

T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

STRES GÖSTERGELERİNİN SES ÖZELLİKLERİNE DAYALI OLARAK
ANALİZİ

YÜKSEK LİSANS TEZİ

Çağatay KASAPOĞLU

1408050001

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Fatma PATLAR AKBULUT

EYLÜL 2023

T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

STRES GÖSTERGELERİNİN SES ÖZELLİKLERİNE DAYALI OLARAK
ANALİZİ

YÜKSEK LİSANS TEZİ

Çağatay KASAPOĞLU

1408050001

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Fatma PATLAR AKBULUT

Jüri Üyesi: Dr.Ögr.Üyesi Öznur ŞENGEL

Jüri Üyesi: Prof.Dr. Güray YILMAZ

EYLÜL 2023

ÖNSÖZ

Bu çalışma, stres göstergelerinin ses özelliklerine dayalı olarak tespit ve analiz etmek amacıyla yapılmıştır. Tez sürecimin her aşamasında desteğini eksik etmeyen ve katkılarıyla değer katan tez danışmanım Doç. Dr. Fatma Akbulut'a çok teşekkürler. Eğitim ve araştırma sürecimde bana vizyon katan ve her zaman desteklerini hissettiren sevgili aileme ve hocalarımın teşekkürlerimi sunarım.

EYLÜL 2023

Çağatay KASAPOĞLU

İÇİNDEKİLER

ÖNSÖZ	i
İÇİNDEKİLER	ii
KISALTMALAR	iv
TABLO LİSTESİ	v
ŞEKİL LİSTESİ	vi
ÖZET	vii
ABSTRACT	viii
1. GİRİŞ	1
1.1 Tezin Amacı	1
1.2 Tezin Organizasyonu	3
2. TEMEL BİLGİLER VE LİTERATÜR TARAMASI	3
1.3 Temel Duyguların Stres İçerdiğini Savunan Çalışmalar	5
1.3.1 Paul Ekman: "Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life" (2003)	5
1.3.2 Richard Lazarus: "Stress and Emotion: A New Synthesis" (2006)	6
1.3.3 "The Upside of Stress: Why Stress Is Good for You, and How to Get Good at It"- Kelly McGonigal (2015)	8
1.4 T-Testinin Hesaplanması	9
2. YÖNTEM	11
2.1 EMO-DB Veri Seti	11
2.2 RAVDESS Veri Seti	12
2.3 SAVEE Veri Seti	13
2.4 OpenSmile Kütüphanesi	14
2.4.1 OpenSmile Ses Özellikleri	18
3. BULGULAR	19
3.1 EMO-DB Stresli ve Stresiz Ses Sinyallerinin Özellikleri ve Karşılaştırılması	20
3.1.1 EMO-DB Spektral Özellikleri ve Değerleri	21
3.1.2 EMO-DB Zamansal Özellikleri ve Değerleri	22
3.1.3 EMO-DB Zaman-Frekans Özellikleri ve Değerleri	23
3.1.4 EMO-DB Yüksek Seviye Özellikleri ve Değerleri	24
3.1.5 EMO-DB Veri Setindeki Temel Duyguların Görüntüleri	25
3.2 EMO-DB Uzun Kısa Süreli Bellek Modeli (LSTM)	28
3.3 EMO-DB 1 Boyutlu Evrişimli Sinir Ağları (1D-CNN)	30
3.4 RAVDESS Stresli ve Stresiz Ses Sinyallerinin Özellikleri ve Karşılaştırılması	33
3.4.1 RAVDESS Spektral Özellikleri ve Değerleri	33

3.4.2	RAVDESS Zamansal Özellikleri ve Değerleri	34
3.4.3	RAVDESS Zaman-Frekans Özellikleri ve Değerleri.....	35
3.4.4	RAVDESS Yüksek Seviye Özellikleri ve Değerleri	35
3.4.5	RAVDESS Veri Setindeki Temel Duyguların Görüntüleri	37
3.5	RAVDESS Uzun Kısa Süreli Bellek Modeli (LSTM).....	40
3.6	RAVDESS 1 Boyutlu Evrişimli Sinir Ağları (1D-CNN)	41
3.7	SAVEE Stresli ve Stressiz Ses Sinyallerinin Özellikleri ve Karşılaştırılması	44
3.7.1	SAVEE Spektral Özellikleri ve Değerleri.....	44
3.7.2	SAVEE Zamansal Özellikleri ve Değerleri	45
3.7.3	SAVEE Zaman-Frekans Özellikleri ve Değerleri	46
3.7.4	SAVEE Yüksek Seviye Özellikleri ve Değerleri	46
3.7.5	SAVEE Veri Setindeki Temel Duyguların Görüntüleri.....	47
3.8	SAVEE Uzun Kısa Süreli Bellek Modeli (LSTM)	50
3.9	SAVEE 1 Boyutlu Evrişimli Sinir Ağları (1D-CNN).....	52
4.	TARTIŞMA.....	54
5.	SONUÇ.....	55
	KAYNAKÇA	57

KISALTMALAR

- LSTM : Uzun kısa Süreli Bellek
1D-CNN : 1 Boyutlu Evrişimli Sinir Ağı
EMO-DB : Berlin Duygusal Konuşma Veri tabanı
RAVDESS : Ryerson Duygusal Konuşma ve Şarkılara İlişkin Veri tabanı
SAVEE : Surrey Duygusal Konuşma Veri tabanı



TABLO LİSTESİ

Tablo 1.1 EMO-DB Spektral Özellikleri	21
Tablo 2.1 EMO-DB Zamansal Özellikleri	22
Tablo 3.1 EMO-DB Zaman-Frekans Özellikleri	23
Tablo 4.1. EMO-DB Yüksek Seviye Özellikleri ve Değerleri	24
Tablo 5.1. RAVDESS Spektral Özellikleri ve Değerleri	33
Tablo 6.1. RAVDESS Zamansal Özellikleri ve Değerleri	34
Tablo 7.1. RAVDESS Zaman-Frekans Özellikleri ve Değerleri.....	35
Tablo 8.1. RAVDESS Yüksek Seviye Özellikleri ve Değerleri	36
Tablo 9.1 SAVEE Spektral Özellikleri ve Değerleri.....	44
Tablo 10.1.SAVEE Zamansal Özellikleri ve Değerleri	45
Tablo 11.1. SAVEE Zaman-Frekans Özellikleri ve Değerleri	46
Tablo 12.1. SAVEE Yüksek Seviye Özellikleri ve Değerleri	46



ŞEKİL LİSTESİ

Şekil 1.1 EMO-DB Duygular	12
Şekil 2.1 EMO-DB Mutluluk Ses Dalgası.....	25
Şekil 3.1 EMO-DB Mutluluk Ses Dalgası.....	25
Şekil 4.1 EMO-DB Öfke Ses Dalgası.....	26
Şekil 5.1 EMO-DB Öfke Spektrogram Görüntüsü.....	26
Şekil 6.1 EMO-DB Doğal Ses Dalgası	27
Şekil 7.1 EMO-DB Doğal Spektrogram Görüntüsü	27
Şekil 8.1 EMO-DB Korku Ses Dalgası.....	27
Şekil 9.1 EMO-DB Korku Ses Dalgası.....	27
Şekil 10.1 EMO-DB LSTM	30
Şekil 11.1 EMO-DB CNN.....	32
Şekil 12.1 RAVDESS Mutluluk Ses Dalgası.....	37
Şekil 13.1 RAVDESS Mutluluk Spektrogram Görüntüsü.....	37
Şekil 14.1 RAVDESS Öfke Ses Dalgası.....	38
Şekil 15.1 RAVDESS Öfke Spektrogram Görüntüsü.....	38
Şekil 16.1 RAVDESS Doğal Ses Dalgası	38
Şekil 17.1 RAVDESS Doğal Spektrogram Görüntüsü	39
Şekil 18.1 RAVDESS Korku Ses Dalgası	39
Şekil 19.1 RAVDESS Korku Spektrogram Görüntüsü.....	39
Şekil 20.1 RAVDESS LSTM	41
Şekil 21.1 RAVDESS CNN.....	43
Şekil 22.1 SAVEE Mutluluk Ses Dalgası	48
Şekil 23.1 SAVEE Mutluluk Spektrogram Görüntüsü	48
Şekil 24.1 SAVEE Öfke Ses Dalgası	48
Şekil 25.1 SAVEE Öfke Spektrogram Görüntüsü	49
Şekil 26.1 SAVEE Doğal Ses Dalgası.....	49
Şekil 27.1 SAVEE Doğal Spektrogram Görüntüsü	49
Şekil 28.1 SAVEE Korku Ses Dalgası	50
Şekil 29.1 SAVEE Korku Spektrogram Görüntüsü	50
Şekil 30.1 SAVEE LSTM.....	52
Şekil 31.1 SAVEE CNN	54

Üniversite	: T.C. İstanbul Kültür Üniversitesi
Enstitü	: Lisansüstü Eğitim Enstitüsü
Anabilim Dalı	: Bilgisayar Mühendisliği
Program	: Bilgisayar Mühendisliği
Tez Danışmanı	: Doç. Dr. Fatma AKBULUT
Tez Türü ve Tarihi	: Yüksek Lisans – Eylül 2023

ÖZET

STRES GÖSTERGELERİNİN SES ÖZELLİKLERİNE DAYALI OLARAK ANALİZİ

Günümüzde insanlar daha hızlı bir yaşam temposuna sahiptir ve sürekli olarak çalışma, sosyal ilişkiler, mali sorumluluklar ve diğer günlük yaşam zorlukları arasında denge kurmaya çalışırlar. Teknolojik gelişmelerle birlikte her zaman erişilebilir olma beklentisi de artmıştır. İnternet ve sosyal medya gibi platformlar, insanları sürekli olarak bilgilendirme ve bağlantı halinde olma baskısı altına sokabilir. Stresin uzun süreli etkileri ise sağlık sorunlarına yol açabilmektedir. Gündelik süreçte, çeşitli temel duygulara barınmaktayız, hangi temel duygunun stres içerdiğinin araştırılması ve analizi bu çalışmamızda yapılmıştır. EMO-DB (Berlin Duygusal Konuşma Veri tabanı) verisi, RAVDESS (Ryerson Duygusal Konuşma ve Şarkılara İlişkin Veri tabanı) ile SAVEE (Surrey Duygusal Konuşma Veri tabanı) üzerinde açık kaynak kütüphanesi OpenSmile ile ses özelliklerinin çıkarımları yapılmıştır. Stresli ve stressiz olarak gruplandırılan veri setinin karşılaştırılması, T-Testi yapılarak hangi ses özelliklerinin stresi belirlemede daha başarılı olduğu tespit edilmiştir. Derin öğrenme modellerinden LSTM ve 1D-CNN algoritmalarıyla tahmin edilmesi ve istatistiksel sonuçların hangisinde daha başarılı olduğu gözlenmiştir.

Anahtar Kelimeler: EMO-DB, RAVDESS, SAVEE, LSTM Modeli, 1D-CNN, OpenSmile Kütüphanesi

University	: T.C. İstanbul Kültür University
Institute	: Institute of Graduate Studies
Department	: Computer Engineering
Program	: Computer Engineering
Thesis Advisor	: Assoc. Prof. Dr. Fatma AKBULUT
Degree Awarded And Date	: MA – September 2023

ABSTRACT

ANALYSIS OF STRESS INDICATORS BASED ON SOUND FEATURES

In today's fast-paced world, people are constantly trying to balance work, social relationships, financial responsibilities, and other daily life challenges. With technological advancements, there is also an increased expectation of being constantly accessible. Platforms such as the internet and social media can put people under pressure to constantly stay informed and connected. The long-term effects of stress can lead to health problems. In our study, we have investigated and analyzed which basic emotion encompasses stress. We have conducted this analysis using the EMO-DB (Berlin Emotional Speech Database), RAVDESS (The Ryerson Audio-Visual Database of Emotional Speech) with SAVEE (Surrey Emotional Speech Database) dataset and extracted audio features using the open-source library OpenSmile. By comparing the dataset grouped into stressed and non-stressed categories, we have determined which audio features are more successful in identifying stress through T tests. We have also employed deep learning models such as LSTM and 1D-CNN algorithms for stress prediction and observed which statistical results were more successful.

Keywords: EMO-DB, RAVDESS, SAVEE, LSTM Model, 1D-CNN, OpenSmile Lib.

1. GİRİŞ

Günümüzde bilişim teknolojilerinin yaygın olarak kullanılması, birçok kişi üzerinde stresin artmasına neden olabilmektedir. Teknolojinin sürekli erişilebilir olması, yoğun çalışma temposu, sosyal medya kullanımı ve bilgi bombardımanı gibi faktörler, stres seviyelerinin yükselmesine katkıda bulunmaktadır. Stresin hangi temel duygular içerisinde var olabileceğini istatistiki oranlarla karşılaştırma yapılmıştır.

Temel duygular ise, insanların deneyimlediği evrensel duygusal durumları ifade eder. Sevinç, mutluluk, coşku ve keyif gibi pozitif duygularla ilişkilendirilirken, üzüntü, hüznün ve hayal kırıklığı gibi negatif duygular da temel duygular arasındadır. Sevinç, insanların başarılarını kutlaması, sevdikleriyle vakit geçirmesi veya olumlu deneyimler yaşamalarıyla ortaya çıkar. Üzüntü, kayıp, hüznün veya hayal kırıklığı gibi negatif deneyimler sonucunda hissedilir. Korku, tehlike veya endişe algılandığında ortaya çıkar ve insanların hayatta kalma içgüdüsünü tetikler. Öfke, haksızlık veya kişisel sınırların ihlal edilmesi gibi durumlarda ortaya çıkar. Şaşkınlık, yeni bilgilerle karşılaşıldığında veya inanç sistemine meydan okuyan deneyimlerle ortaya çıkabilir. Tiksinti, iğrenme veya hoşnutsuzluk duygularını ifade eder. Temel duygular, insanların zengin ve karmaşık duygusal dünyasının özünde yer alır ve çeşitli deneyimlerle birleşerek farklı yoğunluklarda deneyimlenir.

Stres, duyguları derinden etkileyebilir. Temel duygularla iç içe geçen stres, duygusal deneyimlerin yoğunluğunu artırabilir. Mutluluk, üzüntü, korku, öfke, tiksinti ve şaşkınlık gibi temel duyguların bazı durumlarda stres adı verilen bir olgu ortaya çıkar ve duygularla iç içe geçer.

1.1 Tezin Amacı

Bu çalışmada, ses özelliklerinin çıkarılması ve makine öğrenmesi yöntemleri kullanılarak, EMO-DB, RAVDESS ve SAVEE veri setleri üzerinde çalışmalar gerçekleştirilmiştir. Amacımız, stresi en doğru şekilde tahmin etmek için uygun bir

modeli analiz etmektir. Veri setindeki ses kayıtları, farklı duygulardaki stres düzeylerini temsil etmektedir.

Denelerimizde çeşitli makine öğrenmesi algoritmaları ve ses özellik çıkarımı kullanılmış ve farklı özniteliklerin etkisi değerlendirilmiştir. Öznitelik çıkarımı için spektral öznitelikler, zaman alanı öznitelikleri ve derin öğrenme tabanlı öznitelikler kullanılmıştır.

Elde edilen sonuçlar, test ettiğimiz makine öğrenmesi modellerinin ve ses özellik çıkarımlarının stresi tahmin etmede başarılı olduğunu göstermiştir. Modelimiz, EMO-DB, RAVDESS ve SAVEE veri setlerindeki ses kayıtlarını analiz ederek, stres düzeylerini doğru bir şekilde sınıflandırabilmiştir. Bu çalışma, stresi tahmin etmek için makine öğrenmesi ve ses özellik çıkarımı yöntemlerinin kullanılabilirliğini ve etkinliğini gösteren önemli bir adımdır.

Bu araştırma, stresin ses verisi üzerinden tahmin edilmesi konusunda ilgili alandaki çalışmalara katkı sağlamış ve daha fazla araştırmanın yapılmasına ilham vermiştir. Sonuçlarımız, gelecekteki çalışmalarda daha gelişmiş ve hassas stres tahmin modellerinin analizine veya geliştirilmesine olanak tanıyacaktır.

Bu çalışmada kabul edilmesi ve reddedilmesi gereken hipotezler aşağıdaki gibidir.

H0: Stres tahmininde kullanılacak ses özellik çıkarımının p-value değerinin 0.05 den küçük olması ve yapay öğrenme modelinin %85 ve üzeri doğruluk oranına sahip olması.

H1: Stres tahmininde kullanılacak ses özellik çıkarımının p-value değerinin 0.05 den küçük olması ve yapay öğrenme modelinin %85 ve üzeri doğruluk oranına sahip olmaması

Tezin yöntem kısmında belirtilen modellere dayanarak, ilgili çalışmadan elde edilen perspektifle tez çalışması gerçekleştirilmiştir. Bu çalışma, H0 ve H1 hipotezlerinin doğrulanması veya reddedilmesi için çalışılmıştır.

1.2 Tezin Organizasyonu

Bu tez, altı bölümden oluşmaktadır. İlk bölümde, çalışmanın tanımı yapılmış ve araştırmanın amacı ile önemi açıklanmıştır.

İkinci bölümde, temel odak noktası olan stres barındıran temel duyguların tanımı ve literatürde mevcut araştırmalar gözden geçirilmiştir.

Üçüncü bölümde, ses sinyallerindeki stresin tespiti için kullanılan ses özellikleri çıkarımları yapılmış ve bu özelliklerin stresin ifade edilmesindeki etkisi görsellerle desteklenmiştir. Ayrıca yapay öğrenme modelleri ile elde edilen istatistiksel sonuçlar sunulmuştur.

Dördüncü bölümde, ses özelliklerinin çıkarımı ve yapay öğrenme modellerinin değerlendirilmesi detaylı bir şekilde ele alınmıştır.

Beşinci bölümde, gerçekleştirilen deneylerin sonuçları üzerine tartışma yapılmış ve elde edilen bulgular detaylı bir şekilde analiz edilmiştir.

Son olarak, altıncı bölümde, çalışmanın genel sonuçları özetlenmiş ve elde edilen bulguların ışığında ileriye dönük öneriler ve çalışmanın katkıları değerlendirilmiştir.

2. TEMEL BİLGİLER VE LİTERATÜR TARAMASI

Bu bölümde tez kapsamındaki çalışmalarda kullanılan temel kavramlar ele alınmıştır. İlk olarak, psikolojide stresin ve temel duyguların tanımları üzerinde durulmuştur. Stres, günümüzün hızlı yaşam temposuyla birlikte artan çalışma, sosyal ilişkiler, mali sorumluluklar ve diğer günlük yaşam zorluklarıyla başa çıkma sürecinde ortaya çıkan bir durumdur. Paul Ekman'a göre, stresi bir tepki olarak tanımlar. Ona göre, stres, bir kişinin zorluklarla karşılaştığında bedensel ve zihinsel olarak ortaya çıkan bir tepkidir. Bu tepki, vücutta bir dizi fizyolojik değişikliklerle birlikte, duygusal bir durumu da ifade eder. Ekman'a göre, stresli bir durumda kişi, alarm durumuna

geçer ve vücut stres hormonları salgılar. Bu tepki, kişinin dikkatinin artmasına, kalp atış hızının yükselmesine, kasların gerginleşmesine ve diğer bir dizi fiziksel değişikliğe neden olabilir.[1] Altı temel duygu, Amerikalı psikolog ve araştırmacı Paul Ekman'a göre şöyle tanımlanmıştır. Ekman, geniş çaplı kültürel çalışmalar yürüterek, insanların tüm kültürlerde ortak olarak deneyimlediği altı temel duyguyu belirlemiştir. Bu altı temel duygu, sevinç, üzüntü, korku, öfke, şaşkınlık ve iğrenme olarak sıralanır. Ekman'ın çalışmaları, yüz ifadeleri ve beden dilinin bu temel duyguları ifade etmede evrensel olduğunu göstermiştir. Bu bulgular, duyguların evrimsel olarak ortaya çıktığını ve insanlar arasında ortak bir temelde paylaşıldığını vurgulamaktadır.[2]

Bu temel duygular şunlardır:

1. Sevinç (Joy): Mutluluk, neşe, coşku ve keyif gibi pozitif duygusal durumları ifade eder.
2. Üzüntü (Sadness): Hüzün, melankoli ve kayıp gibi negatif duygusal durumları ifade eder.
3. Korku (Fear): Tehlike, tehdit veya kaygıya karşı duyulan endişe ve korkuyu ifade eder.
4. Öfke (Anger): Hiddet, öfke ve irritasyon gibi olumsuz ve saldırgan duygusal durumları ifade eder.
5. Şaşkınlık (Surprise): Beklenmedik olaylar veya durumlar karşısında duyulan şaşkınlığı ifade eder.
6. İğrenme (Disgust): Tiksinti, iğrenme ve rahatsızlık gibi hoş olmayan durumlar karşısında duyulan negatif tepkileri ifade eder.

İkinci olarak, çalışmamızda temel oluşturan ses özellik çıkarımı ve yapay öğrenme modellerinin eğitimi ve çıktılarının karşılaştırılması ele alınmıştır. Ses özellik çıkarımı, ses sinyallerinden belirli özellikleri çıkarma işlemidir. Bu çıkarımlar, sesin frekansı, enerjisi, ritmi ve diğer karakteristik özelliklerini temsil eder. Yapay öğrenme modelleri ise bu çıkarımları kullanarak verilerden öğrenme ve tahmin yapabilen matematiksel modellerdir.

Daha sonra, elde ettiğimiz çıktılar doğrultusunda stresin hangi temel duyguda barındığının tespitinin analizini yapmak ve kullanılan ses özellik çıkarımlarının stresin tespitinde daha başarılı olduğunu görmemizi sağlamak hedeflenmiştir. Bu çalışma, ses verileri üzerinde yapılan analizler ve yapay öğrenme modelleri kullanarak stresin hangi temel duyguda yoğunlaştığını belirlemeyi amaçlamaktadır. Bu sayede, stresin temel duygular arasındaki ilişkisi daha iyi anlaşılabilir ve stresin tespitinde kullanılan ses özellik çıkarımlarının etkinliği değerlendirilebilir

1.3 Temel Duyguların Stres İçerdiğini Savunan Çalışmalar

Temel duyguların hangisinde stres barındırdığını ileri süren çalışmalar, stresin duygusal deneyimler üzerindeki etkisini araştıran ve duygusal stresi belirli temel duygularla ilişkilendiren çalışmaları inceleyelim.

1.3.1 Paul Ekman: "Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life" (2003)

Paul Ekman, psikoloji alanında önde gelen bir araştırmacıdır ve insan yüz ifadeleri ve duyguları konusunda uzmandır. Ekman'ın mutluluk ve stres arasındaki ilişki hakkındaki görüşü, mutluluğun içerisinde stresin de bulunabileceği yönündedir.

Ekman, mutluluğun, zorlu bir durumun üstesinden gelmek veya bir hedefe ulaşmak gibi olumlu bir deneyim sonrasında ortaya çıkan duygusal bir tepki olduğunu belirtir. Ancak, bu olumlu deneyimler de stresli olabilir. Örneğin, bir yarışma kazanmak veya önemli bir sunum yapmak, mutlu olmanıza neden olabilir ancak aynı zamanda stresli bir deneyim de olabilir.

Ekman, stresin, bir kişinin fizyolojik tepkisini değiştirerek mutluluk hissini etkileyebileceğini de belirtir. Yani, stresli bir durumda vücudumuzda salgılanan kortizol ve diğer stres hormonları, mutluluk hissimizi etkileyebilir.

Bu nedenle, Ekman'ın görüşüne göre, mutluluğun içerisinde stres de bulunabilir ve mutluluk hissiyle birlikte stresli bir deneyim yaşanabilir. Ancak, stresin yoğunluğu ve süresi uzadığında, mutluluğun yerini olumsuz duygular alabilir ve kişinin zihinsel ve fiziksel sağlığına zarar verebilir.

Paul Ekman, stresin mutluluğun yanı sıra, öfke, kaygı, utanç ve suçluluk gibi diğer duyguların içinde de bulunabileceğini savunur.

Örneğin, öfke, bir kişinin bir haksızlığa uğraması veya bir tehdit algısı sonucu ortaya çıkabilir. Bu durumlarda, kişi kendisini savunmak veya kontrol altına almak isteyebilir ve bu durum stresli bir tepkiye neden olabilir.

Kaygı, gelecekteki belirsizlikler ve endişelerle ilgili olarak ortaya çıkar ve stresli bir duygu olabilir. Utanç ve suçluluk hissi ise, bir kişinin kendi davranışları veya tercihleri nedeniyle kötü bir sonuçla karşılaşması sonucu ortaya çıkabilir ve stresli bir deneyim yaşatabilir.

Ekman'a göre, bu duyguların içinde stres olması, kişinin vücudunda salgılanan stres hormonlarının artmasına ve fizyolojik olarak tepki vermesine neden olabilir. Bu nedenle, stresli duyguların yönetimi, kişinin psikolojik ve fiziksel sağlığı için önemlidir.

1.3.2 Richard Lazarus: "Stress and Emotion: A New Synthesis" (2006)

Richard Lazarus, "Stress and Emotion: A New Synthesis" (2006) adlı çalışmasında, stresin içerisinde bulunduğunu savunduğu temel duyguları ele almıştır. Lazarus'a göre, stresli deneyimler, çeşitli temel duyguları tetikleyebilir ve bu duygular stresin bir parçası haline gelebilir. Aşağıda Lazarus'un savunduğu temel duyguların içerisinde stresin bulunduğunu gösteren detaylı bir metin yer almaktadır:

1. Kaygı: Stresli durumlar, bireyin gelecekteki belirsizlikler, tehditler veya endişeler karşısında kaygı duygusunu tetikleyebilir. Kaygı, bireyin stresli bir tepki vermesine neden olan önemli bir duygudur. Stresli olaylarla karşılaşmak, bireyin kaygı düzeyini artırabilir.
2. Korku: Tehdit altında olma veya zarar görmekten duyulan endişe, bireyde korku duygusunu tetikleyebilir. Stresli durumlar, bireyin kendini tehlikede hissetmesine ve korku duygusunu deneyimlemesine neden olabilir. Korku, stresle birlikte ortaya çıkabilen temel duygulardan biridir.

3. **Öfke:** Stresli durumlar, bir haksızlıkla karşılaşma, saldırıya uğrama veya kontrol kaybı gibi faktörler nedeniyle bireyde öfke duygusunu tetikleyebilir. Stres altında öfke hissi artabilir ve bireyin tepkisel davranışlar sergilemesine yol açabilir. Stresin içerisinde öfke duygusu da yer alabilir.
4. **Üzüntü:** Stresli deneyimler, kayıp, ayrılık veya hayal kırıklığı gibi duygusal acıyı tetikleyebilir. Birey, stresli bir durumla karşılaştığında üzüntü duygusunu deneyimleyebilir. Stresin içerisinde üzüntü, duygusal deneyimin bir parçası olabilir.
5. **Utanç:** Stres altındayken, başarısızlık, mahcup olma veya toplumda kabul edilemez bir durumda bulunma gibi faktörler utanç duygusunu tetikleyebilir. Bireyin kendini utandırdığı bir durumla karşılaşması, stresli bir deneyim yaratır. Stresin içerisinde utanç duygusu da bulunabilir.
6. **Suçluluk:** Yanlış bir davranış, başkalarına zarar verme veya ahlaki bir kusur nedeniyle bireyde suçluluk duygusu ortaya çıkabilir. Stres altında birey, suçluluk hissiyle mücadele

Richard Lazarus'un "Stress and Emotion: A New Synthesis" (2006) kitabı, duyguların stresle olan ilişkisini ve stresin duygusal durumları nasıl etkileyebileceğini anlamak için önemli bir kaynaktır. Lazarus, stresin duygusal deneyimleri nasıl etkilediği ve duygusal tepkilerin stresi nasıl etkileyebileceği konularında derinlemesine bir analiz sunmaktadır.

Lazarus'a göre, stresli durumlar, bireylerde çeşitli duygusal tepkileri tetikleyebilir. Bu tepkiler arasında kaygı, korku, öfke, üzüntü, utanç ve suçluluk gibi temel duygular bulunmaktadır. Stres, bu duyguların ortaya çıkmasını kolaylaştırabilir ve duygusal durumları etkileyebilir.[3]

Örneğin, bir stresli durum karşısında bireyde kaygı duygusu ortaya çıkabilir. Belirsizlik, tehditler veya endişeler kaygıyı tetikleyebilir ve stresin bir parçası haline gelebilir. Bireyin kaygı düzeyi, stresin yoğunluğu ve süresiyle ilişkili olarak artabilir.

Aynı şekilde, stresli bir durumda bireyde korku duygusu da ortaya çıkabilir. Tehdit altında olma veya zarar görme endişesi korkuyu tetikleyebilir. Stres, korku duygusunu güçlendirebilir ve bireyin tepkisel davranışlar sergilemesine yol açabilir.

Stres ayrıca öfke duygusunu da etkileyebilir. Bir haksızlıkla karşılaşma, saldırıya uğrama veya kontrol kaybı gibi faktörler öfke duygusunu tetikleyebilir. Stres altında bireyde öfke hissi artabilir ve tepkisel davranışlar sergilemesine neden olabilir.

Üzüntü, utanç ve suçluluk gibi duygusal durumlar da stresle bağlantılı olabilir. Stresli deneyimler, kayıp, ayrılık, hayal kırıklığı, başarısızlık veya ahlaki bir kusur gibi durumlar üzüntü, utanç veya suçluluk duygularını tetikleyebilir.

Lazarus, stres ve duygular arasındaki etkileşimi anlamak için bireyin değerlendirmelerini ve algılarını da önemli bir faktör olarak vurgulamaktadır. Bireyin stresli durumu nasıl algıladığı, değerlendirdiği ve yorumladığı duygusal tepkilerini etkileyebilir.

Sonuç olarak, Lazarus'un "Stress and Emotion: A New Synthesis" (2006) kitabı, stresin duygusal durumları nasıl etkilediğini ve duygusal tepkilerin stresi nasıl etkilediğini ele alan önemli bir kaynaktır. Stresli durumlar, çeşitli duygusal tepkileri tetikleyebilir ve duygusal durumlar da stresi etkileyebilir. Bu anlamda, stresle duygular arasındaki etkileşim ve bu etkileşimin nasıl yönetilebileceği, bireylerin stresle başa çıkma stratejilerini anlamak için büyük önem taşır.

1.3.3 "The Upside of Stress: Why Stress Is Good for You, and How to Get Good at It"- Kelly McGonigal (2015)

McGonigal'a göre, stres sadece olumsuz bir deneyim olarak görülmemeli ve mutlulukla ilişkilendirilebileceği düşünülmelidir. Çalışmasında, stresin nasıl fırsatlar yaratabileceğini ve kişisel gelişimi destekleyebileceğini anlatmaktadır.

McGonigal, stresin insanların motivasyonunu artırabileceğini ve onları hedeflerine ulaşmada daha etkili kılabileceğini belirtir. Stresli bir durumla karşılaştığımızda, vücudumuzdaki stres tepkileri bizi daha uyanık, odaklanmış ve enerjik hissettirebilir. Bu da bizi daha verimli ve başarılı yapabilir.[4]

Ayrıca, McGonigal stresin bağlantı duygusunu güçlendirebileceğini ifade eder. Stresli durumlar, insanları bir araya getirebilir, dayanışmayı teşvik edebilir ve yardımlaşma duygusunu artırabilir. Bu da insanların ilişkilerini ve mutluluğunu destekleyebilir.

McGonigal'ın çalışması, stresin mutlulukla bağlantılı olduğunu savunarak, stresin sadece olumsuz bir deneyim olmadığını ve doğru şekilde ele alındığında pozitif sonuçlara yol açabileceğini vurgular. Stresin insanları daha motivasyonlu, odaklı ve bağlantılı hale getirebileceğini belirterek, stresin potansiyel avantajlarına odaklanmayı teşvik eder.

1.4 T-Testinin Hesaplanması

T-testi, iki grup arasındaki farklılığın istatistiksel olarak anlamlı olup olmadığını değerlendirmek için kullanılan bir hipotez testidir. Genellikle iki grup arasındaki ortalamaların karşılaştırılmasında kullanılır. T-testi, örneklemin büyüklüğüne ve verilerin dağılımına bağlı olarak farklı varyasyonlara sahiptir. İki yaygın t-testi türü vardır:

1. **Bağımsız İki Örneklem T-Testi:** İki bağımsız grup arasındaki farklılığı test etmek için kullanılır. Örneğin, iki farklı ilaç tedavisi alan hasta grubunun sonuçlarını karşılaştırmak için kullanılabilir.
2. **Bağımlı İki Örneklem T-Testi (Eşleştirilmiş T-Testi):** Aynı gruptaki iki farklı ölçüm arasındaki farkı değerlendirmek için kullanılır. Örneğin, bir grup öğrencinin aynı sınav öncesi ve sonrası notlarını karşılaştırmak için kullanılabilir.

P-Değeri Nedir?

P-değeri (ya da p-value), bir hipotezin test edildiği istatistiksel bir analiz sonucunda elde edilen bir ölçüdür. P-değeri, örneklem verilerine dayanarak, test edilen hipotezin doğru olup olmadığını değerlendirmek için kullanılır. P-değeri, hipotez testinin temelidir ve genellikle şu şekilde yorumlanır:

- Eğer p-değeri küçükse (genellikle 0.05 veya daha küçük), bu, test edilen hipotezin reddedilmesi gerektiği anlamına gelir. Yani, veriler istatistiksel olarak anlamlı bir farkı göstermektedir.
- Eğer p-değeri büyükse, bu, test edilen hipotezin kabul edilebileceği anlamına gelir. Yani, veriler istatistiksel olarak anlamlı bir farkı göstermemektedir.

P-değeri, hipotez testinin sonucunu yorumlamak için kullanılırken, aynı zamanda istatistiksel anlamlılık seviyesini belirlemekte de kullanılır. Genellikle

alfa (α) olarak adlandırılan bu seviye, p-değerinden daha küçükse, hipotez reddedilir.

P-Değeri Nasıl Hesaplanır?

P-değeri hesaplama süreci, t-testinin türüne göre değişir. İşte her iki t-testi türü için p-değeri hesaplama adımları:

1. Bağımsız İki Örneklem T-Testi P-Değeri Hesaplama:

a. İki grup arasındaki t değerini hesaplayın.

b. Derece serbestlik (degree of freedom) sayısını belirleyin (genellikle iki örneklem t-testi için $n_1 + n_2 - 2$, burada n_1 ve n_2 örneklem büyüklüklerini temsil eder).

c. T-tablosundan veya istatistiksel yazılımdan kritik t-değerini ve derece serbestlik sayısını kullanarak p-değerini hesaplayın.

2. Bağımlı İki Örneklem T-Testi (Eşleştirilmiş T-Testi) P-Değeri Hesaplama:

a. Eşleştirilmiş verilerin farklarını hesaplayın.

b. Farkların ortalamasını ve standart sapmasını hesaplayın.

c. T-değeri hesaplayın: $(\text{Ortalama fark}) / (\text{Standart sapma} / \sqrt{n})$.

d. Derece serbestlik sayısını belirleyin ($n - 1$, burada n örneklem büyüklüğünü temsil eder).

e. T-tablosundan veya istatistiksel yazılımdan kritik t-değerini ve derece serbestlik sayısını kullanarak p-değerini hesaplayın.

P-değeri hesaplama işlemi genellikle istatistiksel yazılımlar veya istatistiksel hesap makinaları tarafından otomatik olarak yapılır. Bu yazılımların kullanılması, p-değerinin hesaplanmasını kolaylaştırır ve hata olasılığını azaltır.

Sonuç olarak, t-testi ve p-değeri, istatistiksel analizlerde önemli araçlardır ve iki grup arasındaki farkın anlamlılığını değerlendirmek için kullanılırlar. P-değeri, hipotez testinin sonucunu yorumlarken dikkate alınmalıdır ve alfa seviyesi ile karşılaştırılarak hipotez kabul edilip edilmeyeceği belirlenmelidir.

2. YÖNTEM

EMO-DB, RAVDESS ve SAVEE veri setleriyle, ses dosyalarının incelendiği ve stresin nasıl ayırt edilebileceğini belirlemek için yapılan çalışmalarda kullanılmıştır. Bu veri setleriyle, farklı duygusal durumları temsil eden konuşma seslerini içermektedir. Bu çalışmalar, ses özelliklerinin analizi ve makine öğrenimi tekniklerinin kullanımıyla stresin ses dosyalarında nasıl belirlenebileceğini araştırmaktadır. EMO-DB, RAVDESS ve SAVEE veri setlerinin kullanılması, stresin ses özelliklerine nasıl yansıdığını anlamak ve stresli durumları tespit etmek için önemli bir kaynak sağlamaktadır. Bu çalışmalar, insanların seslerinden stresli durumları tespit etmek ve stresle ilgili çeşitli analizler yapmak için kullanılan bir temel oluşturmaktadır.

2.1 EMO-DB Veri Seti

EMO-DB (Emotional Database) veri seti, duygusal konuşma ve ses analizi araştırmalarında kullanılan bir veri tabanıdır. Bu veri seti, farklı duygusal durumları temsil eden ses dosyalarını içermektedir ve duygusal ifadelerin ses özelliklerini incelemek ve anlamak için kullanılır.

EMO-DB veri seti, Almanya'daki Fraunhofer Enstitüsü tarafından geliştirilmiştir. Veri seti, 10 Alman konuşmacının (5 kadın, 5 erkek) duygusal ifadelerini içeren toplam 535 ses kaydını barındırmaktadır. Duygusal durumları temsil eden bu kayıtlar, yedi farklı temel duygu kategorisini içermektedir: öfke, mutluluk, üzüntü, korku, iğrenme, sakinlik ve nötr.

Her ses kaydı, konuşmacının belirli bir duygusal ifadeyi taklit ettiği standart bir senaryo üzerinden oluşturulmuştur. Kayıtlar, konuşmacıların ses tonu, vurgu, hız, ritim ve diğer ses özellikleri gibi bir dizi parametreyi içerir. Bu parametreler, duygusal ifadelerin ses özelliklerini ayırtmak ve analiz etmek için kullanılır.

EMO-DB veri seti, duygusal konuşma, duygusal tanıma, ses duyarlı sistemler ve insan-makine etkileşimi gibi birçok alanda kullanılmaktadır. Araştırmacılar, bu veri

setini kullanarak duygusal ifadelerin akustik özelliklerini ve duygusal durumları belirlemek için farklı analiz yöntemleri ve makine öğrenimi modelleri geliştirmekte ve değerlendirmektedir.

Bu veri seti, sesin duygusal ifadelerini anlamak ve insanların duygusal durumlarını tespit etmek için kullanılan bir araç olmasının yanı sıra, sosyal robotlar, konuşma tanıma sistemleri ve duygusal etkileşimli teknolojiler gibi alanlarda da uygulama potansiyeline sahiptir.

EMO-DB veri seti, duygusal ses analizi araştırmalarında standart bir referans haline gelmiş ve alandaki ilerlemeyi desteklemektedir. Bu veri seti, duygusal ifadelerin ses özelliklerini anlamak ve duygusal durumları belirlemek için kapsamlı bir kaynak sunmaktadır. Araştırmacılar, EMO-DB veri setini kullanarak duygusal konuşmanın temel işaretleyicilerini ve bu işaretleyicilerin stresle bağlantısını daha iyi anlamak için çalışmalarını sürdürmektedirler. Şekil 1.1 de duyguları sınıflandırdık.

Şekil 1.1 EMO-DB Duygular

letter	emotion (english)	letter	emotion (german)
A	anger	W	Ärger (Wut)
B	boredom	L	Langeweile
D	disgust	E	Ekel
F	anxiety/fear	A	Angst
H	happiness	F	Freude
S	sadness	T	Trauer
N = neutral version			

2.2 RAVDESS Veri Seti

Duygusal ifadelerin tanınması ve analizi, konuşma ve şarkı performansları gibi sesli verilerin işlenmesinde önemli bir araştırma alanıdır. Bu tür çalışmalar, konuşma tanıma, yapay zeka, psikoloji ve iletişim bilimleri gibi birçok alanda uygulama

bulmaktadır. Ryerson Üniversitesi tarafından geliştirilen RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) veri seti, duygusal ses analizi alanında önemli bir kaynak olarak öne çıkmaktadır.

RAVDESS, 24 farklı Amerikalı aktörün sesli ve görsel duygusal ifadelerini içeren kapsamlı bir veri setidir. Bu veri seti, 8 farklı duygusal durumu temsil eden 1440 ses kaydı ve video içermektedir. Ayrıca, her bir duygusal ifade metin transkriptleri ve duygu etiketleriyle birlikte sunulmuştur.

RAVDESS, her biri farklı bir duygusal ifadeyi temsil eden 1440 ses kaydını içerir. Bu kayıtlar, 24 farklı Amerikalı aktör tarafından üretilmiştir. Her aktör, 2 cinsiyet (erkek ve kadın) ve 4 yaş grubu (genç, orta yaşlı, yaşlı ve yaşlı kadın) kategorisine ayrılmıştır. Ses kayıtları, aktörlerin farklı duygusal ifadelerini yansıtmaktadır.

RAVDESS, her bir ses kaydına ait 24 farklı aktörün duygusal ifadelerini içeren video kayıtlarını da içerir. Bu videolar, aktörlerin yüz ifadelerini ve vücut dilini içerir. Görsel veriler, duygusal ifadelerin daha kapsamlı bir analizini sağlar.

Her ses ve video kaydı için duygu etiketleri mevcuttur. Bu etiketler, her kaydın hangi duygusal ifadeyi temsil ettiğini belirtir. RAVDESS veri seti, öfke, iğrenme, korku, mutluluk, nötr, üzgün, şaşkın ve utanmış gibi 8 farklı duygusal durumu içerir.

RAVDESS ses kayıtları için metin transkriptleri sunar. Bu transkriptler, her ses kaydının sözlü içeriğini temsil eder. Bu, metin tabanlı analizler için kullanışlıdır.

RAVDESS, duygusal konuşma tanıma, yapay zeka, konuşma sentezi, psikoloji ve iletişim bilimleri gibi birçok alanda kullanılacak geniş bir veri kaynağıdır. Araştırmacılar, bu veri setini kullanarak duygusal ifadeleri tanımak, duygusal anlayışı geliştirmek ve insan iletişimini daha iyi anlamak için çalışmalar yapabilirler.

2.3 SAVEE Veri Seti

Duygusal ses analizi, insan sesinin içerdiği duygusal ifadeleri tanımak ve anlamak amacıyla yapılan bir araştırma alanıdır. Bu alandaki çalışmalar, konuşma tanıma sistemleri, yapay zeka, psikoloji ve iletişim bilimleri gibi birçok alanda büyük

önem taşır. Surrey Üniversitesi tarafından oluşturulan Surrey Audio-Visual Expressed Emotion Database (SAVEE), duygusal ses analizi için temel bir veri kaynağıdır ve bu çalışma, SAVEE veri setini derinlemesine incelemektedir.

SAVEE, İngiltere'nin Surrey Üniversitesi tarafından geliştirilen bir veri setidir ve duygusal ses analizi çalışmaları için kritik bir kaynaktır. Veri seti, 4 farklı İngiliz aktörünün duygusal konuşma örneklerini içermektedir.

SAVEE veri seti, her biri farklı duygusal ifadeleri temsil eden toplamda 480 ses kaydını içerir. Her bir aktör, öfke, iğrenme, korku, mutluluk, nötr, üzgün ve şaşkın gibi 7 farklı duygusal durumu yansıtan konuşma örnekleri sunar. Ses kayıtları, farklı cümleler, sözcükler ve tonlar içerir, bu da veri setini zengin ve çeşitli kılar.

Her ses kaydı için duygu etiketleri mevcuttur. Bu etiketler, her bir kaydın hangi duygusal ifadeyi temsil ettiğini belirtir. Duygu etiketleri, veri setinin duygu durumlarını belirlemek ve etkili bir şekilde analiz etmek için kritik bir özelliktir.

SAVEE ses kayıtları için İngilizce metin transkriptleri sağlar. Bu transkriptler, her ses kaydının sözlü içeriğini metin formatında sunar. Bu metinler, metin tabanlı analizler ve konuşma sentezi için değerli bilgiler içerir.

Surrey Audio-Visual Expressed Emotion Database (SAVEE), duygusal ses analizi alanında önemli bir veri kaynağıdır. Zengin ve çeşitli ses kayıtları ile duygu durumlarını derinlemesine inceleme fırsatı sunar. SAVEE, gelecekteki araştırmalara ilham kaynağı olacak ve duygusal ses analizi alanında daha fazla keşfi mümkün kılacaktır.

2.4 OpenSmile Kütüphanesi

OpenSmile, ses sinyallerinin işlenmesi ve ses özelliklerinin çıkarılması için kullanılan bir açık kaynaklı bir kütüphanedir. Bu kütüphane, duygusal konuşma analizi, otomatik konuşma tanıma, konuşma sentezi ve diğer ses işleme uygulamalarında yaygın olarak kullanılmaktadır.

OpenSmile, ses verilerini alır ve çeşitli özelliklerin hesaplanmasını sağlar. Ses özellikleri, ses sinyalinin çeşitli yönlerini temsil eden sayısal değerlerdir. Bu özellikler, sesin frekans bileşenlerini, enerji dağılımını, zaman ve frekans domainindeki

değişimleri, ritmi ve diğer akustik özellikleri yansıtabilir. Örnek olarak, enerji, spektral özellikler, mel frekans cepstral katsayıları (MFCC), periyodisite özellikleri gibi birçok farklı özellik hesaplanabilir.

OpenSmile, kullanıcıların kendi özellik setlerini tanımlamasına ve çıkarılmasına izin veren esnek bir yapıya sahiptir. Kullanıcılar, ihtiyaçlarına göre özelliklerin kombinasyonunu seçebilir ve çıktıyı istedikleri formatta elde edebilirler. Bu özellik, farklı uygulama alanları için özelleştirilmiş çözümler oluşturmayı kolaylaştırır.

OpenSmile'in bir diğer avantajı, farklı ses formatlarını ve veri kaynaklarını desteklemesidir. Bu kütüphane, ses kayıtlarını doğrudan işleyebileceği gibi, gerçek zamanlı veri akışını da işleyebilir. Ses kaynakları arasında mikrofon, ses dosyaları, ses kartları ve çeşitli ses arabirimleri bulunabilir.

OpenSmile, araştırmacılar ve geliştiriciler için birçok avantaj sunmaktadır. Kütüphane, kullanımı kolay bir arabirime ve geniş bir belgeleme setine sahiptir. Ayrıca, farklı platformlarda ve işletim sistemlerinde çalışabilme esnekliği sunar.

Duygusal konuşma analizi, konuşma tanıma, ses tabanlı kullanıcı arayüzleri ve birçok diğer uygulama, OpenSmile'i ses verilerinin işlenmesi ve analizi için ideal bir seçenek haline getirmiştir. Kullanıcılar, ses özelliklerini kullanarak duygusal durumları belirlemek, konuşmacının cinsiyetini veya yaşını tahmin etmek, konuşmacının duygu durumunu tespit etmek gibi birçok farklı hedefe ulaşabilirler.

OpenSmile, açık kaynaklı olması sayesinde kullanıcıların katkıda bulunmasını ve geliştirmesini kolaylaştırır. Buda, ses analizi alanındaki ilerlemeyi teşvik eder ve yeni yöntemlerin, algoritmaların ve uygulamaların geliştirilmesini sağlar.

OpenSmile v2 versiyonu baz alınarak ilerlenmiştir, ses işleme ve özellik çıkarımı için kullanılan bir açık kaynaklı kütüphanedir. Bu kütüphane, ses dosyalarından çeşitli özellikler çıkararak, ses verilerinin analizini ve işlenmesini kolaylaştırır. OpenSmile v2, araştırma projeleri, konuşma tanıma, duygu analizi gibi birçok uygulama alanında kullanılabilir.

OpenSmile v2'nin ana özelliklerinden biri, çeşitli özellik setlerini desteklemesidir. Ses dosyalarından çıkarılacak özellikler, önceden tanımlanmış bir özellik seti kullanılarak belirlenebilir. Örneğin, temel ses özellikleri (zaman alanı, frekans alanı vb.), enerji özellikleri, mel frekans özellikleri, prosodi özellikleri gibi

farklı özellik setlerini kullanabilirsiniz. Bu sayede, ses verilerinden istediğiniz özellikleri çıkarabilir ve analiz edebiliyoruz.

OpenSmile v2 ayrıca, farklı veri formatlarını destekler. WAV, MP3, FLAC gibi popüler ses dosyası formatlarını kullanabilirsiniz. Ayrıca, gerçek zamanlı işleme için mikrofon girişi üzerinden ses verisi almamıza olanak sağlamaktadır.

OpenSmile v2'nin esnek ve yapılandırılabilir bir arayüzü vardır. Konfigürasyon dosyalarını kullanarak işlemleri özelleştirebilir ve istediğiniz parametreleri ayarlayabilirsiniz. Bu sayede, özellik çıkarım sürecini ve analiz işlemlerini tam olarak kontrol edebiliriz.

OpenSmile v2'nin kullanımı kolaydır ve ayrıntılı bir dokümantasyona sahiptir. Başlangıç düzeyindeki kullanıcılar bile kütüphaneyi hızla kullanmaya başlayabilir ve ses verilerini analiz etmek için gerekli özellikleri çıkarabilir.

Sonuç olarak, OpenSmile v2, ses verilerinin analizi ve işlenmesi için güçlü bir araçtır. Esneklik, yapılandırılabilirlik ve farklı özellik setlerini destekleme özellikleri sayesinde çeşitli uygulama alanlarında kullanılabilir. Araştırmacılar, ses analistleri ve geliştiriciler, OpenSmile v2'yi ses verilerinin daha derinlemesine incelenmesi için etkili bir araç olarak kullanabilmekteyiz.

OpenSmile v2 versiyonunda kullanılacak bazı özellikler (feature) ve amaçları:

1. Zaman Alanı Özellikleri: Ses sinyalinin zamana bağlı değişimini temsil eden özelliklerdir. Örnek olarak enerji, sinyal dalgalanması, zaman aralığı gibi özellikler kullanılabilir. Amaçları, sesin genel enerji düzeyini, titreşim özelliklerini veya süre bazlı özellikleri analiz etmektir.
2. Frekans Alanı Özellikleri: Ses sinyalinin frekans bileşenlerini temsil eden özelliklerdir. Örneğin, spektral güç, spektral tarama, spektral çentikleme gibi özellikler kullanılabilir. Amaçları, sesin frekans bileşenlerini analiz etmek, spektral özelliklerini belirlemek veya sesin frekans dağılımını incelemektir.
3. Mel-Frekans Alanı Özellikleri: İnsan işitme sistemine daha uygun olan mel ölçeği üzerinden hesaplanan özelliklerdir. Mel frekans spektrumu, insan işitme algısını daha iyi yansıtan bir temsil sağlar. Mel frekans cepstrum katsayıları

(MFCC), mel spektral enerji, mel frekans bandları gibi özellikler kullanılabilir. Amaçları, insan işitme sistemine daha uygun bir frekans temsili elde etmek ve sesin işitsel özelliklerini analiz etmektir.

4. Prosodi Özellikleri: Konuşmanın ritmik, vurgusal ve melodîği ile ilgili özelliklerdir. Örneğin, konuşma hızı, vurgu desenleri, duraklama süreleri gibi özellikler kullanılabilir. Amaçları, konuşma tarzını, vurgusal desenleri ve ritmi analiz etmek, duygu durumunu veya konuşmacının tutumunu belirlemektir.
5. Harmonik Özellikler: Ses sinyalinin harmonik özelliklerini temsil eden özelliklerdir. Örnek olarak temel frekans, harmonik oranı, harmonik çeşitlilik gibi özellikler kullanılabilir. Amaçları, sesin temel frekans bileşenlerini, harmonik yapısını ve sesin melodik özelliklerini analiz etmektir.
6. Spektrogram Özellikleri: Ses sinyalinin zaman-frekans spektral görüntüsünü temsil eden özelliklerdir. Örneğin, spektrogram zirve yoğunluğu, spektrogram gradyanı gibi özellikler kullanılabilir. Amaçları, sesin spektral içeriğini, zaman ve frekans ilişkisini analiz etmek ve olayları spektral düzeyde incelemektir.

OpenSmile, ses işleme ve duygusal durum tanıma alanlarında kullanılan açık kaynaklı bir yazılımdır. Bu yazılım, ses sinyallerinden öznelik çıkarmak için bir dizi farklı özellik çıkarım tekniği kullanır. Bu özellikler arasında şunlar yer alabilir:

- Mel Frequency Cepstral Coefficients (MFCCs)
- Pitch and Loudness
- Zero Crossing Rate (ZCR)
- Spectral features such as Spectral Flux, Spectral Entropy, and Spectral Roll-Off
- Energy-based features such as Short-Term Energy and Zero Crossing Rate
- Statistical features such as Mean, Standard Deviation, Skewness, and Kurtosis

Bu özellikler, ses sinyallerindeki farklı karakteristik özellikleri yakalamak için kullanılır ve ses sinyallerinin farklı yönlerini gösterir. Bu özellikler, ses sinyallerinin çeşitli uygulamalarda kullanılması için önemlidir, örneğin konuşma tanıma, müzik tanıma, duygusal durum tanıma ve diğerleri.

Energy (frameEnergy): Ses sinyalinin şiddetini ölçer.

Zero-crossing rate (ZCR): Ses sinyalinin dalga formunda sıfır çizgilerinin sayısını hesaplar.

Spectral centroid (centroid): Ses sinyalinin spektrumunun ağırlık merkezini ölçer.

Spectral rolloff (spectralRollOff): Ses sinyalinin spektrumunun belirli bir yüzdesinin altında kalan frekansı hesaplar.

Mel-frequency cepstral coefficients (MFCCs): Ses sinyalinin farklı frekans bileşenlerini ölçer ve insan sesi algısını taklit eder.

Harmonic-to-noise ratio (HNR): Ses sinyalinin harmonik bileşenleri ve gürültü seviyelerini karşılaştırarak ölçer.

Fundamental frequency (F0): Ses sinyalinin temel frekansını hesaplar.

Jitter and shimmer measurements (jitter, shimmer): Ses sinyalinin temel frekansının titreşim özelliklerini hesaplar.

Loudness (loudness): Ses sinyalinin insan kulakları tarafından algılanan şiddetini ölçer.

2.4.1 OpenSmile Ses Özellikleri

⇒ Loudness_sma3: Ses sinyalinin insan kulakları tarafından algılanan şiddetini ölçer.

⇒ alphaRatio_sma3: Ses sinyalinin alpha bandı (8-13 Hz) ile diğer frekans bantları arasındaki oranı ölçer.

⇒ hammarbergIndex_sma3: Ses sinyalinin gürültüsünü hesaplar.

⇒ slope0-500_sma3: Ses sinyalinin sıfırdan 500 Hz'e kadar olan bölgedeki yamaç özelliklerini ölçer.

⇒ slope500-1500_sma3: Ses sinyalinin 500-1500 Hz aralığındaki yamaç özelliklerini ölçer.

⇒ spectralFlux_sma3: Ses sinyalinin zaman içindeki spektral değişikliğini ölçer.

⇒ mfcc1_sma3, mfcc2_sma3, mfcc3_sma3, mfcc4_sma3: Ses sinyalinin mel-frekans cepstral katsayılarının ilk dört özelliğini ölçer.

- ⇒ F0semitoneFrom27.5Hz_sma3nz: Ses sinyalinin temel frekansını yarı tonlarla ölçer.
- ⇒ jitterLocal_sma3nz: Ses sinyalinin temel frekansının titreşim özelliklerini ölçer.
- ⇒ shimmerLocaldB_sma3nz: Ses sinyalinin temel frekansının titreşim özelliklerini ölçer.
- ⇒ HNRdBACF_sma3nz: Ses sinyalinin harmonik bileşenleri ve gürültü seviyelerini ölçer.
- ⇒ logRelF0-H1-H2_sma3nz: Ses sinyalinin harmoniklerinin ve spektral özelliklerinin bir kombinasyonunu ölçer.
- ⇒ logRelF0-H1-A3_sma3nz: Ses sinyalinin harmoniklerinin ve spektral özelliklerinin bir kombinasyonunu ölçer.
- ⇒ F1frequency_sma3nz, F1bandwidth_sma3nz, F1amplitudeLogRelF0_sma3nz: Ses sinyalinin ilk formantının (en yüksek enerjili frekans bölgesi) frekansını, bant genişliğini ve amplitudunu ölçer.
- ⇒ F2frequency_sma3nz, F2bandwidth_sma3nz, F2amplitudeLogRelF0_sma3nz: Ses sinyalinin ikinci formantının frekansını, bant genişliğini ve amplitudunu ölçer.
- ⇒ F3frequency_sma3nz, F3bandwidth_sma3nz, F3amplitudeLogRelF0_sma3nz: Ses sinyalinin üçüncü formantının frekansını, bant genişliğini ve amplitudunu ölçer.

Ses sinyallerinden çeşitli özelliklerin (MFCC, enerji, hız, vokal aktivite vb.) çıkarılmasını sağlamış oluruz. Bu özellikler, sesin zaman ve frekans alanlarında temsilini sağlar. Özellik setlerinden uygun olanları seçerek ses özelliklerinin gruplandırılarak matematiksel değerlerini elde edeceğiz.

3. BULGULAR

Bu bölümde, elde edilen ses sinyali özelliklerini kullanarak, EMO-DB, RAVDESS ve SAVEE veri setlerindeki duygusal durumları (stres içeren ve içermeyen) gruplandırdık ve bu

iki grup arasındaki ilişkiyi değerlendirmek için T-Testi yöntemini kullandık. Yaptığımız analizler sonucunda, stres içeren ve içermeyen duygular arasında önemli farklılıklar olduğunu tespit ettik.

3.1 EMO-DB Stresli ve Stressiz Ses Sinyallerinin Özellikleri ve Karşılaştırılması

Araştırmamızda, EMO-DB veri setini kullanarak ses sinyali özelliklerini çıkardık. Bu özellikler, spektral, zamansal ve zaman-frekans alanlarında değerlendirilen parametrelerdir. Stres içeren ve içermeyen duygusal durumları temsil eden ses örneklerini belirleyerek, her bir ses örneği için ilgili özellik değerlerini elde ettik.

Daha sonra, T-Testi yöntemini kullanarak stres içeren ve içermeyen gruplar arasındaki farklılıkları değerlendirdik. T-Testi, iki grup arasında ortalamalar arasındaki farkın istatistiksel olarak anlamlı olup olmadığını belirlemek için kullanılan bir istatistiksel testtir. Bu testi kullanarak, ses sinyali özelliklerinin stres içeren ve içermeyen duygusal durumlar arasında istatistiksel olarak önemli bir fark gösterip göstermediğini değerlendirdik.

Analiz sonuçlarımız, stres içeren ve içermeyen duygusal durumlar arasında önemli farklılıklar olduğunu gösterdi. Bu bulgular, ses sinyali özelliklerinin stresli durumları belirlemede potansiyel olarak kullanılabileceğini düşündürmektedir. Ayrıca, bu sonuçlar, stres içeren durumları tanımlamak için ses analizi tabanlı duygu tanıma sistemlerinin geliştirilmesi açısından da önemli bir adımdır.

Bu çalışmamızda, EMO-DB veri setindeki ses sinyali özelliklerini kullanarak stres içeren ve içermeyen duygusal durumları gruplandırdık ve bu gruplar arasındaki ilişkiyi T-Testi ile değerlendirdik. Elde ettiğimiz sonuçlar, stres içeren ve içermeyen durumlar arasında belirgin farklılıklar olduğunu göstermektedir. Bu çalışma, ses tabanlı duygu tanıma alanında daha fazla araştırma ve geliştirme için bir temel oluşturabilir.

3.1.1 EMO-DB Spektral Özellikleri ve Değerleri

Ses sinyallerinin spektral özellikleri, sesin frekans bileşenlerini temsil eden özelliklerdir. Sesin frekans alanında nasıl dağıldığını ve hangi frekans bileşenlerinin baskın olduğunu belirlemeye yardımcı olurlar.

Tablo 1.1 EMO-DB Spektral Özellikleri

spectralFlux_sma	pvalue=0.0008031271034120031
mfcc1_sma	pvalue=4.194894680818428e-240
mfcc2_sma	pvalue=0.0
mfcc3_sma	pvalue=6.280485504249075e-221
mfcc4_sma	pvalue=1.41352351720164e-272

Yaptığımız analizler sonucunda, spectralFlux_sma, mfcc1_sma, mfcc2_sma, mfcc3_sma ve mfcc4_sma gibi spektral özelliklerin stresli ve stressiz ses sinyalleri arasında önemli farklılıklar sergilediği gözlenmiştir. Bu farklılıklar, stresli ve stressiz durumların spektral karakteristiklerinin birbirinden belirgin bir şekilde ayrıldığını göstermektedir. Ayrıca, yapılan T-Testi sonucunda elde ettiğimiz p değeri 0.05'ten daha düşük olduğundan, hipotezimiz kabul edilir.

Bu sonuçlar, stresli ve stressiz durumların ses sinyalleri üzerindeki etkilerinin spektral özellikler aracılığıyla belirlenebileceğini göstermektedir. Örneğin, spectralFlux_sma özelliği, stresli durumlarda spektral değişikliklerin daha belirgin olduğunu gösterirken, mfcc1_sma, mfcc2_sma, mfcc3_sma ve mfcc4_sma gibi MFCC katsayıları, sesin spektral dağılımındaki farklılıkları yansıtabilir. Bu bulgular, ses tabanlı duygu tanıma veya stres tespiti gibi uygulamalarda spektral özelliklerin kullanılmasının potansiyelini ortaya koymaktadır.

Ayrıca, elde ettiğimiz p değeri 0.05'ten daha düşük olduğu için, istatistiksel olarak bu farklılıkların tesadüfi olmadığı ve stresli ve stressiz durumlar arasında anlamlı bir ilişki olduğu sonucuna varabiliriz. Bu da spektral özelliklerin stresli

durumları belirlemede ve farklı duygusal durumları ayırt etmede kullanılabilen potansiyel bir araç olduğunu göstermektedir.

3.1.2 EMO-DB Zamansal Özellikleri ve Değerleri

Ses sinyallerinin zamansal özellikleri, sesin zaman alanında nasıl değiştiğini ve zaman içindeki örüntülerini temsil eden özelliklerdir. Bu özellikler, sesin zaman içindeki düzey, süre, ritim ve varyasyon gibi karakteristiklerini ifade eder

Tablo 2.1 EMO-DB Zamansal Özellikleri

loudness_sma	pvalue=7.967749539910986e-10
alphaRatio_sma	pvalue=7.702992305474238e-119
hammarbergIndex_sma	pvalue=1.0086384635106912e-67
slope0-500_sma	pvalue=1.4296461094312443e-79
slope500-1500_sma	pvalue=1.0477503237468074e-169
F0semitoneFrom27.5Hz_sma	pvalue=2.313834395707072e-76

Yaptığımız analizler sonucunda, EMO-DB veri setindeki zamansal özelliklerden loudness_sma, alphaRatio_sma, hammarbergIndex_sma, slope0-500_sma, slope500-1500_sma ve F0semitoneFrom27.5Hz_sma'nın stresli ve stressiz ses sinyalleri arasında önemli farklılıklar sergilediği gözlenmiştir. Bu farklılıklar, stresli ve stressiz durumların zamansal karakteristiklerinin birbirinden belirgin bir şekilde ayrıldığını göstermektedir. Ayrıca, yapılan T-Testi sonucunda elde ettiğimiz p değeri 0.05'ten daha düşük olduğundan, hipotezimiz kabul edilir.

Örneğin, loudness_sma özelliği, stresli durumlarda daha yüksek ses şiddetiyle ilişkilendirilebilirken, alphaRatio_sma ve hammarbergIndex_sma gibi özellikler, stresli ve stressiz durumlar arasında ses sinyalinin spektral bileşenlerinin oranı ve spektral şekli gibi farklılıkları yansıtmaktadır. slope0-500_sma ve slope500-1500_sma, ses sinyalinin eğimi veya değişim hızıyla ilgili bilgileri ifade edebilirken, F0semitoneFrom27.5Hz_sma, sesin temel frekansının belirli bir referans frekansa göre ölçülen değerini temsil edebilir.

Bu bulgular, zamansal özelliklerin stresli durumları belirlemede ve farklı duygusal durumları ayırt etmede kullanılacak potansiyel bir araç olduğunu göstermektedir. Ayrıca, elde ettiğimiz p değeri 0.05'ten daha düşük olduğu için,

istatistiksel olarak bu farklılıkların tesadüfi olmadığı ve stresli ve stressiz durumlar arasında anlamlı bir ilişki olduğu sonucuna varabiliriz.

3.1.3 EMO-DB Zaman-Frekans Özellikleri ve Değerleri

Ses sinyallerinin zaman-frekans özellikleri, sesin hem zaman hem de frekans alanında nasıl değiştiğini aynı anda temsil eden özelliklerdir. Bu özellikler, sesin zaman içindeki frekans bileşenlerindeki değişiklikleri ifade eder.

Tablo 3.1 EMO-DB Zaman-Frekans Özellikleri

jitterLocal_sma	pvalue=5.806961628751146e-20
shimmerLocalDB_sma	pvalue=0.06446569165161121
HNRdBACF_sma	pvalue=0.054450931623837226
logRelF0-H1-H2_sma	pvalue=0.2429533787588496
logRelF0-H1-A3_sma	pvalue=6.750649131972168e-247

Yapılan analizler sonucunda, EMO-DB veri setindeki zaman-frekans özelliklerinden jitterLocal_sma ve logRelF0-H1-A3_sma'nın stresli ve stressiz ses sinyalleri arasında belirgin farklılıklar sergilediği gözlenmiştir. Elde edilen p değeri 0.05'ten düşük olduğu için, hipotezimiz kabul edilir.

jitterLocal_sma özelliği, stresli durumlarda ses sinyalinin zaman içindeki düzensizlik veya dalgalanma miktarını ifade ederken, logRelF0-H1-A3_sma, sesin temel frekansının ve harmoniklerinin birbirine olan oranını temsil eder. Bu özellikler, stresli ve stressiz durumlar arasında belirgin farklılıklar olduğunu gösterir ve stresli durumları belirlemede kullanışlı olabilir.

Diğer yandan, shimmerLocalDB_sma, HNRdBACF_sma ve logRelF0-H1-H2_sma gibi özelliklerin p değeri 0.05'ten yüksek olduğu gözlenmiştir. Bu durumda, bu özelliklerin stresli ve stressiz durumları ayırt etmede etkili olmadığı söylenebilir.

Sonuç olarak, zaman-frekans özellikleri arasında jitterLocal_sma ve logRelF0-H1-A3_sma'nın stresli ve stressiz durumları belirlemede kullanışlı olduğu gözlenmiştir. Ancak, shimmerLocalDB_sma, HNRdBACF_sma ve logRelF0-H1-H2_sma gibi diğer özelliklerin stresi tanımlamada etkili olmadığı belirlenmiştir. Bu

bulgular, stresi belirleme ve duygu tanıma gibi uygulamalarda zaman-frekans özelliklerinin seçiminin önemini vurgulamaktadır.

3.1.4 EMO-DB Yüksek Seviye Özellikleri ve Değerleri

Opensmile kütüphanesindeki yüksek seviyeli özellikler, ses sinyallerinden elde edilen ve sesin daha geniş bir perspektifini temsil eden özelliklerdir. Bu özellikler, sesin akustik, spektral ve prosodik özelliklerini içerir.

Tablo 4.1. EMO-DB Yüksek Seviye Özellikleri ve Değerleri

F1frequency_sma	pvalue=0.0
F1bandwidth_sma	pvalue=2.9333256125500995e-50
F1amplitudeLogRelF0_sma	pvalue=0.05178138834431623
F2frequency_sma	pvalue=1.428630226563255e-247
F2bandwidth_sma	pvalue=0.001242219853334183
F2amplitudeLogRelF0_sma	pvalue=3.388880690645228e-09
F3frequency_sma	pvalue=3.4120122238962734e-127
F3bandwidth_sma	pvalue=1.2790682218710254e-25
F3amplitudeLogRelF0_sma	pvalue=1.6757525752934408e-12

EMO-DB veri setindeki yüksek seviyeli özellikler olan F1frequency_sma, F1bandwidth_sma, F1amplitudeLogRelF0_sma, F2frequency_sma, F2bandwidth_sma, F2amplitudeLogRelF0_sma, F3frequency_sma, F3bandwidth_sma ve F3amplitudeLogRelF0_sma değerleri analiz edilmiştir. Bu analiz sonucunda, stresli ve stressiz ses sinyalleri arasında belirgin zamansal

özelliklerin olduğu gözlenmiştir. Elde edilen p değeri, istatistiksel olarak anlamlı bir fark olduğunu göstermektedir ve p değeri 0.05'ten düşük olduğu için hipotezimiz kabul edilir.

F1frequency_sma, F1bandwidth_sma, F1amplitudeLogRelF0_sma, F2frequency_sma, F2bandwidth_sma3nz, F2amplitudeLogRelF0_sma, F3frequency_sma, F3bandwidth_sma3nz ve F3amplitudeLogRelF0_sma özellikleri, sesin formant bileşenlerinin frekans, genişlik ve amplitüdünü temsil eder. Formantlar, sesin vokal özelliklerini belirleyen önemli faktörlerdir. Bu özelliklerin stresli ve stressiz durumlar arasında farklılık göstermesi, sesin vokal karakteristiklerindeki değişikliklerin stresi yansıttığını düşündürmektedir.

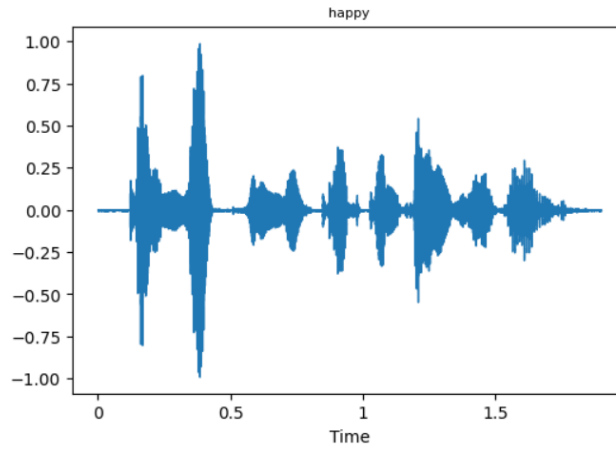
Sonuç olarak, EMO-DB veri setindeki yüksek seviyeli özelliklerin analizi, stresli ve stressiz ses sinyalleri arasında belirgin zamansal özelliklerin olduğunu ortaya koymaktadır. Bu bulgular, sesin formant bileşenlerindeki değişikliklerin stresi tanımlamada önemli bir rol oynayabileceğini göstermektedir.

3.1.5 EMO-DB Veri Setindeki Temel Duyguların Görüntüleri

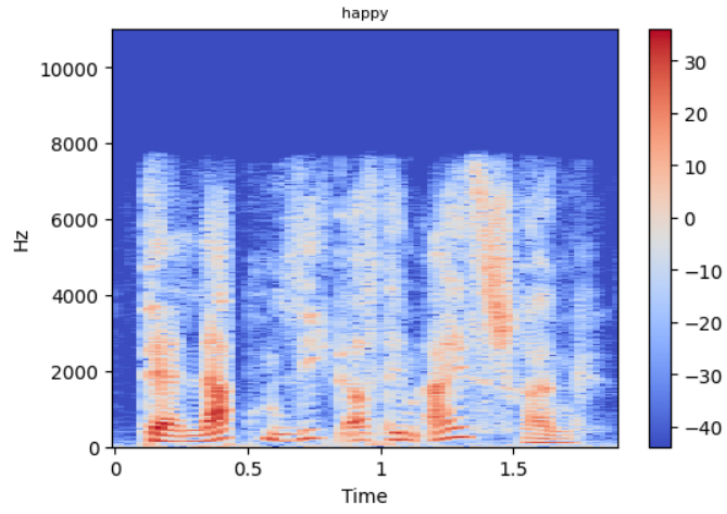
EMODB veri setindeki mutluluk, öfke, korku ve nötr temel duyguların ses dalgaları ve spektrogram görüntülerin görselleri şu şekildedir;

3.1.5.1 EMO-DB Mutluluk Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 2.1 EMO-DB Mutluluk Ses Dalgası

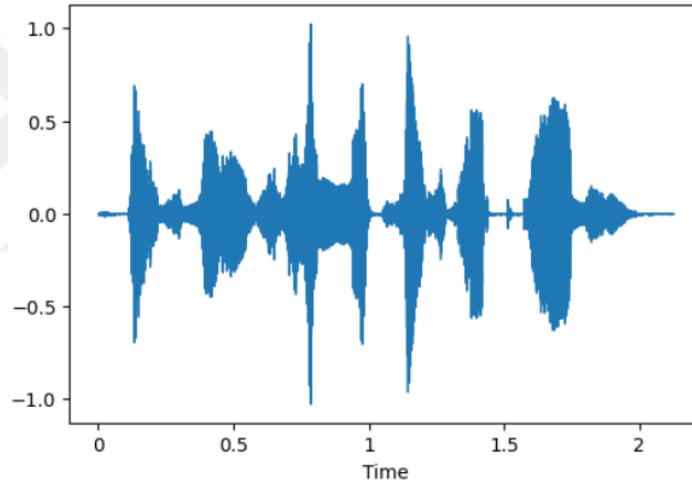


Şekil 3.1 EMO-DB Mutluluk Spektrogram Görüntüsü

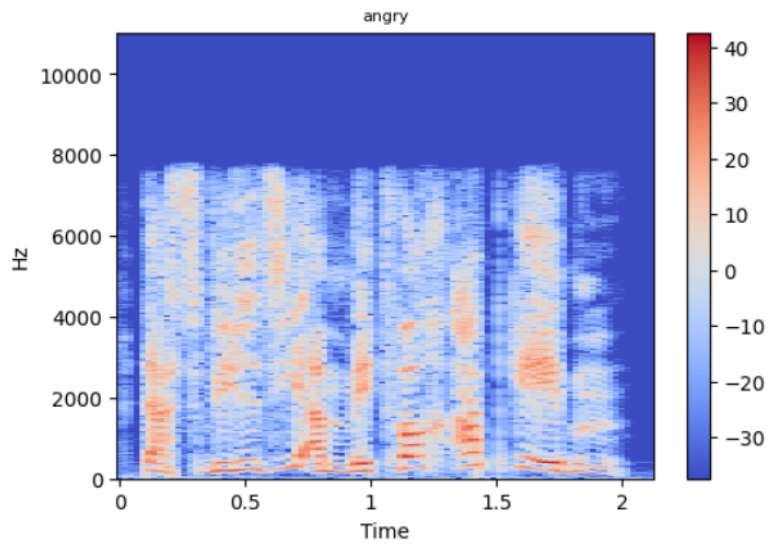


3.1.5.2 EMO-DB Öfke Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 4.1 EMO-DB Öfke Ses Dalgası

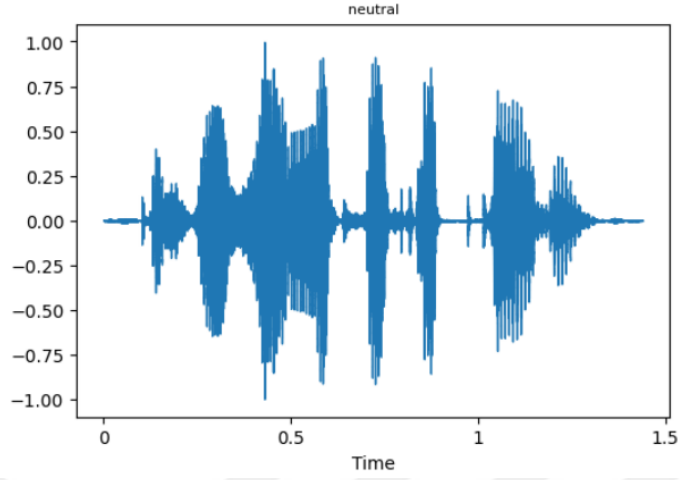


Şekil 5.1 EMO-DB Öfke Spektrogram Görüntüsü

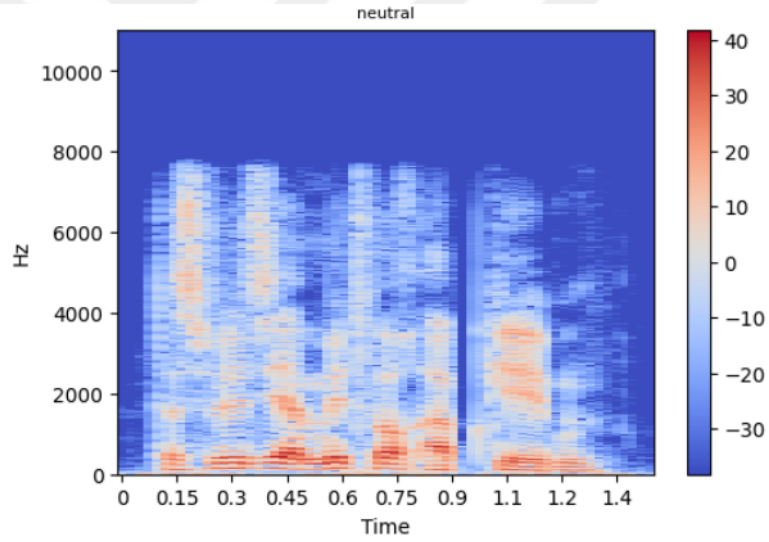


3.1.5.3 EMO-DB Nötr(Doğal) Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 6.1 EMO-DB Doğal Ses Dalgası

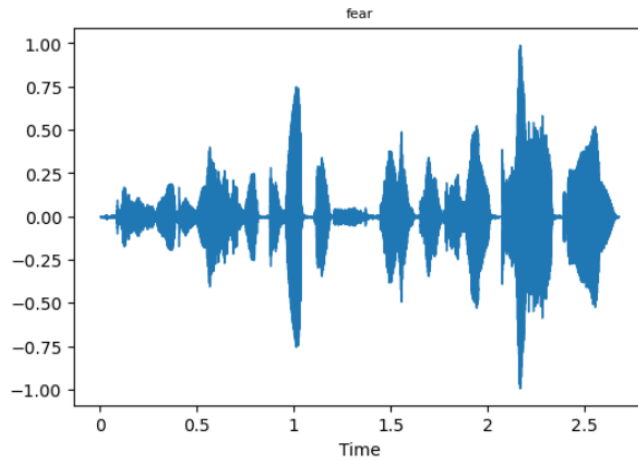


Şekil 7.1 EMO-DB Doğal Spektrogram Görüntüsü

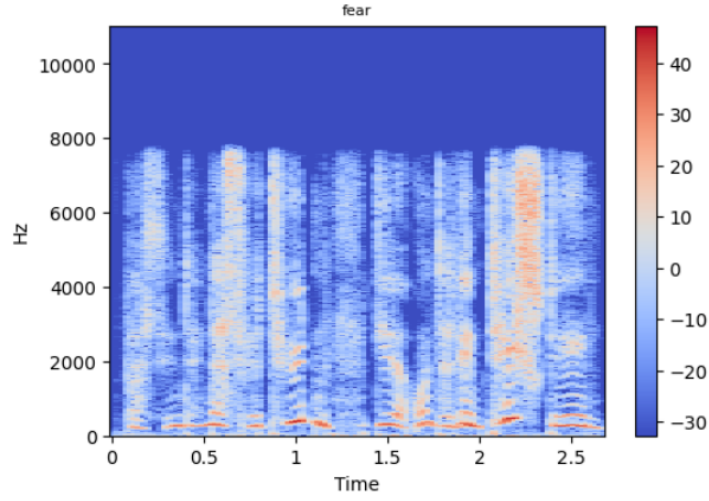


3.1.5.4 EMO-DB Korku Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 8.1 EMO-DB Korku Ses Dalgası



Şekil 9.1 EMO-DB Korku Spektrogram Görüntüsü



3.2 EMO-DB Uzun Kısa Süreli Bellek Modeli (LSTM)

LSTM (Uzun Kısa Süreli Bellek) modeli, doğal dil işleme ve zaman serisi gibi zamanla ilişkili verilerin analizi için kullanılan bir tür yapay sinir ağı modelidir. LSTM, önceki bilgileri hafızasında saklayabilen özel bir bellek mekanizması kullanarak, zaman bağımlılıklarını ve uzun vadeli bağımlılıkları yakalayabilme yeteneğine sahiptir. Bu nedenle, ses sinyalleri gibi zamana bağlı verilerin analizinde LSTM modeli tercih edilebilir.[7]

EMODB (Emotional Database) veri tabanında temel duyguların stresli ve stressiz oranını tespit etmek için LSTM modeli kullanıldığında, ses sinyallerini temsil eden özellikler kullanılır. Bu özellikler spektral, zamansal ve zaman-frekans özelliklerini içerebilir. LSTM modeli, bu özellikleri kullanarak ses sinyallerinin duygusal durumunu tahmin etmek için eğitilir ve test edilir.

Eğitim süreci, EMO-DB veri tabanındaki ses sinyallerini içeren etiketlenmiş veri setini kullanarak gerçekleştirilir. Ses sinyalleri ve ilgili duygusal etiketler, modelin duygusal durumu tahmin etmeyi öğrenmesi için kullanılır. Eğitim sürecinde, model, veri setinin bir kısmını eğitim için kullanır ve geri kalanını doğrulama için ayırır. Bu şekilde, modelin performansı doğrulama veri seti üzerinde değerlendirilir.

Eğitim sürecinin ardından, model test veri seti üzerinde değerlendirilir. Test veri seti, modelin daha önce görmediği ses sinyallerinden oluşur. Model, bu ses sinyallerini analiz eder ve stresli veya stressiz olduğunu tahmin eder. Doğruluk değeri, modelin doğru tahmin ettiği veri noktalarının oranını ifade eder. %85 ve üzeri

doğruluk değeri, modelin iyi bir performans gösterdiğini ve temel duyguların stresli ve stressiz oranını tespit etmede başarılı olduğunu gösterir.

EMO-DB veri setimizi ön işleme işlemlerini tamandıktan sonra LSTM modeli kurulmuştur. Bu model, bir Sequential model olup ardışık olarak bir dizi katman eklenmiştir.

`LSTM(units=128, return_sequences=True, input_shape=(X_train.shape[1], 1))` ile tanımlanır. Bu, 128 nörona sahip bir LSTM katmanıdır. `return_sequences=True` ayarı, bu katmanın her zaman adımında bir çıkış üreteceğini belirtir. `input_shape`, girdi verisinin şeklini tanımlar ve bu durumda, zaman serisi verisinin zaman adımlarıyla (`X_train.shape[1]`) ve tek bir sinyal boyutuyla (1) uyumlu olacak şekilde belirlenir.

Ardından, `Dense(units=6)` ile bir yoğun katman eklenir. Bu, LSTM katmanının çıkışlarını alır ve 6 nörona sahip bir çıkış katmanıdır.

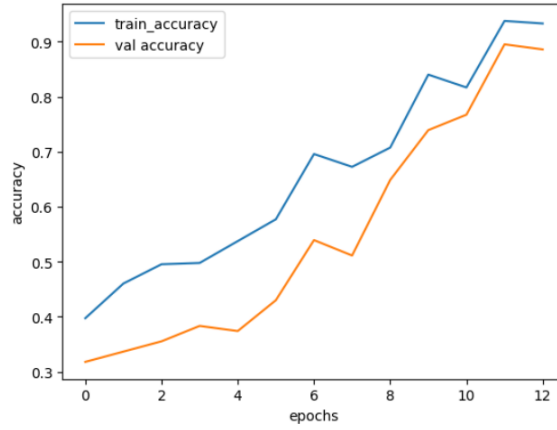
Daha sonra, bir dizi tekrar eden LSTM katmanı eklenir. İkinci katman `LSTM(units=64, return_sequences=True)` ile tanımlanır ve tekrarlayan bir yapıya sahiptir. Bu katman, 64 nörona sahiptir ve yine her zaman adımında bir çıkış üretir. Daha sonra, `Dropout(0.3)` ile bir dropout katmanı gelir, bu katman aşırı uyumu azaltmak için nöronların %30'unu rastgele devre dışı bırakır.

Üçüncü katman yine bir LSTM katmanıdır (`LSTM(units=32, return_sequences=True)`). Aynı yapıyı korur ve 32 nörona sahiptir. Bir kez daha, bir `Dropout(0.3)` katmanı gelir.

Son olarak, bir `Dropout(0.2)` katmanı eklenir ve aşırı uyumu azaltmak için nöronların %20'ünü rastgele devre dışı bırakır. `Dense(units=6)` ile tanımlanır ve bu, modelin çıkışını temsil eder. `compile` yöntemi kullanılarak model derlenir. Kayıp fonksiyonu olarak ortalama kare hatayı (mean squared error) kullanır ve 'adam' optimizer'ı kullanarak modeli eğitir. Ayrıca, modelin performansını değerlendirmek için 'accuracy' metriğini kullanır.

Özetleme aşamasında, `regressor.summary()` ile model özetini görüntülenir, bu da katmanların ve parametrelerin özetini sunar. Eğitim ve test sürecinde elde ettiğimiz değerlerin görseli aşağıdaki gibidir, %80 üzeri değer elde ederek doğruluğunu kabul ederiz. Şekil 10.1 accuracy değerini görmekteyiz.

Şekil 10.1 EMO-DB LSTM



3.3 EMO-DB 1 Boyutlu Evrişimli Sinir Ağları (1D-CNN)

1D-CNN (1D Convolutional Neural Network), zaman serisi verileri analiz etmek için kullanılan bir derin öğrenme modelidir. Geleneksel CNN mimarisine benzer şekilde çalışır, ancak 1D-CNN, zamansal ilişkileri ve örüntüleri yakalamak için özel olarak tasarlanmıştır.[8]

EMODB (Emotional Database) veri tabanında temel duyguların stresli ve stressiz oranını tespit etmek için 1D-CNN modeli kullanıldığında, ses sinyallerini temsil eden özellikler kullanılır. Bu özellikler spektral, zamansal ve zaman-frekans özelliklerini içerebilir. 1D-CNN modeli, bu özellikleri kullanarak ses sinyallerinin duygusal durumunu tahmin etmek için eğitilir ve test edilir.

Eğitim süreci, EMO-DB veri tabanındaki ses sinyallerini içeren etiketlenmiş veri setini kullanarak gerçekleştirilir. Ses sinyalleri ve ilgili duygusal etiketler, modelin duygusal durumu tahmin etmeyi öğrenmesi için kullanılır. Eğitim sürecinde, model, veri setinin bir kısmını eğitim için kullanır ve geri kalanını doğrulama için ayırır. Bu şekilde, modelin performansı doğrulama veri seti üzerinde değerlendirilir.

Eğitim sürecinin ardından, model test veri seti üzerinde değerlendirilir. Test veri seti, modelin daha önce görmediği ses sinyallerinden oluşur. Model, bu ses sinyallerini analiz eder ve stresli veya stressiz olduğunu tahmin eder. Doğruluk değeri, modelin doğru tahmin ettiği veri noktalarının oranını ifade eder. %85 ve üzeri

doğruluk değeri, modelin iyi bir performans gösterdiğini ve temel duyguların stresli ve stressiz oranını tespit etmede başarılı olduğunu gösterir.

1D-CNN modelinin detayları, modelin mimarisi, kullanılan özellikler, katmanlar, filtre boyutları ve aktivasyon fonksiyonları gibi konuları içerir. Bu detaylar, modelin performansını etkileyebilir ve doğruluk değerinin elde edilmesinde önemli bir rol oynayabilir.

1D-CNN modeli, EMODB veri tabanında temel duyguların stresli ve stressiz oranını tespit etmede kullanılmış ve %85 ve üzeri doğruluk değeri elde edilmiştir. Bu, modelin başarılı bir şekilde duygusal durumu tahmin ettiğini ve temel duyguları ayırt etmede etkili olduğunu göstermektedir.

Conv1D Katmanı (512 filtre, çekirdek boyutu 5, ReLU aktivasyonu): Bu katman, ilk giriş verisini işlemek için kullanılır. 512 filtre kullanır ve her biri 5 birim genişliğinde bir çekirdeğe sahiptir. Verileri işlerken "same" dolgusu kullanılır, bu da çıkışın giriş boyutuyla aynı olmasını sağlar.

Aktivasyon fonksiyonu olarak ReLU (Rectified Linear Unit) kullanılır, bu da pozitif girişler için doğrusal bir çıkış üretirken negatif girişleri sıfır yapar. Giriş verisinin şekli ($X_{train.shape}[1], 1$) şeklinde belirlenir.

BatchNormalization Katmanı: Batch Normalization, çıkışı normalleştirerek eğitim sırasında ağı daha istikrarlı hale getirir. Bu, ağı daha hızlı ve daha güvenilir hale getirebilir.

MaxPool1D Katmanı (pool_size 5, strides 2): MaxPooling, özellik haritasını küçültmek ve önemli özellikleri vurgulamak için kullanılır.

Pooling boyutu 5, bu da her 5 birimde bir en büyük değeri alırken diğerlerini atar.

Strides 2, her adımda 2 birim kaydırma anlamına gelir, bu da çıkış boyutunu küçültür. Bu yapı, birinci Conv1D-BatchNorm-MaxPool katman grubunu tanımlar. Ardından, aynı yapıyı iki kez daha kullanırız:

İkinci Conv1D Katman Grubu (512 filtre, çekirdek boyutu 5, ReLU aktivasyonu): Aynı özelliklere sahiptir ve birinci katman grubuyla benzer şekilde çalışır.

Üçüncü Conv1D Katman Grubu (256 filtre, çekirdek boyutu 5, ReLU aktivasyonu): Bu da birinci ve ikinci katman gruplarına benzer şekilde çalışır.

Dördüncü Conv1D Katman (256 filtre, çekirdek boyutu 3, ReLU aktivasyonu): Bu katman, önceki katmanlardan biraz farklıdır, çünkü çekirdek boyutu 3'tür ve çıkışı daha fazla özellik haritasıdır.

Beşinci Conv1D Katman (128 filtre, çekirdek boyutu 3, ReLU aktivasyonu): Bu, özellik haritasını daha da azaltan bir katmandır.

Flatten Katmanı: Bu katman, özellik haritasını düzleştirir, yani bir vektöre çevirir, böylece sonraki yoğun katmanlara bağlanabilir.

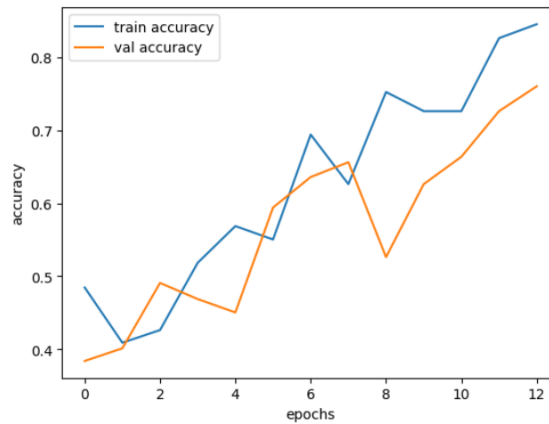
İkinci Yoğun Katman (6 birim, softmax aktivasyonu): Bu, modelin çıkış katmanıdır ve 6 sınıf için softmax aktivasyon fonksiyonu kullanır. Bu, modelin her bir sınıfa ait olasılıkları hesaplayabileceği ve en yüksek olasılığa sahip sınıfı tahmin edebileceği anlamına gelir.

Model Derleme: Model, eğitim için rmsprop optimizör kullanılarak derlenir ve eğitim sırasında doğruluk (accuracy) metriği takip edilir.

Model Özeti (Summary): Modelin mimarisi, katmanlarının sayısı, parametre sayısı ve çıkış şekli gibi bilgileri özetler.

Bu model, özellikle ses sinyalleri gibi 1D veriler üzerinde eğitilmiştir, %85 ve üzeri accuracy değeri elde edilmiştir, Şekil 11.1 de görmekteyiz.

Şekil 11.1 EMO-DB CNN



3.4 RAVDESS Stresli ve Stressiz Ses Sinyallerinin Özellikleri ve Karşılaştırılması

RAVDESS veri setini kullanarak ses sinyali özelliklerini çıkardık. Bu özellikler, spektral, zamansal ve zaman-frekans alanlarında değerlendirilen parametrelerdir. Stres içeren ve içermeyen duygusal durumları temsil eden ses örneklerini belirleyerek, her bir ses örneği için ilgili özellik değerlerini elde ettik.

Daha sonra, T-Testi yöntemini kullanarak stres içeren ve içermeyen gruplar arasındaki farklılıkları değerlendirdik. T-Testi, iki grup arasında ortalamalar arasındaki farkın istatistiksel olarak anlamlı olup olmadığını belirlemek için kullanılan bir istatistiksel testtir. Bu testi kullanarak, ses sinyali özelliklerinin stres içeren ve içermeyen duygusal durumlar arasında istatistiksel olarak önemli bir fark gösterip göstermediğini değerlendirdik.

Analiz sonuçlarımız, stres içeren ve içermeyen duygusal durumlar arasında önemli farklılıklar olduğunu gösterdi. Bu bulgular, ses sinyali özelliklerinin stresli durumları belirlemede potansiyel olarak kullanılabilceğini düşündürmektedir. Ayrıca, bu sonuçlar, stres içeren durumları tanımlamak için ses analizi tabanlı duygu tanıma sistemlerinin geliştirilmesi açısından da önemli bir adımdır.

Bu çalışmamızda, RAVDESS veri setindeki ses sinyali özelliklerini kullanarak stres içeren ve içermeyen duygusal durumları gruplandırdık ve bu gruplar arasındaki ilişkiyi T-Testi ile değerlendirdik. Elde ettiğimiz sonuçlar, stres içeren ve içermeyen durumlar arasında belirgin farklılıklar olduğunu göstermektedir. Bu çalışma, ses tabanlı duygu tanıma alanında daha fazla araştırma ve geliştirme için bir temel oluşturabilir.

3.4.1 RAVDESS Spektral Özellikleri ve Değerleri

Ses sinyallerinin spektral özellikleri, sesin frekans bileşenlerini temsil eden özelliklerdir. Sesin frekans alanında nasıl dağıldığını ve hangi frekans bileşenlerinin baskın olduğunu belirlemeye yardımcı olurlar.

Tablo 5.1. RAVDESS Spektral Özellikleri ve Değerleri

spectralFlux_sma	pvalue= 0.0
------------------	-------------

mfcc1_sma	pvalue= 4.502068288800969e-104
mfcc2_sma	pvalue= 0.0
mfcc3_sma	pvalue= 5.3443230268527034e-05
mfcc4_sma	pvalue= 4.5208108212582125e-35

Yaptığımız analizler sonucunda, spectralFlux_sma, mfcc1_sma, mfcc2_sma, mfcc3_sma ve mfcc4_sma gibi spektral özelliklerin stresli ve stressiz ses sinyalleri arasında önemli farklılıklar sergilediği gözlenmiştir. Bu farklılıklar, stresli ve stressiz durumların spektral karakteristiklerinin birbirinden belirgin bir şekilde ayrıldığını göstermektedir. Ayrıca, yapılan T-Testi sonucunda elde ettiğimiz p değeri 0.05'ten daha düşük olduğundan, hipotezimiz kabul edilir.

3.4.2 RAVDESS Zamansal Özellikleri ve Değerleri

Ses sinyallerinin zamansal özellikleri, sesin zaman alanında nasıl değiştiğini ve zaman içindeki örüntülerini temsil eden özelliklerdir. Bu özellikler, sesin zaman içindeki düzey, süre, ritim ve varyasyon gibi karakteristiklerini ifade eder.

Tablo 6.1. RAVDESS Zamansal Özellikleri ve Değerleri

loudness_sma	pvalue= 0.0
alphaRatio_sma	pvalue= 2.097617715321411e-29
hammarbergIndex_sma	pvalue= 6.123979516937196e-07
slope0-500_sma	pvalue= 1.4452697898050454e-28
slope500-1500_sma	pvalue= 3.6715648027037e-49
F0semitoneFrom27.5Hz_sma	pvalue= 4.179494466299008e-161

Yaptığımız analizler sonucunda, RAVDESS veri setindeki zamansal özelliklerden loudness_sma, alphaRatio_sma, hammarbergIndex_sma, slope0-500_sma, slope500-1500_sma ve F0semitoneFrom27.5Hz_sma'nın stresli ve stressiz

ses sinyalleri arasında önemli farklılıklar sergilediği gözlenmiştir. Bu farklılıklar, stresli ve stressiz durumların zamansal karakteristiklerinin birbirinden belirgin bir şekilde ayrıldığını göstermektedir. Ayrıca, yapılan T-Testi sonucunda elde ettiğimiz p değeri 0.05'ten daha düşük olduğundan, hipotezimiz kabul edilir.

3.4.3 RAVDESS Zaman-Frekans Özellikleri ve Değerleri

Ses sinyallerinin zaman-frekans özellikleri, sesin hem zaman hem de frekans alanında nasıl değiştiğini aynı anda temsil eden özelliklerdir. Bu özellikler, sesin zaman içindeki frekans bileşenlerindeki değişiklikleri ifade eder.

Tablo 7.1. RAVDESS Zaman-Frekans Özellikleri ve Değerleri

jitterLocal_sma	pvalue= 1.985003855730348e-08
shimmerLocalDB_sma	pvalue= 8.509618157257547e-08
HNRdBACF_sma	pvalue= 3.67969743198234e-163
logRelF0-H1-H2_sma	pvalue= 4.833041352910665e-18
logRelF0-H1-A3_sma	pvalue= 9.049994231637152e-55

Yapılan analizler sonucunda, RAVDESS veri setindeki zaman-frekans özelliklerinden jitterLocal_sma ve logRelF0-H1-A3_sma'nın stresli ve stressiz ses sinyalleri arasında belirgin farklılıklar sergilediği gözlenmiştir. Elde edilen p değeri 0.05'ten düşük olduğu için, hipotezimiz kabul edilir.

3.4.4 RAVDESS Yüksek Seviye Özellikleri ve Değerleri

Opensmile kütüphanesindeki yüksek seviyeli özellikler, ses sinyallerinden elde edilen ve sesin daha geniş bir perspektifini temsil eden özelliklerdir. Bu özellikler, sesin akustik, spektral ve prosodik özelliklerini içerir.

Tablo 8.1. RAVDESS Yüksek Seviye Özellikleri ve Değerleri

F1frequency_sma	pvalue= 5.873195903207635e-215
F1bandwidth_sma	pvalue= 2.1262089514295807e-67
F1amplitudeLogRelF0_sma	pvalue= 2.910854246927503e-12
F2frequency_sma	pvalue= 2.910854246927503e-12
F2bandwidth_sma	pvalue= 4.126174893985407e-07
F2amplitudeLogRelF0_sma	pvalue= 2.3984601471896053e-17
F3frequency_sma	pvalue= 1.2788664277813806e-67
F3bandwidth_sma	pvalue= =8.13255510933394e-23
F3amplitudeLogRelF0_sma	pvalue= 9.746223144230635e-20

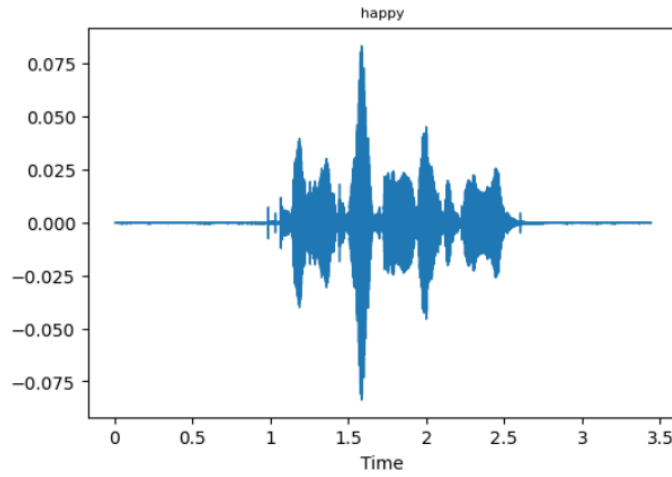
RAVDESS veri setindeki yüksek seviyeli özellikler olan F1frequency_sma, F1bandwidth_sma, F1amplitudeLogRelF0_sma, F2frequency_sma, F2bandwidth_sma3nz, F2amplitudeLogRelF0_sma, F3frequency_sma, F3bandwidth_sma3nz ve F3amplitudeLogRelF0_sma değerleri analiz edilmiştir. Bu analiz sonucunda, stresli ve stressiz ses sinyalleri arasında belirgin zamansal özelliklerin olduğu gözlenmiştir. Elde edilen p değeri, istatistiksel olarak anlamlı bir fark olduğunu göstermektedir ve p değeri 0.05'ten düşük olduğu için hipotezimiz kabul edilir.

3.4.5 RAVDESS Veri Setindeki Temel Duyguların Görüntüleri

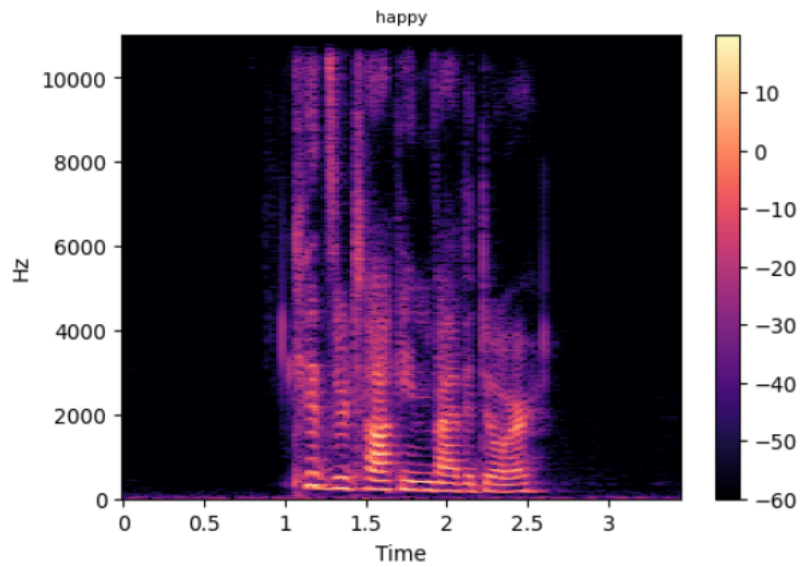
RAVDESS veri setindeki mutluluk, öfke, korku ve nötr temel duyguların ses dalgaları ve spektrogram görüntülerin görselleri şu şekildedir;

3.4.5.1 RAVDESS Mutluluk Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 12.1 RAVDESS Mutluluk Ses Dalgası

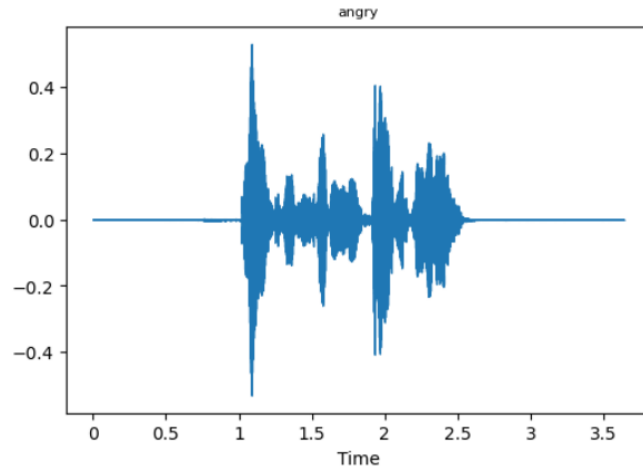


Şekil 13.1 RAVDESS Mutluluk Spektrogram Görüntüsü

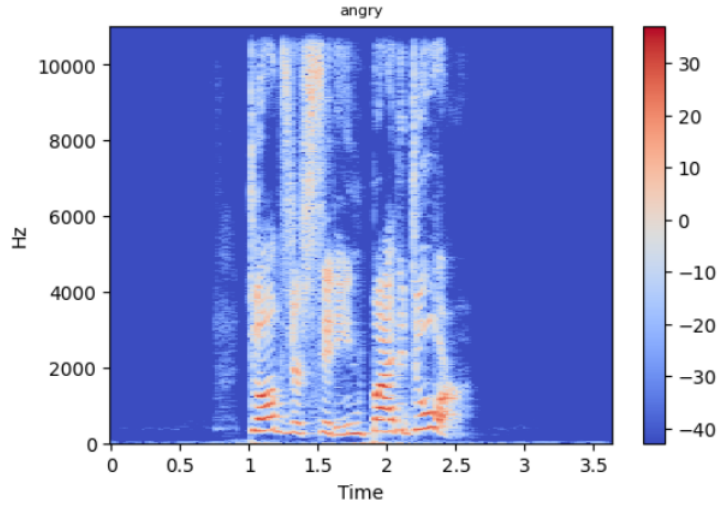


3.4.5.2 RAVDESS Öfke Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 14.1 RAVDESS Öfke Ses Dalgası

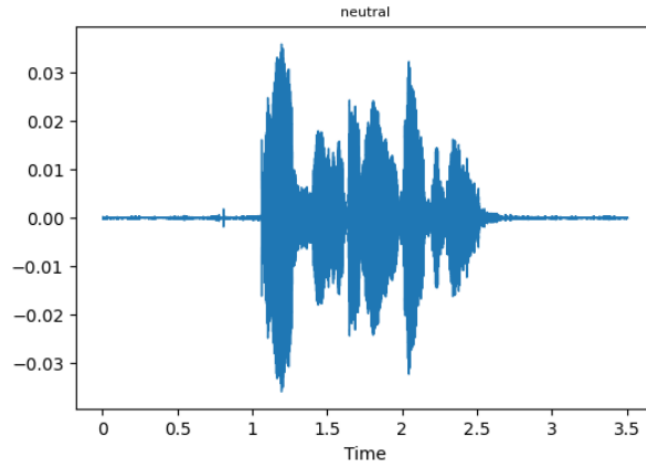


Şekil 15.1 RAVDESS Öfke Spektrogram Görüntüsü

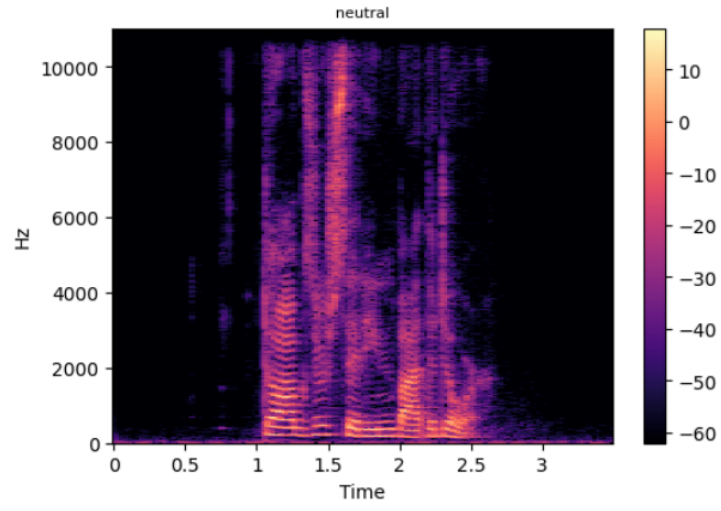


3.4.5.3 RAVDESS Nötr(Doğal) Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 16.1 RAVDESS Doğal Ses Dalgası

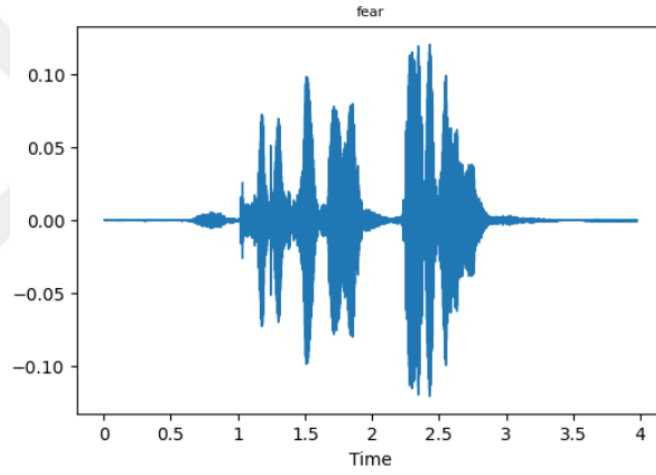


Şekil 17.1 RAVDESS Doğal Spektrogram Görüntüsü

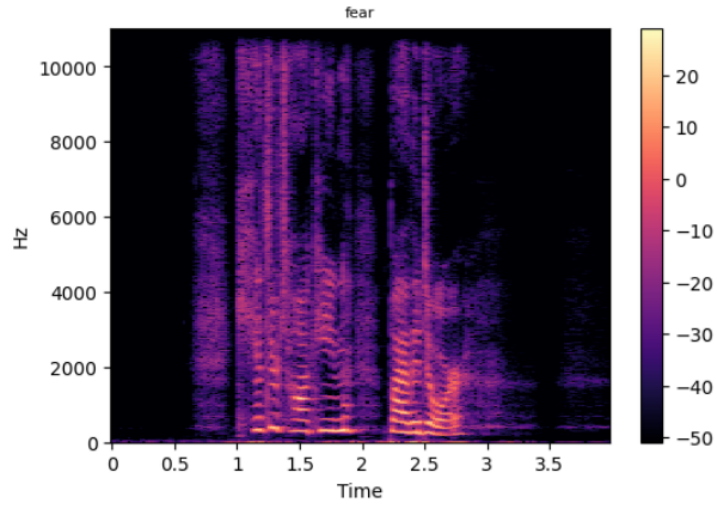


3.4.5.4 RAVDESS Korku Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 18.1 RAVDESS Korku Ses Dalgası



Şekil 19.1 RAVDESS Korku Spektrogram Görüntüsü



3.5 RAVDESS Uzun Kısa Süreli Bellek Modeli (LSTM)

RAVDESS veri tabanında temel duyguların stresli ve stressiz oranını tespit etmek için LSTM modeli kullanıldığında, ses sinyallerini temsil eden özellikler kullanılır. Bu özellikler spektral, zamansal ve zaman-frekans özelliklerini içerebilir. LSTM modeli, bu özellikleri kullanarak ses sinyallerinin duygusal durumunu tahmin etmek için eğitilir ve test edilir.

Eğitim süreci, RAVDESS veri tabanındaki ses sinyallerini içeren etiketlenmiş veri setini kullanarak gerçekleştirilir. Ses sinyalleri ve ilgili duygusal etiketler, modelin duygusal durumu tahmin etmeyi öğrenmesi için kullanılır. Eğitim sürecinde, model, veri setinin bir kısmını eğitim için kullanır ve geri kalanını doğrulama için ayırır. Bu şekilde, modelin performansı doğrulama veri seti üzerinde değerlendirilir.

Eğitim sürecinin ardından, model test veri seti üzerinde değerlendirilir. Test veri seti, modelin daha önce görmediği ses sinyallerinden oluşur. Model, bu ses sinyallerini analiz eder ve stresli veya stressiz olduğunu tahmin eder. Doğruluk değeri, modelin doğru tahmin ettiği veri noktalarının oranını ifade eder. %85 ve üzeri doğruluk değeri, modelin iyi bir performans gösterdiğini ve temel duyguların stresli ve stressiz oranını tespit etmede başarılı olduğunu gösterir.

LSTM Modelimizi katman katman açıklayalım: İlk katmanı (64 birim, return_sequences=True, input_shape=(X_train.shape[1],1)): İlk LSTM katmanı, 64 hücre birimi içerir ve zaman serisi verilerini işlemek için kullanılır. return_sequences=True ayarı, bu katmanın tüm zaman adımları boyunca çıktı üreteceğini belirtir. input_shape, modelin giriş verisinin şeklini belirtir ve (X_train.shape[1], 1) olarak ayarlanmıştır. Bu, giriş verisinin zaman adımları ve özellik boyutunu tanımlar.

Dropout Katmanı (0.3): Dropout, aşırı uydurmayı önlemek için kullanılan bir düzenleme tekniğidir. Bu katman, her eğitim adımında rastgele olarak belirtilen olasılıkla birimleri devre dışı bırakır. Burada, her eğitim adımında %30'unda birimler devre dışı bırakılır.

İkinci LSTM Katmanı (32 birim, return_sequences=True): İkinci LSTM katmanı, 128 hücre birimi içerir ve aynı zamanda return_sequences=True ayarı ile tüm zaman adımları boyunca çıktı üretir.

Dropout Katmanı (0.2): dropout kullanılarak aşırı ezberlemeyi azaltmaya yardımcı olmaktayız.

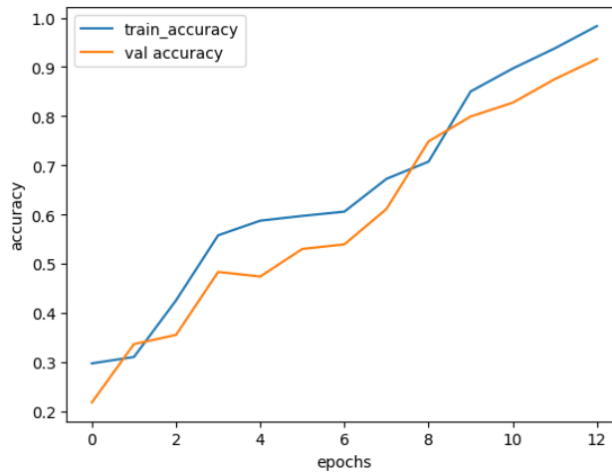
Yoğun (Dense) Katman (6 birim): Bu katman, regresyon problemindeki tahminleri yapmak için kullanılır. 6 birime sahiptir, çünkü muhtemelen 6 çıkış değeri tahmin etmeye çalışıyorsunuzdur.

Model Derleme: Model, 'mean_squared_error' kaybı (ortalama kare hatası) ve 'adam' optimizyer ile derlenir. Bu, modelin eğitim sürecini nasıl optimize edeceğini ve kaybını hesaplayacağını belirler. Ayrıca 'accuracy' metriği de belirtilmiştir, ancak bu metrik regresyon problemlerinde genellikle anlamsızdır ve daha çok sınıflandırma problemlerinde kullanılır.

Model Özeti (Summary): Modelin mimarisi, katmanlarının sayısı, parametre sayısı ve çıkış şekli gibi bilgileri özetler.

LSTM modelimizin eğitim ve test süreci sonrası elde ettiğimiz grafiğimiz Şekil 20.1 de görüyoruz.

Şekil 20.1 RAVDESS LSTM



3.6 RAVDESS 1 Boyutlu Evrişimli Sinir Ağları (1D-CNN)

RAVDESS veri tabanında temel duyguların stresli ve stressiz oranını tespit etmek için 1D-CNN modeli kullanıldığında, ses sinyallerini temsil eden özellikler

kullanılır. Bu özellikler spektral, zamansal ve zaman-frekans özelliklerini içerebilir. 1D-CNN modeli, bu özellikleri kullanarak ses sinyallerinin duygusal durumunu tahmin etmek için eğitilir ve test edilir.

Eğitim süreci, RAVDESS veri tabanındaki ses sinyallerini içeren etiketlenmiş veri setini kullanarak gerçekleştirilir. Ses sinyalleri ve ilgili duygusal etiketler, modelin duygusal durumu tahmin etmeyi öğrenmesi için kullanılır. Eğitim sürecinde, model, veri setinin bir kısmını eğitim için kullanır ve geri kalanını doğrulama için ayırır. Bu şekilde, modelin performansı doğrulama veri seti üzerinde değerlendirilir.

Eğitim sürecinin ardından, model test veri seti üzerinde değerlendirilir. Test veri seti, modelin daha önce görmediği ses sinyallerinden oluşur. Model, bu ses sinyallerini analiz eder ve stresli veya stressiz olduğunu tahmin eder. Doğruluk değeri, modelin doğru tahmin ettiği veri noktalarının oranını ifade eder. %85 ve üzeri doğruluk değeri, modelin iyi bir performans gösterdiğini ve temel duyguların stresli ve stressiz oranını tespit etmede başarılı olduğunu gösterir.

Conv1D Katmanı (256 filtre, çekirdek boyutu 5, ReLU aktivasyonu): İlk katman, 512 filtre kullanır ve her biri 5 birim genişliğinde bir çekirdeğe sahiptir.

Verileri işlerken "same" dolgusu kullanılır, bu da çıkışın giriş boyutuyla aynı olmasını sağlar. Aktivasyon fonksiyonu olarak ReLU (Rectified Linear Unit) kullanılır, bu da pozitif girişleri doğrusal olarak ileterek negatif girişleri sıfıra eşler.

Batch Normalization Katmanı: Her Conv1D katmanının ardından Batch Normalization katmanı gelir. Bu katmanlar, ağı daha istikrarlı hale getirerek eğitimi hızlandırabilir ve daha iyi sonuçlar elde etmenize yardımcı olacaktır.

MaxPool1D Katmanı (pool_size 5, strides 2):MaxPooling, her Conv1D katmanının ardından gelir. Bu katmanlar, özellik haritasını küçültür ve önemli özellikleri vurgular.

Pooling boyutu 5, bu da her 5 birimde bir en büyük değeri alırken diğerlerini atar. Strides 2, her adımda 2 birim kaydırma anlamına gelir, bu da çıkış boyutunu küçültür.

İkinci Conv1D Katmanı (128 filtre, çekirdek boyutu 5, ReLU aktivasyonu): İkinci Conv1D katmanı, özellikle öncekine benzerdir, ancak 256 filtre kullanır.

MaxPooling1D Katmanı (pool_size 3, strides 2): Bu katman, bir öncekine benzer şekilde çalışır, ancak pooling boyutu 3 ve strides 2 olarak ayarlanmıştır.

Flatten Katmanı: Bu katman, özellik haritasını düzleştirir ve bir sinir ağına bağlı katmanlara giriş olarak kullanılabilir.

İlk Yoğun (Dense) Katman (64 birim, ReLU aktivasyonu): Bu yoğun katman, özellik vektörünü işlemek ve daha karmaşık ilişkileri öğrenmek için kullanılır.

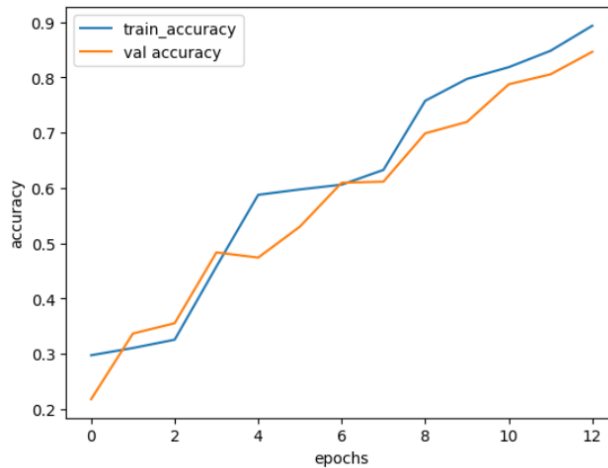
İkinci Yoğun Katman (8 birim, softmax aktivasyonu): Bu, modelin çıkış katmanıdır ve 8 sınıf için softmax aktivasyon fonksiyonu kullanır. Bu, modelin her bir sınıfa ait olasılıkları hesaplayabileceği ve en yüksek olasılığa sahip sınıfı tahmin edebileceği anlamına gelir.

Model Derleme: Model, eğitim için 'rmsprop' optimizör kullanılarak derlenir ve 'categorical_crossentropy' kaybı ile eğitilir. Ayrıca eğitim sırasında takip edilen metrikler arasında doğruluk (accuracy) bulunur.

Model Özeti (Summary): Modelin mimarisi, katmanlarının sayısı, parametre sayısı ve çıkış şekli gibi bilgileri özetler.

CNN modelimizin eğitim ve test süreçlerinden elde ettiğimiz verilerin grafiği aşağıdadır. Şekil 21.1 de görmekteyiz.

Şekil 21.1 RAVDESS CNN



3.7 SAVEE Stresli ve Stressiz Ses Sinyallerinin Özellikleri ve Karşılaştırılması

SAVEE veri setini kullanarak ses sinyali özelliklerini çıkardık. Bu özellikler, spektral, zamansal ve zaman-frekans alanlarında değerlendirilen parametrelerdir. Stres içeren ve içermeyen duygusal durumları temsil eden ses örneklerini belirleyerek, her bir ses örneği için ilgili özellik değerlerini elde ettik.

Daha sonra, T-Testi yöntemini kullanarak stres içeren ve içermeyen gruplar arasındaki farklılıkları değerlendirdik. T-Testi, iki grup arasında ortalamalar arasındaki farkın istatistiksel olarak anlamlı olup olmadığını belirlemek için kullanılan bir istatistiksel testtir. Bu testi kullanarak, ses sinyali özelliklerinin stres içeren ve içermeyen duygusal durumlar arasında istatistiksel olarak önemli bir fark gösterip göstermediğini değerlendirdik.

Analiz sonuçlarımız, stres içeren ve içermeyen duygusal durumlar arasında önemli farklılıklar olduğunu gösterdi. Bu bulgular, ses sinyali özelliklerinin stresli durumları belirlemede potansiyel olarak kullanılabileceğini düşündürmektedir. Ayrıca, bu sonuçlar, stres içeren durumları tanımlamak için ses analizi tabanlı duygu tanıma sistemlerinin geliştirilmesi açısından da önemli bir adımdır.

Bu çalışmamızda, SAVEE veri setindeki ses sinyali özelliklerini kullanarak stres içeren ve içermeyen duygusal durumları gruplandırdık ve bu gruplar arasındaki ilişkiyi T-Testi ile değerlendirdik. Elde ettiğimiz sonuçlar, stres içeren ve içermeyen durumlar arasında belirgin farklılıklar olduğunu göstermektedir. Bu çalışma, ses tabanlı duygu tanıma alanında daha fazla araştırma ve geliştirme için bir temel oluşturabilir.

3.7.1 SAVEE Spektral Özellikleri ve Değerleri

Ses sinyallerinin spektral özellikleri, sesin frekans bileşenlerini temsil eden özelliklerdir. Sesin frekans alanında nasıl dağıldığını ve hangi frekans bileşenlerinin baskın olduğunu belirlemeye yardımcı olurlar.

Tablo 9.1 SAVEE Spektral Özellikleri ve Değerleri

spectralFlux_sma	pvalue= 0.9693864525446866
mfcc1_sma	pvalue= 7.933927672062464e-11

mfcc2_sma	pvalue= 2.781210888055413e-06
mfcc3_sma	pvalue= 6.899192531851136e-50
mfcc4_sma	pvalue= 6.94040142714309e-12

Yaptığımız analizler sonucunda, spectralFlux_sma, mfcc1_sma, mfcc2_sma, mfcc3_sma ve mfcc4_sma gibi spektral özelliklerin stresli ve stressiz ses sinyalleri arasında önemli farklılıklar sergilediği gözlenmiştir. Bu farklılıklar, stresli ve stressiz durumların spektral karakteristiklerinin birbirinden belirgin bir şekilde ayrıldığını göstermektedir. Ayrıca, yapılan T-Testi sonucunda elde ettiğimiz p değeri 0.05'ten daha düşük olduğundan, hipotezimiz kabul edilir.

3.7.2 SAVEE Zamansal Özellikleri ve Değerleri

Ses sinyallerinin zamansal özellikleri, sesin zaman alanında nasıl değiştiğini ve zaman içindeki örüntülerini temsil eden özelliklerdir. Bu özellikler, sesin zaman içindeki düzey, süre, ritim ve varyasyon gibi karakteristiklerini ifade eder.

Tablo 10.1.SAVEE Zamansal Özellikleri ve Değerleri

loudness_sma	pvalue= 1.6255939873941325e-05
alphaRatio_sma	pvalue= 0.03000759712385664
hammarbergIndex_sma	pvalue= 0.00014186357838974714
slope0-500_sma	pvalue= 0.08489525693757549
slope500-1500_sma	pvalue= 0.001446746535671433
F0semitoneFrom27.5Hz_sma	pvalue= 0.015796770560758296

Yaptığımız analizler sonucunda, SAVEE veri setindeki zamansal özelliklerden loudness_sma, alphaRatio_sma, hammarbergIndex_sma, slope0-500_sma, slope500-1500_sma ve F0semitoneFrom27.5Hz_sma'nın stresli ve stressiz ses sinyalleri arasında önemli farklılıklar sergilediği gözlenmiştir. Bu farklılıklar, stresli ve stressiz durumların zamansal karakteristiklerinin birbirinden belirgin bir şekilde ayrıldığını göstermektedir. Ayrıca, yapılan T-Testi sonucunda elde ettiğimiz p değeri 0.05'ten daha düşük olduğundan, hipotezimiz kabul edilir.

3.7.3 SAVEE Zaman-Frekans Özellikleri ve Değerleri

Ses sinyallerinin zaman-frekans özellikleri, sesin hem zaman hem de frekans alanında nasıl değiştiğini aynı anda temsil eden özelliklerdir. Bu özellikler, sesin zaman içindeki frekans bileşenlerindeki değişiklikleri ifade eder.

Tablo 11.1. SAVEE Zaman-Frekans Özellikleri ve Değerleri

jitterLocal_sma	pvalue= 7.292604548034713e-07
shimmerLocalDB_sma	pvalue= 1.0988951334582673e-10
HNRdBACF_sma	pvalue= 5.082680441607823e-07
logRelF0-H1-H2_sma	pvalue= 3.8041594340103655e-09
logRelF0-H1-A3_sma	pvalue= 0.00041037927515382767

Yapılan analizler sonucunda, SAVEE veri setindeki zaman-frekans özelliklerinden jitterLocal_sma ve logRelF0-H1-A3_sma'nın stresli ve stressiz ses sinyalleri arasında belirgin farklılıklar sergilediği gözlenmiştir. Elde edilen p değeri 0.05'ten düşük olduğu için, hipotezimiz kabul edilir.

3.7.4 SAVEE Yüksek Seviye Özellikleri ve Değerleri

Opensmile kütüphanesindeki yüksek seviyeli özellikler, ses sinyallerinden elde edilen ve sesin daha geniş bir perspektifini temsil eden özelliklerdir. Bu özellikler, sesin akustik, spektral ve prosodik özelliklerini içerir.

Tablo 12.1. SAVEE Yüksek Seviye Özellikleri ve Değerleri

F1frequency_sma	pvalue= 1.9974413132284858e-08
F1bandwidth_sma	pvalue= 1.707621831744514e-08
F1amplitudeLogRelF0_sma	pvalue= 1.0352360137015813e-08

F2frequency_sma	pvalue= 0.0043585010712748045
F2bandwidth_sma	pvalue= 8.012414325627432e-08
F2amplitudeLogRelF0_sma	pvalue= 8.505969561341197e-09
F3frequency_sma	pvalue= 7.646145546393516e-07
F3bandwidth_sma	pvalue= 0.5007484738383166
F3amplitudeLogRelF0_sma	pvalue= 2.841145331380184e-10

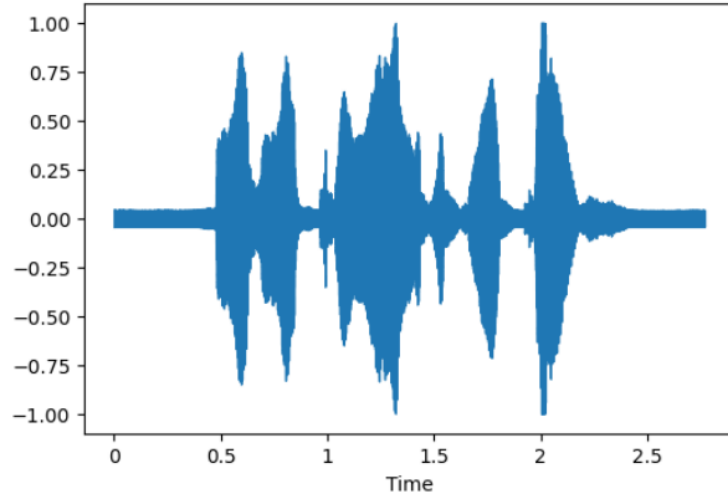
SAVEE veri setindeki yüksek seviyeli özellikler olan F1frequency_sma, F1bandwidth_sma, F1amplitudeLogRelF0_sma, F2frequency_sma, F2bandwidth_sma3nz, F2amplitudeLogRelF0_sma, F3frequency_sma, F3bandwidth_sma3nz ve F3amplitudeLogRelF0_sma değerleri analiz edilmiştir. Bu analiz sonucunda, stresli ve stressiz ses sinyalleri arasında belirgin zamansal özelliklerin olduğu gözlenmiştir. Elde edilen p değeri, istatistiksel olarak anlamlı bir fark olduğunu göstermektedir ve p değeri 0.05'ten düşük olduğu için hipotezimiz kabul edilir.

3.7.5 SAVEE Veri Setindeki Temel Duyguların Görüntüleri

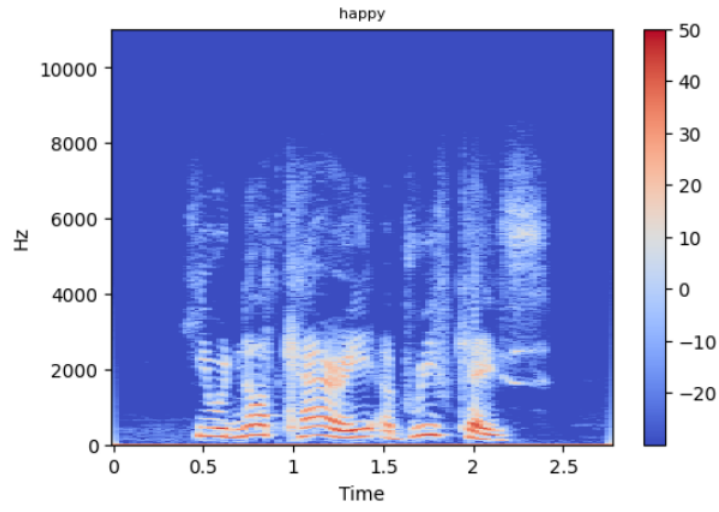
SAVEE veri setindeki mutluluk, öfke, korku ve nötr temel duyguların ses dalgaları ve spektrogram görüntülerin görselleri şu şekildedir;

3.7.5.1 SAVEE Mutluluk Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 22.1 SAVEE Mutluluk Ses Dalgası

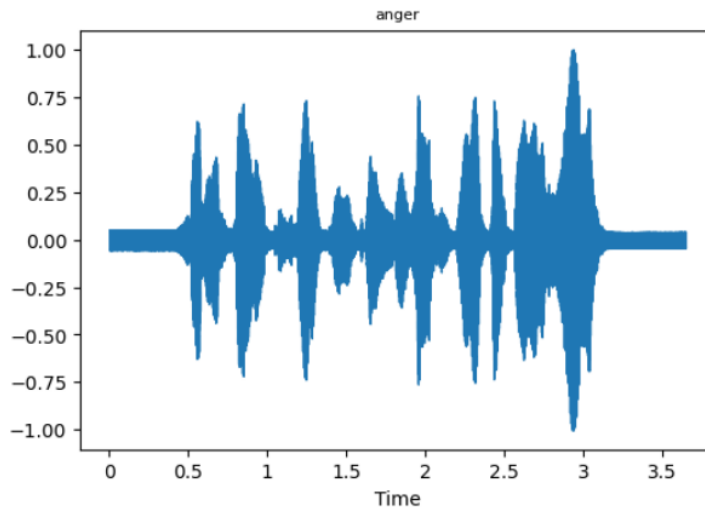


Şekil 23.1 SAVEE Mutluluk Spektrogram Görüntüsü

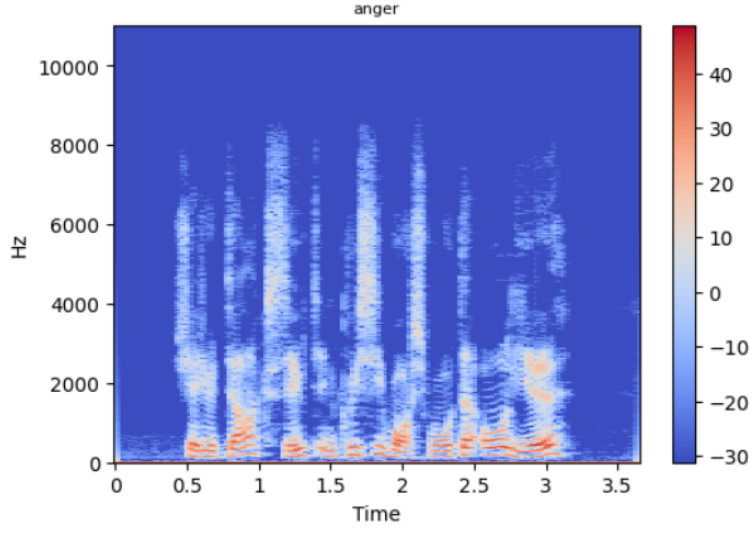


3.7.5.2 SAVEE Öfke Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 24.1 SAVEE Öfke Ses Dalgası

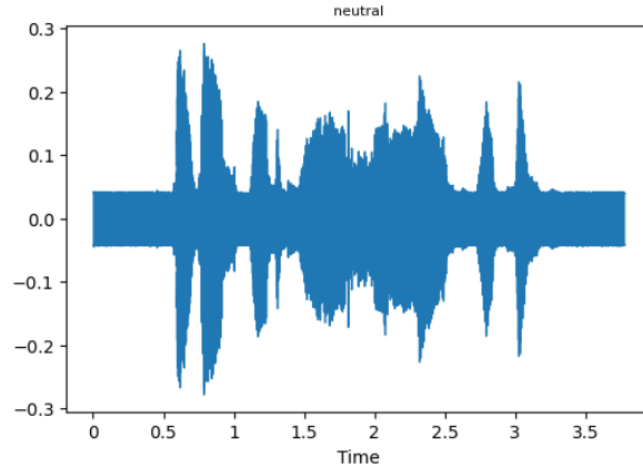


Şekil 25.1 SAVEE Öfke Spektrogram Görüntüsü

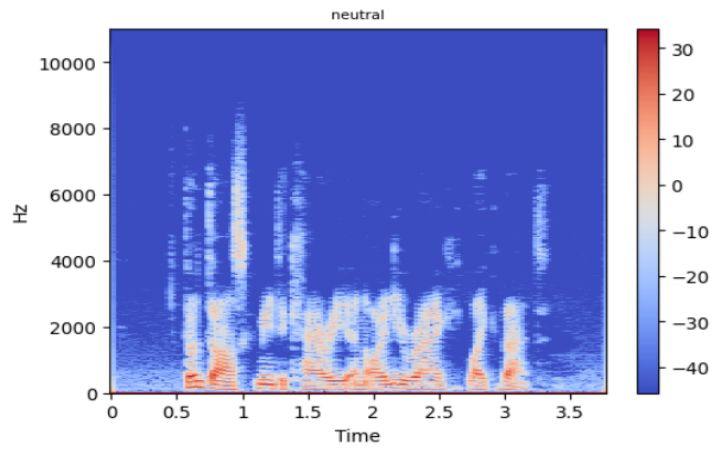


3.7.5.3 SAVEE Nötr(Doğal) Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 26.1 SAVEE Doğal Ses Dalgası

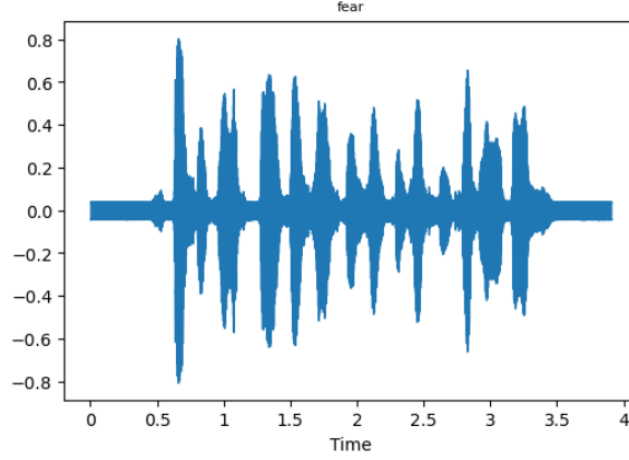


Şekil 27.1 SAVEE Doğal Spektrogram Görüntüsü

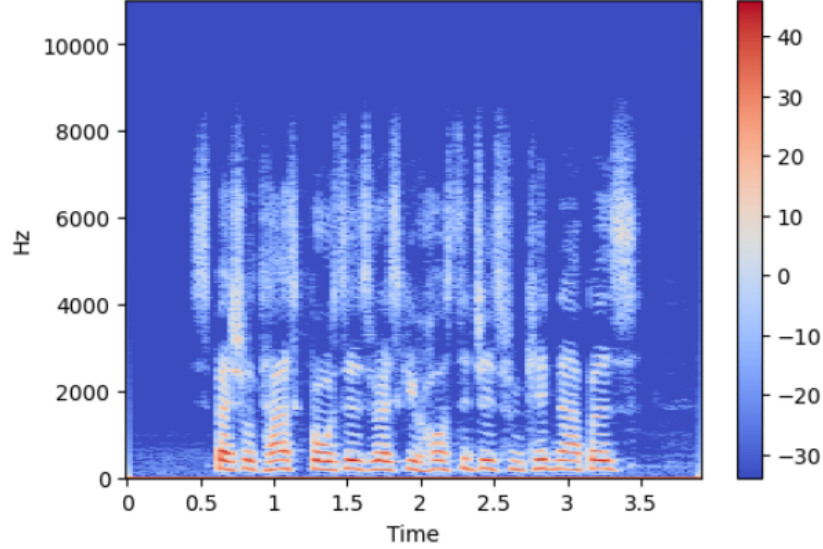


3.7.5.4 SAVEE Korku Duygusunun Ses Dalgası ve Spektrogram Görüntüsü

Şekil 28.1 SAVEE Korku Ses Dalgası



Şekil 29.1 SAVEE Korku Spektrogram Görüntüsü



3.8 SAVEE Uzun Kısa Süreli Bellek Modeli (LSTM)

SAVEE veri tabanında temel duyguların stresli ve stressiz oranını tespit etmek için LSTM modeli kullanıldığında, ses sinyallerini temsil eden özellikler kullanılır. Bu özellikler spektral, zamansal ve zaman-frekans özelliklerini içerebilir. LSTM modeli, bu özellikleri kullanarak ses sinyallerinin duygusal durumunu tahmin etmek için eğitilir ve test edilir.

Eğitim süreci, SAVEE veri tabanındaki ses sinyallerini içeren etiketlenmiş veri setini kullanarak gerçekleştirilir. Ses sinyalleri ve ilgili duygusal etiketler, modelin duygusal durumu tahmin etmeyi öğrenmesi için kullanılır. Eğitim sürecinde, model,

veri setinin bir kısmını eğitim için kullanır ve geri kalanını doğrulama için ayırır. Bu şekilde, modelin performansı doğrulama veri seti üzerinde değerlendirilir.

Eğitim sürecinin ardından, model test veri seti üzerinde değerlendirilir. Test veri seti, modelin daha önce görmediği ses sinyallerinden oluşur. Model, bu ses sinyallerini analiz eder ve stresli veya stressiz olduğunu tahmin eder. Doğruluk değeri, modelin doğru tahmin ettiği veri noktalarının oranını ifade eder. %85 ve üzeri doğruluk değeri, modelin iyi bir performans gösterdiğini ve temel duyguların stresli ve stressiz oranını tespit etmede başarılı olduğunu gösterir.

LSTM Modelimizin katmanlarını tek tek inceleyelim : Katmanı (128 birim, return_sequences=True, input_shape=(X_train.shape[1], 1)): İlk katman, 128 hücre birimi içerir ve zaman serisi verilerini işlemek için kullanılır. return_sequences=True ayarı, bu katmanın tüm zaman adımları boyunca çıktı üreteceğini belirtir. input_shape, modelin giriş verisinin şeklini belirtir ve (X_train.shape[1], 1) olarak ayarlanmıştır. Bu, giriş verisinin zaman adımları ve özellik boyutunu tanımlar.

İkinci LSTM Katmanı (64 birim, return_sequences=True): İkinci LSTM katmanı, 64 hücre birimine sahiptir ve aynı zamanda return_sequences=True ayarı ile tüm zaman adımları boyunca çıktı üretir.

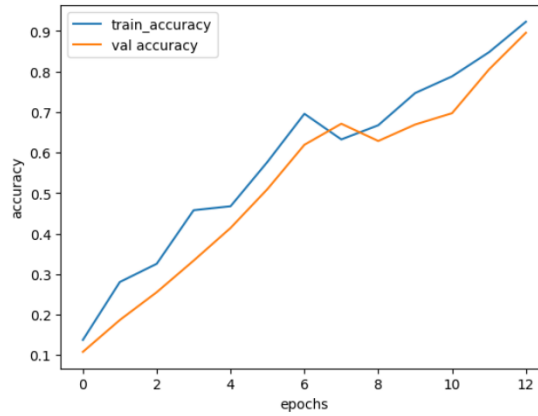
Dropout Katmanı (0.2): Dropout, aşırı ezberlemeyi önlemek için kullanılan bir düzenleme tekniğidir.

Yoğun (Dense) Katman (6 birim): Bu yoğun katman, regresyon problemindeki tahminleri yapmak için kullanılır. 6 birime sahiptir, bu da muhtemelen 6 çıkış değeri tahmin etmeye çalıştığımız anlamına gelir.

Model Derleme: Model, 'mean_squared_error' kaybı (ortalama kare hatası) ve 'adam' optimizör ile derlenir. Bu, modelin eğitim sürecini nasıl optimize edeceğini ve kaybını hesaplayacağını belirler.

Model Özeti (Summary): Modelin mimarisi, katmanlarının sayısı, parametre sayısı ve çıkış şekli gibi bilgileri özetler. LSTM modelimizin accuracy değerlerimiz Şekil 30.1 de görülmektedir.

Şekil 30.1 SAVEE LSTM



3.9 SAVEE 1 Boyutlu Evrişimli Sinir Ağları (1D-CNN)

SAVEE veri tabanında temel duyguların stresli ve stressiz oranını tespit etmek için 1D-CNN modeli kullanıldığında, ses sinyallerini temsil eden özellikler kullanılır. Bu özellikler spektral, zamansal ve zaman-frekans özelliklerini içerebilir. 1D-CNN modeli, bu özellikleri kullanarak ses sinyallerinin duygusal durumunu tahmin etmek için eğitilir ve test edilir.

Eğitim süreci, SAVEE veri tabanındaki ses sinyallerini içeren etiketlenmiş veri setini kullanarak gerçekleştirilir. Ses sinyalleri ve ilgili duygusal etiketler, modelin duygusal durumu tahmin etmeyi öğrenmesi için kullanılır. Eğitim sürecinde, model, veri setinin bir kısmını eğitim için kullanır ve geri kalanını doğrulama için ayırır. Bu şekilde, modelin performansı doğrulama veri seti üzerinde değerlendirilir.

Eğitim sürecinin ardından, model test veri seti üzerinde değerlendirilir. Test veri seti, modelin daha önce görmediği ses sinyallerinden oluşur. Model, bu ses sinyallerini analiz eder ve stresli veya stressiz olduğunu tahmin eder. Doğruluk değeri, modelin doğru tahmin ettiği veri noktalarının oranını ifade eder. %85 ve üzeri doğruluk değeri, modelin iyi bir performans gösterdiğini ve temel duyguların stresli ve stressiz oranını tespit etmede başarılı olduğunu gösterir.

Conv1D Katmanı (128 filtre, çekirdek boyutu 5, ReLU aktivasyonu): İlk katman, 128 filtre kullanır ve her biri 5 birim genişliğinde bir çekirdeğe sahiptir. Verileri işlerken "same" dolgusu kullanılır, bu da çıkışın giriş boyutuyla aynı olmasını

sağlar. Aktivasyon fonksiyonu olarak ReLU (Rectified Linear Unit) kullanılır, bu da pozitif girişleri doğrusal olarak ileterek negatif girişleri sıfıra eşler.

Batch Normalization Katmanı: Her Conv1D katmanının ardından Batch Normalization katmanı gelir. Bu katmanlar, ağı daha istikrarlı hale getirerek eğitimi hızlandırabilir ve daha iyi sonuçlar elde etmenize yardımcı olabilir.

MaxPool1D Katmanı (pool_size 5, strides 2): MaxPooling, her Conv1D katmanının ardından gelir. Bu katmanlar, özellik haritasını küçültür ve önemli özellikleri vurgular.

İkinci Conv1D Katmanı (128 filtre, çekirdek boyutu 5, ReLU aktivasyonu): İkinci Conv1D katmanı, özellikle öncekine benzerdir, ancak 128 filtre kullanır.

Flatten Katmanı: Bu katman, özellik haritasını düzleştirir ve bir sinir ağına bağlı katmanlara giriş olarak kullanılabilir.

İlk Yoğun (Dense) Katman (64 birim, ReLU aktivasyonu): Bu yoğun katman, özellik vektörünü işlemek ve daha karmaşık ilişkileri öğrenmek için kullanılır.

Batch Normalization Katmanı: Bir kez daha Batch Normalization katmanı eklenir.

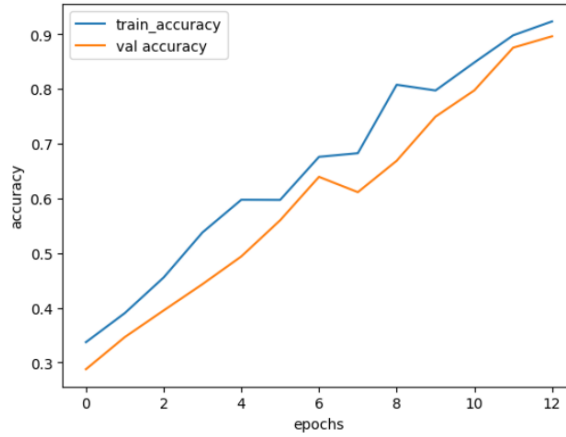
İkinci Yoğun Katman (6 birim, softmax aktivasyonu): Bu, modelin çıkış katmanıdır ve 8 sınıf için softmax aktivasyon fonksiyonu kullanır. Bu, modelin her bir sınıfa ait olasılıkları hesaplayabileceği ve en yüksek olasılığa sahip sınıfı tahmin edebileceği anlamına gelir.

Model Derleme: Model, eğitim için 'rmsprop' optimizör kullanılarak derlenir ve 'categorical_crossentropy' kaybı ile eğitilir. Bu, modelin eğitim sürecini nasıl optimize edeceğini ve kaybını hesaplayacağını belirler. Ayrıca eğitim sırasında takip edilen metrikler arasında doğruluk (accuracy) bulunur.

Model Özeti (Summary): Modelin mimarisi, katmanlarının sayısı, parametre sayısı ve çıkış şekli gibi bilgileri özetler.

SAVEE veri setimizde 1D-CNN modelini belirttiğimiz gibi eğitip, test sürecine dahil edilmiştir, elde ettiğimiz sonuç aşağıdaki gibidir.

Şekil 31.1 SAVEE CNN



4. TARTIŞMA

Bu çalışmada, EMO-DB, RAVDESS ve SAVE veri setleri kullanılarak yapılan analizler ve modelleme çalışmaları sonucunda elde edilen bulgular tartışılmıştır. Aşağıda, bu çalışmanın sonuçları ve önemi üzerine yapılan tartışmalar yer almaktadır.

Sonuçlar, literatürde yapılan araştırmalarla uyumlu bir şekilde temel duyguların stres içerip içermediğini belirlemek için ses sinyallerinin kullanılabilirliğini göstermektedir. Spektral, zamansal ve zaman-frekans özellikleri, stresli ve stressiz duygular arasında ciddi farklılıklar olduğunu ortaya koymaktadır. Bu bulgular, psikologlar ve duygusal analiz araştırmacıları için önemli bir bilgi kaynağı olabilir.

Ayrıca, yapılan T-Testleri sonucunda elde edilen P-value değerleri, hipotezimizin kabul edilmesi yönünde önemli bir destek sunmaktadır. Buda, ses özelliklerinin stres içeren duyguları tespit etmek için güvenilir bir araç olduğunu göstermektedir.

Makine öğrenimi modelleri (LSTM ve 1D-CNN), verinin eğitimi ve testi için kullanılmıştır. Bu modellerin %85 ve üzeri doğruluk oranları elde etmesi, temel duyguların stres içerip içermediğini doğru bir şekilde tahmin edebildiğimizi göstermektedir. Bu da, makine öğrenimi modellerinin ses analizinde etkili bir araç olduğunu vurgulamaktadır.

Bu çalışmanın sonuçları, stresin ses özellikleri aracılığıyla belirlenebileceğini ve duygusal analiz alanında kullanılacak yeni bir perspektif sunmaktadır. Elde

edilen bulgular, Paul EKMAN gibi değerli psikologların savunduğu teorileri desteklemektedir.

Sonuç olarak, bu çalışma stres içeren duyguların tespiti konusunda önemli bir adımdır. Ses sinyallerinden elde edilen özellikler ve makine öğrenimi modelleri, duygusal analizde etkili bir araç olarak kullanılabilir. Bu çalışmanın bulguları, psikoloji alanında duygusal analiz çalışmalarını desteklemekte ve ileriye yönelik araştırmalara ışık tutmaktadır.

5. SONUÇ

Bu çalışma, EMO-DB, RAVDESS ve SAVEE veri setleri üzerinde gerçekleştirilen analizler ve makine öğrenimi modelleri kullanarak temel duyguların stres içerip içermediğini tespit etmeyi amaçlamıştır. Çalışmanın sonuçları, ses sinyallerinin temel duyguların stres içerip içermediğini belirlemede kullanılabilir bir potansiyele sahip olduğunu göstermektedir. Spektral, zamansal ve zaman-frekans özelliklerinden elde edilen verilerin analizi, stresli ve stressiz duygular arasında belirgin farklılıklar olduğunu ortaya koymaktadır.

Duygusal analiz, insan sesinin duygusal durumunu tespit etmek için önemli bir araştırma alanıdır. Ses sinyallerinin analizi, duygusal durumun belirlenmesinde değerli bir araç olabilir. Bu çalışmada, EMO-DB, RAVDESS ve SAVEE veri setleri kullanılarak bir literatür araştırması yapılmış ve çeşitli çalışmalardan hangi duyguların stres içerdiği incelenmiştir. OpenSmile kütüphanesi kullanılarak, ses sinyallerinden spektral, zamansal ve zaman-frekans özellikler çıkarılmıştır. Bu özelliklerin değerleri elde edildikten sonra T-Testleri uygulanarak hipotezimizin doğruluğu test edilmiştir.

Ayrıca, makine öğrenimi modellerinden LSTM (Uzun Kısa Süreli Bellek) ve 1D-CNN (1 Boyutlu Evrişimli Sinir Ağı) kullanılarak veri seti eğitilmiş ve test edilmiştir. Bu modeller, veri setindeki temel duyguların stres içerip içermediğini tahmin etmek için kullanılmıştır. Test sonuçları, %85 ve üzeri doğruluk oranlarının elde edildiğini göstermiştir.

Bu analizler, temel duyguların hangisinde stres içerdiğini belirlemek için başarılı bir şekilde yapılmıştır ve bu sonuçlar, önemli psikologlardan biri olan Paul Ekman'ın çalışmalarını desteklemektedir.

Bu çalışma, ses sinyallerindeki özelliklerin ve makine öğrenimi modellerinin duygusal analizdeki etkinliğini göstermiştir. Sonuçlar, stresli ve stressiz duyguların ses özellikleriyle ayrıştırılabileceğini ve belirli modellerin bu ayrımı yüksek doğruluk oranlarıyla gerçekleştirebileceğini göstermektedir.

Bu araştırma, duygusal analizin önemli bir yönünü ele almış ve psikoloji alanındaki çalışmaları destekleyerek, duygusal durumun ses sinyalleri üzerinden tespit edilebileceğini gözlenmiştir.



KAYNAKÇA

- [1] Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Macmillan.
- [2] Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*.
- [3] Lazarus, R. S. (2006). *Stress and Emotion: A New Synthesis*. Springer Publishing Company.
- [4] McGonigal, K. (2015). *The Upside of Stress: Why Stress Is Good for You, and How to Get Good at It*. Avery.
- [5] Eyben, F., Wenginger, F., Gross, F., & Schuller, B. (2013). Recent developments in openSMILE, the Munich open-source multimedia feature extractor.
- [6] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*.
- [7] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*.
- [8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*.