

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



DERİN ÖĞRENME VE SES TABANLI DUYGU TANIMA

Aslı DEMİR

Yüksek Lisans Tezi

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

Elektrik Elektronik Mühendisliği Teknolojileri Programı

Telekomünikasyon Bilim Dalı

ŞUBAT 2021

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Elektrik Elektronik Mühendisliği Anabilim Dalı
Elektrik Elektronik Mühendisliği Teknolojileri Programı

Yüksek Lisans Tezi

DERİN ÖĞRENME VE SES TABANLI DUYGU TANIMA

Tez Yazarı
Aslı DEMİR

Danışman
Prof. Dr. Abdulkadir ŞENGÜR

ŞUBAT 2021
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Elektrik Elektronik Mühendisliği Anabilim Dalı
Elektrik Elektronik Mühendisliği Teknolojileri Programı

Yüksek Lisans Tezi

Başlığı: Derin Öğrenme ve Ses Tabanlı Duygu Tanıma
Yazarı: Aslı DEMİR
İlk Teslim Tarihi: 19.01.2021
Savunma Tarihi: 02.02.2021

TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

Danışman:	Prof. Dr. Abdulkadir ŞENGÜR Fırat Üniversitesi, Teknoloji Fakültesi	<i>İmza</i> Onayladım
Başkan:	Doç. Dr. Erkan DENİZ Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Prof. Dr. Abdulkadir ŞENGÜR Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Dr. Öğr. Üyesi Ömer Faruk ALÇİN Malatya Turgut Özal Üniversitesi, Mühendislik Fakültesi	Onayladım

Bu tez, Enstitü Yönetim Kurulunun/...../20..... tarihli toplantısında tescillenmiştir.

İmza
Doç. Dr. Kürşat Esat ALYAMAÇ
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Derin Öğrenme ve Ses Tabanlı Duygu Tanıma ” Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

02.02.2021

Ash DEMİR



ÖNSÖZ

Konuşma, bireylerin duygularının dışavurumunda önemli bir rol oynayan temel iletişim aracıdır. Dolayısıyla konuşma sinyalleri aracılığıyla duygu durumunu belirlemek araştırmacılar için önemli bir görevdir. Bir konuşmadaki duygusal içeriğin tanımlanması amacıyla farklı sistemler, algoritmalar ve sınıflandırıcılar geliştirilmiştir. Bu tez çalışmasında konuşma sinyallerinden duygu analizi yapan derin öğrenme ve ses tabanlı bir duygu tanıma sistemi önerilmiştir.

Bu tez çalışmasının hazırlanma aşamasında bilgi birikimi, tecrübesi ve yardımlarını esirgemeyen ayrıca akademisyen kişiliğinden ve insaniyetinden çok şey öğrendiğim değerli danışman hocam Prof. Dr. Abdulkadir ŞENGÜR 'e ve hayatımın her alanında beni destekleyen aileme teşekkürlerimi sunarım

Ash DEMİR
ELAZIĞ, 2021

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vii
ABSTRACT	viii
ŞEKİLLER LİSTESİ	ix
TABLolar LİSTESİ	x
SİMGELER VE KISALTMALAR	xi
1. GİRİŞ	1
1.1. Literatür Özeti	2
1.2. Tezin Organizasyonu	5
2. DUYGU KAVRAMI.....	6
2.1. Duyguların Sınıflandırılması	6
3. SES ÖZELLİKLERİ	7
3.1. Prosodik Özellikler	7
3.2. Spektral Özellikler	7
3.3. Ses Kalitesi Özellikleri	8
4. MATERYAL VE METOT.....	9
4.1. Duygu Tanıma Veri Setleri.....	9
4.2. RAVDESS Veri Seti.....	10
4.3. SAVEE Veri Seti	11
4.4. RML Veri Seti	11
4.5. Önerilen Sistem	11
4.6. Ses Aktivite Algılaması (Voice Activity Detection, VAD)	12
4.7. Özellik Çıkarma.....	12
4.7.1. Mel Frekansı Kepstral Katsayıları.....	13
4.7.2. Lineer Öngörülü Kodlama.....	16
4.7.3. Lineer Öngörü Kepstral Katsayıları	16
4.7.4. Yansıma Katsayıları	17
4.7.5. Log- Alan Oranları Katsayıları.....	17
4.7.6. Spektral Alt-bant Ağırlık Merkezi	17
4.7.7. Log- Enerji	17
5. YAPAY ZEKÂ, MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME	18
5.1. Yapay Sinir Ağlarının Yapısı	18
5.1.1. İleri Beslemeli Yapay Sinir Ağları	20
5.1.2. Geri Beslemeli Yapay Sinir Ağları.....	21
5.2. Derin Öğrenmenin Tarihçesi	21
5.3. Tekrarlayan Sinir Ağları	23
5.4. Uzun-Kısa Süreli Bellek Ağları (UKSBA).....	24
6. BULGULAR.....	27
6.1. Hata matrisi ve Tanıma Doğruluğu	27
6.2. Rml Veri Seri İçin Sınıflandırma Sonuçları	28

6.3. Ravdess Veri Seti İçin Sınıflandırma Sonuçları	30
6.4. Savee Veri Seti İçin Sınıflandırma Sonuçları	32
7. SONUÇLAR VE TARTIŞMA.....	34
KAYNAKLAR	
ÖZGEÇMİŞ	



ÖZET

Derin Öğrenme ve Ses Tabanlı Duygu Tanıma

Ash DEMİR

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü

Elektrik Elektronik Mühendisliği Anabilim Dalı
Elektrik Elektronik Mühendisliği Teknolojileri Programı
Şubat 2021, Sayfa: xi + 39

Duygu, bireyin ruh halinde içsel ve çevresel faktörlerin etkileşiminden doğan kompleks psikofizyolojik bir değişimdir. Duygular, insanların dış dünyayla olan etkileşimlerinde, aldığı kararlarda ve eylemlerinde önemli bir rol oynar. Dolayısıyla duygu tanıma ve yorumlama özellikle insan bilgisayar etkileşimi için önemlidir ve bilgisayarları mantıksal işlem yapan makinelerin ötesine taşımaktadır. Duyguların tespit edilmesinde görsel, işitsel, dokunsal ve diğer biyometrik sinyaller kullanılmaktadır. Bu sinyaller arasında insanlar arasındaki iletişimin en yaygın aracı olan konuşma önemli bir yere sahiptir. Bu durum konuşma tabanlı duygu tanıma uygulamalarının, giderek artan bir ilgiyle ilerletilen bir araştırma alanı olmasına imkân tanımıştır. Bu tez çalışmasında ses işleme tabanlı duygu tanıma uygulaması gerçekleştirilmiştir. SAVEE, RML ve RAVDESS olmak üzere 3 farklı veri seti kullanılmıştır. Mevcut ses dosyalarına ön işlem ve özellik çıkarımı uygulandıktan sonra Uzun- Kısa Süreli Bellek Ağları ile sınıflandırılmıştır. Sistemin duygu tanıma doğruluk oranları SAVEE veri seti için %72,13 RML veri seti için %70,35, RAVDESS veri seti için %68,8 olarak bulunmuştur.

Anahtar Kelimeler: : Otomatik duygu tanıma, derin öğrenme, UKSBA

ABSTRACT

Deep Learning and Audio Based Emotion Recognition

Ash DEMİR

Master's Thesis

FIRAT UNIVERSITY

Graduate School of Natural and Applied Sciences

Department of Electrical and Electronics Engineering

Electrical and Electronics Engineering Technologies

February 2021, Pages: xi + 39

Emotion is a complex psycho-physiological changes in the mood of an individual arising from the interaction of internal and environmental factors. Emotions play an important role in people's interactions with the outside world, in their decisions and actions. Therefore, emotion recognition and is especially important for human-computer interaction, and it takes computers beyond machines that do logical processing. Visual, auditory, tactile and other biometric signals are used to detect emotions. Among these signals, speech, which is the most common means of communication between people, has an important place. This has enabled speech-based emotion recognition applications to become a field of research that has been developed with increasing interest. In this thesis, voice processing-based emotion recognition application was carried out. Three different data sets are used as SAVEE, RML and RAVDESS. After pre-processing and feature extraction are applied to existing audio files, they are classified with Long Short-Term Memory. The emotion recognition accuracy rates of the system were found 72.13% for the SAVEE data set, 70.35% for the RML data set, and 68.8% for the RAVDESS dataset.

Keywords: Automatic emotion recognition, deep learning, LSTM

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 2.1.	2 boyutlu duygu uzayı..... 6
Şekil 4.1.	Önerilen sistemin akış şeması 12
Şekil 4.2.	Ses aktivite algılama işlemi 12
Şekil 4.3.	Özellik matrisinin elde edilişi 13
Şekil 4.4.	MFKK katsayılarının elde edilişi 14
Şekil 4.5.	Çerçeveleme işlemi..... 14
Şekil 4.6.	LÖK katsayılarının elde edilişi 16
Şekil 4.7.	Spektral alt bant ağırlık merkezi hesaplanışı 17
Şekil 5.1.	Biyolojik sinir hücresi 18
Şekil 5.2.	Yapay sinir ağı hücresi 19
Şekil 5.3.	İleri beslemeli sinir ağı 20
Şekil 5.4.	Geri beslemeli sinir ağı 21
Şekil 5.5.	TSA mimarisi..... 23
Şekil 5.6.	TSA mimarisinin açık gösterimi 23
Şekil 5.7.	UKSBA hücresi ve işlemler 24
Şekil 6.1.	Hata matrisi..... 27
Şekil 6.2.	RML veri seti kullanılarak yapılan çalışmalar 29
Şekil 6.3.	RAVDESS veri seti kullanılarak yapılan çalışmalar 31
Şekil 6.4.	SAVEE veri seti kullanılarak yapılan çalışmalar 33

TABLÖLAR LİSTESİ

	Sayfa
Tablo 4.1. Yaygın olarak kullanılan duygu veri tabanları.....	9
Tablo 4.2. RAVDESS, SAVEE ve RML veri setlerinin içeriği.....	11
Tablo 5.1. Biyolojik sinir ağı elemanlarının yapay sinir ağlarındaki karşılıkları	19
Tablo 6.1. RML veri seti için hata matrisi (%)	28
Tablo 6.2. RAVDESS veri seti için hata matrisi (%).....	30
Tablo 6.3. SAVEE veri seti için hata matrisi (%)	32

SİMGELER VE KISALTMALAR

Kısaltmalar

AFD	: Ayrık Fourier Dönüşümü
DVM	: Destek Vektör Makineleri
ESA	: Evrişimsel Sinir Ağları
HFD	: Hızlı Fourier Dönüşümü
HMI	: Human Machine Interaction (İnsan -makine etkileşimi)
LÖK	: Lineer Öngörülü Kodlama
LÖKK	: Lineer Öngörülü Kodlama Katsayıları
MFKK	: Mel Frekansı Kepstral Katsayıları
RAVDESS	: Ryerson Audio-Visual Database of Emotional Speech and Song
SAVEE	: Surrey Audio-Visual Expressed Emotion
SDT	: Ses-tabanlı Duygu Tanıma
UKSBA	: Uzun- Kısa Süreli Bellek Ağları
VAD	: Voice Activity Detection

1. GİRİŞ

Duygu, bireylerin yaşantısının her anına dahil olan bir olgudur. Duygu durumunun ifade edilme şekli her birey için karışık ve farklı olabilmektedir. Yüz ifadesi, kalp atışları ve seste meydana gelen değişimler duygunun dışavurum yollarındandır. Bunlar içerisinde ses önemli bir yere sahiptir. Çünkü insanlar arası sosyal iletişimin en temel aracıdır.

Son yıllarda yapay zekâ alanındaki gelişmeler, insan-makine etkileşimi (Human-Machine Interaction (HMI)) içeren teknolojilerin artmasını sağlamıştır. HMI, insanlar ve otomatik sistemler arasındaki iletişimi geliştirmeye yönelik çalışmaları kapsayan multi disiplinler bir alandır. İnsan-makine etkileşimi uygulamaları genellikle görüntü, ses ve sensor tabanlı veya bu modellerin kombinasyonları ile oluşturulmuş yapılardır [1]. Bu modeller içerisindeki ses tabanlı HMI çalışmaları: ses tanıma, konuşmacı tanıma ve duygu tespiti ana başlıkları altında sürdürülmektedir. Bu çalışmalar içerisinde ses tabanlı duygu tanıma modeli daha zor bir problem olarak karşımıza çıkmaktadır. Çünkü duygu, oluşumu kadar ifade edilişi kadar da karmaşık bir olgudur. Ayrıca aynı ifade birden fazla duyguyu içerebilmekte ve her duygu, sözlü ifadenin farklı bir kısmına karşılık gelebilmektedir. Bu kısımlar arasındaki sınırları belirlemek oldukça zordur. Ses tabanlı duygu tanımadaki bir diğer zorluk ise, duygunun ifade edilme biçiminin genellikle konuşmacıya, konuşmacının kültürüne ve çevresine göre değişmesidir [2].

Ses-Tabanlı Duygu Tanıma (SDT) sistemleri, karmaşık ve zorlu yapısına rağmen birçok farklı alanda kullanılabildiği için bu konu hakkında yapılan çalışmalar artan bir ilgiyle sürdürülmektedir. Bu çalışmaların kullanıldığı sistemler, aşağıdaki gibidir.

Otizmlili bireylerin yüz ifadelerini anlamada ve duygu tanıma becerilerinde eksiklikler olduğu bilinmektedir. Bu durum sosyal iletişim ve etkileşimlerini olumsuz olarak etkilemektedir. Bir robota entegre edilen gerçek zamanlı bir SDT sistemi konuşmacıdan gelen ses sinyallerini analiz ederek duygu sınıflandırma yapabilir. Bu durum otizmlili bireylere kolaylık sağlayabilir [3].

SDT sistemleri özellikle davranış ve depresyon saptama analizi yoluyla tanı koymak için yararlı bir tıbbi araç olabilir [4].

Mobil bankacılık ve otomatik çağrı merkezlerinde kullanılan SDT sistemleri müşterilerin olumsuz duygularını tespit ederek konuşmayı çağrı merkezi elemanına yönlendirebilir [5].

Cep telefonlarında bulunan sanal asistan (örneğin: Apple Siri [6]) uygulamalarına entegre edilen bir SDT sistemi, kullanıcı ve sanal asistan arasındaki iletişimi geliştirebilir.

Çoğunlukla sesli komutlarla çalışan otonom araçlar için, kullanıcının içinde bulunduğu olağandışı durumları tespit etmede SDT sistemleri iyi bir alternatif olabilir.

1.1. Literatür Özeti

Ses tabanlı duygu tanıma yaklaşımı bir sınıflandırma problemidir ve örüntü tanıma yaklaşımlarından faydalanılarak çözülebilir. Bu yaklaşım özellik çıkarma ve sınıflandırma olarak bilinen iki temel aşamayı kapsar. SDT sistemleri için sınıflandırma model geliştirme süreci, makine öğrenimi algoritmalarına dayanan çalışmalarla başlayıp derin öğrenme çerçevesinde sürdürülmektedir. Literatürde, makine öğrenimi ve derin öğrenme algoritmalarının kullanıldığı çalışmalar aşağıda özetlenmiştir.

[7] nolu çalışmada Berlin Emotions veri setinden alınan ses sinyallerine kısa süreli fourier dönüşümü uygulanmıştır. Elde edilen spektrogramlar Evrişimsel Sinir Ağı (ESA) ve Uzun Kısa Süreli Bellek Ağları (UKSBA) ve bu iki ağın sentezlendiği 3 farklı mimaride sınıflandırılarak karşılaştırılmıştır. En iyi sonuç ESA ve UKSBA'nın birlikte kullanıldığı sistemde elde edilmiştir. Sınıflandırılan 7 duygu için ortalama %86,65 doğruluk hesaplanmıştır.

[8] nolu çalışmada akustik özellikler ve sözcüksel özelliklerle sentezlenerek özellik vektörü oluşturulmuştur. Akustik özellik seti formant frekansı ve spektral özelliklerden oluşmuştur. Bu özelliklere farklı istatistiksel işlemler uygulanmış ve özellik setine eklenmiştir. Ayrıca duygu vektörü adlı yeni bir sözcüksel özellik yaklaşımı önerilmiştir. Bu yaklaşımda, belirli duyguları ifade etme eğilimi gösteren kelimelere ağırlık ataması yapılmıştır. Sınıflandırıcı olarak Destek Vektör Makineleri (DVM), veri seti olarak IEMOCAP kullanılmıştır. Sonuçlar 4 duygu sınıfı için %69,2 olarak bulunmuştur.

[9] nolu çalışmada SAVEE ve RML veri seti üzerinde Mel Frekansı Kepstral Katsayıları (MFKK), enerji, spektral entropi özelliklerinin de dahil olduğu 34 ses özelliği çıkarılarak destek vektör makinesi ve K-En Yakın Komşu (K-Nearest Neighbor, KNN) ile sınıflandırmıştır. SAVEE ve RAVDESS veri setinin kullanıldığı çalışmada, kadın konuşmacılar ve erkek konuşmacılar ayrı incelenerek cinsiyete bağlı duygu sınıflandırması gerçekleştirilmiştir.

[10] nolu çalışmada ses sinyallerinin gürültü ve diğer bozucu etkilerden arındırıldığı bir sistem önerilmiştir. Bu amaçla Butterworth filtresi ve sinyal yumuşatma teknikleri uygulanmıştır. Daha sonra sıfır geçiş oranı, enerji, spektral entropi ve perde özellikleri hesaplanmıştır. Elde edilen özellik matrisi; karar ağaçları, rastgele orman ve DVM ile sınıflandırılmıştır. Yapılan sınıflandırmada lineer destek vektör makinesinin ortalama en iyi performansa sahip olduğu görülmüştür.

[11] nolu çalışmada MFKK ve ayrık dalgacık dönüşümünden elde edilen ses özellikleri sentezlenerek normalize edilmiştir. Sınıflandırma mutluluk, kızgınlık, korkma ve nötr duygularının her birinden 10 ses örneği içeren veri seti ile gerçekleştirilmiştir. Sınıflandırıcı olarak DVM kullanılan bu çalışma, sesli yanıt ile gerçekleştirilen bankacılık işlemlerinde güvenlik sağlama amacıyla oluşturulmuştur.

[12] nolu çalışmada Rastgele Orman (RO) ile birlikte evrimsel sinir ağını temel alan yeni bir ağ modeli (ESA-RO) önerilmiştir. Öncelikle, normalize edilmiş spektrogramlardan ESA ile ses özellikleri çıkarılmış ve ses özelliklerini sınıflandırmak için rastgele orman sınıflandırma algoritması uygulanmıştır. ESA-RO modelinin geleneksel ESA modelinden daha üstün olduğunu görülmüştür. ESA-RO modeli Nao robotuna uygulanmış ve Nao robotunun, bir insanın psikolojisini duygularını tanıma yoluyla anlamaya çalışması sağlanmıştır

[13] nolu çalışmada spektral ses özellikleri ve prosodik ses özelliklerinin duygu tanımadaki performansı karşılaştırılmıştır. Yalnızca spektral özellikleri kullanan sistemin performansının, sadece prosodik özellikleri kullanan sistemden daha yüksek olduğu anlaşılmıştır. Spektral ve prosodik özelliklerin kombinasyonunu kullanan sistemde ise en yüksek doğruluk elde edilmiştir. Sınıflandırıcı olarak destek vektör makinesi kullanılmış ve konuşma örnekleri Berlin duygusal veri tabanından alınmıştır.

[14] nolu çalışmada herhangi bir ön işlem uygulanmamış ses sinyalleri üzerinde çalışan evrimsel sinir ağları ve uzun kısa süreli bellek ağlarının birlikte kullanıldığı bir sistem önerilmiştir. Spontane duyguların tanınması amaçlanmış ve bu amaca uygun olan RECOLA veri seti kullanılmıştır. Duygu sınıflandırma, iki boyutlu duygu uzayı yaklaşımına uygun olarak uyarılma-değerlik parametrelerine göre yapılmıştır.

[15] nolu çalışmada Fourier Parametresi (FP) özellikleri adı verilen yeni bir ses özelliği önerilmiştir. FP ve MFKK özelliklerinin birlikte kullanılmasının, duygu tanıma sisteminin performansını yükselttiği anlaşılmıştır. Sınıflandırıcı olarak DVM kullanılmış, konuşma örnekleri EMO-DB, CASIA ve EESDB veri setlerinden alınmıştır.

[16] nolu çalışmada ses sinyallerinden alınan MKFF, mel-ölçekli spektrogram, chroma, tonnetz gösterimi ve spektral kontrast özellikleri tek boyutlu evrimsel sinir ağının girişi olarak kullanılmıştır. Önerilen bu yeni mimari RAVDESS ve EMO-DB veri setlerinde denenmiştir. RAVDESS veri setinde 8 duygu için %71,61, EMO-DB veri setinde 7 duygu sınıfı için %86,1 doğruluk elde edilmiştir.

[17] nolu çalışmada işitsel ve görsel özelliklerin kullanıldığı bir duygu tanıma sistemi önerilmiştir. Bu amaçla AFEW veri seti kullanılmıştır. Görsel özellikleri çıkarmak için bir ESA ve iki yönlü uzun kısa süreli bellek ağı önerilmiştir. Spektral özellikler ve konuşmanın sesli-sessiz bölümlerden çıkarılan özellikler kullanılmıştır. Ek olarak, özellik matrisinde boyut azaltmak için temel bileşen analizi uygulanmıştır. Özellik düzeyinde füzyon ve model düzeyinde füzyon ile önerilen sistemin doğruluğu hesaplanmıştır.

[18] nolu çalışmada zaman-frekans katsayılarına dayanan yeni bir özellik çıkarma yöntemi önerilmiştir. Bu katsayılar MFKK ile kullanılarak özellik matrisi elde edilmiştir. Daha sonra özellik seçme algoritması kullanılmıştır. Üç veri seti için deneysel sonuçlar incelendiğinde doğruluk oranları EMO-DB için %97,21 SAVEE %80 ve PDREC veri kümeleri için, %87,3 şeklindedir.

[19] nolu çalışmada evrişimsel sinir ağlarının farklı boyutlarından faydalanılmıştır. 1 boyutlu ESA ile doğrudan ses kliplerinden ve 2 boyutlu ESA ile log-mel spektrogramlarından özellik çıkarımı yapılmıştır. Bu iki mimariye UKSBA katmanı eklenmiştir. ESA- UKSBA kombinasyonu ile tasarlanan bu ağ; Berlin ve IEMOCAP veri setleri için değerlendirilmiştir. 2 boyutlu ESA'ların 1 boyutlu ESA'lara göre ortalama doğrulukta daha üstün olduğu sonucuna varılmıştır.

[20] nolu çalışmada perde, formant, sıfır geçisi, MFKK ve istatistiksel parametrelerden oluşan bir özellik matrisi elde edilmiştir. Farklı DVM çekirdek fonksiyonlarının performansları her duygu için ayrı ayrı karşılaştırılmıştır. Polish veri setinin kullanıldığı çalışmada tüm duygular için ortalama en yüksek doğruluk radyal tabanlı DVM için elde edilmiştir.

[21] nolu çalışmada ses sinyalleri ön işlemden geçirildikten sonra; Log-Mel spektrogramı, mel frekansı kepsstral katsayıları, perde ve enerji özellikleri elde edilmiştir. Bu özelliklerin duygu sınıflandırmadaki etkileri UKSBA, ESA ve saklı markov modellerine uygulanarak karşılaştırılmıştır. 14 sınıflı (2 cinsiyet $\times$ 7 duygu) sınıflandırma için mel spektrogram özellikleri kullanılarak 4 katmanlı 2 boyutlu bir ESA ile %68' lik bir doğruluk elde edilmiştir.

[22] nolu çalışmada AM-FM modülasyon modeline dayalı, modülasyon spektral özellikleri ve modülasyon frekansı özellikleri adlı yeni ses özellikleri sunulmuştur. Önerilen sistemin etkinliğini değerlendirmek için Berlin ve Spanish veri tabanları kullanılarak deneysel çalışma yapılmıştır. Önerilen SDT sisteminde, İspanyol veri tabanı için tekrarlayan sinir ağları kullanılmış ve %91,16 doğruluk elde edilmiştir. Berlin veri tabanı için ise DVM kullanılarak %86,22 doğruluk elde edilmiştir.

[23] nolu çalışmada RML veri setindeki ses sinyallerinden MFKK, sıfır geçiş hızı, harmonik gürültü oranı ve teager enerji operatörü içeren özellik vektörü çıkarılmıştır. Daha sonra özellik boyutu azalmak amacıyla autoencoder kullanılmış ve autoencoder kullanılmayan özellik seti ile karşılaştırma yapılmıştır. Sınıflandırıcı yöntem olarak destek vektör makineleri kullanılmış ve autoencoder uygulanan özellik setini daha yüksek tanıma performansı gösterildiği anlaşılmıştır.

[24] nolu çalışmada ön işlem uygulanmamış ses sinyallerinden, doğrudan özellik çıkaran bir ESA ve 2 katmanlı uzun kısa süreli bellek ağlarından oluşan ve uçtan uca eğitilen bir duygu tanıma sistemi önerilmiştir. RECOLA veri setinin kullanıldığı çalışmada duygular, boyutsal duygu modeline göre sınıflandırılmıştır.

[25] nolu çalışmada korelasyon analizi ve Fisher kriterine dayalı özellik seçme yönteminin kullanıldığı konuşma tabanlı duygu tanıma sistemi tasarlanmıştır. Aşırı öğrenme makinesi ve karar ağacına dayalı bir sınıflandırma önerilmiş, sınıflandırıcının performansı destek vektör makinesi ile karşılaştırılmıştır. Aşırı öğrenme makinesi ve karar ağacına dayalı sistemin DVM'ye üstünlüğü deneysel sonuçlardan anlaşılmıştır.

1.2. Tezin Organizasyonu

Bu tez çalışmasının ilk bölümünde; ses tabanlı duygu tanıma sistemleri hakkında genel bilgiler, uygulama alanları, derin öğrenme ve makine öğrenmesinin kullanıldığı literatür çalışmaları verilmiştir. 2.Bölümde, duygu kavramı ve duygu sınıflandırma konusuna değinilmiştir.3.Bölümde, duygu tanıma amacına uygun ses özellikleri açıklanmıştır. 4. Bölümde, önerilen duygu tanıma sisteminin yapısı, çalışmada kullanılan veri setleri ve önerilen SDT sisteminin özellik çıkarma aşaması hakkında bilgi verilmiştir. 5.Bölümde derin öğrenmenin temeli sayılan yapay sinir ağları ve sınıflandırma işleminde kullandığımız uzun kısa süreli bellek ağları anlatılmıştır. 6.Bölümde sınıflandırma sonuçları verilerek sistemin başarımı değerlendirilmiş ve literatürdeki diğer çalışmalardan elde edilen sonuçlarla karşılaştırılmıştır.



2. DUYGU KAVRAMI

Duygu, kişinin ruhsal durumunu ve çevre ile etkileşim kurma şeklini içeren psikolojik ve fizyolojik olguları içerir. İletişimde, konuşmacı genel olarak duygusal bilgilerini dinleyiciye ulaştırmak için sözel veya sözel olmayan iletişim kanallarını kullanmaktadır.

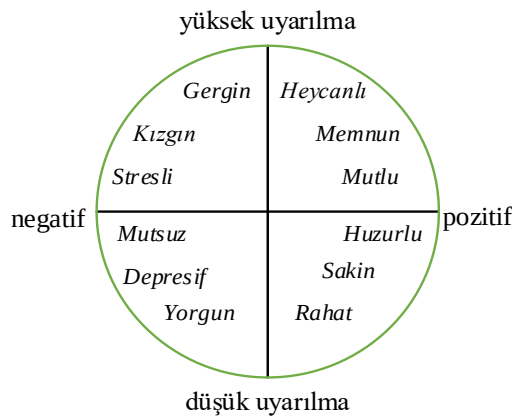
Bir kişi, konuşma aracılığıyla başkalarına duygusal bir gönderme yapacaksa kelimeleri belirli şekillerde düzenleyerek duyguyu sözel olarak aktarabilir. Örneğin, açıkça “Bugün çok üzuldüm” gibi belirli bir duygu işaret eden ifadelerin kullanımı yaygındır. Ancak bazen mevcut duygu durumu ifade edilen duygunun aksine olabilmektedir

Duygu durumunu ifade etmenin bir diğer yolu da sözel olmayan yollardan geçer. İnsan iletişiminde duygular için sözel olmayan temel ipuçları; yüz ifadesi, seslendirme ve beden dilidir. Duygu durumunu aktarmada özellikle ses etkili bir ipucudur. Örneğin bir sesli mesajda sesimiz; mutluluk, üzüntü, panikleme veya başka bir duygusal durumda olup olmadığımızı çoğu zaman açıkça ortaya koymaktadır.

2.1. Duyguların Sınıflandırılması

Duygu sınıflandırma konusunda boyutsal ve kategorik sınıflandırmayı kapsayan farklı yaklaşımlar mevcuttur. Paul Ekman; evrensel olarak tanınabilecek kızma, iğrenme, korku, mutluluk, üzüntü ve şaşırma duygularını “büyük altı” olarak adlandırmıştır. Bu 6 duyguyu temel duygu durumları olarak tanımlayarak bunların kategorik olduğu görüşünü savunmuştur [26].

Duygu sınıflandırmasında yaygın olarak kullanılan bir diğer teori, iki boyutlu yaklaşımdır. En iyi bilinen yapı, döngüsel model olarak da bilinen uyarılma-değerlik eksenlerini içeren modeldir. Şekil 2.1'de gösterildiği gibi eksenler uyarılma ve değerlik özelliklerini kullanır. Burada uyarılma eksen, duygusal deneyim için kullanılan enerjiyi temsil etmektedir. Değerlik eksen ise olumludan olumsuzu doğru giden ve duygu durumunun nasıl hissettirdiğini temsil eden bir ölçektir.



Şekil 2.1. 2 boyutlu duygu uzayı

3. SES ÖZELLİKLERİ

Konuşma sinyali; duygusal durumu, cinsiyeti ve konuşmacının kimliğini yansıtan bilgileri içerir. Her konuşmanın ve konuşmacının, konuşma ifadelerinde gizlenmiş bireysel nitelikleri vardır. Bir SDT sisteminin genel performansında özellik çıkarma işlemi önemli bir role sahiptir. Özellik çıkarma yönteminde öncelik, ses sinyalinin önemli özelliklerini yakalayıp performansa etkisi olmayacak özelliklerin atılmasıdır.

Duygu tanıma amaçlı olarak ses sinyallerinden özellik çıkarımında analiz bölgesine göre iki farklı yaklaşım bulunur. Bazı araştırmacılar konuşma sinyallerini çerçeve adı verilen küçük aralıklara ayırarak yerel bir özellik vektörünü çıkarırken diğer araştırmacılar, konuşma sinyalinin tamamını ele alarak global özelliklerin çıkarımını tercih etmektedir.

SDT sistemlerinde kullanılan yerel ve global özellikleri üç grupta incelenir: Prosodik özellikler, spektral özellikler ve ses kalitesi özellikleri.

3.1. Prosodik Özellikler

Prosodik özellikler, sesteki tonlama veya ritim gibi dinleyici tarafından fark edilebilen özelliklerdir. Yaygın olarak kullanılan prosodik özellikler; temel frekans, enerji ve zamana bağlı olarak çıkarılmaktadır.

Temel frekans, sesli harflerin meydana getirdiği ses telindeki titreşimler sayesinde oluşur. Sesin kalınlığını veya inceliğini gösterir. Enerji, konuşma sinyallerinde zaman içerisindeki genlik değişimine etki eder. Araştırmalarda, yüksek uyarılmanın gerçekleştiği duygularda (öfke, mutluluk veya şaşkınlık) yüksek enerjinin, düşük uyarılmanın gerçekleştiği iğrenme ve üzüntünün gibi duygularda ise düşük enerjinin olduğu anlaşılmıştır [27]. Zamana bağlı özelliklere; konuşma hızı, duraksama bölümlerinin süresi, sesli ve sessiz harflerin sürelerinin oranı örnek verilebilir.

3.2. Spektral Özellikler

Zaman bölgesindeki ses sinyallerinin Fourier Dönüşümü kullanılarak frekans alanına dönüştürülmesiyle elde edilen özelliklerdir. Perde ve enerji gibi zamana bağlı özelliklere ek olarak, spektral özellikler de genellikle konuşma sinyalinin kısa süreli temsili olarak seçilir. Bir ifadenin duygusal içeriğinde frekansın, konuşma aralığı boyunca spektral enerjinin dağılımında bir etkisi olduğu kabul edilmektedir [28]. Örneğin mutluluk duygusu içeren bir seslendirmede yüksek frekans aralığında yüksek enerjiye sahip iken, üzüntü içeren ifadelerin ise aynı aralıkta küçük enerjiye sahip olduğu bilinmektedir. Spektral özelliklere örnek olarak MFKK, LÖKK, LÖK verilebilir.

3.3. Ses Kalitesi Özellikleri

Ses kalitesi özellikleri ses yolunun fiziksel özelliklerine bağı olarak oluş an jitter, shimmer ve harmonik / gürültü oranı örnek verilebilir.

Jitter, birbirini izleyen titreşim döngüleri arasındaki temel frekansın değ işkenliğı shimmer ise genliğ in değ işkenidir. Jitter, frekanstaki kararsızlığ ının shimmer ise genlikte kararsızlığ ın ölçüsüdür. Harmonik Gürültü Oranı, sesli harflerin frekans spektrumu baz alınarak hesaplanan gürültü seviyesinin ölçümüdür.



4. MATERYAL VE METOT

Bu bölümde literatürdeki duygu veri setleri ve bu tez çalışmasında kullanılan SAVEE, RML ve RAVDESS veri setleri açıklanmıştır. Ses dosyalarının duygulara göre dengeli olarak dağılması, farklı dillerde seslendirilen konuşmaların bulunması ve veri setinin birden fazla kişi tarafından seslendirilmesi bu veri setlerinin tercih nedeni olmuştur.

4.1. Duygu Tanıma Veri Setleri

Duygu tanıma sisteminin performansını değerlendirmek için uygun bir veri seti üzerinde çalışmak oldukça önemlidir. Literatürde ses tabanlı duygu tanıma uygulamalarında kullanılan veri setleri 2 türdür:

Doğal duygu veri setleri: Çağrı merkezlerinden alınan konuşmalar, normal olmayan koşullarda alınan kokpit ses kayıtları, Talk Showlar gibi duyguların spontane geliştiği ses veya videolardan oluşmaktadır.

Taklit edilmiş duygu veri setleri: Genellikle profesyonel konuşmacıların belirli bir cümleyi farklı duygu durumları içinde seslendirdiği ses veya videolardan oluşmaktadır. Bu veri setlerinde farklı bir yaklaşım da mevcuttur. Bu yaklaşımda konuşmacılardan istenen duyguyu tamamen yapaylıktan uzaklaştırmak amacıyla duyguyu tetikleyecek etkiler yaratılır. Örneğin: [31] nolu çalışmada konuşmacılara ilk olarak bir senaryo anlatılmış ve konuşmacıdan anlatılan senaryodaki duygu duruma kendini kaptırması istenmiştir. Daha sonra konuşmacılar anlatılan senaryoyla bağlantılı olan cümleleri istenen duygu durumları içerisinde seslendirmiştir.

Tablo 4.1 seslendirme dil, duygu türleri, veri setinin doğallığı ve erişilebilirlik açısından tanımlanmış ve yaygın olarak kullanılan duygusal veri setlerinin bir listesini içermektedir.

Tablo 4.1. Yaygın olarak kullanılan duygu veri tabanları

Veri Seti	Duygu tipleri	Dil	Erişim	Doğallık	Seslendirme
Berlin Emotions Databases [29]	Öfke, mutluluk, sıkılma, nötr, iğrenme, mutsuzluk, panik	Almanca	Açık	Taklit	5 kadın 5 erkek
Danish Emotional Speech database [30]	Öfke, mutluluk, şaşırma, mutsuzluk	Danca	Açık	Taklit	2 kadın 2 erkek

Tablo 4.1. (Devamı)

eINTERFACE'05 [31]	Öfke, mutluluk, şaşırma, korkma, iğrenme, mutsuzluk	İngilizce	Açık	Taklit	34 erkek 8 kadın
RML [32]	Öfke, mutluluk, şaşırma, korkma, iğrenme, mutsuzluk	İngilizce, İtalyanca, Urduca, Çince, Farsça	Açık	Taklit	8 erkek
SAVEE [33]	Nötr, mutsuzluk, mutluluk, şaşırma, korkma, iğrenme	İngilizce	Açık	Taklit	4 erkek
RAVDESS [34]	Nötr, mutsuzluk, öfke, sakinlik, mutluluk, şaşırma, korkma, iğrenme	İngilizce	Açık	Taklit	12 erkek 12 kadın
SUSAS [35]	Stres altında kategorize edilmiş duygular	İngilizce	Ücretli	Doğal	13 erkek 19 kadın
IEMOCAP: [36]	Öfke, mutluluk, mutsuzluk, nötr hayal kırıklığı	İngilizce	Lisans sözleşmesi	Taklit	5 kadın 5 erkek
RECOLA [37]	Uyarılma-değerlik	İngilizce	Açık	Doğal	19 erkek 27 kadın
AFEW [38]	Nötr, mutsuzluk, öfke, mutluluk, şaşırma, korkma, iğrenme	İngilizce	Açık	Doğal	Kadın ve erkeklerden oluşan 300 seslendirici

4.2. RAVDESS Veri Seti

Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) veri seti; “Kids are talking by the door.” “Dogs are sitting by the door.” İfadelerinin kaydını içerir. Veri setini, 12 erkek 12 kadın konuşmacı seslendirmiştir ve 1440 ses dosyası mevcuttur. RAVDESS, cinsiyet açısından dengelidir ve ses dosyaları duygu sınıfları arasında eşit dağılımdadır. Nötr, kızgın, sakin, mutlu, şaşırma, korku,

iğrenme ve üzgün olmak üzere 8 farklı duyguyu içermektedir. Konuşmacılar ifadeleri, güçlü duygu yoğunluğu ve normal duygu yoğunluğuyla iki farklı varyasyonda seslendirmiştir.

4.3. SAVEE Veri Seti

Surrey Audio-Visual Expressed Emotion (SAVEE) veri seti 4 erkek konuşmacının İngilizce olarak seslendirdiği 480 videodan oluşmaktadır. Metinler TIMIT veri tabanından alınan 15 farklı cümleden oluşmaktadır. Seslendirmeler, farklı fonetik yapılarda olan cümlelerden oluşmaktadır. Kızgınlık, mutluluk, şaşırma, korku, iğrenme, üzüntü ve nötr olmak üzere 7 farklı duyguyu içermektedir. Nötr durumu için her bir konuşmacı 30 cümle diğer duygular için 15 cümle seslendirmiştir.

4.4. RML Veri Seti

RML görsel-işitsel duygu veri seti İngilizce, İtalyanca, Urduca, Çince, Pençapça ve Farsça dillerinde seslendirilmiş 720 videodan oluşmaktadır. Konuşmalar; kızgınlık, mutluluk, şaşırma, korku, iğrenme ve üzüntü olmak üzere 6 farklı duyguyu içermektedir.

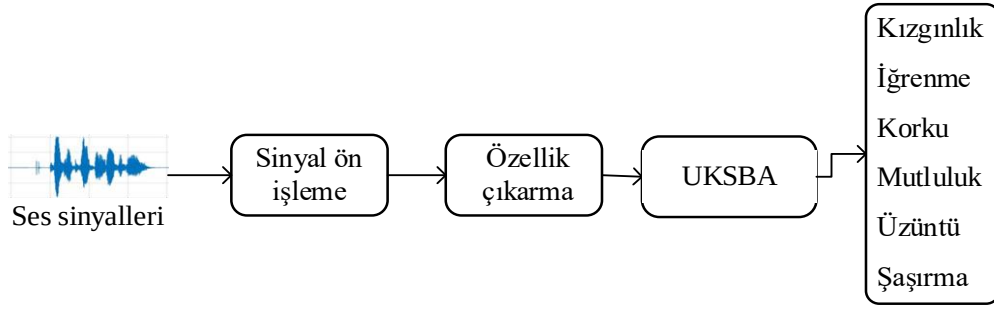
Tablo 4.2 'de RAVDESS, SAVEE ve RML veri setlerinde bulunan ses dosyalarının sayısı, duygu durumu etiketlerine göre verilmiştir.

Tablo 4.2. RAVDESS, SAVEE ve RML veri setlerinin içeriği

Veri Seti	Kızgınlık	İğrenme	Mutluluk	Korku	Üzüntü	Şaşırma	Nötr	Sakin
RAVDESS	192	192	192	192	192	184	96	192
SAVEE	60	60	60	60	60	60	120	-
RML	120	120	120	120	120	120	-	-

4.5. Önerilen Sistem

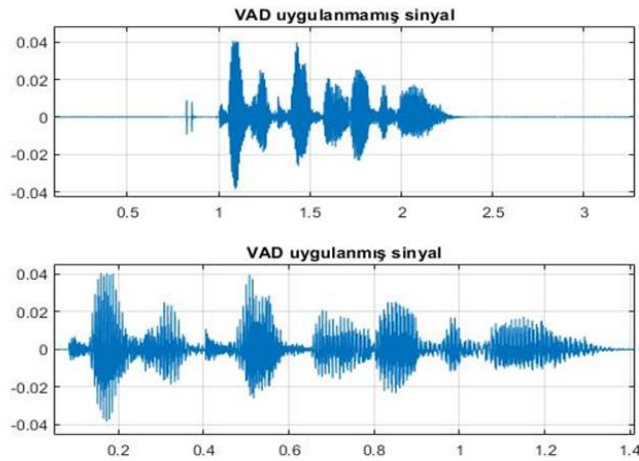
Önerilen SDT sisteminde, veri setinden alınan ses sinyallerine ilk olarak ses aktivite algılaması adı verilen bir ön işlem uygulanarak özellik çıkarma aşamasına hazırlanmıştır. Ses verilerinden, farklı duygu sınıflarını karakterize edebilen 113 ses özelliği çıkarılmıştır. Verilerin %90'ı ağız eğitimi için %10'u ise test amaçlı kullanılmıştır. Sınıflandırma Uzun- Kısa Süreli Bellek Ağlarıyla yapılmıştır. Önerilen yöntem SAVEE, RML ve RAVDESS veri setleri için ayrı ayrı gerçekleştirilmiş ve performansları değerlendirilmiştir. Şekil 4.1'de önerilen sistemin akış şeması verilmiştir.



Şekil 4.1. Önerilen sistemin akış şeması

4.6. Ses Aktivite Algılaması (Voice Activity Detection, VAD)

Ses sinyallerinde genellikle sessiz bölümler bulunur. Konuşmanın olmadığı ve bilgi içermeyen bu kısımların sinyalden çıkarılması gerekmektedir. Bu amaçla ses sinyalleri 200ms'lik çerçevelere bölündükten sonra belirlenen çerçeve süresince ses sinyalinin genliğinin sıfır geçiş sayısını kullanan ZCR (Zero Cross Rate) algoritması kullanılarak sessiz bölümler çıkartılmıştır. Bu işleme, Ses aktivite algılaması (Voice Activity Detection, VAD) denilmektedir. VAD uygulamasının ardından sinyallere normalizasyon işlemi uygulanarak özellik çıkarma aşamasına geçilmiştir. Şekil 4.2 'de VAD uygulaması yapılmış bir ses sinyali verilmiştir.

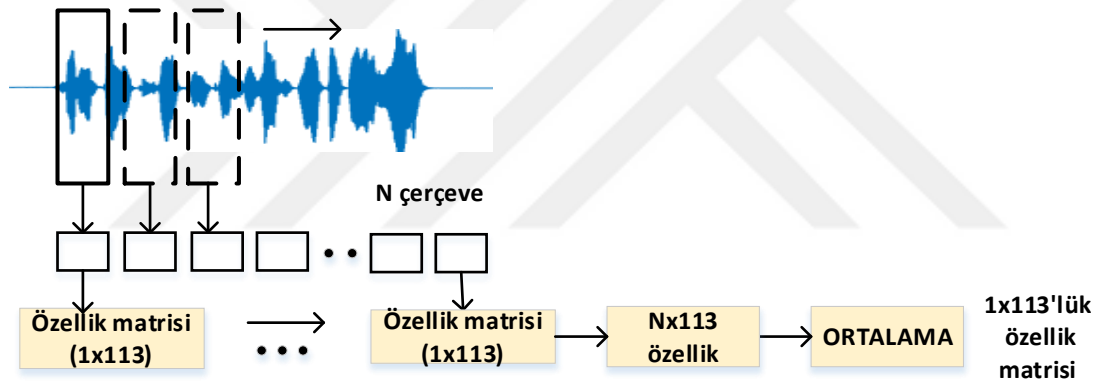


Şekil 4.2. Ses aktivite algılama işlemi

4.7. Özellik Çıkarma

SDT ile ilgili çalışmalarda, kesin ve ayırt edici bir sınıflandırma için genel kabul görmüş bir özellik dizisi bulunmamaktadır [39]. Konuşma özellikleri genellikle araştırmacıların deneyimlerine göre seçilmektedir.

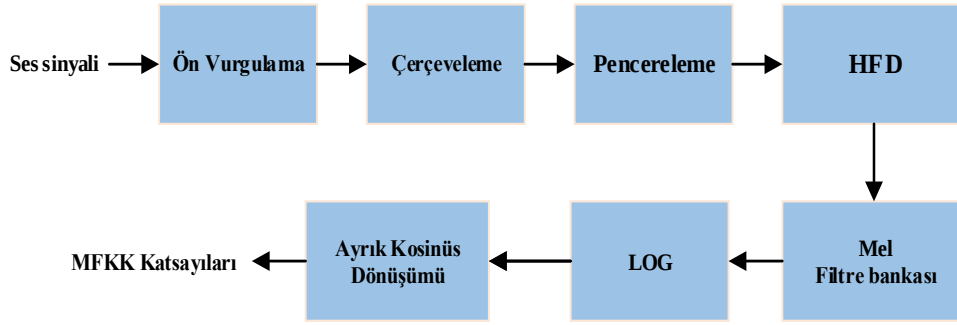
Frekansı veya spektral içeriği zamana göre değişmeyen sinyaller durağan olarak düşünülür. Ancak bir konuşma sinyali birçok farklı frekans içeriğine sahiptir ve bu bileşenler zamanla sürekli olarak değişir. Bu nedenle konuşma sinyalleri, konuşma sinyalinin durağan hale geldiği kısa bir süre boyunca (genellikle 10-30ms) incelenir. Bu süre, çalışmamızda 25ms olarak alınmıştır. Çalışma kapsamında, Mel Frekanslı Kepstral Katsayıları (MFKK), Lineer Öngörülü Kodlama (LÖK), spektral alt-bant ağırlık merkezi, Lineer Öngörü Kepstral Katsayıları (LÖKK), yansıma katsayıları, log- alan oranları katsayıları ve log-enerji özellikleri kullanılmıştır. Bir ses dosyası için, N çerçeve sayısı olmak üzere, 13 MFKK, 26 spektral alt-bant ağırlık merkezi, 12 doğrusal öngörülü kodlama, 12 doğrusal öngörü kepstral katsayıları, 26 log enerji, 12 log-alan oranları katsayıları ve 12 yansıma katsayıları olmak üzere $N \times 113$ 'lük bir özellik matrisi elde edilmiştir. Daha sonra çerçevelerden elde edilen özelliklerin ortalaması alınarak 1×113 'lük özellik matrisi elde edilmiştir. Bu işlem kullanılan 3 veri setindeki tüm ses dosyaları için gerçekleştirilerek, sınıflandırma işlemine geçilmiştir. Önerilen özellik çıkarma yöntemi şekil 4.3 'te verilmiştir.



Şekil 4.3. Özellik matrisinin elde edilişi

4.7.1. Mel Frekanslı Kepstral Katsayıları

Mel Frekanslı Kepstral Katsayıları (MFKK), konuşmacı tanıma sistemlerinde ve SDT sistemlerinde yaygın olarak kullanılan özellik çıkarma tekniklerinden biridir. İnsan işitme sistemini taklit eder ve Hızlı Fourier Dönüşümü (Fast Fourier Transform, FFT) tabanlı olarak hesaplanır. Bir ses sinyalinden MFKK katsayılarının elde edilmesi için izlenmesi gereken yol şekil 4.4.'te verilmiştir.



Şekil 4.4. MFCC katsayılarının elde edilişi

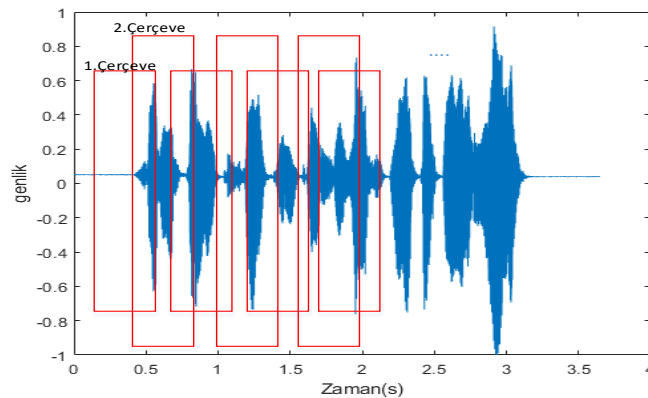
Ön Vurgulama

Ön vurgu filtresinin amacı, yüksek frekanslardaki ses sinyalinin enerjisini arttırmaktır. Ön vurgu filtresinin zaman bölgesindeki transfer fonksiyonu aşağıdaki eşitlikte gösterilen FIR (Finite Impulse Response) filtresi ile yapılır. Giriş ses sinyali $x(n)$ çıkış $y(n)$ çıkış olmak üzere ön vurgulama hesabı eşitlik (4.1)'de verilmiştir.

$$y(n)=x(n)-0.95x(n-1) \quad (4.1)$$

Çerçeveleme

Bu aşamada, ses sinyalleri genellikle 10 ile 200 milisaniye arasında değişen uzunluklarda çerçevelere bölünürler. Her çerçevenin, kendinden önceki çerçevenin bir bölümünü içerdği örtüşmeli çerçeveler kullanılarak çerçeveler arası geçişlerde oluşacak süreksizliğin etkisi düşürülür. Şekil 4.5'te bir ses sinyaline çerçeveleme işlemi gösterilmiştir.



Şekil 4.5. Çerçeveleme işlemi

Pencereleme

Ses sinyali çerçevelere bölündükten sonra giriş veri sinyalinin her çerçevesinin kenar bölümlerinde bazı süreksizlikler oluşur. Bu süreksizlikten kaçınmak için pencereleme işlemi yapılmalıdır. Bu amaçla bütün çerçeveler Hamming penceresi ile çarpılmıştır. $x(n)$ giriş işareti $w(n)$ pencere fonksiyonu ve $y(n)$ çıkış işareti olmak üzere çıkış işareti aşağıda verilen eşitlik ile hesaplanır.

$$y(n) = x(n)y(n) \quad (4.2)$$

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-7}\right) , \quad 1 \leq n \leq N \quad (4.3)$$

Hızlı Fourier Dönüşümü

Bu aşamada, ses sinyalinin spektral bilgisi elde edilmektedir. Ayırık Fourier Dönüşümü (AFD), ayırık frekans bantlarında ayırık zamanlı bir sinyal için spektral bilginin çıkarılması amacıyla kullanılan bir dönüşüm tekniğidir. Hızlı fourier dönüşümü ise AFD'yi hesaplamak için yaygın olarak kullanılan bir yöntemdir. N örnek sayısı olmak üzere AFD'nin matematiksel ifadesi aşağıdaki eşitlikte verilmiştir.

$$X_n = \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi kn}{N}} \quad (4.4)$$

Mel Filtre Bankası

İnsan kulağı, ses frekanslarını 1 kHz'in altı içim doğrusal, 1 kHz'in üzeri için logaritmik olarak algılar. Mel frekans ölçeği ise insan duyma sisteminden esinlenerek tasarlanmıştır. Hz. cinsinden verilen frekansı mel frekansına dönüştürmek için, (4.5) eşitliği kullanılır. Mel filtre bankası, bant genişliği mel ölçeğine göre değişen üçgen bir bant geçiren filtredir.

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (4.5)$$

Ayrık Kosinüs Dönüşümü

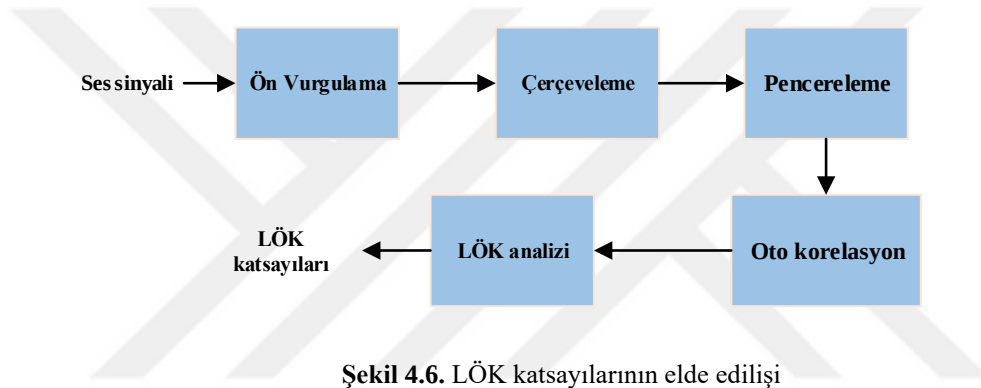
Mel spektrum katsayılarının logaritması (4.6) bağıntısında verilen ayırık kosinüs dönüşümü eşitliğiyle zaman bölgesine geçirilir. Sonuçta mel frekansı kepsral katsayıları elde edilmiş olur.

$$C_n = \sum_{k=1}^K (\log D_k) \cos \left[m\left(k - \frac{1}{2}\right) \frac{\pi}{K} \right] , \quad m=1, 2, \dots, K \quad (4.6)$$

Burada D_k , $k=1, 2, \dots, K$ mel spektrum katsayılarıdır. Dönüşüm sonucu elde edilen ilk bileşen C_0 ortalama logaritmik enerjiyi temsil ettiği için genellikle öznitelik vektöründen çıkarılır. Bu çalışmada öznitelik vektörü olarak ilk 13 MFKK katsayısı kullanılmıştır.

4.7.2. Lineer Öngörülü Kodlama

İnsan sesi üretimi sisteminin modellenmesinde en popüler yaklaşımlardan biri Lineer Öngörülü Kodlamadır. LÖK hesabında izlenmesi gereken adımlar şekil 4.6 ‘da verilmiştir. Ön vurgulama, çerçeveleme ve pencereleme işlemine MFKK hesabında değinilmişti. Pencerelemiş sinyalin her bir çerçevesine oto korelasyon uygulandıktan sonra LÖK analizi yapılarak LÖK katsayıları elde edilir.



Şekil 4.6. LÖK katsayılarının elde edilişi

Oto Korelasyon ve LÖK Analizi

Pencereleme işlemi uygulanmış sinyale (4.7) eşitliği ile verilen Oto korelasyon uygulanarak LÖK katsayıları hesaplanmıştır. Burada p LÖK analizinin derecesidir.

$$r(m) = \sum_{n=0}^{N-1-M} x(n)x(n+m) , \quad m = 0, 1, \dots, p \quad (4.7)$$

4.7.3. Lineer Öngörü Kepstral Katsayıları

LÖKK katsayıları, LÖK katsayılarının (4.8) eşitliği kullanılarak kepsral katsayılar dönüşürmüş halidir. Burada m kepsral katsayıların sırasıdır.

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k} \quad 1 \leq m \leq p \quad (4.8)$$

4.7.4. Yansıma Katsayıları

Yansıma katsayıları, LÖK katsayılarından geriye doğru özyineleme yoluyla (4.9) eşitliğiyle elde edilebilir.

$$k(n) = a_n(n)$$

$$a_{n-1}(m) = \frac{a_n(m) - k(n)a_n(n-m)}{1 - k(n)^2} \quad m = 1, 2, 3, \dots, n-1 \quad (4.9)$$

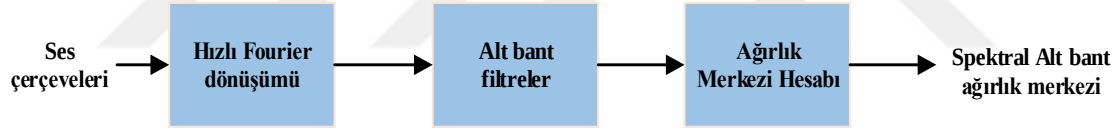
4.7.5. Log- Alan Oranları Katsayıları

Log- Alan oranları katsayıları yansıma katsayılarından (4.10) eşitliği kullanılarak elde edilmektedir. r_k =yansıma katsayısı olmak üzere,

$$A_k = \ln \left(\frac{1-r_k}{1+r_k} \right) \quad (4.10)$$

4.7.6. Spektral Alt-bant Ağırlık Merkezi

Spektral alt-bant ağırlık merkezinin hesaplanma aşamaları şekil 4.7 de verilmiştir.



Şekil 4.7. Spektral alt bant ağırlık merkezi hesaplanışı

$S[k]$: Bir çerçevenin HFD genlik spektrumu

$k=1, \dots, N$: Ayırık frekans indeksi ($k=N$ örnekleme frekansının yarısına karşılık gelir)

$g_l(m), g_h(m) \in [1, N]$: düşük ve yüksek kesim frekansı

γ : dinamik aralık parametresi olmak üzere m . alt bant ağırlık merkezi olmak üzere

$$C_m = \frac{\sum_{k=g_l(m)}^{g_h(m)} kW_m[k]S^\gamma[k]}{\sum_{k=g_l(m)}^{g_h(m)} W_m[k]S^\gamma[k]} \quad (4.11)$$

Eşitliği ile spektral alt bant ağırlık merkezi elde edilir.

4.7.7. Log- Enerji

Ses sinyalinin enerjinin logaritmasıdır. Log-enerji (4.12) eşitliği kullanılarak hesaplanmaktadır.

$$E_s = 10 \log_{10} \left(\frac{1}{N} \sum_{n=1}^N s^2(n) \right) \quad (4.12)$$

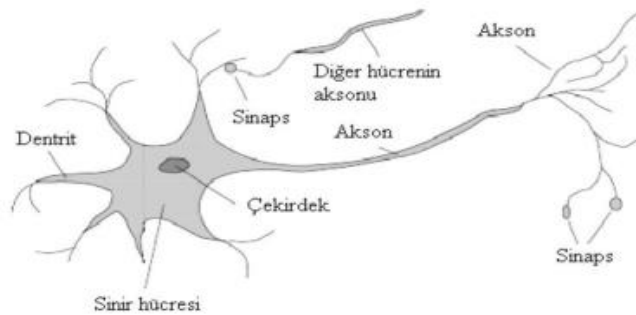
5. YAPAY ZEKÂ, MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME

Yapay zekâ, minimum insan müdahalesiyle akıllı davranışları modellemek için bir bilgisayarın kullanılmasını ifade eden genel bir terimdir. Makine öğrenmesi, tahminlerde bulunmak veya kararlar vermek için "eğitim verileri" olarak bilinen örnek verilere dayalı bir matematiksel model oluşturan yapay zekâ alt kümesidir [40]. Makine öğrenmesinin temel amacı, sağlanan verilere ve algoritmalara dayanarak makineleri eğitmektir. Öğrenme tabiri, makine öğrenmesi algoritmalarının hataları en aza indirgeyerek ve tahminin doğru olma olasılığını en üste çıkarmaya çalışmasıdır.

Son on yıldan beri makine öğrenmesinin gelişimi için çalışmalar yapılmaktadır. Bu alandaki beklentileri yükselten derin öğrenme, bu girişimlerden biridir. Yapay sinir ağlarından geliştirilen ve makine öğreniminin bir alt alanı olan derin öğrenme kavramı ilk olarak 2006 yılında ortaya atılmıştır [41]. Derin kelimesi yapay sinir ağındaki çoklu saklı katmanı ifade etmektedir. Geleneksel sinir ağları yalnızca 2- 3 gizli katman içerirken, derin ağlar ise 150'ye kadar gizli katmana sahip olabilir. Derin öğrenme, evrimsel sinir ağları ve tekrarlayan sinir ağları gibi çeşitli mimarilerden oluşur. Görüntü tanıma, nesne tanıma, konuşma tanıma, müzik sınıflandırma ve otomatik duygu tanıma bu mimarilerin kullanıldığı uygulama alanlarına örnektir.

5.1. Yapay Sinir Ağlarının Yapısı

Yapay sinir ağları (YSA); insan beyninin öğrenme yolunu taklit ederek beynin öğrenme, hatırlama, genelleme yapma yolu ile topladığı verilerden yeni veri üretebilme gibi temel işlevlerin gerçekleştirildiği bilgisayar yazılımlarıdır [42]. YSA deneyimlerden faydalanarak öğrenebilmekte ve bilgiler arasında bağlantılar kurabilmektedir. Yapay Sinir Ağları, insan beyninin sinir sisteminden esinlenerek geliştirilmiştir. Beyinde milyarlarca sinir hücresi ve bu hücrelerin bağlantılarını sağlayan sinapsisler bulunur. YSA'ya yönelik ilk çalışmalar 1943'te McCulloch ve Pitts'in sinir hücrelerinin yani nöronların matematiksel modelinin çıkarılmasıyla başlamıştır [43].



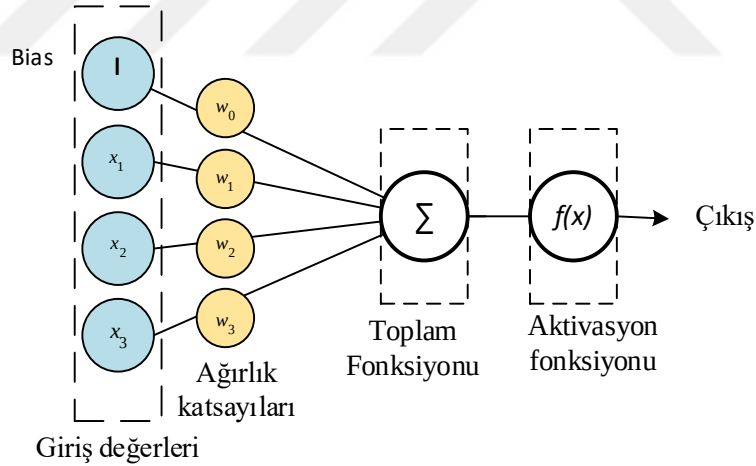
Şekil 5.1. Biyolojik sinir hücresi [44]

Biyolojik sinir sistemini oluşturan nöronların yapısı şekil 5.1’de görüldüğü üzere dört ana bölümden oluşmaktadır. Dendritler, nöronlardan gelen sinyallerin çekirdeğe iletimini sağlamaktadır. Çekirdekte toplanan sinyaller aksonda işlendikten sonra nöronun diğer ucundaki bağlantı noktalarına aktarılır. Bu sistemin yapay sinir ağındaki benzetimi Tablo 5.1’ de verilmiştir.

Tablo 5.1. Biyolojik sinir ağı elemanlarının yapay sinir ağılarındaki karşılıkları

BİYOLOJİK SİNİR AĞLARI	YAPAY SİNİR AĞLARI
Nöron	İşlemci eleman
Aksonlar	Eleman çıkışı
Dendrit	Toplama fonksiyonu
Hücre Gövdesi	Aktivasyon fonksiyonu
Sinapslar	Ağırlık

Şekil 5.2’de görüldüğü üzere bir Yapay sinir ağında beş ana eleman bulunmaktadır. Bunlar girdiler, ağırlıklar, toplam fonksiyonu, aktivasyon fonksiyonu ve çıkıştır.



Şekil 5.2. Yapay sinir ağı hücresi

Girdiler: Bir yapay sinir ağında dış dünyadan veya diğer bir yapay sinir hücresinden bilgi alınan kısımdır.

Ağırlıklar: Ağırlık, girdinin çıktı üzerinde ne kadar etkisi olacağına karar verir. Örneğin; düğüm 1’den düğüm 2’ye kadar olan ağırlığın daha büyük olması, nöron 1’in nöron 2 üzerinde daha büyük etkiye sahip olduğu anlamına gelir. Negatif ağırlıklar, bu girdinin artırılmasının çıktıyı azaltacağı anlamına gelir.

Toplam Fonksiyonu: Hücreye gelen net girdinin hesaplandığı yerdir. Ağırlıklandırılmış girdi değerlerine eşik değerinin eklenmesiyle hesaplanmaktadır.

$$\sum_{i=1}^n x_i w_i \quad (5.1)$$

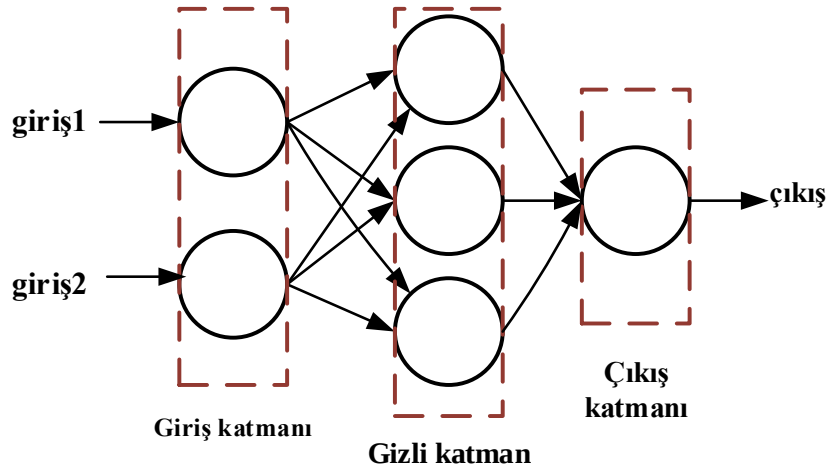
Aktivasyon Fonksiyonu: YSA'da nöronun çıkış değerini belirlenen değerlere sınırlandırmak için kullanılır. Elde edilen değeri, kullanılan fonksiyona bağlı olarak genellikle (0-1) veya (-1 1) arasına sıkıştırır. Aktivasyon Fonksiyonu, genellikle doğrusallık göstermez. En sık kullanılan fonksiyonlar: Doğrusal fonksiyon, tek kutuplu basamak fonksiyonu çift kutuplu basamak fonksiyonu, sigmoid fonksiyonu ve hiperbolik tanjant fonksiyonudur.

Çıktı: Aktivasyon fonksiyonunun sonucunun dış dünyaya gönderildiği katmandır.

Yapay sinir ağları, hücrelerin bağlantı şekillerine göre ileri beslemeli ağ ve geri beslemeli ağ şeklinde sınıflandırılmaktadır.

5.1.1. İleri Beslemeli Yapay Sinir Ağları

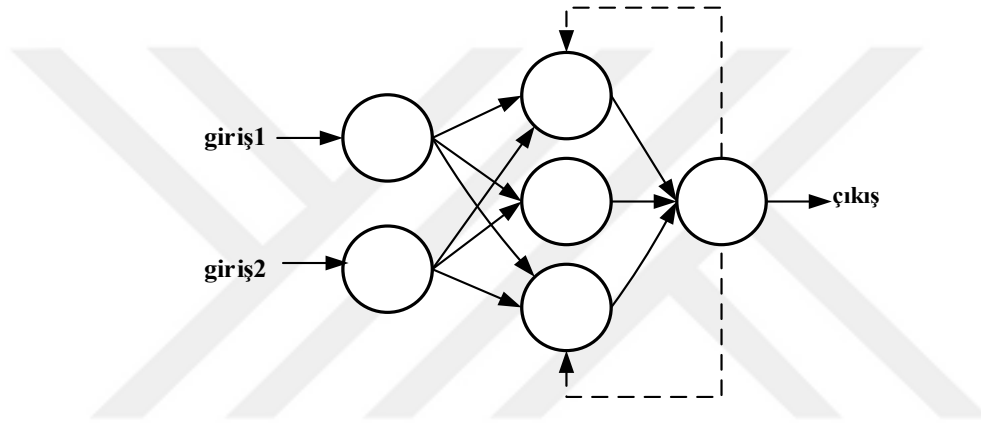
İleri beslemeli yapay sinir ağı yapısında nöronların yalnızca bir sonraki katmanda bulunan nöronlarla bağlantısı bulunur. Alınan bir çıkışın gerideki katmanlara giriş olması mümkün değildir. Şekil 5.3 'te görüldüğü üzere aynı katmandaki nöronların birbirleriyle bağlantılı olmaması ağın ileriye doğru iletimine olanak sağlamıştır. Giriş katmanındaki bilgiler gizli katmanlara değişim olmadan aktarılır ve çıkış katmanı işlenen bilgi dış dünyaya aktarılır.



Şekil 5.3. İleri beslemeli sinir ağı

5.1.2. Geri Beslemeli Yapay Sinir Ağları

Geri beslemeli yapay sinir ağlarında, ileri beslemeli sinir ağlarından farklı olarak bir nöronun çıkışı yalnızca kendinden sonraki katmana değil önceki katmanda veya kendi katmanında bulunan bir nörona giriş olarak bağlanabilir. Şekil 5.4’de görüldüğü gibi giriş değerleri hem ileri hem de geri yönde aktarılabilir. Bir ağı yinleme eklemek, özellikle öğrenme sürecini güçlendirmek için dahili bir bellek gerektiren zaman serilerinde veya örüntü tanımayı içeren problemlerde fayda sağlamaktadır. Geri beslemeli ağların çoğu, Elman ve Hopfield ağları gibi tek katmandır. Ancak tekrarlayan çok katmanlı sinir ağları gibi yinelenen çok katmanlı bir ağ oluşturmak mümkündür.



Şekil 5.4. Geri beslemeli sinir ağı

5.2. Derin Öğrenmenin Tarihçesi

Derin Öğrenme, iki veya daha fazla gizli katman içeren çok katmanlı sinir ağlarını kullanan bir makine öğrenimi tekniğidir. Son yıllarda popüler olmasına rağmen ortaya çıkışı uzun bir geçmişe dayanmaktadır.

1943’te sinir ağları Amerikalı bir matematikçi olan Walter Pitts ve bir psikolog olan Warren McCulloch tarafından sinir ağının matematiksel olarak modellenmesi yapay sinir ağlarının ve derin öğrenmenin temelini oluşturmuştur [45].

1957’de Frank Rosenblatt, birçok bilim adamının bu alandaki ilgisine ilham verecek tek katmanlı algılayıcıyı ortaya attı [46].

1965’te Ivakhnenko çok katmanlı algılayıcıyı tasarladı [47].

1969’da Marvin Minsky ve Seymour Papert “algılayıcılar” kitabında, tek katmanlı algılayıcının çeşitli sınırlamalarını gösteren bir dizi matematiksel kanıt sundular [48]. Örneğin:

XOR gibi karışık işlemleri çözemediğini gösterdiler. Bu durum yapay sinir ağlarının 1970'lerde daha az popüler olmasına neden oldu.

Çok katmanlı sinir ağının eğitim problemi 1986'da David Rumelhart, Geoffrey Hinton tarafından geri yayılma algoritması tanıtıldığında çözüldü [49].

Yann LeCun, 1989'da yaygın olarak kullanılan bir derin öğrenme modeli olan evrişimsel sinir ağını önerdi ve el yazısı rakamlarını tanıma görevine başarıyla uyguladı [50].

1980'lerde bilgisayarların çok yavaş olması derin ağların eğitimini güçleştiriyordu. Ayrıca geri yayılma algoritmasıyla eğitim yapılırken, sinir ağını büyütmek ilerlemeyi sınırlayan ortadan kaybolan gradyan sorununa yol atmıştır. 1990'larda destek vektör makinesi gibi algoritmalar önerildi. Sınıflandırma ve regresyon problemlerinde sinir ağlarından daha iyi performans göstermesi, sinir ağı araştırmalarına olan ilginin azalmasına neden oldu.

S. Hochreiter ve J. Schmidhuber, 1997 yılında derin öğrenmede devrim yaratmaya devam edecek tekrarlayan sinir ağı mimarisi olan “Uzun Kısa Süreli Bellek” üzerine bir makale yayınladı [51].

2006 yılında, G. Hinton “derin öğrenme” terimini ilk kez kullandı ve derin öğrenmenin başlangıcı edildi. Yayınladıkları makalede, kaybolan gradyan sorununa ayrıntılı bir çözüm sundular [52].

Bilgisayar donanımlarındaki hızlı ilerlemeler ve GPU'ların geniş kullanımı, bilgi işlem gücünü artırmıştır ek olarak İnternet sayesinde büyük verilerin elde edilmesi mümkün olmuştur. Bu durumlar derin öğrenmeye olan ilginin artmasını sağlamıştır.

2011 yılında, G. Hinton ve Microsoft'un araştırmacıları, derin öğrenmeyi konuşma tanıma problemin uyguladı ve önemli bir dönüm noktasına ulaştı. 2012'de G.Hinton ve ekibi, derin öğrenme modeli AlexNet ile görüntü sınıflandırma yarışmasını kazandı. Derin öğrenmenin konuşma ve görüntü görevlerindeki atılımları, akademi ve endüstriden büyük ilgi görmüştür.

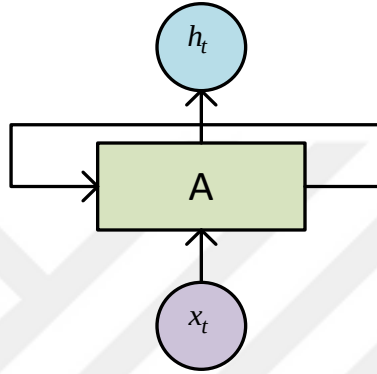
2016 yılında, Google DeepMind tarafından derin öğrenmeye dayalı olarak geliştirilen AlphaGo, Güney Koreli profesyonel go oyuncusu ve dünya şampiyonu karşı galibiyet dünyayı şok etti. Oyunun satrançtan daha karışık olması göz önüne alındığında bu başarı derin öğrenme vaadini yepyeni bir seviyeye taşımıştır.

Derin öğrenme algoritmaları dört temel kategoride ele alınabilir:

- Evrişimsel Sinir Ağları
- Tekrarlayan Sinir Ağları
- Kısıtlanmış Boltzmann Makineleri
- Otomatik Kodlayıcılar

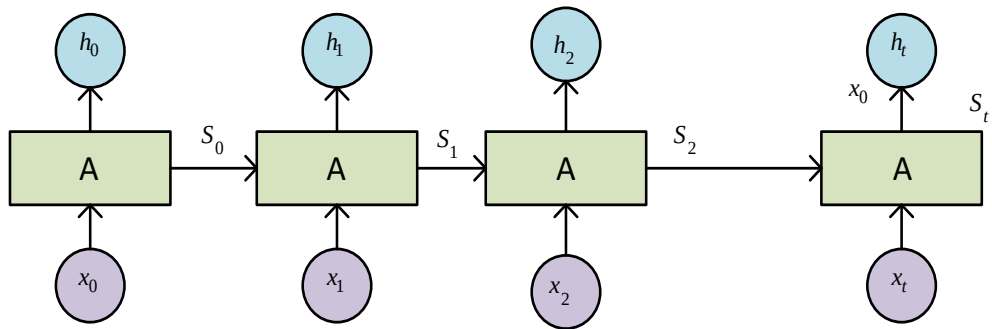
5.3. Tekrarlayan Sinir Ağları

İleri beslemeli sinir ağlarının sıralı verileri işleyememesi, sadece mevcut girişi dikkate alırken önceki girişleri hatırlayamaması Tekrarlayan Sinir Ağlarının (TSA) yaratılmasına neden olmuştur. İleri beslemeli yapay sinir ağlarından farklı olarak bir TSA bir önceki çıktıları giriş olarak ağı aktararak geri besleme sağlar. TSA'lar, sıralı verilerde veya geçmiş durumların hatırlanmasını gerektiren görevlerde kullanılan temel bir derin öğrenme mimarisidir. Bilgilerin kalıcı olmasını sağlayan döngülere sahiptir yani bu ağlar belleğe sahiptir [53].



Şekil 5.5. TSA mimarisi

Şekil 5.5 'te A ağı, t zamanda x_t girişi almış ve h_t çıktısı üretmiştir. Bu mimari, bir önceki zamana ait bilgilerin kullanılmasına imkân sağlamaktadır. TSA yapısındaki bu döngüler anlaşılması güç gibi görünse de açık haline bakıldığında zaman klasik yapay sinir ağına benzediği anlaşılmaktadır.



Şekil 5.6. TSA mimarisinin açık gösterimi

Şekil 5.6 'da TSA yapısının açık hali incelenirse; Dizideki her aşamada, her nöronunu iki şey ürettiği görülür:

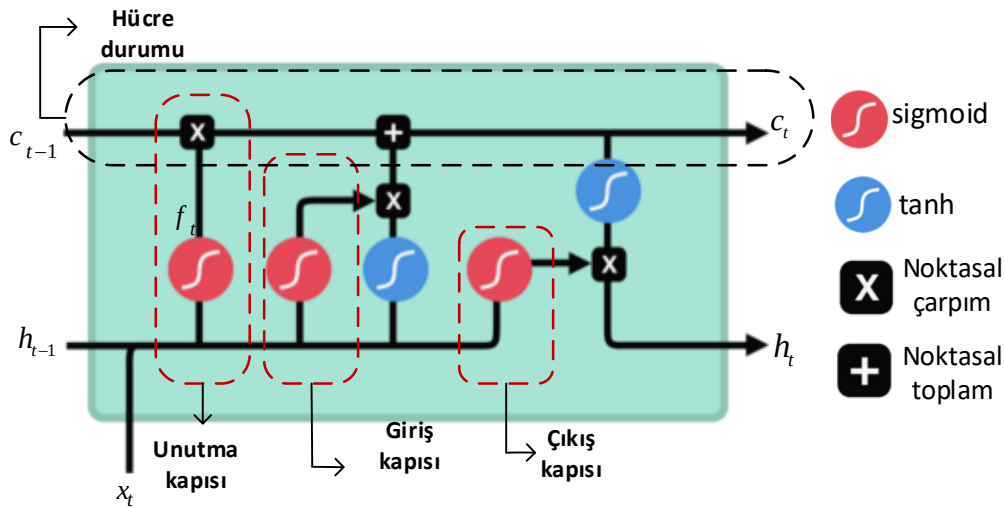
Çıktı ($H_0, H_1, H_2 \dots$): Dizideki sonraki ögenin ne olması gerektiğine dair modelin tahminidir.

Gizli durum ($S_0, S_1, S_2 \dots$): Ağın kısa süreli hafızası olarak düşünülür. Hem önceki adımın gizli durumunu hem de mevcut adımın çıktısını girdi olarak alan bir aktivasyon fonksiyonudur. Modelin önceki adımlardan mevcut adıma bilgi taşımaya izin verir.

TSA'ların belleğe sahip olan yapısı sıralı gerçekleşen olayların birbirleriyle ilişkilendirilmesini sağlar. Ancak eğer mevcut durumu etkileyen önceki durum yakın geçmişte değilse TSA doğru bir tahminde bulunamayabilir. Örneğin 'Ayşe'nin fındık alerjisi var. O fındık ezmesi yiyemez' cümlesinde fındık içeren ürünlerinde alerji kapsamında olduğunu bağlantısı çıkar. Ancak eğer bu bağlam birkaç cümle öncesinde olaydı TSA'nın bu bağlamı kurması zor hatta imkânsız olabilirdi. Bu sorunu düzeltmek için, uzun kısa süreli bellek ağları mimarisi ortaya atılmıştır [54]. Bu ağ yapısında sinir ağının gizli katmanlarında bir giriş kapısı, bir çıkış kapısı ve bir unutma kapısına sahip hücreler bulunur. Bu kapılar, ağıdaki çıktıyı tahmin etmek için gereken bilgi akışını kontrol eder. Örneğin, "o" gibi zamirleri önceki cümlelerde birçok kere tekrar edildiyse bu hücre durumunun dışına bırakılabilir.

5.4. Uzun-Kısa Süreli Bellek Ağları (UKSBA)

Tekrarlayan Sinir Ağlarının en büyük dezavantajının kısa süreli belleğe sahip olması önceki bölümde belirtilmişti. Bu temel problemin nedeni geri yayılım sırasında, kaybolan gradyandır. Gradyanlar, sinir ağlarının ağırlıklarını güncellemek amacıyla kullanılan değerlerdir. Gradyanın zaman içinde geri yayılırken küçülmesi kaybolan gradyan sorununa neden olur. Eğer bir gradyan değeri aşırı küçülürse öğrenmeye katkıda bulunmaz. Dolayısıyla, küçük gradyan güncellemesi alan katmanlarda öğrenme durur. Bu durum genellikle önceki katmanlarda oluşur. Sonuç olarak bu katmanlarda öğrenme sorunu yaşanır ve TSA'lar kısa süreli bir hafızaya sahip olur. Şekil 5.7 'de UKSBA hücresi ve işlemler verilmiştir.



Şekil 5.7. UKSBA hücresi ve işlemler [55]

UKSBA'lar kısa süreli belleğe çözüm olarak oluşturulmuştur. “Kapı” olarak adlandırılan bilgi akışını düzenleyebilen mekanizmalar bulunur. Bu kapılar, hangi verilerin saklanması veya atılmasının önemli olduğunu öğrenebilir. Bu sayede tahminlerde bulunabilmek için gerekli bilgileri uzun dizi zincirinden geçirebilir. Kapılar ve fonksiyonların işlevleri aşağıda verilmiştir.

x_t : t zaman adımındaki giriş verilerini

h_t : gizli durum çıkışı, x_{t-1} onların önceki durum çıkışı

i_t : giriş kapısı çıkışı

f_t : unutma kapısı

o_t : çıkış kapısını

c_t : mevcut hücre durumunu ve W ağırlıkları ve b biası temsil etmektedir.

Tanjant hiperbol (Tanh) aktivasyonu: Ağ üzerinden iletilen değerlerin düzenlenmesine yardımcı olmak için kullanılır. Tanh işlevi, giriş değerlerini her zaman -1 ile 1 arasında olacak şekilde sıkıştırır.

Sigmoid aktivasyonu: Bir sigmoid aktivasyonu, giriş değerlerini 0 ile 1 arasına sıkıştırır. Bu durum verileri güncellemek veya unutmak için faydalıdır çünkü 0 ile çarpım değerlerin kaybolmasına veya “unutulmasına” sebep olur. 1 ile çarpımda ise değer aynı kalır veya “tutulur”.

Unutma Kapısı: Bu kapı, hangi bilgilerin atılması veya saklanması gerektiğine karar verir. Önceki gizli durumdan bilgiler ve mevcut girişten gelen bilgiler sigmoid işlevinden geçirilerek 0 ile 1 arasına sıkıştırılır. 0'a yakınlık bilgiyi unutma, 1'e yakınlık ise bilgiyi tutma anlamına gelir.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (5.2)$$

Giriş Kapısı: Mevcut hücre durumunu güncelleme işlemini gerçekleştirildiği kapıdır. İlk aşamada gizli durum ve mevcut giriş sigmoid aktivasyonuna aktarılarak hangi değerlerin güncelleneceğine karar verilir. 0 önemli olmayan ve 1 önemli olan bilgi anlamına gelir. Daha sonra, tanh katmanı mevcut hücre durumuna eklenebilecek yeni aday değerlerin vektörü ($\tilde{c}_t$) elde edilir. Son aşamada giriş katmanında bu iki değer bir araya gelir saklanacak olan yeni değer oluşturulur.

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5.3)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (5.4)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5.5)$$

Çıkış Kapısı: Çıkış kapısı, bir sonraki gizli durumun ne olması gerektiğine karar verir. Gizli durum, önceki girişler hakkında bilgi içermektedir ve tahminler için kullanılır. Gizli durum için, önceki gizli durum ve mevcut giriş bir sigmoid işlemine aktarılır. Daha sonra yeni değiştirilmiş

hücre durumu tanh fonksiyonundan geçirilir. Yeni hücre durumu ve yeni gizli durum bir sonraki zaman adımına taşınır.

$$o_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + b_o) \quad (5.6)$$

$$h_t = o_t \tanh(c_t) \quad (5.7)$$



6. BULGULAR

Uzun Kısa Süreli Bellek ağının eğitimi için ses verilerinin %90'ı eğitim için ve %10'u test için kullanılarak her veri setinin performansını ayrı ayrı değerlendirilmiştir. Ayrıca, eğitim ve test verileri rastgele seçildiği için, deneysel çalışmalar 100 kez tekrarlatılmış ve ortalama bir doğruluk hesaplanmıştır. Bütün deneysel çalışmalar ve analizler MATLAB programında gerçekleştirilmiştir.

6.1. Hata matrisi ve Tanıma Doğruluğu

Hata matrisi, sınıflandırıcılarının başarısını ölçmek amacıyla kullanılan özel bir tablodur. Hata matrisinin her bir satırında algoritmanın doğru ve yanlış şekilde tahmin ettiği örneklerle ve gerçek değerlerin olduğu dört farklı sayısal endeks bulunmaktadır. Şekil 6.1'de verilen hata matrisinde:

Hedef Sınıf			
Çıkış Sınıfı	DP	YP	
	YN	DN	

Şekil 6.1. Hata matrisi

Doğru pozitif (DP): Doğru şekilde tahmin edilen doğru örnekler

Doğru negatif (DN): Doğru şekilde tahmin edilen yanlış örnekler

Yanlış pozitif (YP): Yanlış tahmin edilen doğru örnekler

Yanlış negatif (YN): Yanlış tahmin edilen yanlış örnekleri temsil eder.

Hata matrisi, sınıflandırıcının karakteristiğini temsil eden farklı başarımların hesaplanmasına olanak sağlamaktadır. Bu tez kapsamında kullandığımız başarımların kriteri “doğruluk” ölçütüdür. Matematiksel eşitliği aşağıda tanımlanmıştır.

$$\text{Doğruluk} = \frac{DP + DN}{DP + YP + YN + DN} \quad (6.1)$$

6.2. Rml Veri Seri İçin Sınıflandırma Sonuçları

Tablo 6.1’de verilen hata matrisinde görüleceği üzere, en iyi doğruluk “Kızgınlık” duygusu için elde edilmiştir. En düşük doğruluk ise “korku” duygusu için elde edilmiştir. “Kızgınlık” duygusu için ortalama doğruluk %81,3 iken “korku” duygusunun ortalama başarımı ise %49,5’tir. “Kızgınlık” duygusu en çok “şaşıрма” duygusu ile karıştırılmıştır. Ayrıca, ikinci en yüksek ortalama doğruluk oranı %75,8 ile “üzüntü” duygusu için elde edilmiştir. Yine “üzüntü” duygusu en çok “korku” duygusu ile karıştırılmıştır. Sistemin en çok karıştırdığı iki duygu “iğrenme” ve “korku” duygusu iken en iyi ayırt ettiği iki duygu “kızgınlık” ve “iğrenme” ile birlikte “kızgınlık” ve “üzüntü” duygu olmuştur. 6 duygu için ortalama doğruluk %70,35 olarak hesaplanmıştır.

Tablo 6.1. RML veri seti için hata matrisi (%)

	Kızgınlık	İğrenme	Korku	Mutluluk	Üzüntü	Şaşıрма
Kızgınlık	81,3	0	2,91	3,67	0	12,12
İğrenme	0	68,3	22,6	2,2	4,6	3,3
Korku	7,6	18,25	49,5	7,65	9,6	7,4
Mutluluk	5,1	7,35	7,75	72,5	1,7	5,6
Üzüntü	0	5,82	10,83	7,55	75,8	0
Şaşıрма	10	3,25	6,47	5,58	0	74,7

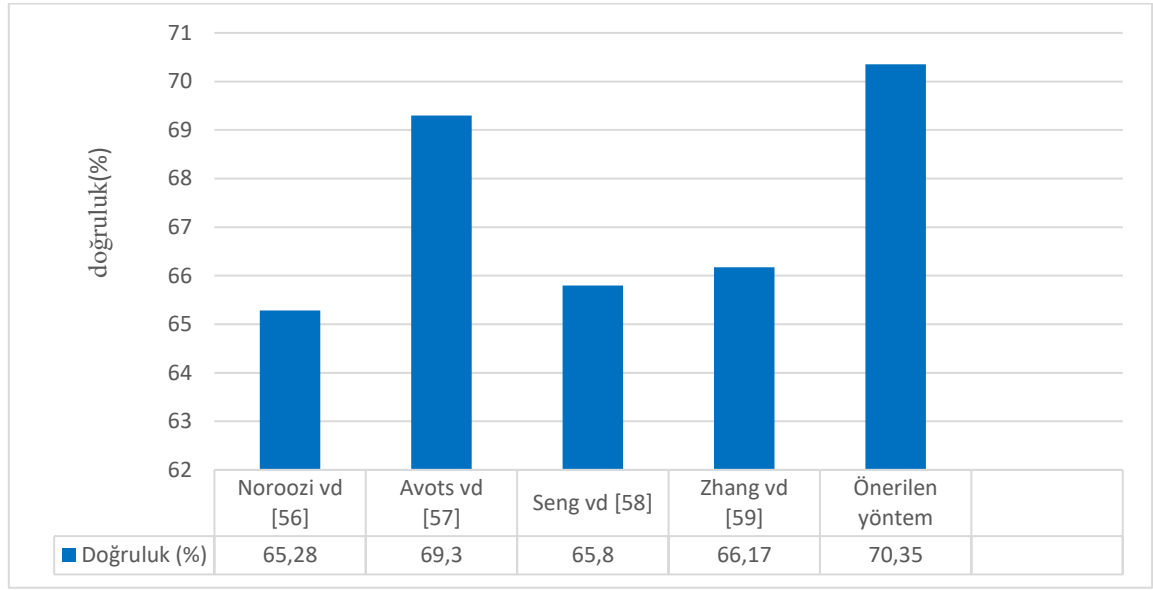
Şekil 6.2’de RML veri seti kullanılarak yapılan literatürdeki çalışmalar gösterilmiştir. Bu çalışmalarda kullanılan metotlar aşağıda detaylandırılmıştır.

Noroozi vd [56] yaptıkları çalışmada işitsel ve görsel verilerin işlenmesine dayanan multimodel bir duygu tanıma sistemini sunmuştur. Ses sinyallerinden MFCC özelliklerinin yanı sıra filtre bankası enerjileri ve prosodik özellikleri içeren 88 özellik çıkarmışlardır. Özellikler destek vektör makinesi ve rastgele orman sınıflandırıcılarında değerlendirilmiştir. 6 duygu için en yüksek ortalama doğruluk %65,28 ile rastgele orman tarafından elde edilmiştir.

Avots vd [57] yerel ve global özelliklerin bir arada kullanıldığı bir özellik çıkarma sistemi sunmuştur. SAVEE, RML ve eNTERFACE’05 veri setlerinden alınan ses dosyaları karışık olarak DVM ‘nin eğitiminde kullanılmıştır. RML veri seti için ortalama doğruluk %69,3 olarak bulunmuştur.

Seng vd [58] görsel-işitsel duygu tanıma sistemi önermektedir. Prosodik özellikleri (perde, log-enerji, sıfır geçiş oranları ve Teager enerji operatörü) ve spektral özellikleri (MFKK) bir arada kullanılmıştır. Ortalama doğruluk %65,8 olarak bulunmuştur.

Zhang vd [59] evrimsel sinir ağları ve derin inanç ağlarından oluşan görsel-işitsel bir SDT modeli önermişlerdir. Ses sinyallerinden çıkarılan spektogramlar evrimsel sinir ağı girişine verilerek özellik çıkarımı yapılmış daha sonra DVM ile sınıflandırılmıştır. Ortalama doğruluk %66,17 olarak bulunmuştur.



Şekil 6.2. RML veri seti kullanılarak yapılan çalışmalar

Şekil 6.2 ‘de görüleceği üzere RML veri seti için önerilen yöntemin doğruluğu önceki çalışmalarda elde edilen doğruluk değerinin üzerindedir.

6.3. Ravdess Veri Seti İçin Sınıflandırma Sonuçları

RAVDESS veri seti için elde edilen hata matrisi Tablo 6.2’de gösterilmiştir. Tablo 6.2’de görüleceği üzere en yüksek ortalama doğruluk oranı “kızgınlık” duygusu için %79,4 olarak elde edilirken en düşük doğruluk %62,3 ile “korku” duygusu için elde edilmiştir. İkinci en başarılı tanınan duygu ise %70,2’lik ortalama doğruluk değeri ile “iğrenme” duygusudur. “Mutluluk” duygusu %70 doğrulukla tahmin edilmiş ve en çok “şaşıрма” duygusu ile karıştırılmıştır. En çok karıştırılan iki duygu “mutluluk” ve “şaşıрма” duygusu iken en iyi ayırt edilen iki duygu “kızgınlık” ve “üzüntü” duygusudur. 6 duygu için ortalama doğruluk %68,8 olarak hesaplanmıştır.

Tablo 6.2. RAVDESS veri seti için hata matrisi (%)

	Kızgınlık	İğrenme	Korku	Mutluluk	Üzüntü	Şaşıрма
Kızgınlık	79,4	7,12	6,8	4,8	0,78	1,1
İğrenme	10	70,2	5,47	1,1	6,85	6,38
Korku	5,73	4,15	62,3	8,1	9,79	9,94
Mutluluk	2,63	4,3	6,92	70	6,57	11,1
Üzüntü	0,78	7,1	8,68	7,1	67,6	8,15
Şaşıрма	2,46	7,1	9,73	7,9	9,41	63,4

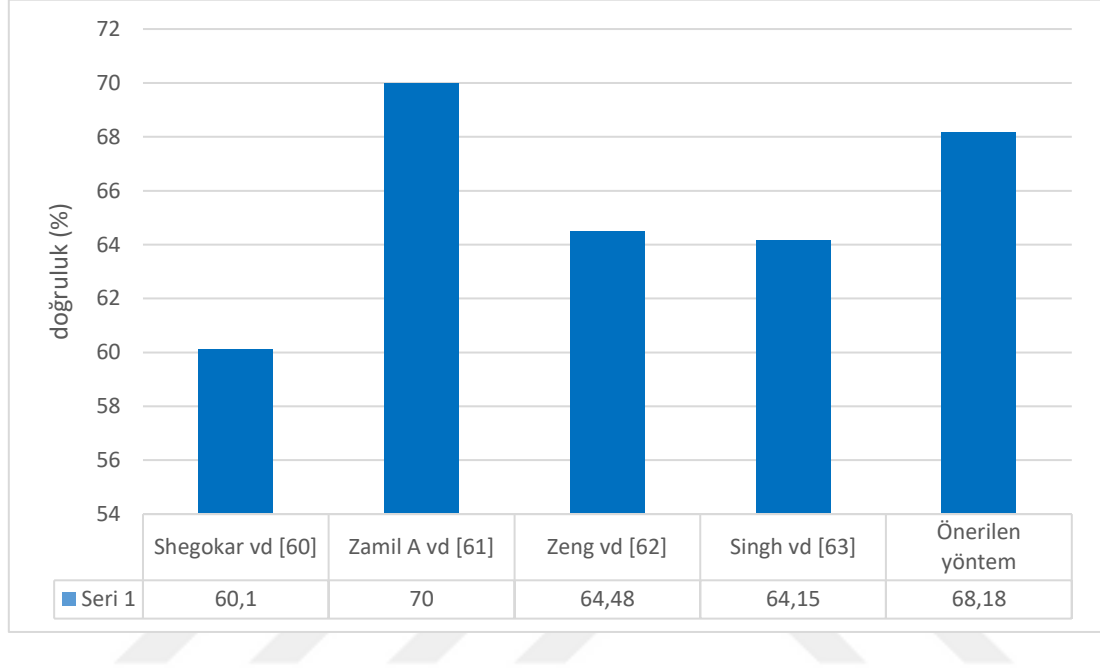
Şekil 6.3’te RAVDESS veri seti kullanılarak yapılan literatürdeki çalışmalar gösterilmiştir. Bu çalışmalarda kullanılan metotlar aşağıda detaylandırılmıştır.

Shegokar vd [60] RMS enerjisi, ZCR ve Entropi gibi prosodik özellikler çıkarmış daha sonra bu özellikleri DVM modellerini eğitmek için kullanmıştır. 5 kat çapraz doğrulama ile 8 duygu içi en yüksek doğruluk %60,1 olarak bulunmuştur.

Zamil A vd [61] lojistik regresyon ve karar ağaçlarının sentezlendiği bir sınıflandırma algoritması önerilmiştir. Ses özelliklerini yalnızca MFKK’dan alınan 13 katsayı oluşturmuştur. “Şaşıрма” duygusunun dahil edilmediği 7 duygu için ortalama %70 doğruluk hesaplanmıştır.

Zeng vd [62] konuşmacının tanınması, konuşmacının aksanı ve duygu tanıma gibi ses sınıflandırma görevleri arasındaki ilişkileri kullanan çok görevli bir derin sinir ağı önermiştir. Spektogramlar kullanılarak oluşturulan modelde doğruluk %64,48 olarak hesaplanmıştır.

Singh vd [63] spektral özellikler, prosodik özellikler ve bu özelliklerin istatistiksel değerlerini kullanarak oluşturdukları özellik setini DVM kullanarak 6 duygu için sınıflandırmışlardır. “Sakin” ve “şaşıрма” duygularının dahil edilmediği çalışmada doğruluk %64,15 olarak bulunmuştur.



Şekil 6.3. RAVDESS veri seti kullanılarak yapılan çalışmalar

RAVDESS veri seti için yapılan çalışmalarda farklı sayılarda duygu sınıflandırıldığı ve 8 duygunun altındaki çalışmalarda, sınıflandırmaya dahil edilmeyen duyguların aynı olmadığı görülmüştür. Önerilen yöntemin doğruluğu; 7 duygunun sınıflandırıldığı [61] nolu çalışmanın altında kalırken, 6 duygunun sınıflandırıldığı [63] nolu ve belirtilen diğer çalışmaların üzerindedir.

6.4. Savee Veri Seti İçin Sınıflandırma Sonuçları

SAVEE veri seti için elde edilen başarımların değerleri Tablo 6.3'te sunulmuştur. Diğer veri setlerinden farklı olarak “üzüntü” duygusu, SAVEE veri setinde en başarılı şekilde tanınan duygu olmuştur. “Üzüntü” için elde edilen ortalama doğruluk değeri %90’dır. Diğer veri setlerinde olduğu gibi “korku” duygusu SAVEE veri setinde de en düşük başarımlar oranı ile sınıflandırılan duygu olmuştur. En çok karıştırılan iki duygu “kızgınlık” ve “korku” olmuştur. Ayrıca, “iğrenme” duygusu ikinci en başarılı sınıflandırılan duygu olarak göze çarpmaktadır. 6 duygu için ortalama doğruluk %72.13 olarak hesaplanmıştır.

Tablo 6.3. SAVEE veri seti için hata matrisi (%)

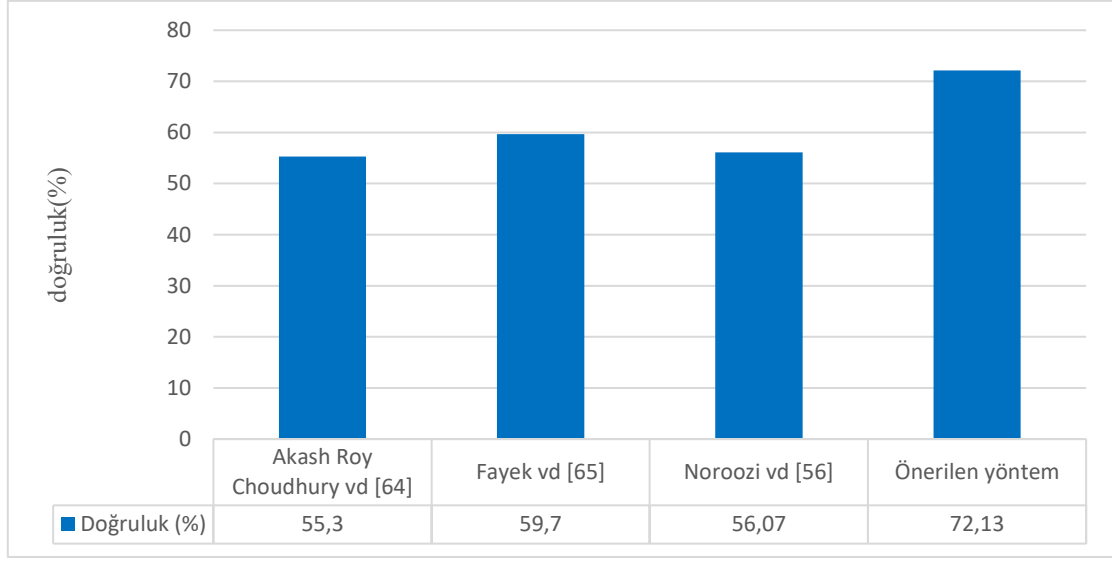
	Kızgınlık	İğrenme	Korku	Mutluluk	Üzüntü	Şaşırma
Kızgınlık	61,75	0	16,57	15,75	0	5,84
İğrenme	4,66	77,5	5,4	2,5	6,27	3,67
Korku	4,64	9,12	60	6	4,5	14,5
Mutluluk	14,26	1,8	6,64	69	0	8,3
Üzüntü	0	5,74	4,26	0	90	0
Şaşırma	1,7	4,16	9,44	10	0	74,7

Şekil 6.4’de SAVEE veri seti kullanılarak yapılan literatürdeki çalışmalar gösterilmiştir. Bu çalışmalarda kullanılan metotlar aşağıda detaylandırılmıştır.

Choudhury vd [64]; enerji, sıfır geçiş oranı, spektral entropi, MFKK özelliklerini kullanarak duygu sınıflandırması yapmıştır. DVM ve rastgele orman sınıflandırıcıları kullanılmış 7 duygu için ortalama doğruluk sırasıyla %55,5 ve %55,3 olarak hesaplanmıştır.

Fayek vd [65] duygu sınıflandırmasını konuşma spektrogramları çerçevesinde gerçekleştiren bir derin sinir ağı sunmuştur. 7 duygu için %59,7 doğruluk elde edilmiştir.

Noroozi vd [56] ses sinyallerinden MFKK özelliklerinin yanı sıra filtre bankası enerjileri ve prosodik özellikleri içeren 88 özellik çıkarmışlardır. Özellikler destek vektör makinesi ve rastgele orman sınıflandırıcılarında sınıflandırılmıştır. 6 duygu için en yüksek ortalama doğruluk %59,07 ile rastgele orman sınıflandırıcısı tarafından elde edilmiştir.



Şekil 6.4. SAVEE veri seti kullanılarak yapılan çalışmalar

Şekil 6.4 'te görüleceği üzere SAVEE veri seti için önerilen yöntemin doğruluğu önceki çalışmalarda elde edilen doğruluk değerinin üzerindedir.

7. SONUÇLAR VE TARTIŞMA

Ses, farklı duyguları açığa çıkaracak önemli öznitelikleri barındırmaktadır. Bu özniteliklerin sinyal işleme yöntemleriyle belirlenmesi ve daha sonra makine öğrenmesi kullanılarak bu özniteliklerin sınıflandırılması önem arz etmektedir. Bu tez çalışmasında ses tabanlı duygu sınıflandırması gerçekleştirilmiştir. 113 adet ses tabanlı öznitelik ve bir derin öğrenme algoritması olan uzun kısa süreli bellek ağları kullanılarak 3 farklı veri setinde duygu sınıflandırılması gerçekleştirilmiştir. Elde edilen sonuçlar incelendiğinde, en başarılı sonuçların SAVEE veri seti için elde edildiği görülmektedir. Bu durumun, SAVEE veri setinin içeriğini ana dili İngilizce olan yalnızca erkek konuşmacılardan elde edilmiş olmasından kaynaklandığı söylenebilir. SAVEE veri setinden elde edilen sonuçlar incelendiğinde en yüksek doğrulukta tanınan duygunun “üzüntü” duygusu, en düşük doğrulukta tanınan duygunun “korku” duygusu olduğu görülmektedir. Diğer taraftan RAVDESS veri seti incelendiğinde en yüksek doğrulukta tanınan duygunun “kızgınlık” duygusu, en düşük doğrulukta tanınan duygunun SAVEE veri setinde olduğu gibi “korku” duygusu olduğu görülmektedir. Ortalama doğruluk incelendiğinde RAVDESS veri setinden elde edilen doğruluk değerinin, SAVEE veri setinden daha düşük olduğu anlaşılmıştır. SAVEE veri setini 4 erkek konuşmacı yalnızca İngilizce seslendirirken, RAVDESS veri seti ise daha fazla sayıda kadın erkek konuşmacı tarafından seslendirilmiştir. Bu durumun, duygu tanıma sisteminin performansı etkilediği düşünülmektedir. RML veri seti incelendiğinde en yüksek doğrulukta tanınan duygunun “kızgınlık” duygusu, en düşük doğrulukta tanınan duygunun “korku” duygusu olduğu görülmektedir.

Gelecek çalışmalarda duygu tanıma başarımını arttıracak yeni ses öznitelikleri ve derin öğrenme mimarileri araştırılarak sistemin başarımı artırılabilir. Ayrıca, derin öğrenme mimarileri için veri sayısının azlığı bir dezavantajdır. Yüksek sayıda veri içeren yeni duygu veri tabanlarının oluşturulması, bu dezavantajı ortadan kaldırarak daha yüksek sınıflandırma performansı elde edilmesine olanak sağlar.

KAYNAKLAR

- [1] Karray, F., Alemzadeh, M., Abou Saleh, J., & Arab, M. N. (2008). Human-computer interaction: Overview on state of the art.
- [2] El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3), 572-587.
- [3] Kim, J. C., Azzi, P., Jeon, M., Howard, A. M., & Park, C. H. (2017, June). Audio-based emotion estimation for interactive robotic therapy for children with autism spectrum disorder. In *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)* (pp. 39-44). IEEE.
- [4] Han, H., Byun, K., & Kang, H. G. (2018, October). A deep learning-based stress detection algorithm with speech signal. In *proceedings of the 2018 workshop on audio-visual scene understanding for immersive multimedia* (pp. 11-15).
- [5] Ramakrishnan, S., & El Emary, I. M. (2013). Speech emotion recognition approaches in human computer interaction. *Telecommunication Systems*, 52(3), 1467-1478.
- [6] Apple. (2018). iOS - Siri. [online] Available at: <https://www.apple.com/ios/siri/>
- [7] W. Lim, D. Jang and T. Lee, "Speech emotion recognition using convolutional and Recurrent Neural Networks," 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, 2016, pp. 1-4, doi: 10.1109/APSIPA.2016.7820699.
- [8] Q. Jin, C. Li, S. Chen and H. Wu, "Speech emotion recognition with acoustic and lexical features," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, 2015, pp. 4749-4753, doi: 10.1109/ICASSP.2015.7178872.
- [9] A. Iqbal and K. Barua, "A Real-time Emotion Recognition from Speech using Gradient Boosting," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-5.
- [10] B. E. Martínez and J. C. Jacobo, "An improved characterization methodology to efficiently deal with the speech emotion recognition problem," 2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, 2017, pp. 1-6.
- [11] S. T. Saste and S. M. Jagdale, "Emotion recognition from speech using MFCC and DWT for security system" 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 701-704.
- [12] L. Zheng, Q. Li, H. Ban and S. Liu, "Speech emotion recognition based on convolution neural network combined with random forest." 2018 Chinese Control And Decision Conference (CCDC), Shenyang, 2018, pp. 4143-4147.
- [13] P. Shen, Z. Changjun and X. Chen, "Automatic Speech Emotion Recognition using Support Vector Machine," *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, Harbin, 2011, pp. 621-625, doi: 10.1109/EMEIT.2011.6023178.
- [14] G. Trigeorgis et al., "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 5200-5204, doi: 10.1109/ICASSP.2016.7472669.

- [15] K. Wang, N. An, B. N. Li, Y. Zhang and L. Li, "Speech Emotion Recognition Using Fourier Parameters," in *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 69-75, 1 Jan.-March 2015, doi: 10.1109/TAFCC.2015.2392101
- [16] ISSA, Dias; DEMIRCI, M. Fatih; YAZICI, Adnan. Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control*, 2020, 59: 101894.
- [17] J. Cai et al., "Feature-Level and Model-Level Audiovisual Fusion for Emotion Recognition in the Wild," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 2019, pp. 443-448, doi: 10.1109/MIPR.2019.00089.
- [18] Langari, S., Marvi, H., & Zahedi, M. (2020). Efficient speech emotion recognition using modified feature extraction. *Informatics in Medicine* 20, 100424.
- [19] Zhao, J., Mao, X., & Chen, L. (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomedical Signal Processing and Control*, 47, 312-323.
- [20] P. P. Dahake, K. Shaw and P. Malathi, "Speaker dependent speech emotion recognition using MFCC and Support Vector Machine," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp. 1080-1084, doi: 10.1109/ICACDOT.2016.7877753.
- [21] Venkataramanan, K., & Rajamohan, H. R. (2019). Emotion Recognition from Speech. *arXiv preprint arXiv:1912.10458*
- [22] Kerkeni, L., Serrestou, Y., Raoof, K., Mbarki, M., Mahjoub, M. A., & Cleder, C. (2019). Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO. *Speech Communication*, 114, 22-35.
- [23] Aouani, H., & Ayed, Y. B. (2020). Speech Emotion Recognition with deep learning. *Procedia Computer Science*, 176, 251-260.
- [24] Tzirakis, P., Zhang, J., & Schuller, B. W. (2018, April). End-to-end speech emotion recognition using deep neural networks. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5089-5093). IEEE.
- [25] Liu, Z. T., Wu, M., Cao, W. H., Mao, J. W., Xu, J. P., & Tan, G. Z. (2018). Speech emotion recognition based on feature selection and extreme learning machine decision tree. *Neurocomputing*, 273, 271-280.
- [26] Ekman P. Friesen, W.V., Constants across cultures in the face and emotion, *Journal of Personality and Social Psychology* 17: 124–129, 1971.
- [27] Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, 116, 56-76.
- [28] El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3), 572-587.
- [29] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In *Ninth European Conference on Speech Communication and Technology*.

- [30] Engberg, I. S., Hansen, A. V., Andersen, O., & Dalsgaard, P. (1997). Design, recording and verification of a Danish emotional speech database. In Fifth European conference on speech communication and technology.
- [31] Martin, O., Kotsia, I., Macq, B., & Pitas, I. (2006, April). The eNTERFACE'05 audio-visual emotion database. In 22nd International Conference on Data Engineering Workshops (ICDEW'06) (pp. 8-8). IEEE.
- [32] RML emotion database. <http://www.rml.ryerson.ca/rml-emotiondatabase>
- [33] S. Haq, P. J. B. Jackson, and J. Edge, "Audio-visual feature selection and reduction for emotion classification." in Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP '08), pp. 185–190, Tangalooma, Australia, 2008
- [34] S. R. Livingstone, K. Peck and F. A. Russo, "Ravdess: The Ryerson audio-visual database of emotional speech and song", Proc. Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, 2012.
- [35] Hansen, J. H., & Bou-Ghazale, S. E. (1997). Getting started with SUSAS: A speech under simulated and actual stress database. In Fifth European Conference on Speech Communication and Technology.
- [36] Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., ... & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. Language resources and evaluation, 42(4), 335.
- [37] Ringeval, F., Sonderegger, A., Sauer, J., & Lalanne, D. (2013, April). Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG) (pp. 1-8). IEEE.
- [38] Kossaifi, J., Tzimiropoulos, G., Todorovic, S., & Pantic, M. (2017). AFEW-VA database for valence and arousal estimation in-the-wild. Image and Vision Computing, 65, 23-36.
- [39] Li, D., Zhou, Y., Wang, Z., & Gao, D. Exploiting the potentialities of features for speech emotion recognition. Information Sciences, 548, 328-343.
- [40] Zhang, Xian-Da. A Matrix Algebra Approach to Artificial Intelligence. Springer, 2020.
- [41] X. Du, Y. Cai, S. Wang and L. Zhang, "Overview of deep learning," 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 2016, pp. 159-164, doi: 10.1109/YAC.2016.7804882.
- [42] Öztürk, Kadir; ŞAHİN, Mustafa Ergin. Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. Takvim-i Vekayî, 2018, 6.2: 25-36.
- [43] McCulloch, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 1943, 5.4: 115-133
- [44] Canan, S. (2006). Yapay sinir ağları ile GPS destekli navigasyon sistemi (Doctoral dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü).
- [45] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4), 115-133.
- [46] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6), 386.

- [47] Ivakhnenko, A. G. and Lapa, V. G. (1965). Cybernetic Predicting Devices. CCM Information Corporation.
- [48] Minsky, M., & Papert, S. A. Perceptrons: An introduction to computational geometry. MIT press.
- [49] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science.
- [50] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.
- [51] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
- [52] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554.
- [53] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine, 29(6), 82-97.
- [54] "Understanding LSTM Networks" [http://colah.github.io/posts/2015-08-](http://colah.github.io/posts/2015-08-Understanding-LSTM-Networks/)
- [55] "LSTM" <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>
- [56] Noroozi, F., Marjanovic, M., Njegus, A., Escalera, S., & Anbarjafari, G. "Audio-visual emotion recognition in video clips." IEEE Transactions on Affective Computing, 2017
- [57] Avots, E., Sapiński, T., Bachmann, M. et al. "Audiovisual emotion recognition in wild". Machine Vision and Applications (2018)
- [58] Seng, K. P., Ang, L. M., & Ooi, C. S. (2016). A combined rule-based & machine learning audio-visual emotion recognition approach. IEEE Transactions on Affective Computing, 9(1), 3-13.
- [59] Zhang, S., Zhang, S., Huang, T., Gao, W., & Tian, Q. (2017). Learning affective features with a hybrid deep model for audio-visual emotion recognition. IEEE Transactions on Circuits and Systems for Video Technology, 28(10), 3030-3043.
- [60] Shegokar, P., & Sircar, P. (2016, December). Continuous wavelet transform based speech emotion recognition. In 2016 10th International Conference on Signal Processing and Communication Systems (ICSPCS) (pp. 1-8). IEEE.
- [61] Zamil, A. A. A., Hasan, S., Baki, S. M. J., Adam, J. M., & Zaman, I. (2019, January). Emotion detection from speech signals using voting mechanism on classified frames. In 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (pp. 281-285). IEEE.
- [62] Zeng, Y., Mao, H., Peng, D., & Yi, Z. (2019). Spectrogram based multi-task audio classification. Multimedia Tools and Applications, 78(3), 3705-3722.
- [63] Singh, R., Puri, H., Aggarwal, N., & Gupta, V. (2020). An Efficient Language-Independent Acoustic Emotion Classification System. Arabian Journal for Science and Engineering, 45(4), 3111-3121.

- [64] Choudhury, Akash Roy, et al. "Emotion Recognition from Speech Signals using Excitation Source and Spectral Features." 2018 IEEE Applied Signal Processing Conference (ASPCON). IEEE, 2018.
- [65] Fayek, H. M., Lech, M., & Cavedon, L. (2015, December). Towards real-time speech emotion recognition using deep neural networks. In 2015 9th international conference on signal processing and communication systems (ICSPCS) (pp. 1-5). IEEE.



ÖZGEÇMİŞ

[REDACTED]

KİŞİSEL BİLGİLER

ARAŞTIRMACI BİLGİLERİ

[REDACTED] [REDACTED]
[REDACTED] [REDACTED]

EĞİTİM BİLGİLERİ

Lisans : Fırat Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği
Bölümü, 2016
Lise : Kaya Karakaya Anadolu Lisesi, Elazığ, 2012

ARAŞTIRMA DENEYİMİ

✓ MATLAB, PROTEUS , AUTOCAD

AKADEMİK FAALİYETLER

Bildiriler:

1. Demir, A., Atıla, O., & Şengür, A. (2019, September). Deep Learning and Audio Based Emotion Recognition. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-6). IEEE.
2. A. Demir, B. N. Akılotu, Z. Kadiroğlu and A. Şengür, "Bitcoin Price Prediction Using Machine Learning Methods," 2019 1st International Informatics and Software Engineering Conference (UBMYK), Ankara, Turkey, 2019, pp. 1-4, doi: 10.1109/UBMYK48245.201