VERİ BİLİMİ DERGİSİ
www.dergipark.gov.tr/veri

Arnavutça Konuşma Verilerini Kullanan Derin Öğrenme Tabanlı Duygu Durum Analizi ve Sınıflandırma

Bahadır KARASULU[1]*, Elif AVCI[1], Tesnima STRAZİMİRİ[1], Betül CENGİZ[1]

[1]*Çanakkale Onsekiz Mart Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Çanakkale, TÜRKİYE*

Makale Bilgisi

Başvuru: 04/10/2024
Kabul: 08/11/2024

Özet

Günümüzde konuşma veya ses verilerinden konuşmacının duygu durumunu analiz edebilen derin öğrenme tabanlı yazılımlar, etkileşimli sesli çağrı yanıtlama sistemlerinin geliştirilmesinde önem kazanmıştır. Duygu analizi, metin, ses ve video gibi veri türlerinden duygusal bilgilerin otomatik olarak elde edilmesini amaçlayan bir araştırma alanıdır. Literatürde Arnavutça konuşma verileri üzerinde duygu analizi üzerine yapılan çalışmaların oldukça sınırlı olduğu görülmektedir. Bu çalışmada, Arnavutça konuşmaların duygu analizini gerçekleştirmek amacıyla, dört farklı duygu sınıfını (angry, happy, sad, surprised) içeren özgün bir veri kümesi oluşturulmuştur. Veri kümesi, ana dili Arnavutça olan dokuz farklı bireyden özel olarak elde edilmiştir. Ses verilerinin spektral ve duyusal analizi için, görüntü tanıma alanında başarılı olan ResNet50 ve Xception modellerinin yanı sıra, klasik bir derin öğrenme modeli olan Evrişimli Sinir Ağı (ESA) birlikte kullanılmıştır. Bu derin öğrenme tabanlı yaklaşım, Arnavutça konuşma verileri için duygu analizi alanına katkı sunmakta ve dilin kendine özgü özelliklerine göre bir analiz imkanı sağlamaktadır. Deneysel sonuçlarda, alıcı işletim karakteristik (AİK) eğrisi altında kalan alan (EAKA) değerleri; angry (öfkeli) sınıfı için 0.76, happy (mutlu) sınıfı için 1.00, sad (üzgün) sınıfı için 1.00 ve surprised (şaşkın) sınıfı için 0.93 olarak elde edilmiştir. Bu çalışma, Arnavutça konuşma verilerinde derin öğrenme tabanlı duygu analizi alanında sınırlı sayıda bulunan çalışmalara katkı sunmakta ve oluşturulan veri kümesi ile bu alandaki gelecekteki araştırmalara referans teşkil edebilmektedir.

***Anahtar Kelimeler**: Derin öğrenme, Duygu durumu, Evrişimli sinir ağı*

Deep Learning Based Emotional State Analysis and Classification Using Albanian Speech Data

Abstract

Nowadays, deep learning-based software that can analyze the speaker's emotional state from speech or audio data has gained importance in the development of interactive voice call answering systems. Sentiment analysis is a research area that aims to automatically extract emotional information from data types such as text, audio and video. In the literature, studies on sentiment analysis on Albanian speech data are quite limited. In this study, a unique dataset containing four different emotion classes (angry, happy, sad, surprised) was created to perform sentiment analysis of Albanian speech. The dataset was specially

[1]1*İletişim e-posta: bahadirkarasulu@comu.edu.tr

**Bu çalışmanın bir kısmı 6th International Conference on Data Science and Applications 2024'de sözlü olarak sunulmuştur.

obtained from nine different native Albanian speakers. For the spectral and sensory analysis of the audio data, a classical deep learning model, the Convolutional Neural Network (CNN), was used in combination with the successful ResNet50 and Xception models for image recognition. This deep learning-based approach contributes to the field of sentiment analysis for Albanian speech data and provides an analysis based on the specific characteristics of the language. In the experimental results, the area under the receiver operating characteristic curve (AUC-ROC) values were 0.76 for the angry class, 1.00 for the happy class, 1.00 for the sad class and 0.93 for the confused class. This study contributes to the limited number of studies in the field of deep learning-based sentiment analysis on Albanian speech data.

Keywords *: Deep Learning, Emotional State, Convolutional Neural Network*

1. Giriş

Ses, insanların hislerini ve ifadelerini yansıtan en önemli iletişim yolu olarak kabul edilmektedir. Bu çalışmada; ses verilerinin hem spektral hem de duygusal açıdan daha ayrıntılı bir analizi yapılarak, bu alandaki derin öğrenme modelleri incelenmiştir [1, 2]. Ses verileri üzerinde yapılan analizler bir yandan Mel spektrogramlarına dönüştürme işlemi yapılırken, bu sayede ses verilerini etkili bir şekilde zaman frekans düzleminde görselleştirilmesine olanak sağlanmıştır.

Duygu durum analizi de aynı ses verileri üzerinde yapılmış olup, duygu (sentiment) analizi sonuçlarını sağlıklı bir şekilde elde etmek için Değerlik Farkındalığında Sözlük ve Duygu Muhakemesi (Valence Aware Dictionary and Sentiment Reasoner, VADER), TextBlob ve Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (Bidirectional Encoder Representations from Transformers, BERT) gibi araçlar kullanılmıştır [3].

Arnavutça gibi düşük kaynaklı dillerin duygu durum analizine odaklanmanın çeşitli önemli gerekçeleri bulunmaktadır. Doğal Dil İşleme (Natural Language Processing, DDİ-NLP) ve duygu analizi çalışmalarının büyük çoğunluğu, genellikle İngilizce gibi yüksek kaynaklı diller üzerinde yoğunlaşmakta, bu durum ise düşük kaynaklı dillerin ve bu dillerdeki kullanıcıların yeterince temsil edilmediği bir eksiklik yaratmaktadır.

Arnavutça özelinde gerçekleştirilen bu çalışma, hem Arnavutça diline yönelik duygu analizi algoritmalarının geliştirilmesine katkı sağlamakta, hem de bu gibi düşük kaynaklı dillerin analizinin DDİ alanında nasıl genişletilebileceğine dair bilgi sunmaktadır. Dilin kendine özgü tonlamaları ve ifade biçimleri, duygu sınıflandırma başarımı açısından dilin zorluklarını ve fırsatlarını inceleme fırsatı sağlamaktadır.

Bu çalışmada kullanılan veri kümesiyle hem Mel spektrogramları'ndan hem de ses verisinden elde edilen duygu durum analizlerinin birleştirilmesiyle, bir derin öğrenme modeli olan Evrişimli Sinir Ağı (Convolutional Neural Network, ESA-CNN) eğitimi uygulanmaktadır [3]. Bunun sonucunda, ses verilerinin spektral yapısı ve duygusal tonlamaları analiz edilebilmektedir.

Günümüzde, duygu temelli metinden ses sentezi, insan ile makine etkileşimini daha etkili hale getirmenin yanı sıra iletişim cihazlarından kullanıcı deneyimini zenginleştirmeyi amaçlayan bir araştırma alanı olarak popülerdir.

Bu çalışmada, metin tabanlı duygu durumu analizi ile ses işleme uygulamaları metin tabanlı duygu analizi yoluyla sentezlenmiş ses sinyalleri kullanarak derin sinir ağları, özellikle Üretken Çekişmeli Ağ (Generative Adversarial Networks, ÜÇA-GAN) gibi güncel derin öğrenme tekniklerinden biriyle metinden ses, senkron vizyon ve ses sentezi dahil olmak üzere bir dizi daha duyarlı ve çağdaş uygulamalar için temel oluşturmak hedeflenmektedir.

Ayrıca, duygu analizi kullanma yaklaşımı, sentezlenmiş ses dosyalarında duygusal kaliteyi artırırken sonuçlarda bir iyileşme sağlamaktadır [4, 5]. Bu çalışma, insan sesinin karmaşıklığını anlama ve gelecekteki çalışmalar için bir temel oluşturma potansiyeline sahip ses analizi ve duygu tanıma fikirlerine dayanan çözümleri hedeflemektedir.

Bu makale beş bölümden oluşmaktadır. İkinci bölümde literatürdeki çalışmalara ve ilgili bilimsel çalışma alanına genel bir bakış sunulmaktadır. Üçüncü bölümde kullanılan metot ve materyallere değinilmektedir. Dördüncü bölümde çalışmadaki deneyler sonucu elde edilen bulgulara yer verilmiştir. Beşinci bölümde ise bilimsel

değerlendirme ve tartışma ışığında varılan sonuçlar sunulmaktadır.

2. Literatür İncelemesi ve Genel Bakış

Duygu analizi, metin, ses ve video gibi veri türlerinden duygusal bilgilerin otomatik olarak elde edilmesini amaçlayan bir araştırma alanıdır. Doğal Dil İşleme (Natural Language Processing, DDİ-NLP), hesaplamalı dilbilim ve metin madenciliği gibi disiplinlerle yakından ilişkilidir. Duygu analizi, müşteri yorumları, sosyal medya gönderileri ve kullanıcı geri bildirimleri gibi çeşitli alanlarda kullanılmaktadır.

Literatürde duygu analizi için çeşitli yöntemler kullanılmaktadır. Bunlara yakından bakacak olursak:

- **Denetimli Öğrenme (Supervised Learning) :** Bu yöntemde, etiketlenmiş veri kümeleri kullanılarak duygu sınıflandırma modelleri eğitilir. Destek Vektör Makineleri (Support Vector Machine, DVM-SVM), Saf Bayes (Naive Bayes, SB-NB) ve derin öğrenme modelleri gibi algoritmalar yaygın olarak kullanılmaktadır. Bu teknikler, belirli veri kümelerinde yüksek doğruluk oranları elde etmek için uygundur [1].
- **Denetimsiz Öğrenme (Unsupervised Learning):** Etiketlenmemiş veriler üzerinde duygu analizi yapmak için kullanılır. Kümeleme teknikleri ve duygu sözlükleri gibi araçları içerir. Duygu sözlükleri, belirli kelimelerin veya ifadelerin duygusal anlamlarını belirlemek için kullanılır. Bu yöntem, özellikle geniş ve çeşitli veri kümelerinde kullanışlıdır [1].
- **Büyük Dil Modelleri (Large Language Models, LLM):** GPT-3, BERT ve diğer büyük dil modelleri, duygu analizi gibi görevlerde üstün başarım göstermiştir. Bu modeller, geniş veri kümelerinde eğitilerek bağlam içindeki duyguları daha doğru bir şekilde sınıflandırabilir. Bu modeller, sıfır örnekli (zero-shot) ve az örnekli (few-shot) öğrenme yetenekleri ile dikkat çekmektedir [3].

Duygu analizi çalışmalarında kullanılan bazı yaygın veri kümeleri şunlardır:

- **IEMOCAP:** İngilizce konuşmaların yer aldığı, duygusal etiketlenmiş büyük bir veri kümesidir [3].
- **EMO-DB:** Alman konuşma verilerini içeren, çeşitli duygusal ifadelerle etiketlenmiş bir veri kümesidir [4].
- **RAVDESS:** İngilizce konuşma ve şarkı verilerini içeren (Ryerson Audio-Visual Database of Emotional Speech and Song), duygusal ifadelerle etiketlenmiş bir veri kümesidir [5].

Arnavutça üzerinde yapılan duygu analizi çalışmaları, genellikle küçük ölçekli veri kümeleri kullanarak geleneksel makine öğrenimi yöntemlerine odaklanmıştır. Örneğin, Besjana Muraku ve Lu Xiao, Arnavutça sosyal medya verileri üzerinde Saf Bayes (Naive Bayes, SB-NB), Destek Vektör Makineleri (Support Vector Machine, DVM-SVM), Lojistik Regresyon, Karar Ağaçları ve Rastgele Orman algoritmalarını kullanarak sahte haber tespiti yapmışlardır [6].

Bir diğer çalışma, Erion Çano tarafından yapılmış olup, Arnavutça film eleştirileri üzerinde duygu analizi gerçekleştirmiştir. İlgili çalışmada AlbMoRe veri derlem kümesini (corpus) kullanmıştır. AlbMoRe, CSV biçeminde 800 kayıttan oluşan, Arnavutça film incelemelerini barındırmaktadır. Her kayıt, IMDb'den alınan ve yazar tarafından Arnavutçaya çevrilen bir metin incelemesini içermektedir. Ayrıca yazar tarafından eklenen 0 (negatif) veya 1 (pozitif) etiketini de içerir. Farklı türden 67 film hakkında 400 olumlu ve 400 olumsuz yorumdan oluşan derlem kümesi tamamen dengelidir. AlbMoRe derlem kümesi CC-BY lisansı altında yayımlanmıştır. DVM, Lojistik Regresyon, Karar Ağaçları ve Rastgele Orman algoritmaları ile yapılan deneylerde, DVM en yüksek doğruluğu %92.5 oranıyla sağlamıştır [7].

Arnavutça Sosyal Medyada Duygu Analizi Teknikleri başlıklı bir makalede [8], Arnavutça sosyal medya metinlerine yönelik duygu analizi çalışmalarını kapsamlı bir şekilde incelemektedir. Çalışmada, Karar Ağaçları, DVM ve Yapay Sinir Ağları gibi çeşitli sınıflandırıcılar kullanılarak deneyler gerçekleştirilmiştir. Bu bağlamda, çalışma, belge, cümle ve varlık düzeylerinde duyguları analiz etmek için kullanılan farklı yaklaşımların (sözlük tabanlı, makine öğrenmesi ve melez yöntemler dahil) etkinliğini ortaya koymaktadır. Araştırmanın bulguları, tüm kriterlere göre en iyi modelin, doğruluk oranı %79.2, F-ölçütü oranı %87.8, dengeli doğruluk

oranı %87.2 ve en yüksek Matthews Korelasyon Katsayısı (Matthews Correlation Coefficient, MKK-MCC) oranı 0,617 ile Uzun Kısa Süreli Bellek (Long Short-Term Memory, UKSB-LSTM) tabanlı Tekrarlayan Sinir Ağı (Recurrent Neural Network, TSA-RNN) olduğunu göstermektedir. UKSB tabanlı TSA'nın başarılı olmasının nedeni, zaman serisi verisi ile etkili bir şekilde çalışarak kelimelerin bağlamını ve uzun dönem bağımlılıklarını öğrenebilmesi, bu sayede metinlerdeki duygusal anlamı daha iyi analiz edebilmesidir. Bu tür tekniklerin Arnavutça gibi daha az çalışılmış dillerdeki uygulamaları, dilin kendine özgü özelliklerini dikkate alarak, duygu analizi alanına önemli katkılar sağlamaktadır [8].

Bu çalışmalar, Arnavutça üzerinde yapılan duygu analizi araştırmalarının başlangıç aşamasında olduğunu ve derin öğrenme yöntemlerinin bu alanda daha fazla kullanılmasının gerekli olduğunu göstermektedir. Literatürde, Arnavutça konuşma verileri ile derin öğrenme tabanlı duygu analizi üzerine yapılan sınırlı sayıda çalışma bulunmaktadır, bu da alanda önemli bir boşluk olduğunu ortaya koymaktadır. Arnavutça üzerinde yapılan duygu analizi çalışmaları genellikle küçük ölçekli veri kümeleri kullanmış ve geleneksel makine öğrenimi yöntemlerine odaklanmıştır. Ancak, derin öğrenme yöntemlerinin uygulanması konusunda sınırlı sayıda çalışma bulunmaktadır.

3. Materyal ve Metot

Çalışmamızda kullanılan veri kümesinin ve kullanılan metotların detayları aşağıda verilmiştir:

- **Ses Dosyaları:** Bu çalışma, Arnavutça olarak 9 farklı kişinin konuşmasından elde edilen belirli duygusal durumları temsil eden etiketli ses dosyaları (Örnekleme frekansı 48 kHz, 32 bit'lik stereo MPEG4 AAC kodlama ile elde edilerek, ses işleme için WAV dosyasına dönüştürülmüş) kullanılarak gerçekleştirilmiştir. Etiketler, mutlu (happy), üzgün (sad), sinirli (angry) ve şaşkın (surprised) duygularını temsil etmektedir.

Bunun yanı sıra, çalışmamızdaki altyapının oluşturulmasında kullanılan yazılım kütüphaneleri ve araçlar aşağıdaki gibidir:

- **Librosa:** Ses dosyalarını yüklemek ve Mel spektrogramlar oluşturmak için kullanılmıştır [9].
- **Matplotlib:** Zamana bağımlı ses verilerinden elde edilen frekans uzayındaki spektrogramların görselleştirilmesi ve çizdirilmesi için kullanılmıştır [10].
- **PIL (Python Imaging Library):** Görselleri işlemek için kullanılmıştır [11].
- **TensorFlow ve Keras:** Derin öğrenme modeli oluşturmak ve eğitmek için kullanılmıştır [12].
- **Skimage:** Görselleri yeniden boyutlandırmak için kullanılmıştır [13].
- **Scikit-learn:** Veri kümesini eğitim ve test amacıyla parçalara ayırmak ve etiketleri işlemek için kullanılmıştır [13].
- **Google Colab:** Kodların çalıştırıldığı platform olarak kullanılmıştır [14].

BERT modeli ile çalışılarak İngilizce veri kümesi üzerinde eğitilerek denenmiş, bu süreçte model, veri kümesindeki duygu durumlarını metin bazlı olarak tespit etmiş ve TextBlob özelliği eklenerek verilen cümleleri İngilizceye çevirip duygu durumlarını tahmin etmiştir. Böylece çalışmamızda Arnavutça verileri kullanarak duyguyu analiz etme amacıyla araştırmaya katkısı incelenmiştir [15].

Çalışmamızda veri hazırlama aşamasında dosya adlarının temizlenmesi ve etiketlenmesi için ses dosyalarının dosya adlarından duygu etiketleri çıkarılmış ve ilgili klasörlere kopyalanmıştır. Bu işlem programatik olarak, "*clean_and_label_files*" fonksiyonu kullanımıyla gerçekleştirilmiştir. Fonksiyon, dosya adlarındaki duygu kodlarını çıkararak uygun klasörlere yerleştirmiştir.

Klasörlere yerleştirilen ses dosyaları, Mel spektrogramlara dönüştürülmüştür. Bunun için kullanılan "*create_spectrogram*" fonksiyonu sayesinde ses dosyalarını yükleyerek Mel spektrogramlar görüntü formatında kaydedilmiştir. Ayrıca, "*create_pngs_from_wavs*" fonksiyonu ise belirli bir klasördeki tüm ses dosyalarını Mel spektrogram görüntülerine dönüştürmüştür. Bunları aşağıdaki detaylı olarak verecek olursak;

- **Veri yükleme ve ön işleme çalışmaları amacıyla;** "*load_images_from_path*" fonksiyonu,

oluşturulan Mel spektrogramlarını yüklemiş ve etiketlemiştir. Yüklenen Mel spektrogramlar ve etiketler ön işlem amacıyla Numpy biçemindeki dizilere dönüştürülmüş ve bu sayede yeniden boyutlandırılmıştır.

- **Model oluşturma ve eğitme aşamasında;** Öğrenim aktarımı (Transfer learning) ile çalışmalar sürdürülmüştür. Önceden eğitilmiş ResNet50 ve Xception modelleri kullanılmıştır [16, 17]. ResNet50, derin öğrenme alanında özellikle görüntü tanıma görevlerinde kullanılan güçlü bir yapay sinir ağıdır. Derin ağların karşılaştığı gradyan kaybolma sorununu çözmek için tasarlanmıştır. Artık (Residual) bağlantılar sayesinde bilgi akışı kolaylaşır ve daha derin katmanlara ulaşır. Bu ağın 50 katmana sahip olması, karmaşık görsel öznitelikleri öğrenme yeteneğini artırır. Darboğaz blokları ise hesaplama maliyetini düşürür. Bu yapı, modelin hem eğitim süresini kısaltmakta hem de daha yüksek doğruluk oranlarına ulaşmasına olanak tanımaktadır. ResNet50'nin bu özellikleri, ses spektral analizinde de etkili bir şekilde kullanılabilir. Ses verilerindeki karmaşık ve yüksek boyutlu öznitelikleri elde etme yeteneğine sahiptir. Xception, görüntü işleme ve derin öğrenme alanında kullanılan bir evrişimli sinir ağı (ESA-CNN) mimarisidir ve "*Extreme Inception*" ifadesinin kısaltmasıdır. Google'ın Inception mimarisinin bir uzantısı olarak François Chollet tarafından geliştirilmiştir. Xception, derin ayrıştırma evrişimlerini kullanarak her bir girdi kanalı için ayrı bir evrişim uygular. Bu sayede hesaplama maliyetini ve modelin parametre sayısını azaltır. Bu yapı, modelin karmaşık öznitelikleri öğrenme kapasitesini artırırken, görüntü sınıflandırma ve nesne tanıma gibi görevlerde yüksek başarım sağlar. Xception, özellikle ImageNet veri kümesi üzerinde yüksek doğrulukta başarım göstererek derin öğrenme modellerinin verimliliğini artıran önemli bir mimari olarak öne çıkmaktadır. Bu modeller, büyük ölçüde görüntü sınıflandırma görevleri için eniyilenmiştir. Kullanılan modellerin üstüne ek katmanlar eklenerek özel sınıflandırma modelleri de oluşturulabilmektedir.

Model, 40 eğitim adımı (epoch) boyunca eğitilmiş ve Uyarlanır Moment Hesaplama (Adaptive Moment Estimation, UMH-ADAM) optimizasyon algoritması kullanılmıştır [17]. Eğitim (training) ve geçerlileme (validation) kümeleri üzerindeki başarım, eğitim süresince izlenmiştir. Xception modelinde önceden eğitimli öznitelik elde eden katmanlarını takip eden sınıflandırma katmanları bulunmamaktadır. Modelin öznitelik elde edilmesi yöntemi kullanarak özel sınıflandırma katmanları eklenmiştir. Model 70 eğitim adımı (epoch) boyunca eğitilmiş ve ADAM optimizasyon algoritması kullanılmıştır [17].

Model başarımını değerlendirmekte şunlar kullanılmıştır:

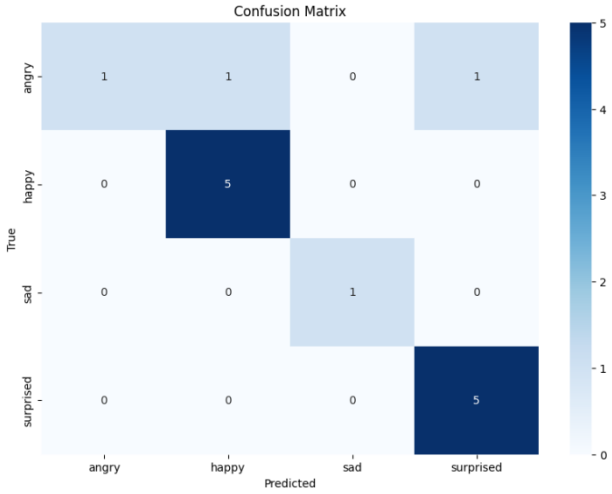
- **Amaç Fonksiyonu Ölçümleri:** Eğitim ve test veri kümeleri üzerindeki doğruluk ve kayıp değerleri hesaplanmıştır.
- **Duyarlılık (Precision), Anma (Recall) ve *F*1 Skoru ölçütleri:** Modelin genel başarımı, duyarlılık, anma ve ve *F*1 skoru ölçütleri ile değerlendirilmiştir [18].
- **Çapraz tahmin tablosu (Confusion Matrix):** Modelin sınıflandırma başarımını temel ölçütler bazında görselleştirmek için kullanılmıştır.
- **Alıcı İşlem Karakteristik (Receiver Operating Characteristics, AİK-ROC) ve Eğri Altında Kalan Alan (Area Under Curve, EAKA-AUC):** Modelin her bir sınıf için AİK eğrisi çizdirilmiş ve EAKA değerleri hesaplanmıştır [18].

4. Bulgular

Çalışmamızda Arnavutça dilinde duygu durum analizinde ResNet50 ve Xception derin öğrenme modellerini kullanılmıştır [16, 17]. Bu sayede her iki modelin başarımını ve genel etkilerini ayrıntılı olarak değerlendirilmiştir.

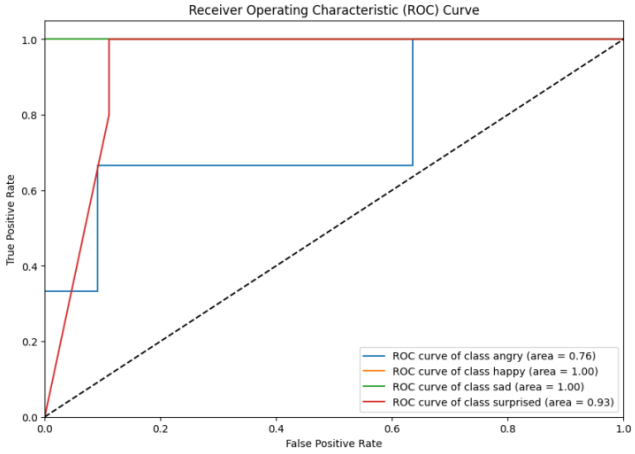
Elde edilen sonuçlar, modellerin Arnavutça dilinde duygu durum analizinde oldukça başarılı olduğunu göstermektedir. Eğitim doğruluğunun %100 olması, modellerinin eğitilmiş veriler üzerinde mükemmel başarım gösterdiğini, ancak test

doğruluğunun %85 olması, modelin genel başarımının iyileştirilmesi gerektiğini ortaya koymaktadır. Gelecekte yapılacak çalışmalar, veri kümesinin genişletilmesi ve daha fazla örnek ile modelin başarımlarını artırılmasını hedeflemelidir. Aşağıdaki Şekil1 ve Şekil 2'de ResNet50 modeli ile elde edilen, sırasıyla, çapraz tahmin tablosu ve AİK tabanlı deney sonucu görülmektedir.



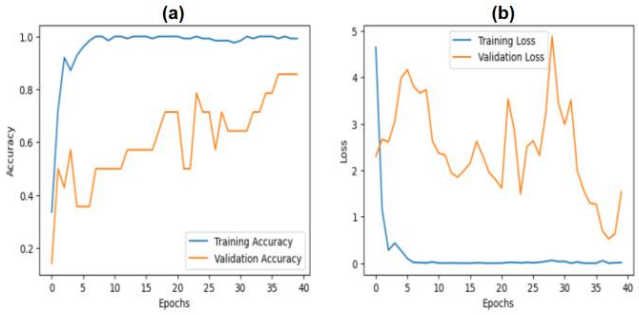
Şekil 1. ResNet50 çapraz tahmin tablosu

Çalışmadaki 'Angry' sınıfında model, 1 doğru tahmin yapmış ve 2 yanlış tahmin yapmıştır; yanlış tahminler 'Happy' ve 'Surprised' olarak sınıflandırılmıştır. 'Happy' sınıfında 5 doğru tahmin yapılmış ve hiç yanlış tahmin yapılmamıştır, bu sınıfta yüksek bir doğruluk oranı sağlanmıştır. 'Sad' sınıfında 1 doğru tahmin yapılmış ve hiç yanlış tahmin yapılmamıştır. 'Surprised' sınıfında ise 5 doğru tahmin yapılmış ve hiç yanlış tahmin yapılmamıştır. Genel olarak model, 'Happy', 'Sad', ve 'Surprised' sınıflarında oldukça başarılı, ancak 'Angry' sınıfında beklenen başarıyı sağlayamamıştır. Bu, modelin bazı duyguları diğerleriyle karıştırdığını ve 'Angry' sınıfında iyileştirmeler yapılması gerektiğini göstermektedir.



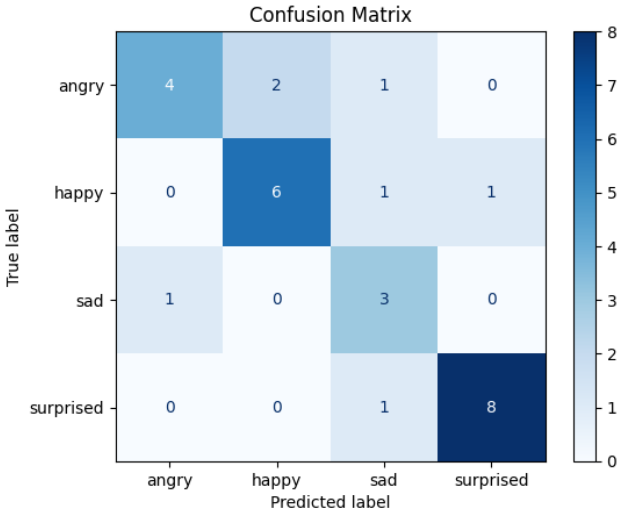
Şekil 2. ResNet50 AİK tabanlı deney sonucu

Grafiğe dayanarak AİK eğrisi altında kalan alan (EAKA) değerlerine göre modelin başarımını değerlendirecek olursak; 'Angry' sınıfında EAKA değeri 0.76 olup, modelin bu sınıfta orta düzeyde bir ayrım yeteneğine sahip olduğunu ve bu sınıfta daha fazla hata yaptığını göstermektedir. 'Happy' sınıfında EAKA değeri 1.00 olup, modelin neredeyse hatasız bir ayrım yapabildiğini ve bu sınıfta mükemmel bir başarım sergilediğini ortaya koymaktadır. 'Sad' sınıfında da EAKA değeri 1.00 olarak gösterilmiş olup, modelin bu sınıfta çok yüksek bir başarımla neredeyse hatasız bir şekilde doğru tahminler yapabildiğini göstermektedir. 'Surprised' sınıfında ise EAKA değeri 0.93 olup, modelin bu sınıfta oldukça iyi bir ayrım yeteneğine sahip olduğunu belirtmektedir. Genel olarak model, 'Happy' ve 'Sad' sınıflarında çok yüksek doğruluk ve ayrım yeteneği sergilerken, 'Angry' sınıfında orta düzeyde bir başarım göstermektedir. 'Surprised' sınıfında ise iyi bir başarım göstermekte, ancak yine de bazı karışıklıklar yaşanabilmektedir. Bu sonuçlar, modelin bazı duygular arasındaki ayrımı yaparken zorlandığını ve özellikle 'Angry' sınıfında ayrım yeteneğinin iyileştirebileceğini göstermektedir. Şekil 3'de ilgili ResNet50 derin öğrenme modeli ile elde edilen, a) eğitim sonucundaki grafik ve b) kayıp fonksiyonu (amaç fonksiyonu) grafiği görülmektedir.



Şekil 3. a) Eğitim grafiği b) Kayıp fonksiyonu grafiği

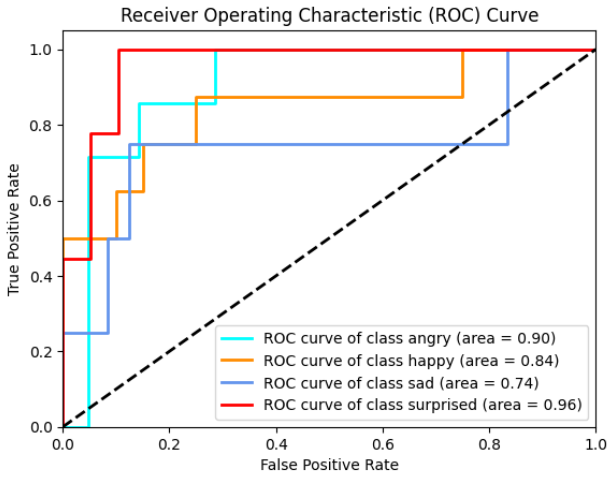
Eğitim sürecinde elde edilen grafikler, modelin başarımını detaylı bir şekilde gözler önüne sermektedir. Eğitim doğruluğu, hızlı bir artış göstererek yaklaşık beşinci eğitim adımı civarında %100'e ulaşmış ve bu seviyede stabil kalmıştır. Bu, modelin eğitim verisi üzerinde çok iyi başarım gösterdiğini ancak aşırı öğrenme ya da ezberleme (overfitting) riskinin olduğunu göstermektedir. Geçerlileme doğruluğu ise başlangıçta dalgalanmalar göstererek düşük seviyelerde başlamış, ancak eğitim ilerledikçe kademeli olarak artarak yaklaşık %70 seviyesine çıkmıştır. Daha sonraki eğitim adımları (epoch) boyunca geçerlileme doğruluğunda sürekli bir artış gözlenmektedir, bu da modelin geçerlileme verisi üzerinde daha iyi başarım göstermeye başladığını işaret eder. Eğitim kaybı, hızlı bir şekilde düşüş göstererek yaklaşık beşinci eğitim adımı (epoch) civarında neredeyse sıfıra inmiş ve süreç boyunca düşük seviyede stabil kalmıştır. Geçerlileme kaybı ise başlangıçta yüksek seviyelerde başlamış, eğitim ilerledikçe azalmış ancak belirli bir seviyede (yaklaşık 1 ilâ 2 arasında) sabit kalmıştır. Eğitim adımları (epoch) boyunca geçerlileme kaybında dalgalanmalar gözlenmesi, modelin geçerlileme verisinde başarımının zaman zaman değiştiğini göstermektedir. Modelimizi değerlendirdiğimizde, eğitim veri kümesi üzerindeki doğruluk oranı %100 olarak elde edildi. Ancak, test veri kümesi üzerindeki doğruluk oranı %85 olarak gerçekleşti. Bu sonuçlar, modelin eğitim verisi üzerinde mükemmel başarım sergilediğini ancak test verisi üzerindeki başarımının biraz daha düşük olduğunu göstermektedir. Çapraz Tahmin Tablosu ve AİK eğrisi analizlerimize göre, modelin 'Angry' ve 'Sad' sınıflarında yüksek doğruluk ve ayrım yeteneğine sahip olduğunu, ancak 'Happy' ve 'Surprised' sınıflarında bazı sınıf ayrıştırma yetersizlikleri oluştuğu anlaşılmaktadır. Özellikle 'Happy' ve 'Angry' sınıfları arasındaki ayrımda zorluklar gözlemlenmiştir. Bu sonuçlar, modelimizin genel olarak duygu sınıflandırma görevinde iyi başarım sergilediğini, ancak daha yüksek doğruluk oranlarına ulaşmak için ek veri işleme ve model iyileştirmeleri yapılması gerektiğini açıkça göstermektedir.



Şekil 4. Xception çapraz tahmin tablosu

Gelecekteki çalışmalar, veri kümesinin çeşitlendirilmesi, modelin ince ayarlarının yapılması ve daha gelişmiş tekniklerin uygulanması ile modelin doğruluğunu artırmaya odaklanmalıdır. Aşağıdaki Şekil 4 ve Şekil 5'de Xception modeli ile elde edilen, sırasıyla, çapraz tahmin tablosu ve AİK tabanlı deney sonucu görülmektedir. Çapraz Tahmin Tablosu (Confusion Matrix) sonuçları, modelimizin dört duygu sınıfındaki (Angry, Happy, Sad, Surprised) başarımını göstermektedir. 'Angry' sınıfında model, 4 doğru tahmin yapmış ve 4 yanlış tahmin yapmıştır; yanlış tahminler 'Happy' ve 'Sad' olarak sınıflandırılmıştır. 'Happy' sınıfında 6 doğru tahmin yapılmış ve 2 yanlış tahmin yapılmıştır. Bu deneye göre, yanlış tahminler 'Surprised' ve 'Sad'

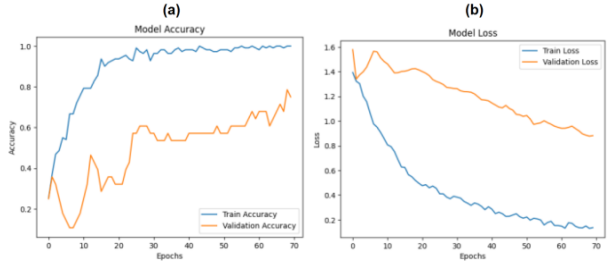
olarak sınıflandırılmıştır. 'Sad' sınıfında ise 3 doğru tahmin yapılmış ve 4 yanlış tahmin yapılmıştır. Buna göre model, 'Sad' sınıfı için düşük değerler vermektedir. Ayrıca, 'Surprised' sınıfında ise 8 doğru tahmin yapılmış ve 2 yanlış tahmin yapılmıştır. Bu sınıf için böylece yüksek bir doğruluk oranı sağlanmıştır.



Şekil 5. Xception AİK tabanlı deney sonucu

Genel olarak model, 'Angry', 'Happy', ve 'Surprised' sınıflarında oldukça başarılı, ancak 'Sad' sınıfında beklenen başarıyı sağlayamamıştır. Bu, modelin bazı duyguları diğerleriyle yeterince ayrıştıramadığını ve 'Sad' sınıfında iyileştirmeler yapılması gerektiğini göstermektedir.

Grafiğe dayanarak AİK eğrisi altında kalan alan (EAKA) değerlerine göre modelin başarımını değerlendirecek olursak; 'Angry' sınıfında EAKA değeri 0.90 olup, modelin bu sınıfta oldukça iyi düzeyde bir ayrım yeteneğine sahip olduğunu ve bu sınıfta fazla hata yapmadığını göstermektedir. .'Happy' sınıfında EAKA değeri 0.84 olup, modelin bu sınıfta da oldukça iyi düzeyde bir ayrım yeteneğine sahip olduğunu ve bu sınıfta fazla hata yapmadığını göstermektedir. 'Sad' sınıfında ise EAKA değeri 0.74 olarak gösterilmiş olup, modelin bu sınıfta kötü bir başarım gösterdiğini söylenebilir. 'Surprised' sınıfında ise EAKA değeri 0.96 olup, modelin bu sınıfta oldukça iyi bir ayrım yeteneğine sahip olduğunu belirtmektedir. Şekil 6'da ilgili Xception derin öğrenme modeli ile elde edilen, a) eğitim sonucundaki grafik ve b) kayıp fonksiyonu (amaç fonksiyonu) grafiği görülmektedir. Genel olarak model, 'Surprised' ve 'Angry' sınıflarında çok yüksek doğruluk ve ayrım yeteneği sergilerken, 'Sad' sınıfında orta düzeyde bir başarım göstermektedir. 'Happy' sınıfında ise iyi bir başarım göstermekte, ancak yine de bazı karışıklıklar yaşanabilmektedir.



Şekil 6. a) Eğitim grafiği b) Kayıp fonksiyonu grafiği

Bu sonuçlar, modelin bazı duygular arasındaki ayrımı yaparken zorlandığını ve özellikle 'Sad' sınıfında ayrım yeteneğinin modelin parametreleri daha uygun ayarlanarak (fine tuning) iyileştirebileceğini göstermektedir.

Eğitim sürecinin analizi, modelin başarımını değerlendiren birkaç kritik bulguyu ortaya koymaktadır. Eğitim sürecinin başlangıç aşamasında, modelin doğruluk oranı düşük olmakla birlikte, zamanla belirgin bir artış gözlemlenmiştir. İlk eğitim adımında (epoch) eğitim doğruluğu %25.23 olarak ölçülürken, 70. eğitim adımında (epoch) bu oran %100'e ulaşmıştır. Eğitim kaybı da benzer bir eğilim göstermiş, başlangıçta 1.3935 değerinde iken, 70. eğitim adımında (epoch) 0.1371'e düşmüştür.

Geçerlileme (Validasyon) kümesine yönelik sonuçlar ise daha değişken bir başarım profili sergilemiştir. Başlangıçta doğruluk oranı %25 değeriyle düşük olup, 20. eğitim adımından (epoch) itibaren artış göstermiş ve 70. eğitim adımında (epoch) %75 seviyesine ulaşmıştır. Geçerlileme kaybı ise başlangıçta yüksek iken, zamanla iyileşmiş ve 70. eğitim adımında (epoch) 0.8818 olarak ölçülmüştür.

Bu bulgular, modelin eğitim kümesi üzerinde mükemmel bir başarım sergilediğini ve genel olarak geçerlileme kümesinde güçlü bir başarım gösterdiğini ortaya koymaktadır.

Eğitim doğruluğunun %100 olması, modelin eğitim verisi üzerinde mükemmel başarım sergilediğini, test doğruluğunun %75 olması, modelin genel başarımının iyi olduğunu fakat test verisi üzerinde daha fazla iyileştirme potansiyelinin mevcut olduğunu işaret etmektedir.

Eğitim ve test sonuçlarına dair başarım ölçütleri arasındaki bu fark, modelin eğitim verisine aşırı uyum (overfitting) sağlamış olabileceğini gösteren bir işarettir [19]. Ancak, test kümesindeki sonuçların güçlü olması, modelin genel genelleme kapasitesinin yüksek olduğunu ve pratik uygulama bağlamında etkili olabileceğini göstermektedir.

Deneylerimizde her iki model de benzer başarımlar göstermiştir. Bunlardan sınıf ayrıştırıcılığına dair belirgin bir fark göstermesi nedeniyle Xception modeli için sınıflandırma başarım ölçütleri aşağıdaki Tablo1'de incelendiğinde, modelin dört farklı duygu sınıfında genel olarak oldukça yeterli bir başarı sergilediğini görülmektedir.

Tablo 1. Xception başarım ölçütleri

Sınıf	Duyarlılık	Anma	*F*1 skoru
Angry	0.80	0.57	0.67
Happy	0.75	0.75	0.75
Sad	0.50	0.75	0.60
Suprised	0.89	0.89	0.89
Ağırlıklı ortalama	*0.77*	*0.75*	*0.75*

Bu çalışmadaki 'Angry' sınıfı için duyarlılık (precision) değeri 0.80 olup, modelin bu sınıfı doğru tahmin etme yeteneğinin yüksek olduğunu, ancak anma (recall) değeri 0.57 olması nedeniyle gerçek 'Angry' sınıfından örneklerini tespit etme oranının nispeten düşük olduğunu göstermektedir. Ayrıca, 'Happy' ve 'Surprised' sınıflarında ise hem duyarlılık hem de anma oranları oldukça dengeli ve yüksek olup, modelin bu sınıfları tanımlamada tutarlı ve doğru bir şekilde başarım gösterdiğini ortaya koymaktadır [18].

Bu deneye göre, ayrıca 'Sad' sınıfında anma değeri 0.75 olarak oldukça yüksek iken, duyarlılık değeri 0.50 olarak bir miktar düşük olmuştur. Bu da modelin, 'Sad' sınıfında yanlış pozitif tahminlerde bulunduğunu ve bazı yanlış sınıflandırmalar yaptığını göstermektedir. Genel doğruluk oranı %75 olup, makro ve ağırlıklı ortalama değerler de sırasıyla 0.73 ve 0.75 civarında olarak, modelin farklı sınıflar arasında dengeli bir başarım sergilediğini görülmektedir.

Sınıflandırma başarımının bazı deneylerde düşük olması, birkaç temel nedeni vardır. Özellikle 'Angry' duygusuna ait ses örneklerinin sayısının diğer duygulara göre belirgin şekilde az olması, modelin bu duyguyu yeterince öğrenememesine neden olmaktadır. Ayrıca, Arnavutça'da belirli ifadelerin ve tonlamaların eksikliği, modelin bu duyguyu ayrıştırmasında zayıf kalmasına yol açmaktadır. 'Angry' duygusunun Arnavutça konuşan bireyler tarafından farklı tonlama ve vurgularla ifade edilebilmesi, sınıflandırma sürecideki sınıf ayrıştırmasını zorlaştırmaktadır.

Ses kaynağındaki çeşitlilik, arka plan gürültüsü ve kayıt koşullarındaki değişkenlik, modelin bu sınıfı doğru bir şekilde ayrıştırarak sınıflandırmasını olumsuz etkileyen unsurlar arasındadır. Ayrıca, 'Angry' sınıfı yoğun duygu ifadelerinde sesin tonlaması ve öznitelik aralıkları açısından diğer duygularla (örneğin 'Sad' veya 'Surprised') sınıfsal olarak karışabilir. Bu sınıflandırma hatasını artırır.

Sınıflandırma hatalarını azaltmak ve modelin 'Angry' duygusunu iyice öğrenerek diğer sınıflardan daha kesin ayrıştırmasını sağlamak için çeşitli stratejiler uygulanması önerilmektedir. İlk olarak, veri artırma ile 'Angry' sınıfına ait ses kayıtlarının sayısını artırmak, modelin bu duygu üzerindeki öğrenme kapasitesini güçlendirecektir. Bunun yanı sıra, modelin üstün parametrelerinin (hyper parameters) eniyilemesi (optimization), öğrenme sürecini iyileştirecektir. Karmaşık derin öğrenme mimarilerinin, örneğin Attention mekanizması veya TSA tipi (UKSB) gibi ağ yapılarının entegrasyonu, modelin duygu ifadelerini daha doğru bir şekilde anlamasını sağlayabilir. Ayrıca, 'Angry' sınıfını daha iyi temsil eden ek ses örnekleri toplanarak modelin eğitim kümesini zenginleştirmek, genel başarımın iyileştirilmesine katkıda bulunacaktır. Bu stratejilerin uygulanması, 'Angry' duygusunun doğru sınıflandırılmasını ve modelin genel duygu tanıma başarımını artıracaktır.

5. Sonuçlar ve Tartışma

Bu çalışmada, Mel spektrogramlarını kullanarak ses verilerinden duygu tanımlamaya yönelik derin öğrenme modelleri geliştirilmiş ve değerlendirilmiştir. Çalışma kapsamında, Xception ve ResNet50 gibi görüntü sınıflandırmada kullanılan gelişmiş derin öğrenme modelleri kullanılarak duygusal durumların sınıflandırılması

hedeflenmiştir. Modeller, ses verilerinden elde edilen Mel spektrogramları görüntüleri üzerinde eğitilmiştir.

Eğitim sürecinde modellerin doğruluk oranları hızlı bir artış göstermiş ve eğitim doğruluğu %100 seviyesine ulaşmıştır. Eğitim kaybı ise neredeyse sıfıra inmiştir. Geçerlileme doğruluğu kademeli olarak artarak Xception modeli için %75, ResNet50 modeli için %85 seviyelerine ulaşmıştır. Geçerlileme kaybı ise belirli bir seviyede sabit kalmıştır.

Bu sonuçlar, kullanılan modellerin Mel spektrogramlarını kullanarak duygu tanımlamada başarılı olduğunu göstermektedir. Özellikle, ResNet50 modelinin geçerlileme doğruluğunun %85 seviyelerine ulaşması, bu modelin Mel spektrogramlarından duygu tanımlamada yüksek başarım gösterdiğini ortaya koymaktadır. Xception modelinin de benzer bir başarı gösterdiği görülmüştür.

Çalışmada geliştirilen modellerin her birinin başarımlarının artırılmasına yönelik öneriler sunulmuş ve gelecekte yapılacak çalışmalar için yol gösterici bilgiler elde edilmiştir. Metin madenciliği alanında doğal dil işleme ile yapılan literatürdeki duygu analizi çalışmalarında Arnavutça metinler üzerinde yapılan deneylerde Roland Vasili ve Endri Xhina 2021 yılındaki çalışmasında doğruluk oranını %79.2 olarak elde etmiştir [8]. Amarildo Rista ve Arbana Kadriu'nun 2022 yılında gerçekleştirdiği konuşma tanıma üzerine çalışması, derin öğrenme tekniklerini kullanarak başarılı sonuçlar elde etmiştir [20]. Bir diğer çalışmada ise Muhamet Kaçuri, Arnavutça dilinde nefret söylemi sınıflandırmasının açıklanabilirliği üzerine deneyler gerçekleştirirken sadece bir duygu durumu (hate) üzerine odaklanmıştır. Buna karşın, bizim çalışmamızda dört farklı duygu durumu için veri kullanılarak deneyler yapılmıştır [21]. Bu çalışmalara kıyasla başarım sonuçlarımız, birçok duygu durumu içeren veri kümesi kullanılmasına rağmen, deneylerimizde oldukça iyi düzeydedir.

Bu bakış açısıyla daha büyük ve çeşitli veri kümelerinin kullanılması, model üstün parametrelerinin eniyilemesi ve farklı derin öğrenme mimarilerinin denenmesi gibi konular gelecekteki çalışmalar için planlanmaktadır.

Teşekkür

Veri kümemizin oluşturulmasında gönüllü olarak katkı sunan değerli katılımcılara teşekkür ederiz.

Kaynaklar

[1] Sudhanshu K., Partha Pratim R., Debi Prosad D., Byung-Gyu K. "A Comprehensive Review on Sentiment Analysis: Tasks, Approaches and Applications". arXiv Prepr. arXiv: 2311.11250v1, 2023.

[2] Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning, USA, MIT Press, 2016.

[3] Wenxuan Z., Yue D., Bing L., Sinno Jialin P., Lidong B. “Sentiment Analysis in the Era of Large Language Models: A Reality Check”. arXiv Prepr. arXiv:2305.15005, 2023.

[4] Burkhardt F, Paeschke A, Rolfes M, Sendlmeier F, Weiss B. “A Database of German Emotional Speech", *9th European Conference on Speech Communication and Technology, INTERSPEECH 2005 - Eurospeech*, 1517-1520, Lisbon, Portugal, 4-8 Eylül, 2005.

[5] Livingstone S. R., Russo F. A. “The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English” *PLOS ONE*, 13(5), .e0196391. 2018.

[6] Muraku B., Xiao L., Meçe E. K. "Toward Detection of Fake News Using Sentiment Analysis for Albanian News Articles". Editors: Barolli L., Advances in Internet, Data & Web Technologies, EIDWT, Lecture Notes on Data Engineering and Communications Technologies, 19, 575–585, Springer, Cham, 2024.

[7] Çano E., “AlbMoRe: A Corpus of Movie Reviews for Sentiment Analysis in Albanian”. *Digital Philology Data Mining and Machine Learning*, arXiv Prepr. arXiv:2306.085262023. University of Vienna, Austria, 2023.

[8] Vasili, R., Xhina, E., Ninka, I., & Terpo, D. "Sentiment Analysis on Social Media for Albanian Language". OALib, 8(1-31), 2021.

[9] McFee B, Raffel C, Liang D, Ellis D, Mcvicar M, Battenberg E, Nieto O. “Librosa: Audio and Music Signal Analysis in Python”. *Proceedings of the 14th Python in Science Conference*, January 2015.

[10] Hunter J.D. “Matplotlib: A 2D Graphics Environment”. *Computing in Science & Engineering*, 9(3), 90-95, 2007.

[11] Clark A. PIL: Python Imaging Library. Pillow (PIL Fork) Documentation, 1999.

[12] Tensorflow web sitesi, 2024. https://www.tensorflow.org/

[13] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. "Scikit-learn: Machine Learning in Python". *Journal of*

Machine Learning Research, 12(85), 2825-2830, 2011.

[14] Bisong, E.. "Google Colaboratory". *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 59-64, Apress, Berkeley, CA. 2019.

[15] Grabas, L.. "NLP-Text Dataset, *A balanced dataset with 5 labels joy, sad, anger, fear, and neutral, 3202 unique values"*. 2024. https://github.com/lukasgarbas/nlp-text-emotion/blob/master/data/data_test.csv

[16] He K, Zhang X, Ren S, Sun J. "Deep residual learning for image recognition", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778, Las Vegas, Nevada, USA, 27-30 Haziran 2016.

[17] Chollet, F. "Xception: Deep Learning with Depthwise Separable Convolutions". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[18] Powers, D. M. W. "Evaluation: from Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation". arXiv Prepr. arXiv:2010.16061, 2020.

[19] Ying, X. "An Overview of Overfitting and its Solutions". *Journal of Physics: Conference Series*, 1168(2), 022022, 2019.

[20] Kadriu, A., Rista, A. "A Model for Albanian Speech Recognition Using End-to-End Deep Learning Techniques". *International Journal of Research and Development*, 9(3), 2022.

[21] Kaçuri, M. Explainability of Hate Speech Classification for Albanian Language Using Rule Based Systems and Neural Networks. Diploma Thesis, Technische Universität Wien. reposiTUm, 2023.