölümü
Bilgisayar Projesi

Danışman: Dr.Öğr.Üyesi Hamza Osman İLHAN

Yorgunluk, gündelik hayatta çeşitli kazalara ve hatalara yol açabilen yaygın bir durumdur. Bu kazalara özellikle trafik kazaları, iş kazaları örnek verilebilir. Bu kazaları ve hataları önlemek için yorgunluk durumunun oluşması durumunda kaza yapma potansiyeline sahip olan yorgun bireyi uyarmak ya da genel anlamda bu konuda veri toplayıp bunları iş saatlerini düzenlemekte kullanmak gibi sistemler mevcuttur.

Yorgunluk, insanın anlık yüz özelliklerine ya da ani olaylara verdiği tepki süresinden anlaşılabilirdiği gibi, insanın sesinden de anlaşılabilir. İnsan sesinin nicel olarak nasıl ifade edilebileceği konusunda da farklı fikirler mevcuttur.

Bu projede, bir derin öğrenme metodu olan CNN modeli kullanılarak insan seslerinin ne kadar yorgunluk belirttiğine dair tahminler yapılmıştır. Bu tahminlerin yapılırken, insan sesinin nicel verilerle ifade edilebilmesi için, seslerin spektrogram görüntülerinin renk değerleri kullanılmıştır. Proje sonucunda ses spektrogramlarının dinç ya da yorgun olarak sınıflandırılmasında yaklaşık %94'lük bir doğruluk oranına ulaşılmıştır.

Anahtar Kelimeler: Yorgunluk ses analizi, ses spektrogramları, transfer öğrenme, derin öğrenme, yapay sinir ağları, ses sınıflandırma

ABSTRACT

FATIGUE ANALYSIS BASED ON VOICE SPECTROGRAM IMAGES USING DEEP LEARNING

Uğur KESKİN

Department of Computer Engineering
Computer Project

Advisor: Assist. Prof. Dr. Hamza Osman İLHAN

Fatigue is a common condition that can lead to various accidents and mistakes in everyday life. Traffic accidents, occupational accidents can be given as examples. In order to prevent these accidents and errors, there are systems such as to alert the tired individual who has the potential to have an accident in case of fatigue or to collect data in this regard and to use them to organize the working hours.

As well as fatigue can be understood from a person's reaction time to sudden events or from her/his current facial features, it can be understood from a person's voice. Also there are plenty options for expressing human voice as quantitative data.

In this project, by using the CNN model, a deep learning method, estimates of how tired the human voices indicate are made. In making these predictions, the color values of the spectrogram images of the voices were used to express the human voice with quantitative data. As a result of the project, an accuracy rate of approximately %94 has been reached in the classification of sound spectrograms as vigorous or tired.

Keywords: Fatigue voice analysis, voice spectrograms, transfer learning, deep learning, neural networks, audio classification

1

Giriş

Çoğu zaman dikkatsizlik, yapılan işte verimsizlik ve hatalara yol açan büyük bir etken olan yorgunluk, gün içinde çoğunlukla rastlanan bir durumdur. Yorgunluk insanın belirli hareketlerinden ve ifadelerinden anlaşılabilir. Bununla birlikte uzun sürüş süresi otoyollarda veya büyük otoyollarda yorgunluğa bağlı kazaların önemli bir nedenidir[1], bu nedenle gün içerisinde beliren yorgunluk trafik kazalarına ve ölümcül iş kazalarına yol açan önemli bir etkidir. Bunu belli eden etkenlerden birisi de insanın sesidir. İnsanlar yorgunken bu yorgunluk sese yansayabilir dolayısıyla insan sesinden onun yorgun olup olmadığını anlamak mümkündür. Uygulama kendi araştırmalarında ses üzerinden yorgunluk analizi özelliğine ihtiyaç duyanların kullanması için geliştirilmiştir.

2000'lerin başından bu yana, CNN modelleri görüntülerdeki nesne ve bölgelerin algılanması, bölümlere ayrılması ve tanınmasında büyük bir başarıyla uygulanmıştır[2]. Bu ağların bir konuda eğitilebilmesi için sayısal veriler gereklidir, metindeki kelimelerden insan sesine kadar birçok şey matematik ile ifade edilebilecek durumdadır. Bu projede derin öğrenme yöntemi kullanılmış, insan sesinin farklı tekniklerle elde edilmiş ses spektrogram görüntülerinden faydalanan bu görüntülerden sesin sahibi olan kişinin yorgun olup olmadığını analiz etmek için, seçilen CNN yapıları optimize edilerek bunların arasından en başarılı derin öğrenme modeli geliştirilmeye çalışılmıştır. Bununla birlikte kullanıcının sınıflandırmak istediği ses spektrogramı görüntüsünü girip bunun geliştirilen model yardımıyla hangi sınıfa ait olduğunu kullanıcıya bildirmesine imkan tanıyacak, arayüze sahip bir uygulama geliştirilmiştir.

2 Ön İnceleme

Yorgunluk analizi çoğunlukla trafik ya da iş kazalarını önlemek amacıyla, arabadaki sürücü asistanı olarak sürücünün yorgunluğunu tespit edip onu uyararak ya da iş yerindeki saha çalışanlarının yorgunluğunu tespit edip yetkililere uyarı vererek sürücülere ve işçilere yarar sağlar. Yabancı kaynaklarda daha çok "Fatigue Analysis" olarak adlandırılan yorgunluk analizi çalışmalarının bazılarında yorgunluk analizi ve tespiti için farklı tip verilerin kullanıldığı görülmektedir. Bu bölümde yorgunluk analizi için yapılan çalışmaların incelemesi yapılmıştır.

2.1 Mevcut Çalışmalar

[3]’de yapılmış olan çalışmada, ses verileri kullanılarak insan yorgunluğunun tahmini hedeflenmiştir. Ses analizi yöntemleri kullanarak yapılan bu çalışmada 3 farklı gruba ayrılmış 31 normal insandan alınan 296 konuşma kayıt altına alınmış ve veri olarak kullanılmıştır. Elde edilen bu ses kayıtları Fourier Dönüşümünden geçirildikten sonra Mel-scale Filterbank Analizi yöntemiyle insan kulağının tepki özelliklerine göre ölçeklendirilmiştir ve Cepstral analizi ile birlikte 2 türevsel süreçten geçirilerek ses verilerinin her karesi için 36 farklı değişken elde edilmiştir. Ses verilerinin uykulu-uykusuz olarak puanlanarak etiketlenmesi için SOL(Uyku Gecikmesi Testi) ve p ("pea" kelimesi içinde) ile t("tea" kelimesi içinde) sözcüklerinin denendiği VCM(Ses Korelasyon Metriği) metriklerinden yararlanılmıştır. Gerçek sonuç ve tahmin edilen değer arasındaki hata R^2 ile ölçülmüş ve sonuç olarak tahminler SOL metriği için %68, VCM("p") için %79 ve VCM("t") için %45 başarı sağlamıştır.

[4]’de sürücülere yönelik yapılmış olan çalışmada, sürücünün yüz ifadeleri izlenerek, elde edilen görüntüleri tahminleyip sürücünün yorgunluğunu tespit eden bir uygulama geliştirilmiştir. Yüzün farklı kısımları için ayrı ayrı analiz yapılan çalışmada, göz hareketi, göz bebeği hareketi, kafa yönü tahmini, göz bakışı doğrultusu saptama, yüz ifadesi analizinden faydalanılmıştır. Faydalanılan yüz hareketlerinden elde edilen veriler, BN yöntemi ile sınıflandırılmış ve yorgunluğun her bir verinin hangi

durumunda ne kadar ihtimalle belirlediği açıklanmıştır. Elde edilen durumlardan ortalama göz kapatma hızının yüksek olduğu ve esneme frekansının düşük olduğu duruma bakılarak yapılan sınıflandırma ile %95.52 başarı oranı elde edilmiştir.

[5]'de yapılmış olan çalışmada, ses kayıtları alınan 12 katılımcıların ses kaydı ayrıntılı olarak incelenmiştir, 2 katılımcının ses kayıtları ses sorunlarından dolayı parçalı olarak analiz edilmiştir. Uykudan alıkoyulan katılımcıların gece her ses kaydı alınmadan hemen önceki 1 saat, KSS yöntemi kullanılarak yorgunluğa göre 1 ve 10 arasında puanlandırılmıştır. Elde edilen ses kayıtları Praat Konuşma Analizi yazılımı ile akustik özelliklerine ayrılmış ve böylece her konuşma örneği için toplam 45,216 değişken elde edilmiştir. Elde edilen veriler 8 makine öğrenme yöntemine ayrı ayrı verilerek sınıflandırılmaya çalışılmıştır. Bu yöntemler MLP, SVM_l, SVM_r, 5-NN, Karar ağacı, PC, Lojistik Regresyon ve LDA modelleridir. Daha iyi bir sonuç elde etmek için Makine Öğrenmesi yöntemlerinin sınıflandırma çıktılarının ortalaması alınarak bir birleşik sınıflandırıcı modeli oluşturulmuştur. Oluşan EC modeli, modelin eğitiminde daha önceden kullanılmamış ses verileri ile test edilmiştir, elde edilen başarı oranı %83 olarak kaydedilmiştir.

Önceki çalışmalardan anlaşıldığı üzere, bu projede önceki çalışmalardan farklı olarak farklı veri tipi ve farklı bir çözüm yöntemi kullanarak yüksek bir performans sağlamayı amaçladım. Kullandığım veriler sesin çeşitli yöntemlerle dönüştürülerek oluşan ses spektrogram görüntüleridir, yorgunluk tespitinin çözümü etiketlenmiş veriler kullanılarak eğitilen derin öğrenme yöntemi kullanılarak bulunmaya çalışılmıştır.

Bu kısımda projenin gereksinimlerini belirlemek adına fizibilite çalışmalarından bahsedilmiştir. Teknik fizibilite kısmında Yazılım ve Donanım Fizibilitesinden söz edilmiştir. Yazılım Fizibilitesinde hangi yazılımsal ortamların, kütüphanelerin ve araçların kullanıldığından, Donanım Fizibilitesinde sistem için ihtiyaç duyulan donanım gereksinimlerinden, Yasal Fizibilitede projenin ihtiyaç duyduğu lisanslardan Zaman Fizibilitesinde projenin gerçekleştirilmesi için gereken süreden, Ekonomik fizibilitede sistem gerçekleşmesi için gereken giderlerden söz edilmiştir.

3.1 Teknik Fizibilite

Teknik Fizibilite Yazılım Fizibilitesi ve Donanım Fizibilitesi olmak üzere 2 ayrı başlıkta açıklanmıştır.

3.1.1 Yazılım Fizibilitesi

Proje geliştirilirken yeniden kullanılabilirliğinin yüksek olması, kaynak bolluğu, yapay zeka kütüphanelerinin getirdiği kolaylıklar ve kolay okunabilirliğinden dolayı Python dili tercih edilmiştir. Derin Öğrenme modellerinin daha anlaşılır ve kolay bir biçimde ifade edilebilmesi ve kullanılabilmesi için Tensorflow ve Tensorflow destekli Keras kütüphanesi kullanılmıştır. Önceden belirlenmiş derin öğrenme modellerinin eğitilmesi ve Python kodu ortamı için ücretsiz GPU desteği sağlaması açısından Google Colabratory kullanılmıştır.

Uygulamanın arayüzü Python'da Tkinter kütüphanesi kullanılarak geliştirilmiştir.

3.1.2 Donanım Fizibilitesi

Uygulamanın çalışması için gerekli sistem Tablo 3.1 daki gibidir.

CNN Modellerinin eğitiminde kullanılan Google Colaboratory 'nin sunduğu bulut

İşlemci	IntelCore i3 U 330
Ekran Kartı	AMD Radeon HD 5500
İşletim Sistemi	Windows 10
RAM	2 GB

Tablo 3.1 Gereken Minimum Özellikler

İşlemci	Intel Xeon CPU
Ekran Kartı	Nvidia Tesla K80
Depolama	359 GB
GPU RAM	12 GB
CPU RAM	13.3 GB

Tablo 3.2 Google Colab Sistem Özellikleri

servisin teknik özellikleri Tablo 3.2 da gösterilmiştir.

3.2 Yasal Fizibilite

CNN modellerinin eğitimi ve test edilmesi için kullanılan veriseti önceden hazırlanmıştır. Projede kullanılan kütüphaneler, lisanslarına uygun bir biçimde kullanılmıştır.

Keras kütüphanesi The MIT License (MIT) ile lisanslıdır[6].

Tensorflow kütüphanesi Apache License 2.0 ile lisanslıdır[7].

OpenCV kütüphanesi The MIT License (MIT) ile lisanslıdır [8]

3.3 Ekonomik Fizibilite

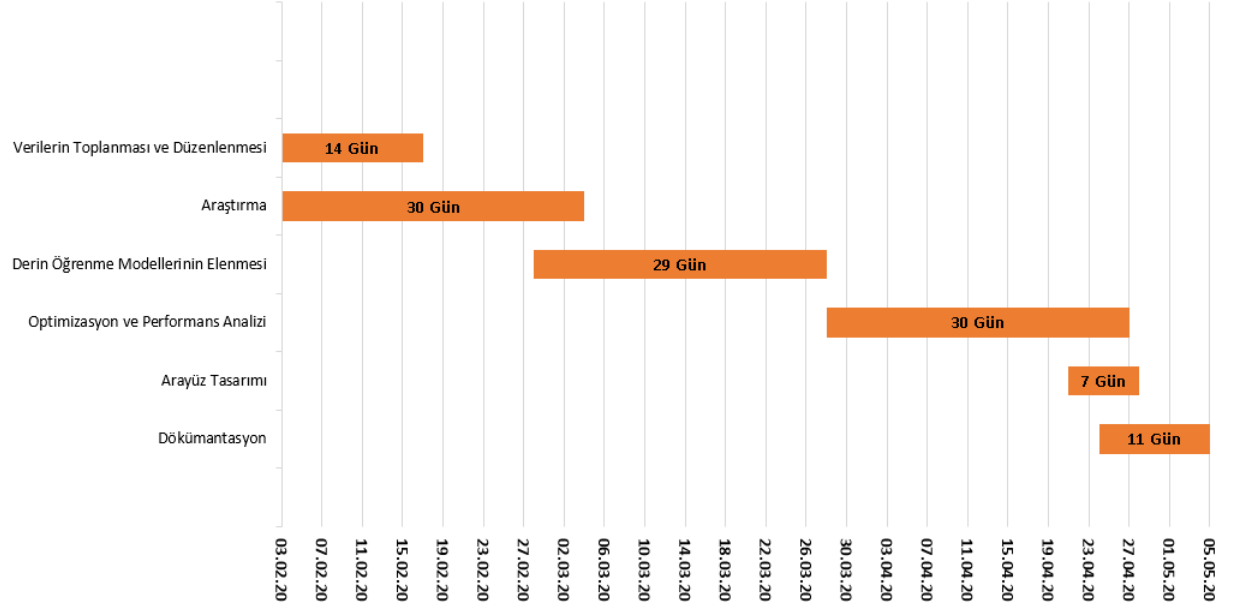
CNN Modeli'nin eğitilmesi, sistemin kurulması durumunda harcanan maliyet Tablo 3.3 deki gibidir.

İş Gücü Maliyeti	4000 TL
Elektrik Maliyeti	50 TL
Toplam Maliyet	4050 TL

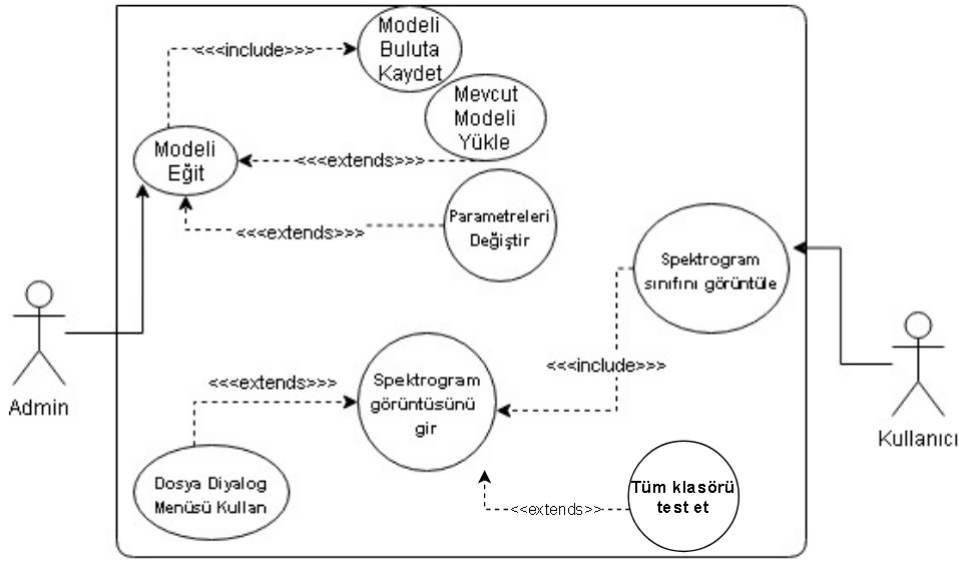
Tablo 3.3 Maliyetler

3.4 Zaman Fizibilitesi

Projenin bahar dönemi içerisinde bitmesi planlanmıştır. Proje tek kişilik olduğu için herhangi bir görev dağılımı olmamıştır. Gant diyagramı Şekil 3.1 de gösterilmiştir.



Şekil 3.1 Gant Diyagramı



Şekil 4.1 Kullanım Senaryosu Diyagramı

Bu bölümde kullanım senaryosu, sistemin ihtiyaç duyduğu veriler ve nesneler, verilerin toplanma ve düzenlenme süreci, sistemin eğitiminde kullanılan yöntem ve bu yöntemin geliştirilme süreci açıklanmıştır. Sistemin içindeki aktörler ve kullanım senaryoları da Şekil 4.1 de gösterilmiştir.

4.1 Verilerin Tanımı

Bu bölümde verilerin toplanma yöntemlerinden ve eğitime hazır hale getirilmesinden bahsedilmiştir.

4.1.1 Verilerin Toplanması ve Düzenlenmesi

Kurulan derin öğrenme modelinin eğitilmesi için gereken veriler ses spektrogramı görüntüleridir. Görüntüler MATLAB ortamında BARK, ERB, LOG, MEL olmak üzere

her 4 spektrogram tipi için ayrı ayrı oluşturulmuştur. Bu görüntüler, oluşturulan insan seslerinin birden fazla kişinin 1-10 arasında puanlamasıyla yorgun veya dinç olarak etiketlenmiştir ve puanlarına göre farklı klasörler oluşturulmuştur.



Şekil 4.2 Dosya Düzeni

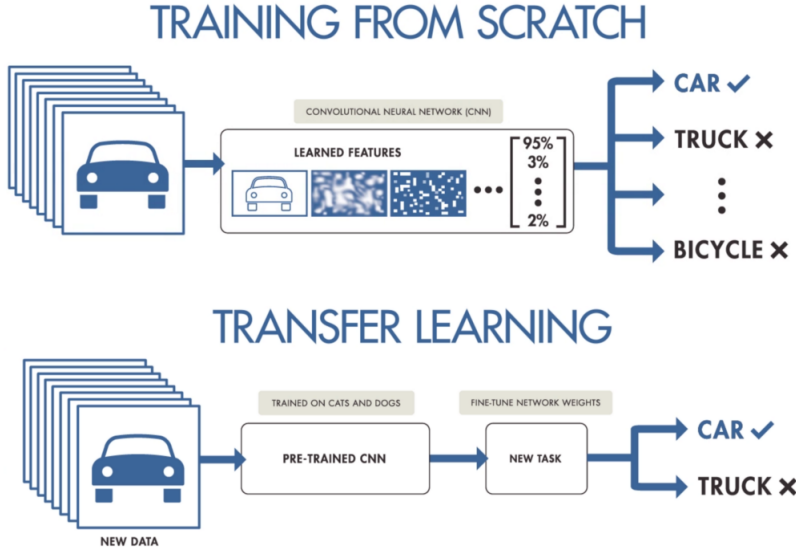
Oluşan dosya düzeni Şekil 4.2 deki gibidir.

4.1.2 Verilerin Düzenlenmesi

Kullanılan her spektrogram oluşturma yöntemi ile puan aralıkları için oluşturulan farklı klasörlerin her biri, eğitimin daha tutarlı ve doğru olabilmesi için 5-fold validation mimarisine uyması açısından içinde 5 klasör barındırmaktadır. Derin Öğrenme modellerinde verinin eğitime verilmeden önce daha iyi sonuçlar elde edebilmek ve overfitting gibi sorunları çözebilmek için verilerin düzenlenmesi gerekmektedir. Bu adımları sağlamak için görüntülerin boyutları ve piksel değerleri seçilen CNN modellerine uyacak şekilde değiştirilmiştir.

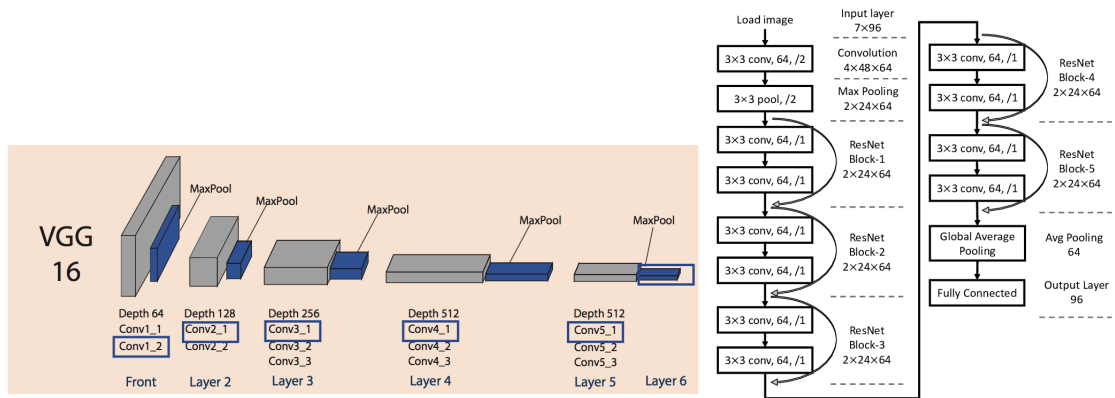
4.2 Transfer Learning ve CNN Modelleri

Bir CNN modelinin ilk önce ImageNet kullanarak ön eğitim yapılması, daha sonra da yeni bir hedefe adapte olması için tekrardan eğitilmesi, bilgisayarlı görü problemlerini çözmek için bir standart haline gelmiştir [9] yazacağımız rapor ara proje raporda hangi kısımlar olacak her kısım olacak mı ve 4.3 görselinde açıklandığı üzere, önceden eğitilmiş bir modelin elde edilip yeni bir probleme adapte edilmesi için transfer learning yönteminden faydalanılır.



Şekil 4.3 Transfer Learning ile Sınıflandırma

Hangi CNN Modelinin çözüme daha iyi ulaşacağını bulmak için bu projede üzerinde çalışılması belirlenmiş 7 yapay sinir ağı modellerinin belirli verisetleri üzerindeki skorlarına bakılmıştır. Bunlar Resnet50, VGG16, VGG19, MobileNet, MobileNetV2, DenseNet201 ve NasNetMobile modelleridir.



Şekil 4.4 VGG16 ve Resnet CNN Modelleri

4.4 görselinden görüldüğü gibi, karmaşık derin öğrenme problemlerini çözmede

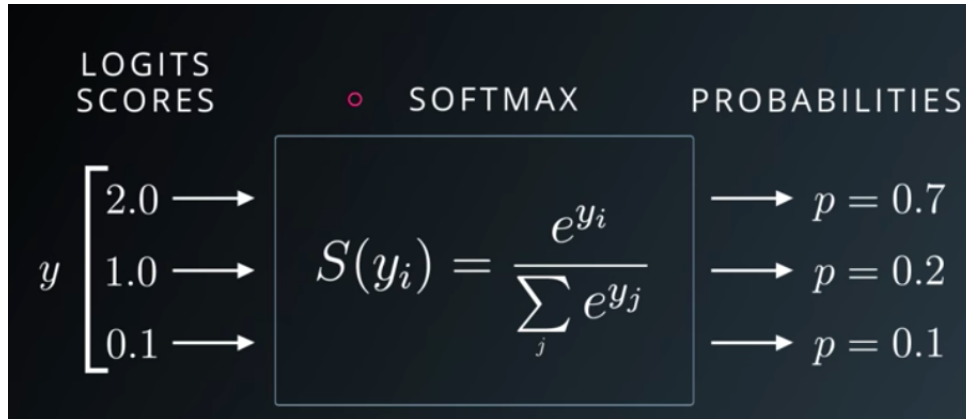
standart haline gelen bu CNN modelleri, farklı katman, katman sayısı, her katman için belirlenen aktivasyon fonksiyonları, ek katmanlar ve kısacası mimarisi yönüyle birbirinden farklıdır.

4.3 Değişkenlerin ve Parametrelerin Optimizasyonu

Derin Öğrenme eğitiminin daha doğru ve tutarlı sonuçlar vermesi için birçok optimizasyon fonksiyonları denenmiştir, bunlar Adam, SGD, RmsProp, Nadam, Adamax ve Adadelta fonksiyonlarıdır. Bu fonksiyonların yanı sıra öğrenme katsayısı, yapay sinir ağındaki eğitilebilir gizli katman sayısı, her katman için belirlenecek aktivasyon fonksiyonları, epoch sayısı, batch boyutu ve loss fonksiyonu gibi değişkenlerin en iyi sonuç verecek şekilde belirlenmesi hedeflenmiştir.

4.4 Aktivasyon ve Hata Fonksiyonları

Bu bölümde kullanılan hata ve aktivasyon fonksiyonlarından söz edilmiştir.



Şekil 4.5 Softmax Fonksiyonu

$$CCE(p, t) = - \sum_{c=1}^C t_{o,c} \log(p_{o,c})$$

Şekil 4.6 Categorical Cross Entropy Hata Fonksiyonu

Geliştirilen CNN modelinde son katmanın aktivasyon fonksiyonu Şekil 4.5’de açıklanan softmax, hata fonksiyonu da veri sınıflandırmaya uygun olması için

Şekil 4.6’de gösterilen Categorical Cross Entropy olarak belirlenmiştir. Bunların dışında aktivasyon fonksiyonu olarak sigmoid fonksiyonu kullanmak, hata fonksiyonu olarak Binary Cross Entropy kullanmak modelin gelişiminde seçilenler kadar etkili olmamıştır.

Bu bölümde daha önce yüzeysel olarak anlatılan CNN modellerinin detaylı açıklaması yapılmış, verilerin işlenmesi, özellikleri , modelin optimizasyonu ve optimizasyonda kullanılan değişkenlerden söz edilmiştir.

5.1 CNN Modellerinin Eğitilmesi

Bu bölümde, üzerinde çalışılmasında karar kılınmış Resnet50, VGG16, VGG19, MobileNet, MobileNetV2, DenseNet201 ve NasNetMobile ağlarının eğitimindeki aşamalar açıklanmıştır.

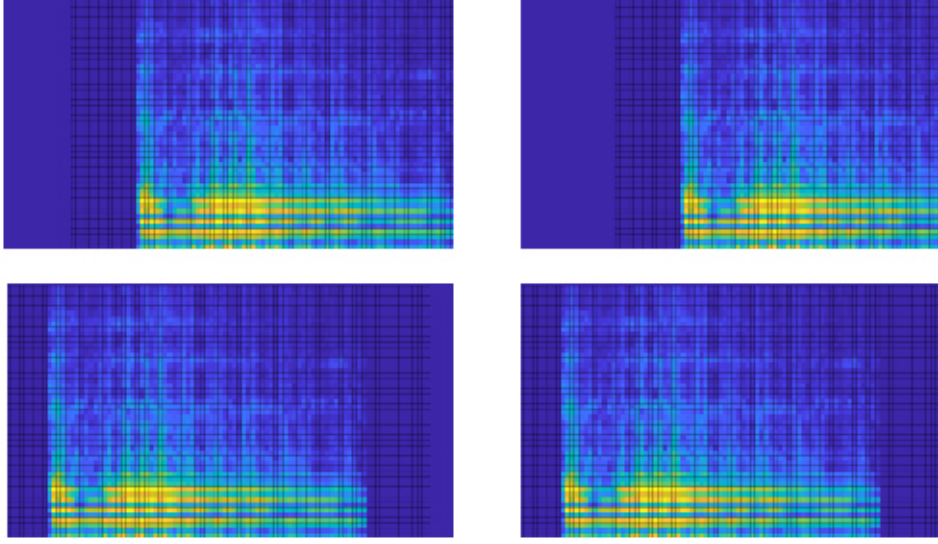
5.1.1 Verilerin Düzenlenmesi

Verilerin eğitim ve test için kullanılabilmesini sağlamak için eğitim ve test için kullanılacak görüntülerin boyutları RGB tipinde oldukları için ve kullanılan sinir ağlarının giriş katmanına uyması için 1800x1013x3 piksel boyutlarından 224x224x3 piksel boyutlarına dönüştürülmüş, 0 ve 255 arasında olan RGB değerleri 255'e bölünerek 0 ve 1 arasına normalize edilmiştir. Veriler "one hot encoding" yöntemi kullanılarak etiketlenmiştir. Böylece "yorgun" ve "dinç" sınıfları çıkış katmanının nöronları için sırasıyla [0 1] ve [1 0] şeklinde etiketlenmiştir.

5.1.2 Görüntü Biçimlendirme ve Çeşitlendirme

Ağın veriyi yanlış bir desen uydurarak tahminlemeye çalışmasını engellemek, overfitting problemini azaltmak için yapılan yöntemlerden birisi de görüntüler üzerinde, beyazlatma, aynalama, kaydırma, çevirme gibi işlemler yaparak görüntüyü çoğaltmak ya da çeşitlendirmektir. Fakat bu biçimleme işleminin görüntülerin anlamını kaybetmesine yol açma ihtimalini de göz önünde bulundurmak gereklidir. Bu neenle ses spektrogram görüntülerinde sadece yatay kaydırma ile görüntüler biçimlendirilmeye çalışılmıştır. Görüntünün hangi oranda kaydırılacağı ve

kaydırıldıktan sonra oluşan boşluğun nasıl doldurulacağını belirleyen parametreler de ağı eğitiminin optimizasyonuna etki eden değişkenlerdendir. Görsel 5.1 de bir ses spektrogram görüntüsünün yatay kaydırma ile biçimlendirilmiş örnekleri görülmektedir.



Şekil 5.1 Belirli Oranlarda Yatay Olarak Kaydırılmış Ses Spektrogramı Görüntüsü

5.1.3 Verilerin CNN Modellerinin Eğitiminde kullanılması

Seçilen ağlar Google Colaboratory üzerinden eğitilmiştir. eğitim ve test için kullanılacak ses spektrogram görüntüleri Google Drive üzerinden çekilmiştir. Bu görüntülerin çözünürlükleri ve dolayısıyla boyutları yüksek olduğu için hepsinin önceden yüklendikten sonra kullanılması disk ve ram değerleri düşünüldüğünde verimli değildir, bu yüzden Python Keras ve Tensorflow kütüphanelerinden yararlanılarak görüntüler eğitim sırasında mevcut oturum boyunca indirilerek kullanılır. Eğitimin daha hızlı olması için Google Colab'ın GPU desteği özelliğinden yararlanılmıştır.

5.1.4 CNN Modellerinin Seçilmesi

Üzerlerinde çalışılmasına karar verilmiş Resnet50, VGG16, VGG19, MobileNet, MobileNetV2, DenseNet201 ve NasNetMobile ağlarının problemin çözümü için hangisinin üzerinde yoğunlaşılacağını belirlenmesi için, her ağ hata fonksiyonu Categorical Cross Entropy, ağı son katmanın aktivasyon fonksiyonu softmax olacak şekilde eğitilmiş, başarının ölçülmesinde metrik olarak test doğruluğu dikkate alınmış,

her ağ için 6 farklı optimizasyon fonksiyonu denenmiş ve eğitim boyunca test başarı oranları kaydedilmiştir.

Kullanılan optimizasyon fonksiyonları Nadam, Adam, Adamax, Adadelta, RMSProp ve SGD'dır Eğitim ve test SPECT_MEL_12_910_harf dosyası üzerindeki her birinde eğitim için 265, test için 65 ses sprektrogram görüntüsü bulunan 5 fold klasörünün kendi eğitim klasöründeki görüntüler ile eğitilip kendi test klasöründeki görüntüler ile test edilmiş, başarı oranları bu 5 klasörün ulaşılan en yüksek test başarısının ortalaması olarak belirlenmiştir. En yüksek doğruluk oranına ulaşan sinir ağı-optimizasyon fonksiyonu çifti, sonradan yapılacak optimizasyon işlemlerinde kullanılmak üzere seçilmiştir.

5.2 CNN Yapısının Optimize Edilmesi

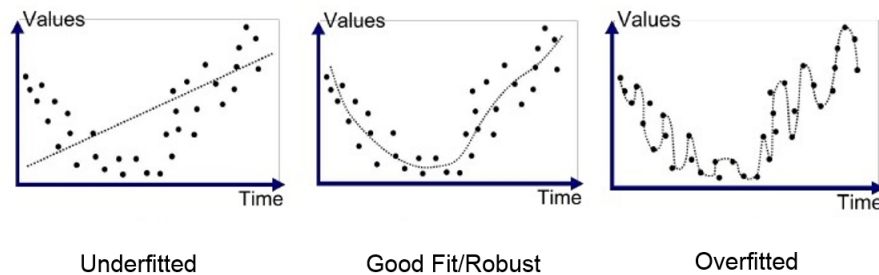
CNN yapılarının eğitiminde karşılaşılan problemlerin çözülmesi ve doğruluk oranının artırılması için çeşitli optimizasyon yöntemleri kullanılmıştır.

5.2.1 Öğrenme Katsayısı

Öğrenme Katsayısı, ağın eğitiminde ağırlıkların ne kadar oranla değişeceğini belirlemede kullanılır. Yüksek öğrenme katsayısı, ağın hata uzayındaki mutlak minimuma ulaşmasını önleyebilir. Düşük öğrenme katsayısı ise ağın öğrenme süresini uzatabilir. Öğrenme katsayısının eğitim sırasında giderek azalması ya da seçilen optimizasyon fonksiyonuna göre bir değer alması tercih edilmiştir.

5.2.2 Batch Boyutu

Batch Boyutu, ağın eğitimde tek seferde kaç tane veri için ağırlıkların değişeceğini belirler, bu yüzden düşük batch seçmek ağın eğitim verileri üzerinde aşırı uyum göstermesine yol açıp test doğruluğunu düşürebilir, sistemin ram'ine uygun yüksek bir batch boyutu seçmek tercih edilmiştir[10].

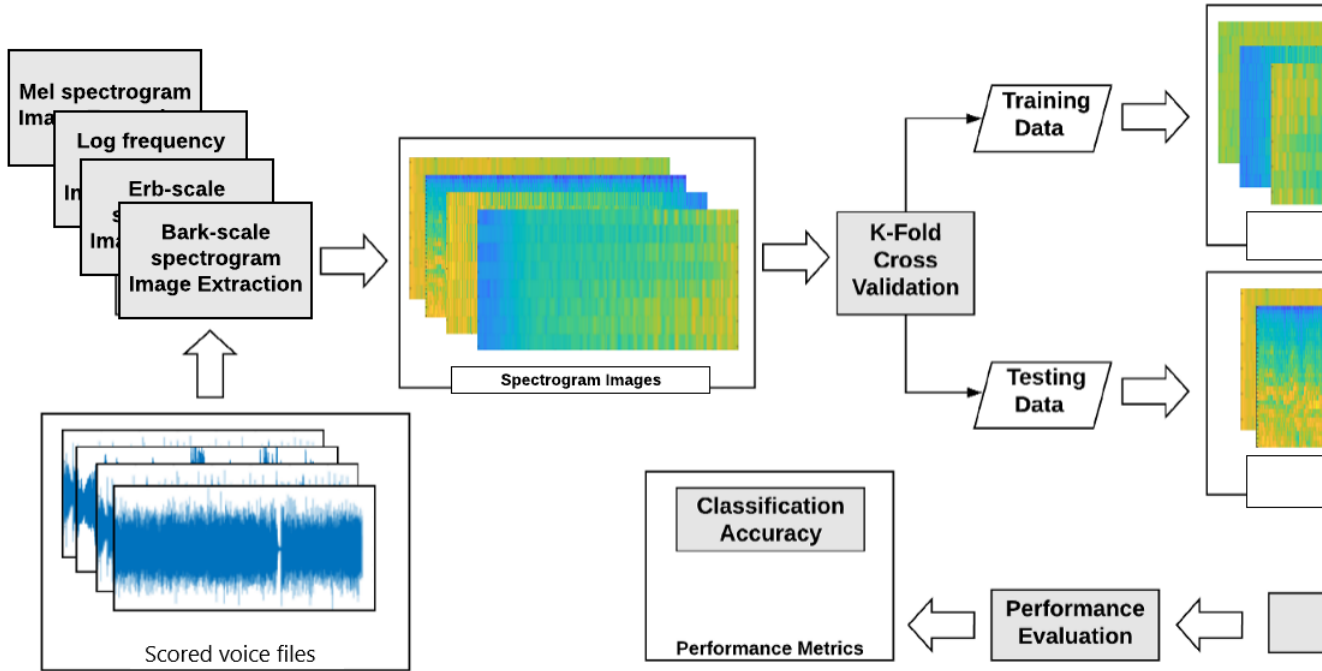


Şekil 5.2 Overfitting ve Underfitting

5.2 görselinde aşırı uyum ve eksik uyum durumları açıklanmıştır.

5.2.3 CNN Katmanları

Transfer öğrenme yönteminde seçilen önceden eğitilmiş CNN modeli, ImageNet görüntüleriyle eğitilmiş ağırlıklarıyla tanımlandıktan sonra modelin hangi katmanlarının dondurulup hangi katmanlarındaki ağırlıklarının eğitilebilir kılınacağına seçilmesi önemli bir performans artışı sağlayabilmektedir. Son katman çıktı katmanı olarak adlandırılmaktadır ve veriler kategorik olarak etiketlendikleri için çıktı katmanındaki her nöron bir sınıfı temsil edecek biçimde ayarlanıp buna göre çıktı katmanındaki nöron sayısı belirlenmiştir.



Şekil 5.3 Verilerin Elde Edilmesi ve Eğitilmesi Aşamaları

Optimizasyon denemelerinde, seçilen ağın son 3, son 2 ve son 1 konvolusyon katmanına kadar olan katmanların ağırlıkları eğitilebilir kılınarak eğitim yapılmıştır. Bunun yanı sıra, pooling katmanı, normal katman, dropout katmanı gibi katmanların

eklenmesi de performansı deęiřtirebilmekte ve belirli ölçütlerde aęın veriler üzerindeki aşırı uyum problemlerini giderebilmektedir [11].

5.3 Modelin Kaydedilmesi ve Yüklenmesi

Eęitilen Model, içerdęi aęırlıkların deęeri, katmanları ile birlikte bütün bir řekilde Google Drive’da eęitilmiş modellere özel açılmış bir klasörde depolanmıştır. Google Drive API’sinin sağladığı imkanla bu klasöre erişilip, model istenildięi zaman yüklenilebilmektedir. Modelin depolandığı dosya yerel depoya indirilebilmektedir.

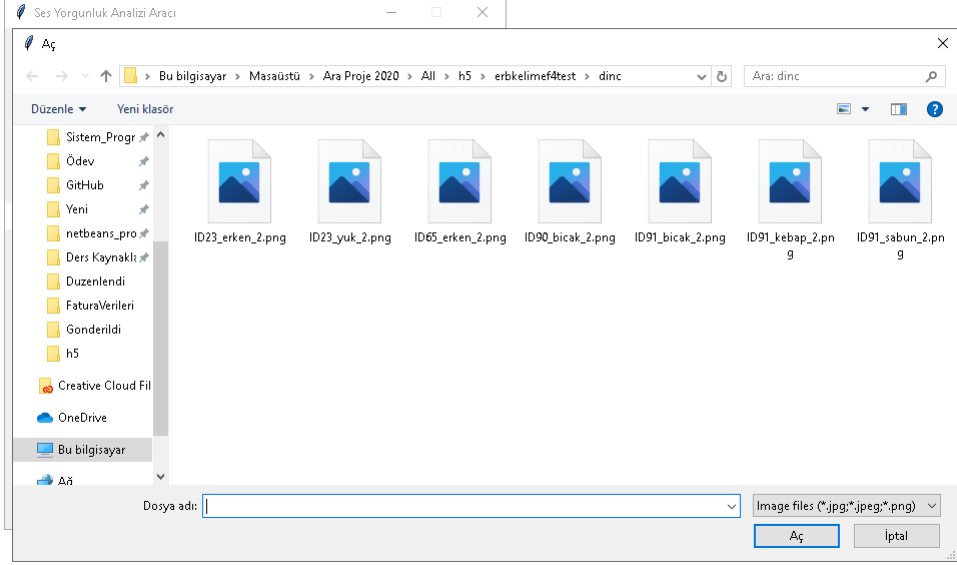
6 Uygulama

Bu bölümde uygulamaya ait ekran görüntülerine yer verilmiştir.

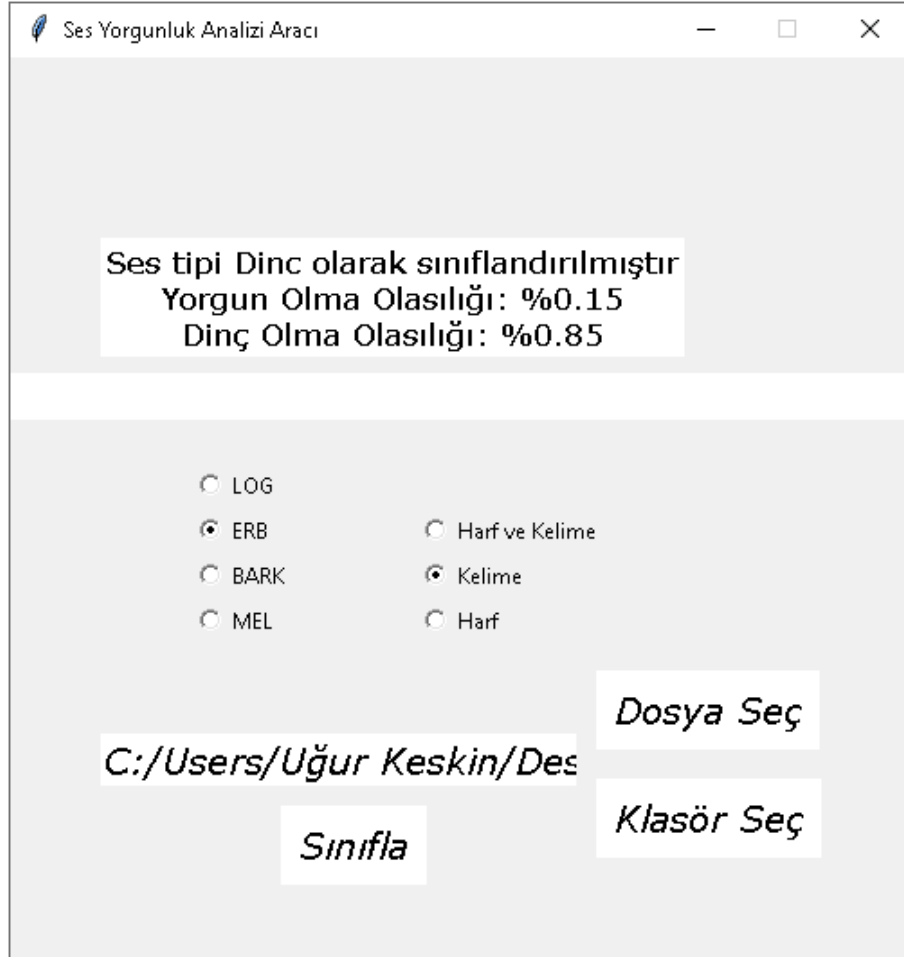


Şekil 6.1 Arayüz Görüntüsü

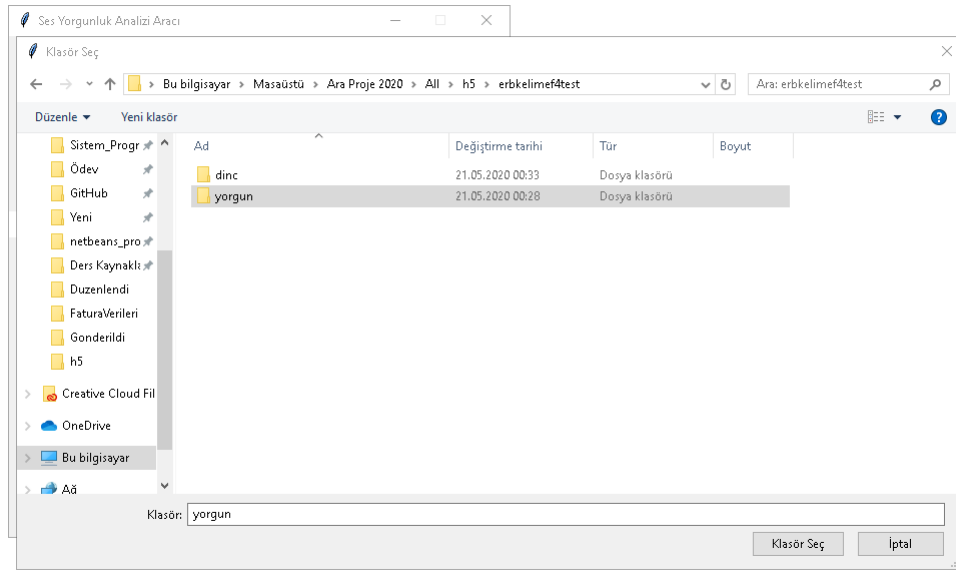
Şekil 6.1 de basit kullanıcı arayüzü görülmektedir. Kullanıcı spektrogram türünü ve eğitilmiş olduğu ses türünü seçer ve sınıflandırmada kullanılacak model belirli bir dosya düzenine göre otomatik olarak seçilir. Kullanıcı tek bir görüntüyü test etmek için "Dosya Seç" butonuna , bir klasördeki görüntüleri test etmek için "Klasör Seç" butonuna tıklayıp açılan dosya diyalogu menüsünden uygun görüntü ya da klasörü seçerler.



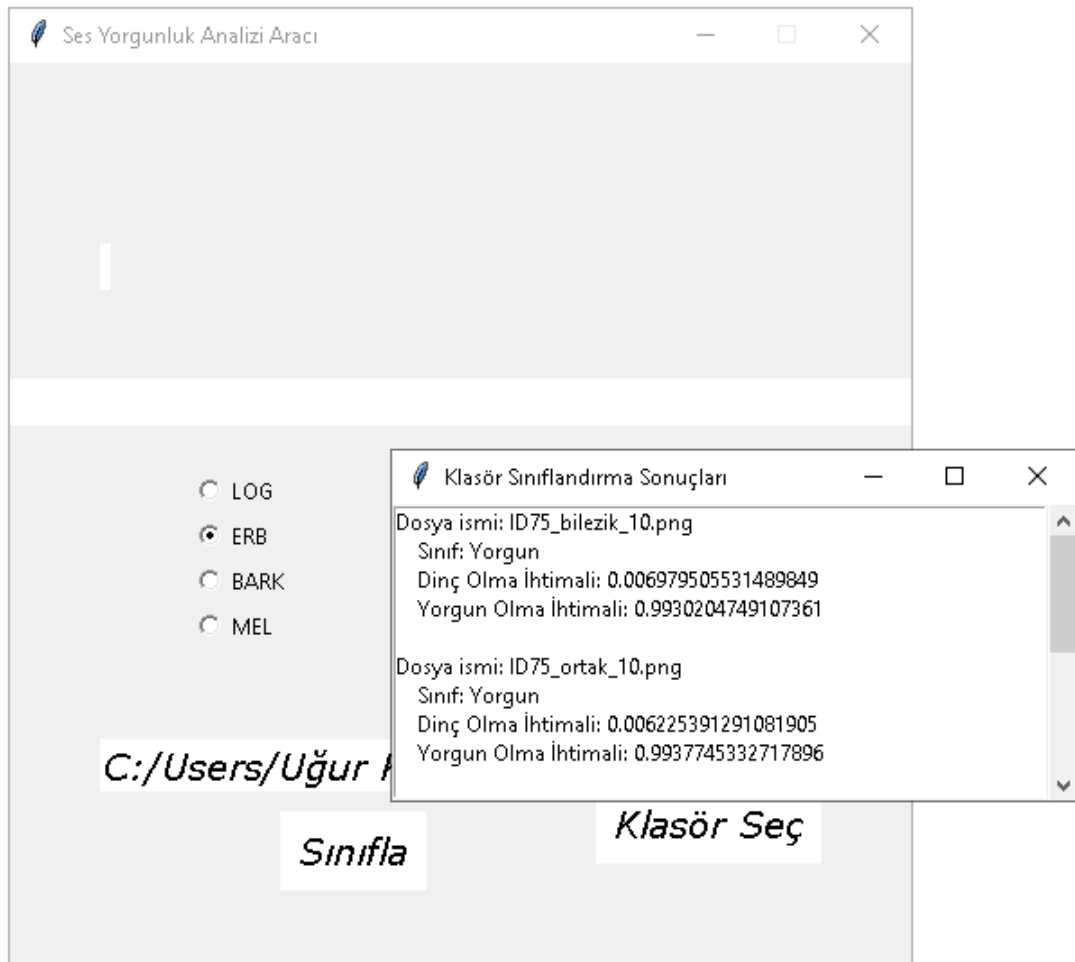
Şekil 6.2 Dosya Diyaloğu Menüsü



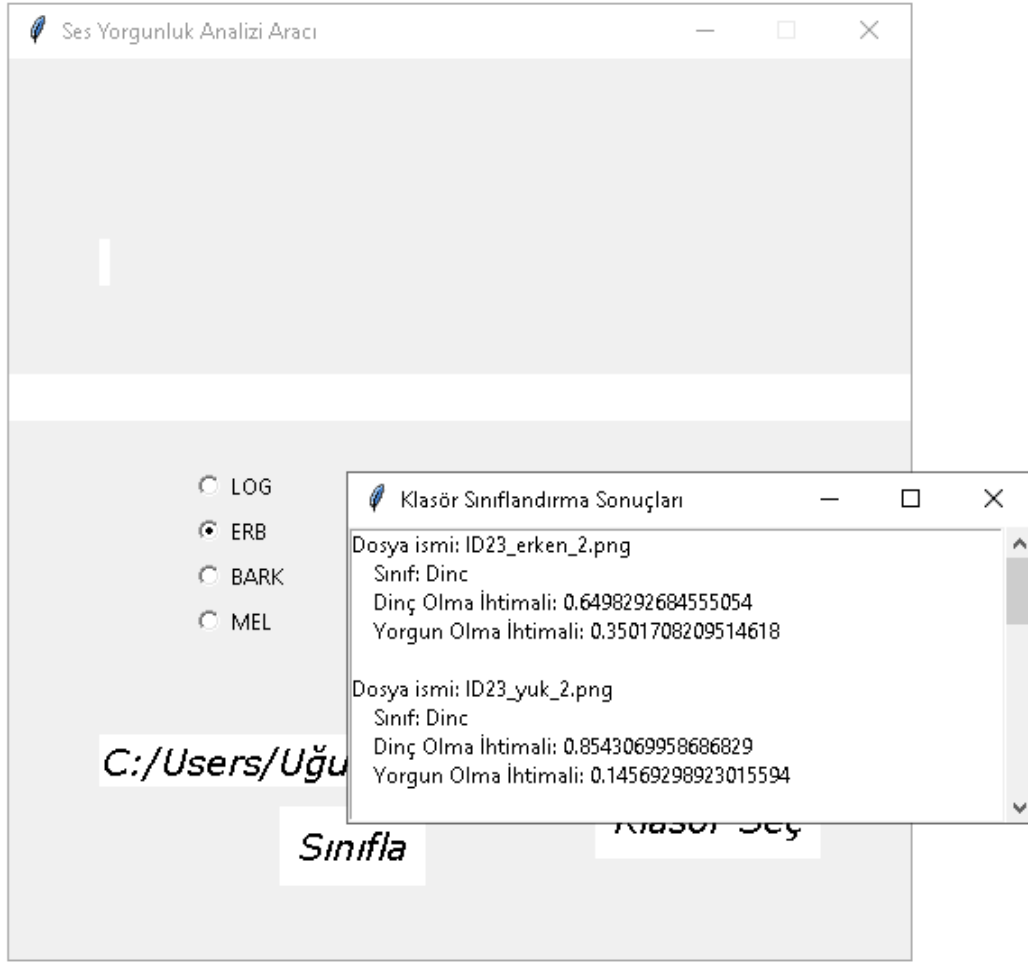
Şekil 6.3 Sınıflandırma Sonuçlarının Gösterilmesi



Şekil 6.4 Test Edilecek Klasörün Seçilmesi



Şekil 6.5 Klasör Sınıflandırma Sonuçlarının Gösterilmesi



Şekil 6.6 Klasör Sınıflandırma Sonuçlarının Gösterilmesi

Eğer kullanıcı tek bir görüntü sınıflandırmayı seçerse "Sınıfla" butonuna bastıktan sonra eğer model mevcut oturuma yüklenmemişse yüklenir ardından modelin sınıf tahmini ve bu sınıfların softmax fonksiyonundan elde edilen ihtimalleri ekranda yazdırılır. Eğer kullanıcı klasör sınıflandırmayı seçmiş ise klasördeki tüm görüntülerin sonuçları sırasıyla kaydırılabilen bir pencerede sıralanır.

7

Deneyisel Sonuçlar

Bu bölümde, geliştirilen modellerin başarı ölçümleri ve bu başarıyı sağlayan etkenlerden söz edilmiştir.

Yapılan ölçümlerde, ele alınan problemin çözümünü sağlayabilecek en iyi transfer öğrenme modelini bulmak için test edilen modeller arasında VGG16 en yüksek performansı göstermiştir ve ulaşılan bu modelin üzerinde yapılan optimizasyon denemelerinden sonra elde edilen son model farklı spektrogram tipleriyle, farklı seslerden üretilmiş görüntüler üzerinde eğitilip test edilmiştir.

Number Of Sample		Yorgunluk Seviyesi								
		1-10			1,2-9,10			1,2,3-8,9,10		
		Dinç	Yorgun	Toplam	Dinç	Yorgun	Toplam	Dinç	Yorgun	Toplam
Harf	BARK	32	87	119	131	197	328	438	396	834
	ERB	32	87	119	131	197	328	438	396	834
	LOG	32	87	119	131	197	328	438	396	834
	MEL	32	87	119	131	197	328	438	396	834
Kelime	BARK	40	112	152	176	254	430	497	428	925
	ERB	23	60	83	175	251	426	596	519	1115
	LOG	41	112	153	171	253	424	597	518	1115
	MEL	43	105	148	173	261	424	584	525	1109
Hepsi	BARK	72	199	271	307	451	758	935	824	1759
	ERB	55	147	202	306	448	754	1034	915	1949
	LOG	73	199	272	302	450	752	1035	914	1949
	MEL	75	192	267	304	458	752	1022	921	1943

Tablo 7.1 Spektrogram tipi, Ses tipi, Yorgunluk seviyesi ve sınıfa göre veri sayıları

Oluşan en son model, 1 veya 10; 1,2,9 veya 10; 1,2,3,8,9 veya 10 olarak puanlandırılmış seslerin spektrogram görüntülerini içeren verisetleri üzerinde ayrı ayrı eğitilip test edilmiş, başarı değerlendirmeleri yapılırken bu görüntüler üzerinde durulmuştur. Başarı ve Hata yüzdeleri, klasörlerin 5 foldunun ortalaması ve ulaşılmış en yüksek foldunun değeri gösterilerek ifade edilmiştir.

1 veya 10 olarak puanlanmış harf seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 96, test için 23 görüntü; 1 veya 10 olarak puanlanmış kelime seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 67, test için 16 görüntü ve bu verisetlerinin birleşimi olan verisetinde eğitim için 163, test için 39 görüntü bulunmaktadır.

1,2,9 veya 10 olarak puanlanmış harf seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 265, test için 65 görüntü; 1,2,9 veya 10 olarak puanlanmış kelime seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 594, test için 124 görüntü ve bu verisetlerinin birleşimi olan verisetinde eğitim için 757, test için 189 görüntü bulunmaktadır.

1,2,3,8,9 veya 10 olarak puanlanmış harf seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 667, test için 167 görüntü; 1,2,3,8,9 veya 10 olarak puanlanmış kelime seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 494, test için 227 görüntü ve bu verisetlerinin birleşimi olan verisetinde eğitim için 1161, test için 394 görüntü bulunmaktadır.

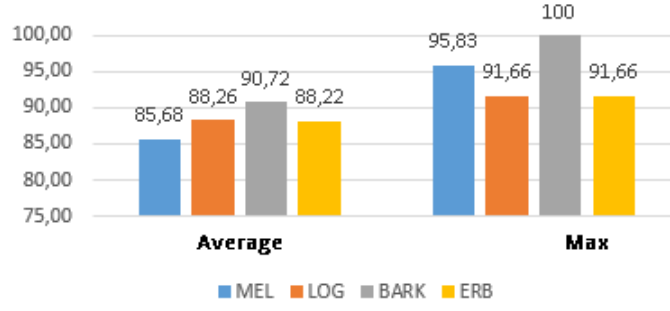
Bütün puanlanmış harf seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 1388, test için 347 görüntü; Bütün puanlanmış kelime seslerinin spektrogram görüntülerinden oluşan verisetinde eğitim için 1016, test için 254 görüntü ve bu verisetlerinin birleşimi olan verisetinde eğitim için 2404, test için 601 görüntü bulunmaktadır.

Model	VGG16
Augmentation: Width Shift Range	0.26
Augmentation: Fill Mode	Constant
Batch Size	64
Learning Rate	8e-06
Image Resize Dimension	224x224 pixel
Dropout Rate	0.2
Global Pooling Type	Max
Optimizer Function	ADAM
Last Layer Function	Softmax
Cost Function	Categorical Cross Entropy

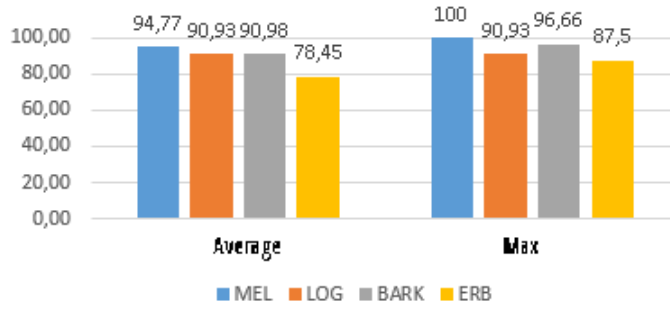
Tablo 7.2 En Yüksek Doğruluk Oranını Sağlayan Parametreler

7.2 tablosunda, optimize edilen değişkenlerin son durumları gösterilmiştir. Yapılan optimizasyonda Değiştirilen model, görüntü biçimlendirme ve çoğaltma parametreleri, katmanların parametreleri ve hiperparametreler yer almıştır.

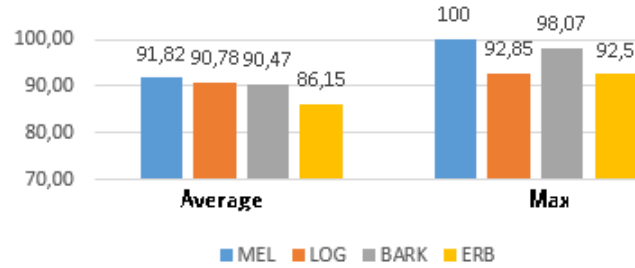
Modelin bütün puanlanmış verisetleri (Yorgun ve Dinç olup olmadığı en az belli olan veriseti) üzerindeki test sonuçları %60 ila %70 arasında bir doğruluk oranına ulaşmıştır.



Şekil 7.1 1 veya 10 olarak puanlandırılmış harf veriseti üzerindeki test sonuçları

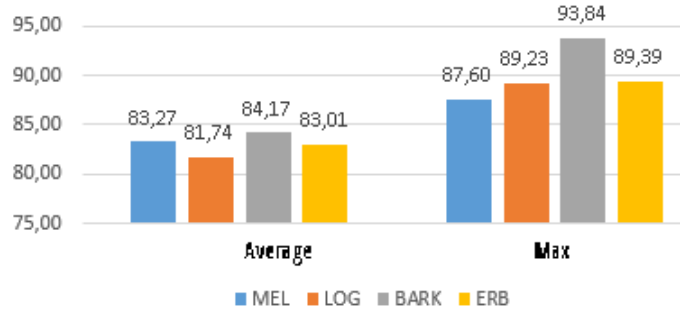


Şekil 7.2 1 veya 10 olarak puanlandırılmış kelime veriseti üzerindeki test sonuçları

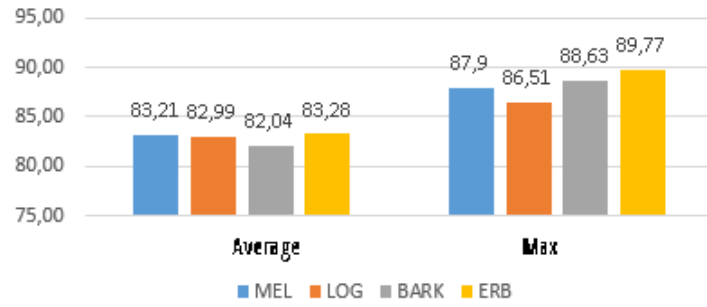


Şekil 7.3 1 veya 10 olarak puanlandırılmış harf ve kelime birleşimi veriseti üzerindeki test sonuçları

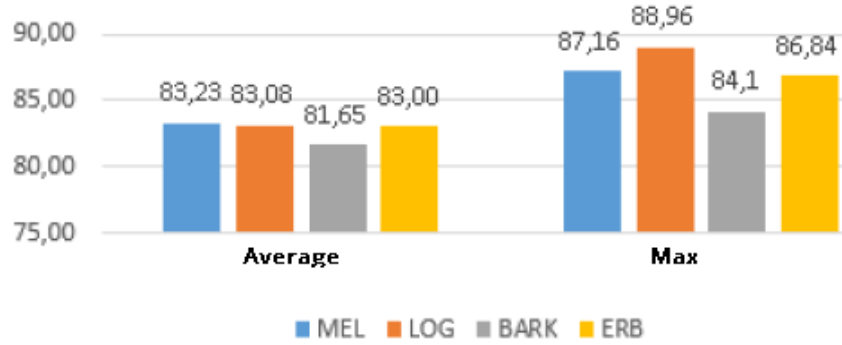
7.1, 7.2 ve 7.3 şekillerinde gösterildiği gibi, tüm klasörler arasındaki en yüksek başarıyı kelime klasörleri üzerindeki test sonuçlarındaki %94,77 ile MEL tipi spektrogram görüntüleri sağlamıştır. Ayrıca ulaşılan en yüksek fold skoru da %100 ile yine MEL tipi spektrogram görüntüleri üzerinden elde edilmiştir.



Şekil 7.4 1, 2, 9 veya 10 olarak puanlandırılmış harf veriseti üzerindeki test sonuçları

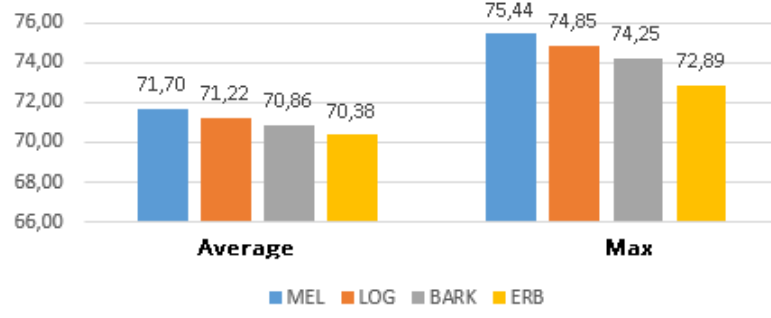


Şekil 7.5 1, 2, 9 veya 10 olarak puanlandırılmış kelime veriseti üzerindeki test sonuçları

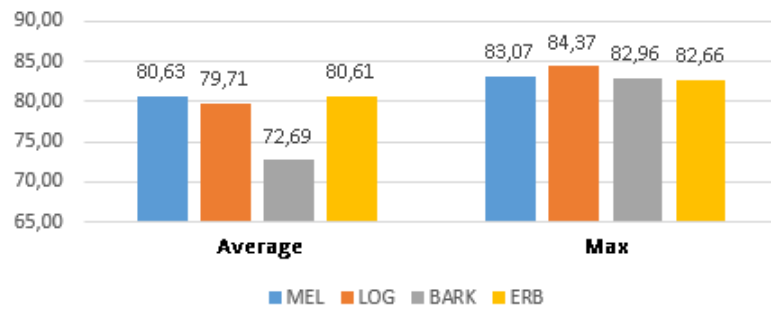


Şekil 7.6 1, 2, 9 veya 10 olarak puanlandırılmış harf ve kelime birleşimi veriseti üzerindeki test sonuçları

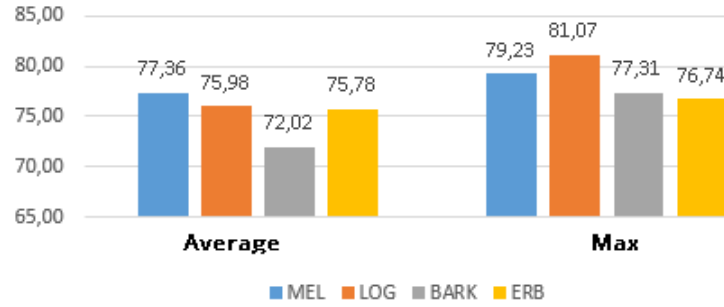
7.4, 7.5 ve 7.6 şekillerinde gösterildiği gibi, tüm klasörler arasındaki en yüksek başarıyı kelime klasörleri üzerindeki test sonuçlarındaki %84,17 ile BARK tipi spektrogram görüntüleri sağlamıştır. Ayrıca ulaşılan en yüksek fold skoru da %93,84 ile yine BARK tipi spektrogram görüntüleri üzerinden elde edilmiştir.



Şekil 7.7 1, 2, 3, 8, 9 veya 10 olarak puanlandırılmış harf veriseti üzerindeki test sonuçları

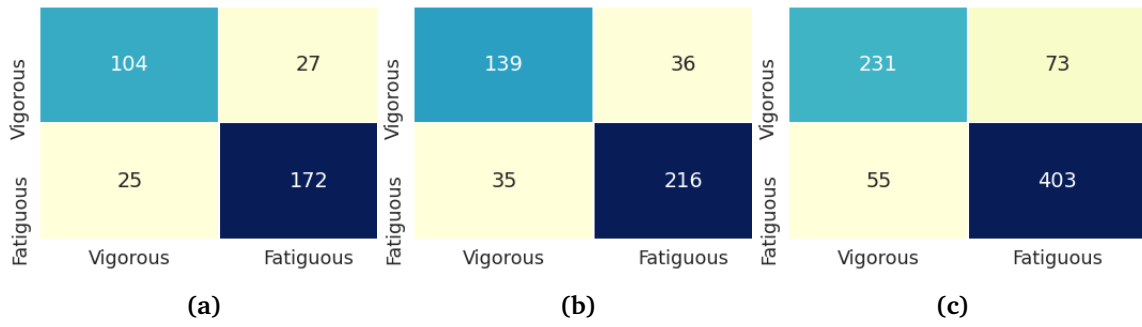


Şekil 7.8 1, 2, 3, 8, 9 veya 10 olarak puanlandırılmış kelime veriseti üzerindeki test sonuçları



Şekil 7.9 1, 2, 3, 8, 9 veya 10 olarak puanlandırılmış harf ve kelime birleşimi veriseti üzerindeki test sonuçları

7.7, 7.8 ve 7.9 şekillerinde gösterildiği gibi, tüm klasörler arasındaki en yüksek başarıyı kelime klasörleri üzerindeki test sonuçlarındaki %80,63 ile MEL tipi spektrogram görüntüleri sağlamıştır. Ayrıca ulaşılan en yüksek fold skoru da %84,37 ile LOG tipi spektrogram görüntüleri üzerinden elde edilmiştir.



Şekil 7.10 Sırasıyla 1, 2, 9 veya 10 olarak sınıflandırılmış Harf, Kelime ve Birleştirilmiş Veriseti Üzerindeki Sonuçlar

	Letter		Word		Combination	
	Vigorous	Fatigued	Vigorous	Fatigued	Vigorous	Fatigued
Precision	0.79	0.87	0.79	0.86	0.76	0.88
Recall	0.81	0.86	0.80	0.86	0.81	0.85
F-Score	0.80	0.87	0.80	0.86	0.78	0.86
Accuracy	0.841		0.833		0.832	
Kappa	0.669		0.655		0.646	

Tablo 7.3 En başarılı modellerin Şekil 7.10 den türetilen performans ölçütleri

Şekil 7.10'in (a) kısmında yaklaşık %84,84 test başarısına ulaşmış olan modelin harf seslerinden BARK spektrogram tipi ile oluşturulmuş görüntüler üzerindeki test sonucunun doğruluk matrisi gösterilmiştir. Şekil 7.10'in (b) kısmında Kelime klasörleri üzerinde en yüksek başarı olan %83,22 test başarısına ulaşmış olan modelin ERB spektrogram tipi ile oluşturulmuş görüntüler üzerindeki test sonucunun doğruluk matrisi gösterilmiştir. Şekil 7.10'in (c) kısmında Birleştirilmiş verisetleri üzerinde en yüksek başarı olan %83,23 test başarısına ulaşmış olan modelin LOG spektrogram tipi ile oluşturulmuş görüntüler üzerindeki test sonucunun doğruluk matrisi gösterilmiştir.

8 Sonuç

Yapılan projede, sesin yorgun olup olmadığını tahmin edebilmek için oluşturulan model ve optimizasyonu üzerinde durulmuştur. Projede önceden eğitilmiş çeşitli CNN modelleri üzerinde ön test yapıldıktan sonra, genel performansı yüksek olan model üzerine optimizasyon yapılarak en yüksek doğruluk oranına ulaşılmaya çalışılmıştır. Bu optimizasyona yapay sinir ağının eğitiminde kullanılan hiper parametrelerin, ağın topolojisinin değiştirilmesi ve iyileştirici fonksiyonların yanı sıra, ses verilerinin ses spektrogramına çevirilme biçimi, sesin formları, görüntünün biçimlendirilmesine kullanılan parametreler gibi verilerin düzenlenmesine ilişkin yöntemler de dahil olmuştur.

Modelin başarısını ölçmek için 5 fold mimarisine göre hazırlanmış her klasör için model o klasörün tek tek her foldunun eğitim klasörü ile eğitilip, aynı foldun test klasörü ile test edilmiştir. Klasörün başarısını kaydetmek için klasörün tüm foldlarının test değerlerinin ortalaması ve ulaşılan en yüksek değeri göz önüne alınmıştır. Test için kullanılan hata fonksiyonu CCE (Categorical Cross-Entropy) fonksiyonudur ve metrik olarak da doğruluk metriği kullanılmıştır.

Önceden imagenet ile eğitilmiş CNN modellerinin arasından üzerinde yoğunlaşılacak ve optimize edilecek modelin seçilmesinde, belirli klasörler üzerinde yapılan test sonuçları etkili olmuştur ve sonuç olarak ön testte en yüksek genel doğruluğa sahip olduğu için CNN modeli olarak VGG16'nın üzerinde yoğunlaşmıştır. Ön testte yaklaşık %73 doğruluk oranına ulaşılan VGG16, yapılan optimizasyonlarla genel olarak tüm klasörler için en az %82 doğruluk oranına ulaşılmıştır.

Sesin puanlarına göre ayrılmış 4, spektrogramların tipine göre ayrılmış 4, sesin türüne göre ayrılmış 3 çeşit klasör seçeneği ve böylece toplamda 48 veriseti üzerinde eğitim ve test yapılmıştır. Elde edilen sonuçlara göre, en yüksek doğruluk oranını %84,17 fold ortalaması ve %93,84 en yüksek fold başarısı ile BARK spektrogram tipi ve harf seslerine sahip spektrogram görüntüleri vermiştir.

- [1] P.-H. Ting, J.-R. Hwang, J.-L. Doong, and M.-C. Jeng, "Driver fatigue and highway driving: A simulator study," *Physiology & behavior*, vol. 94, no. 3, pp. 448–453, 2008.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [3] H. P. Greeley, J. Berg, E. Friets, J. Wilson, G. Greenough, J. Picone, J. rey Whitmore, and T. Nesthus, "Fatigue estimation using voice analysis," *Behavior research methods*, vol. 39, no. 3, pp. 610–619, 2007.
- [4] Z. Zhu and Q. Ji, "Real time and non-intrusive driver fatigue monitoring," in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)*, IEEE, 2004, pp. 657–662.
- [5] J. Krajewski, U. Trutschel, M. Golz, D. Sommer, and D. Edwards, "Estimating fatigue from predetermined speech samples transmitted by operator communication systems," 2009.
- [6] (), [Online]. Available: <https://github.com/keras-team/keras/blob/master/LICENSE>.
- [7] (), [Online]. Available: <https://github.com/tensorflow/tensorflow/blob/master/LICENSE>.
- [8] (), [Online]. Available: <https://github.com/skvark/opencv-python/blob/master/LICENSE.txt>.
- [9] M. Huh, P. Agrawal, and A. A. Efros, "What makes imagenet good for transfer learning?" *arXiv preprint arXiv:1608.08614*, 2016.
- [10] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," *arXiv preprint arXiv:1711.00489*, 2017.
- [11] A. Ashiquzzaman, A. K. Tushar, M. R. Islam, D. Shon, K. Im, J.-H. Park, D.-S. Lim, and J. Kim, "Reduction of overfitting in diabetes prediction using deep learning neural network," in *IT Convergence and Security 2017*, Springer, 2018, pp. 35–43.

BİRİNCİ ÜYE

İsim-Soyisim: Uğur KESKİN

Doğum Tarihi ve Yeri: 25.05.1999, Samsun

E-mail: 11117044@std.yildiz.edu.tr

Telefon: 0545 811 49 60

Staj Tecrübeleri: Oredata Şirketi Veri Bilimi Stajı

Proje Sistem Bilgileri

Sistem ve Yazılım: Windows İşletim Sistemi, Python, Google Colabratory

Gerekli RAM: 2GB

Gerekli Disk: 256MB