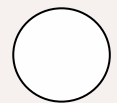




Çanakkale Onsekiz Mart
Üniversitesi,
Bilgisayar Mühendisliği Bölümü

- Prof.Dr. Bahadır Karasulu
- Betül Cengiz
- Elif Avcı
- Tesnim Strazimiri

ARNAVUTÇA KONUŞMA VERİLERİNDEN DERİN ÖĞRENME TABANLI DUYGU DURUM ANALİZİ VE SINIFLANDIRMA



ICONDATA2024 CONFERENCE, 2-6 SEPTEMBER, 2024, KOSOVA, PRISHTINA.

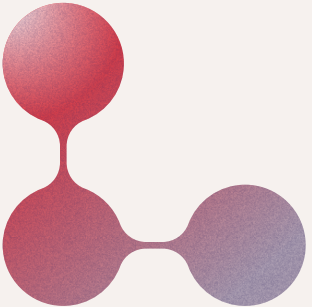
İÇİNDEKİLER

- Ses Verilerinden Duygu Sınıflandırması için Mel Spektrogramları ve Derin Öğrenme Modelleri Kullanımı
- Veri Kümesi
- Neden Arnavutça
- Yöntemler
- Metodoloji
- Model Performansını Değerlendirme
- Verilerin analizi
- Confusion Matrix-ReSNET50
- ROC Eğrisi ve AUC Değerleri-RESNET50
- Confusion Matrix-XCEPTION
- ROC Eğrisi ve AUC Değerleri-XCEPTION
- Sonuç

SES VERİLERİNDEN DUYGU SINIFLANDIRMASI İÇİN MEL SPEKTROGRAMLARI VE DERİN ÖĞRENME MODELLERİ KULLANIMI

Bu çalışmada, ses verilerinden mel spektrogramları kullanarak duygu sınıflandırması yapılması amaçlanmıştır. Duygu tanıma, insan-bilgisayar etkileşiminde önemli bir rol oynamaktadır. Mel spektrogramları, ses sinyallerinin zaman-frekans temsili olup, duygu sınıflandırma modelleri için zengin bir öznitelik kümesi sunmaktadır.

Duygusal durumların ses verileri üzerinden otomatik olarak sınıflandırılması, özellikle doğal dil işleme ve insan-bilgisayar etkileşimi alanlarında büyük önem taşımaktadır. Bu çalışmada, dört temel duygu (mutlu, üzgün, kızgın, şaşırmış) sınıflandırılmak üzere bir sistem geliştirilmiştir. Çalışmanın amacı, mel spektrogramlar ve derin öğrenme modelleri kullanarak bu duyguların doğrulukla tespit edilmesini sağlamaktır.



VERİ KÜMESİ

Çalışmamızda kullanılan veri kümesi, 9 Arnavut bireyden elde edilen ses kayıtlarından oluşmaktadır. Her bir birey, 4 farklı duygu (mutlu, üzgün, sinirli, şaşkın) durumunu temsil etmek için 4'er cümle söylemiştir. Bu ses kayıtları öncelikle mel spektrogramlara dönüştürülmüş ve veri artırma teknikleri ile eğitim kümesi genişletilmiştir. Mel spektrogramlar, ses sinyallerinin zaman-frekans temsili olup, modellerin eğitimi için zengin bir öznitelik kümesi sağlamaktadır.

Çalışmalarımızın başında RAVDESS veri kümesi kullanılarak deneyler gerçekleştirilmiş, ardından Arnavutça veri kümesi ile çalışmalarımıza devam edilmiştir. RAVDESS veri kümesi, 24 profesyonel aktör tarafından ifade edilen 8 farklı duygusal durumu içermekte olup, duygu analizi için yaygın olarak kullanılan bir veri kümesidir. Bu veri kümesi, duygusal durumların sesli ifadelerini analiz etmek için zengin bir kaynak sunmaktadır.



NEDEN ARNAVUTÇA?

Sınırlı Çalışmalar:

Arnavutça üzerinde yapılmış duygu analizi çalışmaları sınırlı olup, bu alanda önemli bir boşluk bulunmaktadır.

Dil Özellikleri:

Arnavutça, geniş çaplı veri kümeleri ve modellerle çalışılmamış bir dil olduğu için, bu araştırma, dilin kendine özgü özelliklerini anlamada ve bu özelliklere uygun modeller geliştirmede önemli katkılar sağlayabilir.

Çalışmanın Amacı:

Arnavutça konuşma verilerini kullanarak duygu durumlarını sınıflandıran bir derin öğrenme modeli geliştirmek.

Bu alanda yapılacak gelecekteki çalışmalara yol gösterebilecek nitelikte bir araştırma sunmak.



YÖNTEMLER

Bu çalışmada kullanılan yöntemler, denetimli öğrenme ve derin öğrenme tekniklerini içermektedir. Denetimli öğrenme, etiketlenmiş veri kümeleriyle modelin eğitildiği ve yeni verilere tahminler yaptığı bir yaklaşımdır. Derin öğrenme, karmaşık veri yapıları üzerinde yüksek düzeyde öznitelik öğrenimi sağlar ve ses spektrumlarının ayrıntılı analizini mümkün kılar. Bu çalışmada, ses dosyaları duygularla etiketlenmiş ve ResNet50 , Xception gibi derin öğrenme tabanlı modellerle işlenmiştir.

Her iki model, etiketlenmiş verilerle eğitim almış ve yüksek doğruluk oranlarıyla performans sergilemiştir.

METODOLOJİ

Yazılım Kütüphaneleri ve Araçlar

- Librosa
- Matplotlib
- PIL (Python Imaging Library)
- TensorFlow ve Keras
- Scikit-learn
- Google Colab

Veri Hazırlama

- Dosya Adlarının Temizlenmesi ve Etiketlenmesi.
- Mel Spektrogramlarının Oluşturulması

Model Oluşturma

- Önceden eğitilmiş ResNet50 ve Xception modelleri kullanıldı.
- Bu modellerin üzerine ek katmanlar eklenerek özel sınıflandırma modelleri oluşturuldu.

Model Eğitimi

- ResNet50: 40 eğitim adımı boyunca eğitildi, ADAM optimizasyon algoritması kullanıldı.
- Xception: 70 eğitim adımı boyunca eğitildi, ADAM optimizasyon algoritması kullanıldı.



MODEL PERFORMANSINI DEĞERLENDİRME

01. Amaç Fonksiyonu Ölçümleri

Eğitim ve test veri kümeleri üzerindeki doğruluk ve kayıp değerleri.

02. Duyarlılık (Precision), Anma (Recall) ve F1 Skoru

Modelin genel performansını değerlendirmek için kullanıldı.

03. Çapraz Tahmin Tablosu (Confusion Matrix):

Sınıflandırma performansını görselleştirmek için kullanıldı.

04. ROC Eğrisi ve AUC:

Her sınıf için ROC eğrisi çizildi ve AUC değerleri hesaplandı.

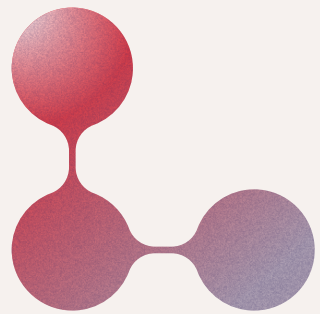


VERİLERİN ANALİZİ

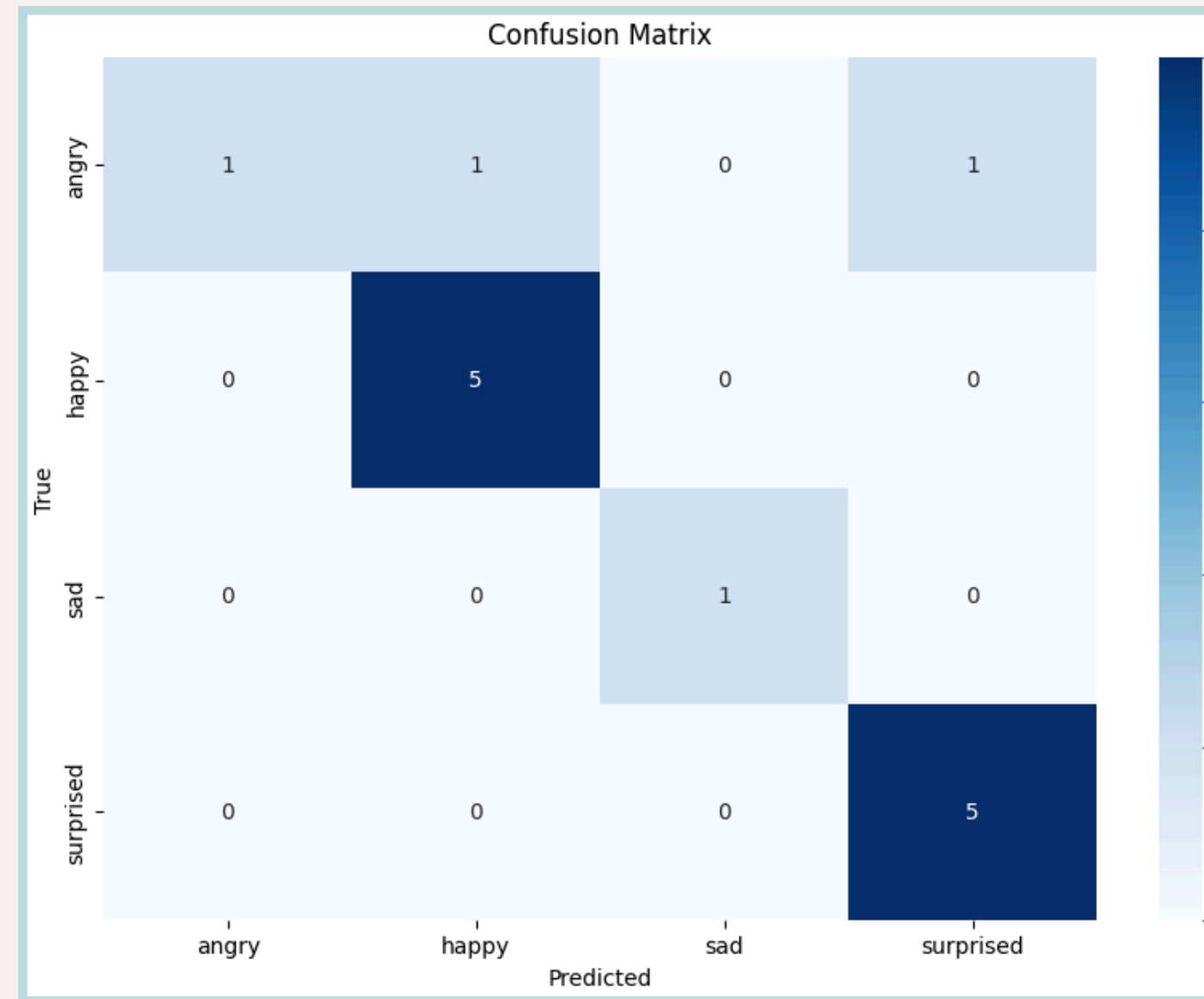
Bu çalışmada, model performansını değerlendirmek için Confusion Matrix ve ROC Eğrisi analizleri yapılmıştır. Analizlerin sonuçları, modelin genel başarı performansını ve sınıf bazında ayırım yeteneğini açık bir şekilde ortaya koymaktadır.

Confusion Matrix, modelin doğru ve yanlış sınıflandırmalarını görselleştirir. Sınıflar arasında doğruluk, hata oranı ve hatalı sınıflandırmaların dağılımını gösterir.

ROC eğrisi ve AUC (Area Under Curve) değerleri, modelin her bir sınıfta ayırım yeteneğini ölçer.



CONFUSION MATRIX-RESNET50

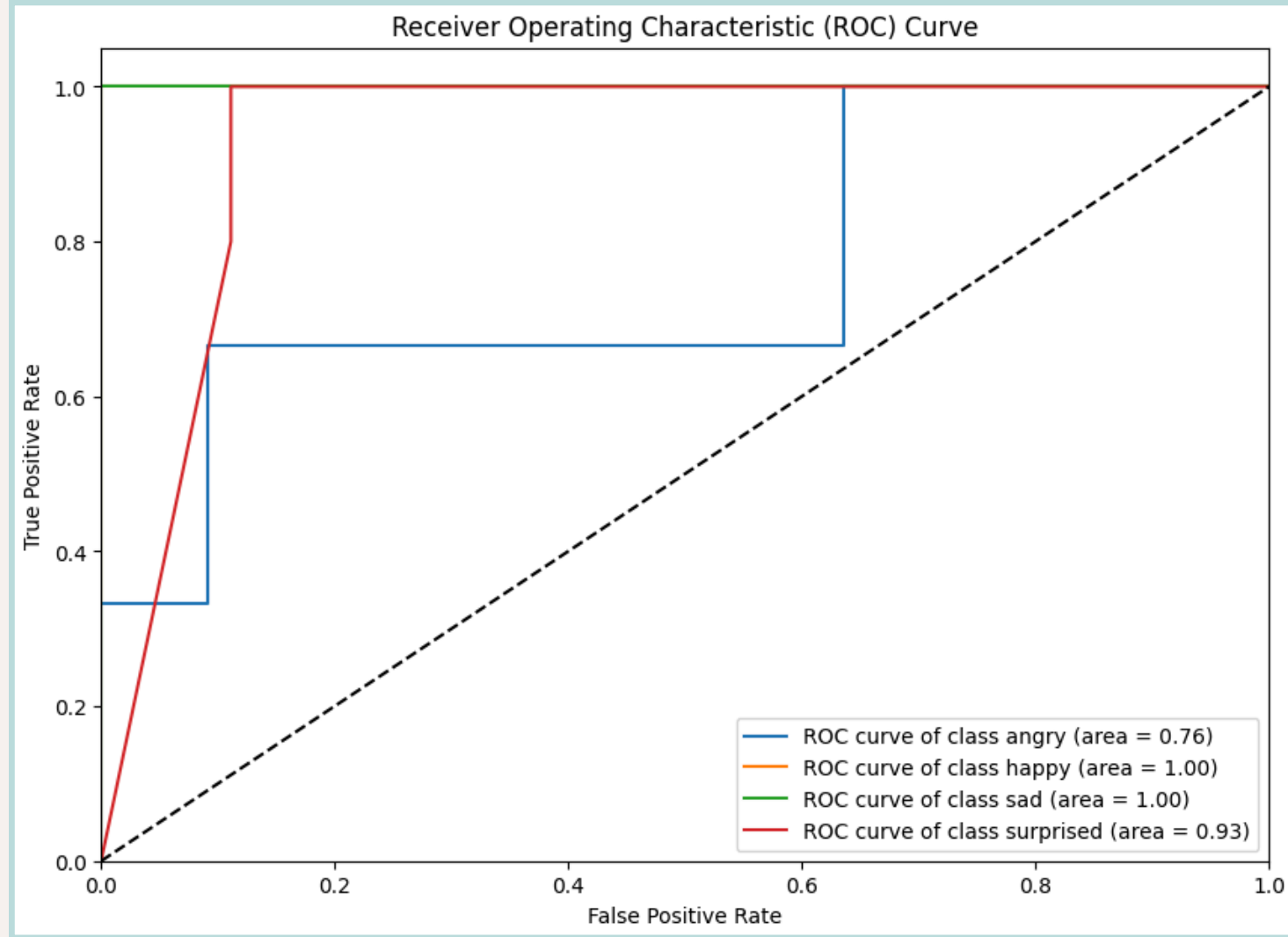


- **Angry:** Model, bu sınıfta %33 recall ile daha düşük performans sergilemiştir. Bu durum, sınıfın ayırımında daha fazla iyileştirme gerektirdiğini göstermektedir.
- **Happy:** Bu sınıfta %100 recall ve %83 precision ile neredeyse mükemmel bir performans elde edilmiştir.
- **Sad:** Model, bu sınıfta %100 recall ve %100 f1-score ile hatasız tahminler yapmıştır.
- **Surprised:** %100 recall ve %83 precision ile yüksek performans gösterilmiştir.

Genel doğruluk oranı %85 olarak belirlenmiştir. Macro ve weighted average metrikleri, modelin genel performansını ve çeşitli sınıflardaki dengeyi yansıtmaktadır.

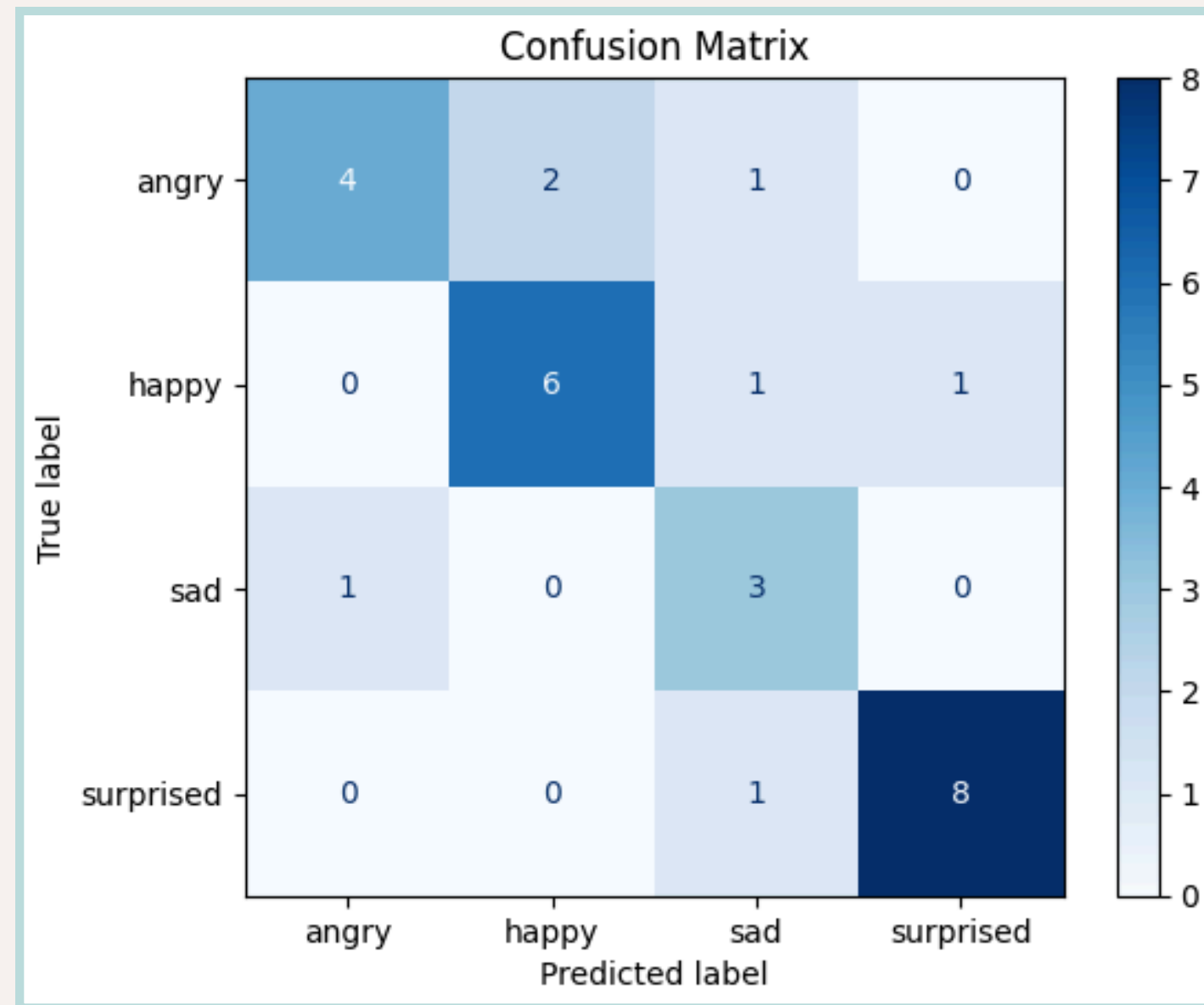


ROC EĞRİSİ VE AUC DEĞERLERİ-RESNET50



- **Angry:** AUC değeri 0.76 olarak bulunmuştur. Bu, modelin bu sınıfta orta düzeyde bir ayırım yeteneğine sahip olduğunu ve iyileştirmeye açık olduğunu göstermektedir.
- **Happy:** AUC değeri 1.00 ile mükemmel bir ayırım yeteneği sağlanmıştır, model bu sınıfta neredeyse hatasız tahminler yapabilmıştır.
- **Sad:** AUC değeri 1.00 ile mükemmel performans sergilenmiştir, model bu sınıfta hatasız tahminler yapabilmıştır.
- **Surprised:** AUC değeri 0.93 olarak belirlenmiştir, model bu sınıfta yüksek ayırım yeteneğine sahip olduğu gözlemlenmiştir.

CONFUSION MATRIX-XCEPTION

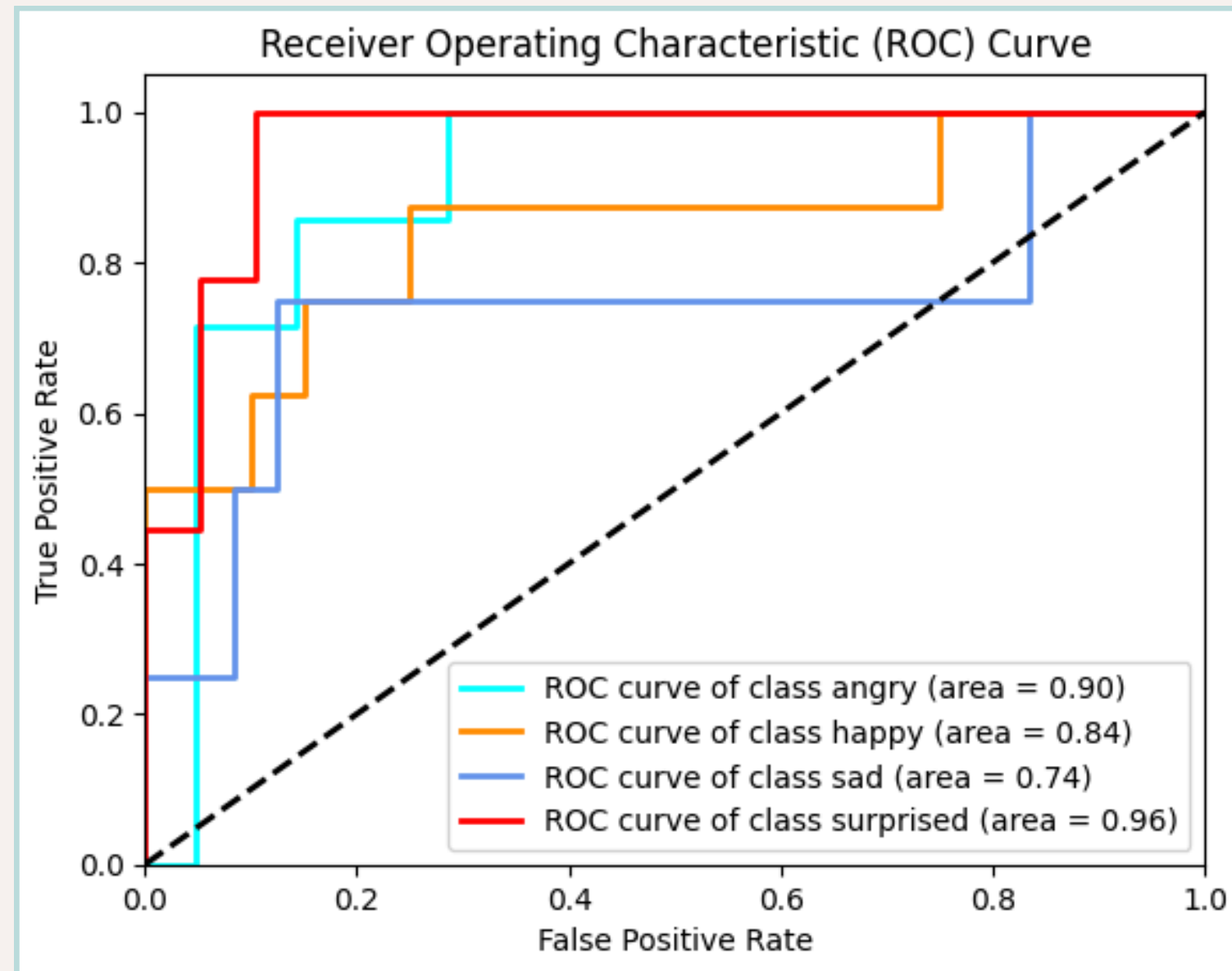


- **Angry:** Model, bu sınıfta %57 recall ve %67 F1-score ile ortalama bir performans göstermiştir, bu da öfkeli durumların tespiti konusunda iyileştirme yapılması gerektiğini gösteriyor.
- **Happy:** Model %75 recall ve %75 precision ile model bu sınıfta dengeli ve iyi bir performans sergilemiştir.
- **Sad:** Model, %75 recall,%60 F1-score ve %50 precision ile bu sınıfta daha dikkatli tahminler yapmıştır, ancak precision değerinin düşük olması, bu kategoride bazı yanlış pozitif sonuçlar olduğunu gösterir.
- **Surprised:** %89 recall ve %89 precision ile model bu sınıfta çok başarılı bir performans göstermiştir.

Genel doğruluk oranı %75 olarak belirlenmiştir. Macro avg ve Weighted avg metrikleri ise, modelin bazı sınıflarda daha iyi performans gösterdiğini, ancak genel olarak başarılı olduğunu gösterir.



ROC EĞRİSİ VE AUC DEĞERLERİ-XCEPTION



- **Angry:** AUC değeri 0.90 olarak bulunmuştur. Bu, modelin bu sınıfta oldukça iyi düzeyde bir ayırım yeteneğine sahip olduğunu ve bu sınıfta fazla hata yapmadığını göstermektedir.
- **Happy:** AUC değeri 0.84 ile modelin bu sınıfta da oldukça iyi düzeyde bir ayırım yeteneğine sahip olduğunu ve bu sınıfta fazla hata yapmadığını göstermektedir.
- **Sad:** AUC değeri 0.74 ile modelin bu sınıfta çok iyi bir performans gösterdiğini söyleyemeyiz.
- **Surprised:** AUC değeri 0.96 olarak belirlenmiştir. Bu, modelin bu sınıfta çok iyi bir ayırım yeteneğine sahip olduğunu belirtmektedir.



SONUÇ

Bu çalışmada, mel spektrogramlarını kullanarak ses verilerinden duygu tanımlamaya yönelik derin öğrenme modelleri başarıyla geliştirilmiş ve değerlendirilmiştir. Xception ve ResNet-50 gibi ileri seviye görüntü sınıflandırma modelleri kullanılarak duygusal durumların sınıflandırılması hedeflenmiş ve bu modellerin etkili bir şekilde çalıştığı görülmüştür.

Eğitim sürecinde, her iki modelin de doğruluk oranları hızlı bir artış göstermiş ve eğitim doğruluğu %100 seviyesine ulaşmıştır. Eğitim kaybı neredeyse sıfıra inmiştir. Doğrulama doğruluğu kademeli olarak artmış; Xception modeli için %75, ResNet-50 modeli için ise %85 seviyelerine ulaşmıştır. Doğrulama kaybı ise belirli bir seviyede sabit kalmıştır. Bu sonuçlar, modellerin mel spektrogramlarını kullanarak duygu tanımlamada başarılı olduğunu göstermektedir.

Bu bakış açısıyla daha büyük ve çeşitli veri kümelerinin kullanılması, model üstün parametrelerinin (hyper parameters) eniyilemesi (optimization) ve farklı derin öğrenme mimarilerinin denenmesi gibi konular ileriki çalışmalar için planlanmaktadır.



TEŞEKKÜRLER...

İletişim:

bahadirkarasulu@comu.edu.tr

elffavci@gmail.com - <https://github.com/elfavci>

tesnimastrazimiri@gmail.com - <https://github.com/Tesnimm>

cengizzbetul@gmail.com - <https://github.com/BetulCengiz>

***Proje için github adreslerimizi
ziyaret edebilirsiniz.***