T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

SES SİNYALLERİNDEN DUYGU TANIMA İÇİN FARKLI YAKLAŞIMLAR

Semiye DEMİRCAN

DOKTORA TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Agustos-2020
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Semiye DEMİRCAN tarafından hazırlanan "Ses Sinyallerinden Duygu Tanıma için Farklı Yaklaşımlar" adlı tez çalışması 04/08/2020 tarihinde aşağıdaki jüri üyeleri tarafından oy birliği / oy çokluğu ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS/DOKTORA TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Prof. Dr. Erkan ÜLKER	……..
Danışman Doç. Dr. Humar KAHRAMANLI ÖRNEK	……..
Üye Doç Dr. Gülay TEZEL	……..
Üye Doç. Dr. Halife KODAZ	……..
Üye Dr. Öğr. Üyesi Ayşe Merve ACILAR	……..

Yukarıdaki sonucu onaylarım.

Prof. Dr. Saadettin Erhan KESEN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Semiye DEMİRCAN
Tarih:

ÖZET

DOKTORA TEZİ

SES SİNYALLERİNDEN DUYGU TANIMA İÇİN FARKLI YAKLAŞIMLAR

Semiye DEMİRCAN

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç Dr. Humar KAHRAMANLI ÖRNEK

2020, 116 Sayfa

Jüri
Danışman: Doç Dr. Humar KAHRAMANLI ÖRNEK
Prof. Dr. Erkan ÜLKER
Doç. Dr. Gülay TEZEL
Doç. Dr. Halife KODAZ
Dr. Öğr. Üyesi Ayşe Merve ACILAR

Duygu tanıma son yıllarda önemi gittikçe artan konulardan bir tanesidir. Duygu tanıma yüz ifadelerinden ve biyomedikal sinyallerden yapılabildiği gibi ses verilerinden de yapılabilmektedir. Özellikle yüz yüze iletişim kurulmadığı durumlarda kişinin duygu durumunun tespitinde ses verisi kullanılmaktadır. Elle özellik çıkarımı ve özellik seçimi klasik yöntemlerle sesten duygu tanımada en önemli aşamalardır. Sesten duygu tanıma alanında en çok çıkarılan özellikler spektral, prozodik ve format özellikleridir. Özellik seçimi için ise oldukça çok yöntem önerilmiştir. Buna rağmen sesten duygu tanıma problemi tam olarak çözülememiş, karmaşık bir problemdir ve tanıma oranını arttırmak için daha iyi tekniklere ihtiyaç duyulmaktadır. Bu nedenle bu tez çalışmasında sesten duygu tanıma işlemi gerçekleştirilmiştir.

Bu çalışmada en çok bilinen ve açık bir veri tabanı olan Berlin duygu veri tabanı (EmoDB) kullanılmıştır. EmoDB yedi farklı duyguya ait ses veri tabanıdır. Bu tez çalışmasında uygulamalar cinsiyet ve kişi bağımsız olarak gerçekleştirilmiştir. Tez çalışmasında üç farklı uygulama gerçekleştirilmiştir. Tüm uygulamalar yedi duygu ve üçerli duygu grupları üzerinde gerçekleştirilmiştir.

Uygulamalarda kullanılacak veri kümeleri ve önişlemlerin açıklanması amacıyla iki farklı ön çalışma gerçekleştirilmiştir. Bu ön çalışmaların ilkinde Uygulama 1 ve Uygulama 2'de kullanılan veri kümeleri oluşturulmuştur. Veri kümeleri farklı sayıda ve farklı özelliklerden oluşturulmuştur. Çıkarılan özellikler Spektral özelliklerden Mel Frekansı Kepstrum Katsayıları (MFCC), Doğrusal Öngörüm Katsayıları (LPC), Ayrık Dalgacık Dönüşümü (DWT) ve Öz Bağımlı Model Parametreleri (AR) özellikleridir. Ayrıca prozodik özelliklerden Temel Frekansa ait özelliklerde çıkarılmıştır. Özellik seçiminde kullanılan spektral, prozodik ve format özellikleri incelenmiş sınıflandırma için etkin özellikler üzerinde çalışma yoğunlaştırılmıştır. Oluşturulan veri kümeleri özellik seçimi yapılmadan sınıflandırılmıştır. Sınıflandırma amacı ile tüm uygulamalarda Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM), k En yakın Komşuluk Algoritması (kNN) ve Sade Bayes (SB) algoritmaları kullanılmıştır. Ön çalışmada ikinci olarak Uygulama 3'de kullanılan veri kümeleri ve ön işlemler açıklanmıştır.

Uygulama 1'de özellik seçimi için önerilen Etmen Tabanlı Otomatik Özellik Seçimi (*Agent Based Automatic Feature Selection - ABAfs)* yaklaşımı sunulmuştur. Ön çalışmalarda belirlenen etkin veri kümeleri üzerinde uygulanan çalışmada seçilen özellikler sınıflandırılmıştır.

Tez çalışmasında gerçekleştirilen ikinci uygulama Bulanık C-Ortalama Algoritması (BCO) ile duygu tanıma işlemidir. Bu uygulamada BCO yöntemi ile boyut azaltımı işlemi gerçekleştirilmiştir.

Oluşturulan özellikler ön çalışmada belirlenen veri kümeleri üzerinde gerçekleştirilmiş ve sınıflandırıcıya verilmiştir.

Bu çalışmadaki üçüncü ve son uygulama Derin Öğrenme Algoritmaları ile duygu tanımadır. İlk iki uygulamada duygu tanıma işlemi yapılmadan önce özellik çıkarımı ve özellik seçimi işlemleri gerçekleştirilmiştir. Bu uygulamada özellik seçimi yapılmadan ham verilerden spektrogram görüntüleri elde edilerek Derin Öğrenme Algoritmalarından AlexNET ile sınıflandırılmıştır. Ayrıca elle çıkarılan özelliklerin derin öğrenme ile sınıflandırma başarısının karşılaştırılabilmesi için etkin özellikler DSA ile sınıflandırılmıştır.

Tez kapsamında gerçekleştirilen tüm çalışmalar değerlendirildiğinde, yedi duygu grubunda en yüksek sınıflandırma doğruluğu (SD) 16 adet MFCC katsayılarından oluşturulan veri kümesi ile %92.98 ile BCO ile özellik seçimi yöntemiyle gerçekleştirilmiş ve literatüre kazandırılmıştır. Ayrıca literatür incelendiğinde bu zamana kadar duygu tanıma alanına hiç uygulanmamış özellik seçim yöntemlerinin bu çalışmada gerçekleştirilmiş olması da tez çalışmasının özgünlüğünü ortaya koymaktadır. Kısacası bu tez çalışması, elde edilen sonuçlar, Duygu Tanıma probleminde hangi özelliklerin daha etkin olduğunun tespiti ve sınıflandırılması konularında gerçekleştirilen çalışmalar açısından literatürde önemli bir yere sahip olacaktır.

Anahtar Kelimeler: AlexNET, BCO, MFCC, Opt-aiNET, Özellik Çıkarımı, Özellik Seçimi, Sesten Duygu Tanıma

ABSTRACT

PhD THESIS

VARIOUS APPROACHES TO EMOTION RECOGNITION FROM SPEECH SIGNALS

Semiye DEMİRCAN

Konya Technical University
Institute of Graduate Studies
Department of Computer

Advisor: Assoc.Prof. Dr. Humar KAHRAMANLI ÖRNEK

2020,116 Pages

Jury
Advisor: Assoc.Prof. Dr. Humar KAHRAMANLI ÖRNEK
Prof. Dr. Erkan ÜLKER
Assoc.Prof. Dr. Gülay TEZEL
Assoc.Prof. Dr. Halife KODAZ
Assist. Prof. Dr. Ayşe Merve ACILAR

Speech emotion recognition from data gains significance in recent years. Emotion recognition has been made from facial expressions and biomedical signals as well as from speech data. Emotion recognition from speech has been used when there is no face-to-face communication. Manually feature extraction and feature selection were the most important steps in traditional speech emotion recognition. Spectral, prosodic and format features are the most frequently extracted features in this area. There were many methods which have been proposed for feature selection. Despite this, the problem of emotion recognition from the speech has not been solved completely and varios techniques are needed to increase the recognition rate. Therefore, in this thesis, emotion recognition from sound was carried out.

Berlin Emotion Database (EmoDB), which is the most known and open access database, was used in the study. EmoDB is speech database consist of seven emotion. In this thesis the applications were performed gender and person independent. In this thesis, three different applications were carried out.

Two different preliminary studies were carried out in order to explain the data sets and pre-processes to be used in the applications. In the preliminary studies, firstly, the data sets used in Application 1 and Application 2 were created. Data sets are made up of from different numbers and different features. The features of Mel Frequency Kepstrum Coefficients (MFCC), Linear Prediction Coefficients (LPC), Discrete Wavelet Transform (DWT) and Autoregressive Parameters (AR) were extracted with different dimensions. In addition, the format features which were format frequency and pitch have been extracted. ANN, DVM, kNN and NB algorithms are used for classification purposes. In the preliminary studies, secondly the data sets used in Application 3 and pre-processing was explained.

First, the Agent Based Automatic Feature Selection (*ABAfs*) approach was proposed for feature selection. The study applied on the effective data sets, determined the features selected in the in the preliminary studies, were classified. The data sets created in the preliminary study were given to the classifier.

The third and last application of the study was emotion recognition with deep learning algorithms. Before performing emotion recognition in the first two applications, feature extraction and feature selection has been performed. In this application, spectrogram images are obtained from raw data without any feature selection. Afterwards it was classified with AlexNET algorithm. In addition, MFCC

attributes are classified with DNN in order to compare the classification success of manually extracted features.

When all the studies conducted within the scope of the thesis were evaluated, the highest classification accuracy (SD) in seven emotion groups was achieved with a data set consisting of 16 MFCC coefficients with a BCO feature selection method with a rate of 92.98%, and it was added to the literature. In addition, when the literature is examined, the fact that feature selection methods, which have not been applied to emotion recognition until now, have been carried out in this study reveals the originality of the thesis study. So this thesis study will have an important place in the literature in terms of the results obtained, the studies conducted on the determination and classification of which features are more effective in the Emotion Recognition problem.

Keywords: AlexNET, FCM, Feature Extraction, Feature Selection, MFCC, Opt-aiNET, Speech Emotion Recognition

ÖNSÖZ

Doktora çalışmamda bana yardımlarını ve desteğini eksik etmeyerek tez çalışmamı başarı ile tamamlamamı sağlayan danışman hocam Sayın Doç. Dr. Humar Kahramanlı ÖRNEK'e;

Tezin gelişmesine yönlendirici görüş ve önerileriyle yardımcı olan ve manevi desteklerini esirgemeyen Sayın hocalarım Prof. Dr. Erkan ÜLKER ve Doç Dr. Gülay TEZEL'e;

Çalışmalarım boyunca ve hayatımın her anında yanımda olan ve maddi, manevi desteklerini esirgemeyen annem Dudu YILMAZ, babam Sadi YILMAZ, eşim Ahmet DEMİRCAN ve kardeşlerim Serap ŞAHİN ile Oğuzhan YILMAZ'a;

Sevgili kızlarım Berra ve Rana'ya;

Tüm içtenliğimle teşekkür ederim.

Semiye DEMİRCAN

KONYA-2020

İÇİNDEKİLER

TEZ BİLDİRİMİ ..... **iii**

ÖZET ..... **iv**

ABSTRACT ..... **vi**

ÖNSÖZ ..... **viii**

İÇİNDEKİLER ..... **ix**

SİMGELER VE KISALTMALAR ..... **xi**

1. GİRİŞ **1**

1.1. Tezin Literatüre Katkısı 5
1.2. Tezin Organizasyonu 6

2. KAYNAK ARAŞTIRMASI **7**

3. MATERYAL **15**

3.1. Berlin Duygu Veri tabanı (EmoDB) 17

4. YÖNTEM **20**

4.1. Çalışmada Kullanılan Özellik Çıkarım Yöntemleri 21
4.1.1. Mel Frekansı Kepstrum Katsayıları (MFCC) 21
4.1.2. Doğrusal Öngörüm Katsayıları (LPC) 24
4.1.3. Temel Periyot (T0), Temel Frekans (F0), Perde Frekansı (Pitch) 26
4.1.4. Ayrık Dalgacık Dönüşümü (DWT) 28
4.1.5. Öz Bağımlı Model Parametreleri (AR) 30
4.2. Özellik Seçimi Amacıyla Kullanılan Algoritmalar 30
4.2.1. OPT-aiNET Optimizasyon Algoritması 30
4.2.2. Önerilen Etmen Tabanlı Modelleme 32
4.2.3. Bulanık C-Ortalama Kümeleme Yöntemi (BCO) 36
4.3. Sınıflandırma Amacıyla Kullanılan Algoritmalar 37
4.3.1. Yapay Sinir Ağları (YSA) 37
4.3.2. Derin Öğrenme 41
4.3.3. Destek Vektör Makineleri (DVM) 45
4.3.4. K - en Yakın Komşu Algoritması (kNN) 48
4.3.5. Sade Bayes Sınıflandırıcı (SB) 49
4.4. Performans Değerlendirme Kriterleri: 50

5. UYGULAMA **51**

5.1. Ön Çalışmalar 52
5.1.1. Uygulama 1 ve Uygulama 2 için Veri Kümelerinin Belirlenmesi ve Ön İşlemler 52
5.1.2. Uygulama 3 için Veri Kümelerinin Belirlenmesi ve Ön İşlemler 66
5.2. Ana Çalışmalar ve Sonuçları 69

5.2.1. Uygulama 1: Etmen Tabanlı Otomatik Özellik Seçimi (*Agent Based Automatic Feature Selection - ABAfs*) Yaklaşımı 69
5.2.2. Uygulama 2: Bulanık C-Ortalama Algoritması (BCO) ile Duygu tanıma Yaklaşımı 78
5.2.3. Uygulama 3: Derin Öğrenme Algoritmaları ile Ses Verilerinden Duygu Tanıma 84

6. TARTIŞMA 91

7. SONUÇLAR 94

KAYNAKLAR 96

ÖZGEÇMİŞ 102

SİMGELER VE KISALTMALAR

ABAfs	: Etmen Tabanlı Otomatik Özellik Seçimi *(**A**gent **B**ased **A**utomatic **F**eature **S**election)*
AR	: Özbağımlı Model Parametreleri *(**A**uto**R**egressive Parameters)*
AR-MA	: Özbağımlı Hareketli Ortalama *(**A**uto**r**egressive **M**oving **A**verage)*
BCO	: Bulanık C-Ortalama Kümeleme Yöntemi*(**F**uzzy **C**-**M**eans Clustering)*
CCR	: Doğru Sınıflandırma Oranı *(**C**orrect **C**lassification **R**ate)*
CWT	: Sürekli Dalgacık Dönüşümü *(**C**ontinous **W**avelet **T**ransform)*
ÇKP	: Çok Katmalı Perseptron *(**M**ulti**l**ayer **P**erceptron)*
DESA	: Derin Evrişimsel Sinir Ağı *(**D**eep **C**onvolutional **N**eural **N**etwork)*
DCT	: Ayrık Kosinüs Dönüşümü (***D**iscrete **C**osine **T**ransform)*
DFT	: Ayrık Fourier Dönüşümü *(**D**iscrete **F**ourier **T**ransform)*
DSA	: Derin Sinir Ağları *(**D**eep **N**eural **N**etworks)*
DRNN	: Derin Tekrarlayan Sinir Ağı *(**D**eep **R**ecurrent **N**eural **N**etworks)*
DVM	**:** Destek Vektör Makineleri *(**S**upport **V**ector **M**achines)*
DWT	: Ayrık Dalgacık Dönüşümü *(**D**iscrete **W**avelet **T**ransform)*
ECAs	: Gömülü Konuşma Etmenleri *(**E**mbodied **C**onversational **A**gents)*
ESA	**:** Evrişimsel Sinir Ağı *(**C**onvolutional **N**eural **N**etwork)*
F0	: Temel Frekans, Perde Frekansı *(Pitch Frequency)*
FAO	: Fisher Ayırıcı Oranı *(**F**isher **D**iscriminant **R**atio)*
FFT	: Hızlı Fourier Dönüşümü *(**F**ast **F**ourier **T**ransform)*
FIR	: Sonlu Dürtü Yanıtı *(**F**inite **I**mpulse **R**esponse)*
GA	**:** Genetik Algoritmalar *(**G**enetic **A**lgorithms)*
GF	**:** Gabor Filtresi *(**G**abor **F**ilter)*
GKM	: Gaus Karışım Modeli *(**G**aussian **M**ixture **M**odels)*
GLCM	: Gri Seviye Eş Oluşum Matrisi *(**G**ray **L**evel **C**o-Occurrence **M**atrix)*
HOG	: Yönlendirilmiş Degradelerin Histogramı *(**H**istogram of **O**riented **G**radients)*
kNN	**:** K- en Yakın Komşu Algoritması *(**k**-**N**earest-**N**eighbor)*
LDA	: Doğrusal Ayrımcı Analizi (**L**inear **D**iscriminant **A**nalysis)
LDC	: Doğrusal Ayrımcılık Analizi *(**L**inear **D**iscriminant **C**lassifiers)*
LPC	: Doğrusal Öngörüm Katsayıları *(**L**inear **P**redictive **C**oding),*
MAS	: Çoklu Etmen Sistemleri *(**M**ulti-**A**gent-**S**ystem)*
MFCC	: Mel Ölçekli Kepstrum Katsayıları *(**M**el**F**requency **C**epstral **C**oefficient),*
MFDWT	: Mel Frekanslı Ayrık Dalgacık Dönüşümü *(**M**el **F**requency **D**iscrete **W**avelet **T**ransform)*
MSFs	: Modüler Spektral Özellikler *(**M**odulation **S**pectral **F**eatures)*
NLDs	: Doğrusal Olmayan Dinamik Özellikler *(**N**onlinear **D**ynamics **F**eatures)*
PLP	: Algısal Doğrusal Öngörü *(**P**erceptual **L**inear **P**rediction)*
RBFNN	: Radyal Taban Fonksiyonu Sinir Ağları *(**R**adial **B**asis **F**unction **N**eural **N**etworks)*
ReLU	**:** **R**ectified **L**inear **U**nit
RNN	: Tekrarlayan Sinir Ağı *(**R**ecurrent **N**eural **N**etworks)*
SB	: Sade Bayes Sınıflandırıcı (*Navie Bayes*)
SD	**:** Sınıflandırma Doğruluğu
SFFS	: Ardışıl İleri Taşma Seçim Özellik Seçim Yöntemi *(**S**equential **F**orward **F**loating **S**election)*
SFS	: Ardışıl İleri Seçim Özellik Seçim Yöntemi *(**S**equential **F**orward **S**election)*
SMM	: Saklı Markov Modeli *(**H**idden **M**arkov **M**odels)*
SMO	: Sıralı Minimal Optimiazsyon *(Sequential Minimal Optimization)*

STFT	: Kısa Zamanlı Fourier Dönüşümü *(**S**hort-**T**ime **F**ourier **T**ransform)*
UA	: Ağırlıksız Doğruluk (**U**nweighted **A**ccuracy)
WA	: Ağırlıklı Doğruluk (**W**eighted **A**ccuracy)
WD	: Dalgacık Ayrışması *(**W**avelet **D**ecomposition)*
VQ	: Vektör Nicelemesi *(**V**ector **Q**uantization)*
YSA	: Yapay Sinir Ağları (*Artificial Neural Network*)

1. GİRİŞ

İnsanların sahip olduğu en önemli özelliklerden birisi duygulanmak ve duygu yaşamaktır. Duygu insanın yaşamında her anına eşlik eder. Kişinin içinde bulunduğu duygu durumunu ifade etme şekli ise değişir. Kimi zaman bir mimik veya farklı bir yüz ifadesi, kasların gerilmesi, kalp atışının değişmesi, el kol hareketleri ve tabii bütün bunların bir sonucu olarak sesteki değişiklik bu ifade şeklini belirler. Ses en önemli iletişim araçlarından bir tanesidir. Ses ile karşımızdaki kişiye sadece söylemek istediklerimiz değil aynı zamanda duygularımız da iletilir.

Ses, insan kulağı tarafından algılanabilinen havadaki basınç değişimleri olarak tanımlanabilir (Başaran, 1981). Ses, insan kulağını etkileyerek işitme duyusu oluşturan hava molekülleri titreşimleri, ya da bunların neden olduğu ufak hava basınç değişimleri olarak da ifade edilebilir. Kulağı etkileyen ve işitme duyusu yaratan belirli bir kaynaktan çıkan ses dalgaları, esas olarak hava ortamında iletilirler. Ses dalgaları 20 °C'de, deniz seviyesinde 344 m/s hızla yayılırlar (Başaran, 1981). Yükseklik ve sıcaklık ses hızının artmasına yol açan etkenlerdir. Sesler, fiziksel açıdan üç özellikleri ile ayırt edilirler (Başaran, 1981). Bunlar sesin şiddeti, sesin frekansı ve sesin tınısıdır.

Ses titreşimlerinin şiddeti sesin şiddetini, saniyedeki titreşim sayısı sesin sıklığını (frekansını) verir. Sesin şiddeti sesin enerjisine ve dalga şekline bağlıdır. Sesin şiddeti birimi desibel (dB)'dir. 0-140 dB arasındaki sesleri insan kulağı normal bir şekilde duymaktadır.

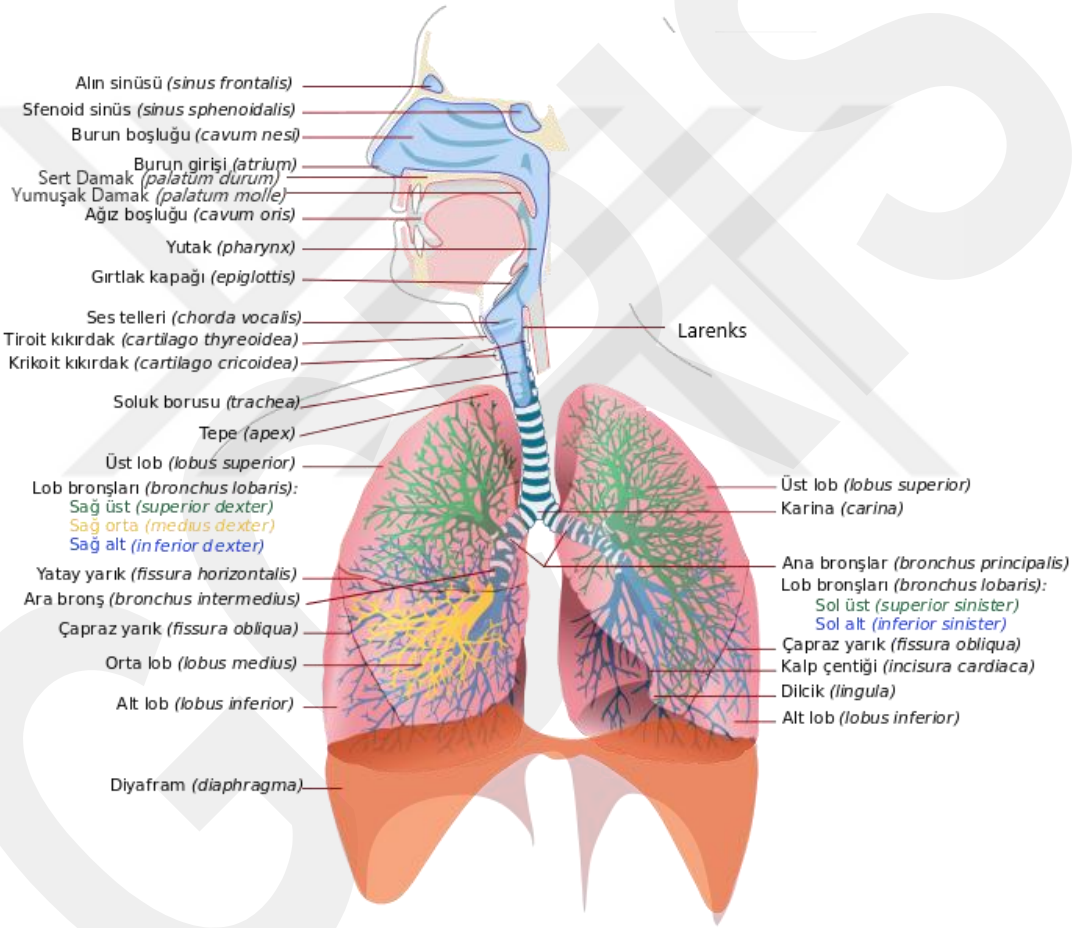
Sesi oluşturan kaynağın bir saniyedeki titreşim sayısına frekans denir. İnsan kulağı frekansı 20-20000 Hz arasındaki sesleri duyabilir. Ancak insanın en iyi duyduğu sesler 1000-3000 Hz frekanslı seslerdir.

Sesin tınısı, seslerin frekansları aynı da olsa farklı kaynaktan olduğunun algılanabilmesidir. Ses kendine özgü bu tınıyı, taşıyıcı bir dalganın üzerine "oranlı" olarak, yan titreşimlerin (harmoniklerin) binmesiyle oluşan karmaşık titreşimlerin sonucunda kazanır. Frekansı oluşturan saniyedeki periyot sayısı, ne kadar fazla olursa ses o kadar tiz; ne kadar az olursa, o kadar pes niteliğe bürünür.

Akciğerler, nefes borusu, gırtlak, yutak (Farink), burun ve ağız boşlukları insan ses organının bileşilenleridir. Şekil 1.1'de görüldüğü gibi bu organların hepsi bir tüple birbirine bağlıdır (Furui, 2000) .

Çene, dil, dudaklar ve diğer giriş bölümleri ile değişkenlik gösteren ve gırtlağın üstünde kalan bölüm ses yolu olarak ifade edilir. Burun boşluğu yutak ile, ağız boşluğu

ise yumuşak damak ve üst damak ile birbirinden ayrılır (Furui, 2000). Ciğerlerden gelen hava larenks bölgesine hava akımı sağlar. Larenks hava akımını yöneterek kısa hızlı nefes alma veya gürültülü hava akımı kaynağı gibi davranarak üst solunum yolu yani ağız, burun ve yutak boşluğuna iletir. Üst solunum yolu da hava akımını spektral olarak biçimlendirerek sese "rengini" verir. Dudakların ürettiği hava basıncı değişikliği dinleyen kişinin kulağına konuşma olarak gelen ses dalgasını oluşturur (Turk ve ark., 2004; Demirci, 2005). Özetle konuşma sesin karmaşık bir fonksiyonu olarak tarif edilebilir.



Şekil 1.1. Sesin oluşumunda rol alan organlar (Villerreal, 2017)

İnsan-bilgisayar etkileşimi açısından bakıldığında son yıllarda ses sinyallerinden konuşmacı, konuşma ve duygu tanıma işleminin de yoğun bir şekilde gerçekleştirildiğini görmekteyiz. Bu tez çalışmasında ses sinyallerinden duygu tanıma işlemi gerçekleştirilmiştir.

Duygu tanıma nabız sayısı, kan basıncı, yüz mimikleri, el ve kol hareketleri, beyin dalgaları, hormonsal değişikliklerden de yapılmaktadır. Fakat özellikle yüz yüze iletişim kurulmadığı durumlarda kişinin duygu durumunun tespitinin yapılması gerekirse ses verilerinden duygu tespiti önem kazanmaktadır.

Duygu tanımada giriş verisi konuşma içeren ses sinyalinden oluşmaktadır. Doğal olarak ses sinyali farklı uzunluklarda olmaktadır. Bu tez çalışmasında Berlin EmoDB duygu veri tabanı kullanılmıştır. EmoDB veri tabanında yedi farklı duyguya ait ses sinyalleri mevcuttur.

Ön işlem adımında verileri aynı seviyeye getirmek için belli zaman çerçevelerine (20, 30 milisaniye gibi) bölünür ve üst üste bindirmeli pencere (10, 15 milisaniye gibi) uygulanır. Ön işlem adımlarında uygulanan bir diğer işlem ise farklı filtrelerdir. Ön işlem adımı bu tez çalışmasında kullanılmamıştır. Ses sinyali içinde gürültü gibi görülen nefes alma veya diğer hırıltılı seslerin kişinin duygu durumunu belirtmede yardımcı olabileceği düşünüldüğünden ve veri bütünlüğünü bozmamak adına ses sinyali direk veri tabanında olduğu gibi kullanılmıştır.

Duygu tanımada en önemli adımlardan bir tanesi özellik çıkarımıdır. En çok kullanılan özellik çıkarım yöntemleri spektral ve prozodik özellikler olarak iki grupta incelenebilir. Spektral özelliklerden en çok bilenen ve kullanılanları, Mel Ölçekli Kepstrum Katsayıları (MFCC) yöntemi (Wu ve ark., 2011; Kotti ve Paternò, 2012; Sun ve ark., 2015), Doğrusal Öngörüm Katsayıları (LPC), (Hanilçi, 2007), Dalgacık Dönüşümü (Wavelet) (Reddy ve Vijayarajan, 2020) ve Öz Bağımlı Model Parametreleri (AR, MA, ARMA) (Milton ve Tamil Selvi, 2014) yöntemleridir. Bu tez çalışmasında MFCC yöntemi, LPC, Temel Frekans değeri, Dalgacık Dönüşümü, Öz Bağımlı Model Parametreleri yöntemlerini kullanarak özellik çıkarımı yapılmıştır.

Teoride doğrudan sinyal bilgisinden duygu çıkarımı mümkün görünse de veri boyutunun çok yüksek olması sebebiyle, veriden farklı yöntemlerle çıkarılan özellikler kullanılarak veri seçimi işlemi yapılmaktadır. Bu açıdan bakıldığında duygu tanımada özellik seçiminin bir diğer adıyla veri azaltmanın büyük önem arz ettiği açıktır. Alışılagelmiş birçok özellik seçimi yöntemi bulunmaktadır. Bunlardan bazıları şu şekildedir (Bhatti ve ark., 2004; Schuller ve ark., 2005; Polat, 2007; Zhang ve ark., 2018a).

- Biri - Diğerleri Yaklaşımı (one – vs - rest)
- Biri - Diğeri Yaklaşımı (one - vs - one)
- Ardışıl İleri Seçim Özellik Seçim Yöntemi (SFS)

- Ardışıl ileri yönlü Özellik Seçim Yöntemi (SFFS)
- En Düşük Kareler Sınırı Özellik Seçim Yöntemi
- Karşılıklı Bilgi Temelli Özellik Seçim Algoritması
- R2W2 Özellik Seçim Algoritması
- Karşılıklı Bilgi Temelli Özellik Seçim Algoritması
- Derin Öğrenme Algoritmaları

Sesten duygu tanıma işleminde son adım olarak sınıflandırma işlemi vardır. Sınıflandırmada makine öğrenmesi algoritmalarından Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM), k – en yakın Komşu Algoritması (kNN) ve Sade Bayes (SB) algoritmaları yoğunlukla kullanılmaktadır (Gilke ve ark.; Park ve ark., 2009a). Bununla birlikte son yıllarda Derin Öğrenme Algoritmaları kullanılarak yapılan duygu tanıma çalışmaları da artmaktadır (Sun ve ark., 2019) .

Literatür incelendiğinde duygu tanıma problemine farklı yaklaşımlar olduğu görülmektedir. Duygu veri tabanlarındaki veriler ve duyguların tamamı değil de bazı duyguların seçilerek kullanıldığını (Harar ve ark., 2017) veya duygulu cümleyi seslendiren kişinin cinsiyetine bağımlı (Shahzadi ve ark., 2015), ya da konuşmacı bağımlı (Karimi ve Sedaaghi, 2016) olarak çalışıldığı görülmektedir. Veri tabanındaki tüm duyguların hiçbir ayrım yapılmadan kullanıldığı çalışmalarda ise sınıflandırma doğruluğunun (SD) henüz istenilen seviyede olmadığı görülmektedir (Albornoz ve ark., 2011; Milton ve Tamil Selvi, 2014).

Yapılan tez çalışmasında bu eksikliğe yönelik olarak EmoDB veri tabanındaki tüm veriler cinsiyet ve kişi bağımsız olarak çalıştırılmıştır. Çalışma veri tabanındaki yedi duygunun tamamında ve frekansları yakın olduğu için birbiriyle en çok karıştırılan duygu grupları (Kızgın – Mutlu – Korku ve Sıkıntılı – Normal – Üzgün) duyguları üzerinde gerçekleştirilmiştir. Bu amaçla bu tez çalışmasında özellik seçim işlemi için aşağıdaki yöntemler önerilmiştir:

- Etmen Tabanlı Modelleme ile OPT-aiNET Optimizasyonu kullanılarak özellik Seçimi
- Bulanık C-Ortalama Kümeleme ile Özellik Seçimi

Önerilen yöntemlerle oluşturulan özellikler ve bu özelliklerden elde dilen veri kümeleri YSA, DVM, kNN ve SB ile sınıflandırılmıştır. Önerilen yöntemlerin yedi

duygu grubundaki duygularda iyi sonuçlar verirken en çok karıştırılan üçerli duygu gruplarında eksik kaldığı gözlemlenmiştir. Bu amaçla Üçüncü uygulama olan Derin öğrenme algoritmaları kullanılarak bu soruna çözüm aranmıştır.

1.1. Tezin Literatüre Katkısı

Bu çalışmada yedi farklı duygu içeren EmoDB verileri kullanılarak üç farklı uygulama gerçekleştirilmiştir.

Birinci uygulama Opt-aiNET Optimizasyon Algoritması kullanılarak özellik seçimi ve sınıflandırılmasıdır.

İkinci uygulama Bulanık C-Ortalama Algoritması (BCO) kullanılarak özellik seçimi yapılarak duygu sınıflandırma işlemidir.

Üçüncü ve son uygulama ise Derin Öğrenme Algoritmaları ile ses verilerinden duygu tanıma işleminin gerçekleştirilmesidir.

Çalışmanın literatüre sağladığı başlıca katkılar şunlardır:

- Literatürde ilk kez uygulanan Etmen Tabanlı Modelleme uygulaması ile özelliklerin akıllı birer etmen gibi ele alınarak, özellik seçimine uygulanabilecek yöntemlerin kapsamı genişletilmiştir.
- Opt-aiNET optimizasyon yöntemi ile özellik seçimi geçekleştirilerek otomatik duygu tanıma sistemi geliştirilmiştir.
- Önerilen etmen tabanlı yaklaşım ve otomatik duygu tanıma sistemi ile elle deneyerek tek tek hesaplanması gereken birçok işlem otomatik olarak gerçekleştirildiğinden zaman kaybı giderilmiş ve hızlı bir şekilde özellik seçimi işlemi gerçekleştirilmiştir.
- Etmen tabanlı modelleme yaklaşımı ile özellik seçimi işlemi bir optimizasyon problemi haline getirilmiş ve «hangi özellikler seçilirse daha yüksek tanıma gerçekleştirilir» sorusu bir maksimizasyon problemi gibi düşünülmüştür.
- BCO kümeleme algoritması uygulanarak özellik azaltma işlemi gerçekleştirilmiş ve gerçekleştiren yüksek tanıma başarısı ile literatüre katkı sağlanmıştır.
- Derin Öğrenme Mimarileri kullanılarak gerçekleştirilen çalışmada duygu tanıma probleminde ayırt edici performans elde edilmiştir.
- Derin Öğrenme mimarisi kullanılarak ses verilerinin resim olarak gösterimi diyebileceğimiz spektrogram görüntülerinin duygu tespitinde önemi vurgulanmıştır.

1.2. Tezin Organizasyonu

Gerçekleştirilen tez çalışması 6 bölümden oluşmaktadır. İlk bölümde tezin konusu tanıtılmış ve literatüre katkısı verilmiştir.

İkinci bölümde literatürde yapılan çalışmalar ile ilgili bilgi verilmiştir.

Üçüncü bölümde tez çalışmasında kullanılan veri tabanı tanıtılmıştır.

Dördüncü bölümde önerilen özellik çıkarma yöntemleri, kullanılan sınıflandırıcılar ve Derin Öğrenme açıklanmıştır.

Beşinci bölümde ilk olarak tez çalışmasında gerçekleştirilen ön işlemler anlatılmıştır. İkinci bölümünde ise tez çalışmasında gerçekleştirilen üç uygulama açıklanmıştır ve elde edilen sonuçlar analiz edilmiştir. Bu uygulamalar şunlardır:

- Uygulama 1: Etmen Tabanlı Otomatik Özellik Seçimi (*Agent Based Automatic Feature Selection - ABAfs*) Yaklaşımı
- İkinci uygulama Bulanık C-Ortalama (BCO) Algoritması ile duygu tanıma yaklaşımıdır.
- Üçüncü uygulama ise Derin Öğrenme Algoritmaları ile ses verilerinden duygu tanımadır.

Altıncı bölümde tartışma verilmiştir. Yedinci ve son bölümde tez çalışmasının sonuçlarının yorumları ve bilgi ve öneriler verilmiştir.

2. KAYNAK ARAŞTIRMASI

Konuşma insanların en temel iletişim yöntemlerinden bir tanesidir. Konuşarak insanlar sadece birbirlerine anlatmak istediklerini değil aynı zamanda duygularını da ses yoluyla aktarırlar. Konuşulan kişinin yüzü görülmese bile muhatabı o kişinin hangi duygu durumunda (üzgün, mutlu, öfkeli vb.) olduğunu konuşma sesinden anlayabilir. Örneğin bazı hastalıkların ve psikolojik durumun teşhisinde ses verilerinin analizi önem arz etmektedir. Depresyon ve intihar etmeye meyilli insanların konuşmalarının akustik özellikleri incelendiğinde, teşhisi ile alakalı ipuçları verdiğine dair kanıtlar bulunmuştur (France ve ark., 2000).

Duygu analizi ifadesini ilk olarak 2003 yılında Nasukawa ve Yi (2003) ele almıştır. Bu çalışmada tüm metin sınıflandırma işlemine tabi tutulması yerine, belirli nesneler için duyguların olumlu veya olumsuz kutup ile ilişkilendirerek duygu çıkarımı yaklaşımı sunulmuştur. Bu çalışmadaki duygu analizinde temel konu, duygunun kelimelerle nasıl ifade edildiğidir. Bir diğer önemli nokta ise duygusal ilişki ile nesnelerin ilişkisinin doğru bir şekilde tanımlanmasıdır. Sentiment Analiz Yöntemi kullanılarak yapılan çalışma web sayfalarında ve haber kaynaklarında duyguları bulmak için uygulanmış ve yüksek doğrulukla (%75 - %95) başarılı olmuştur (Nasukawa ve Yi, 2003).

Bir diğer uygulama alanı insan-makine etkileşiminin bir örneği olan sürücü kontrolü sağlayan çalışmalardır. Schuller ve ark. (2004) araç-içi kontrol sistemlerinde sürücünün güvenliğini sağlanması amacıyla uyarı durumuna geçirebilme özelliği bulunan bir uygulama geliştirmiştir. Bu uygulamada özellikler Doğrusal Ayrımcı Analizi (LDA) ile sıralanarak ilk 33 özellik alınmıştır. Bu özelliklerle sürücünün duygularını kontrol ederek sürücü hatalarını gidermek ve güvenlik sağlamak amaçlanmaktadır (Schuller ve ark., 2004).

Özellik çıkarımında en çok kullanılan diğer bir vektör özelliği ise Doğrusal Öngörüm Katsayılarıdır (LPC). LPC, zaman ortamında ses işaretinin spektral özelliklerini karakterize etmede önemli bir yere sahiptir (Nwe ve ark., 2003). LPC ses yolunu modellerken, MFCC özellikleri insan kulağını modellemektedir MFCC özellikleri süzgeç yöntemi kullanılarak çıkarıldığından gürültülü seslerde LPC özelliklerine göre daha iyi sonuç vermektedir (Hanilçi, 2007).

Çağrı merkezlerine yönelik yapılan çalışmalara bakıldığında da duygu tanıma kullanılarak yapılan uygulamalar görülmektedir. Ses portallarında akustik öfke tespit

algılama işlemi gerçekleştiren bir uygulamada (Burkhardt ve ark., 2009) kapsamlı bir ses portalı veri toplama işlemi gerçekleştirilmiş ve gerçek hayat problemlerine yeni bir bakış açısı getirilmiştir. Manuel etiketlenmenin uygulanarak sınıflandırma başarısının yükseltildiği bu uygulamada perde frekansı ve enerji ile alakalı özelliklerin aksine zaman ölçümlerinin önem arz etmediği görülmüştür. Sınıflandırıcı olarak Gaus Karışım Modeli (GKM) ve DVM karşılaştırıldığı çalışmada DVM ile yapılan sınıflandırmanın daha başarılı olduğu görülmüştür.

Özellikle kriz anında acil çağrı merkezlerine gelen telefonların sayısı fazla olduğunda hangisinin daha önemli olduğu duygu tespiti yapılarak tahmin edilebilir. Bu tarz bir uygulama Lulia ve ark. (2011) tarafından gerçekleştirilmiştir. Bu çalışmada öncelikle sistem bir veri tabanı kullanılarak eğitilmiştir. Gaus Karışım Modeli (GKM), YSA, DVM, Saklı Markov Modeli (SMM) makine öğrenmesi algoritması kullanılarak yapılan bu çalışma konuşmacı bağımsız ve 10 çapraz-doğrulama ile gerçekleştirilmiştir. Uygulama sonunda doğrusal lojistik regresyon kullanarak hata oranının %19.0 dan %4.2'ye iyileştirildiği gözlemlenmiştir (Lulia Lefter, 2011).

Park ve ark. (2009) servis robotları üzerine geliştirdikleri uygulamada sesten duygu tanıma işlemi gerçekleştirerek müşterinin duygusunun negatif veya negatif-olmayan olduğunu tespit etmeyi amaçlamaktadır. Örtüşen özellik vektörlerin ayrımını yapan efektif özellik sınıflandırma mekanizması önerilmiştir. Geleneksel yöntemlerle özellik vektörlerini örtüşen ve örtüşmeyen diye ayırırken önerilen yöntem örtüşen yöntemleri yeniden yapılandırma modeline ekler. Önerilen yöntemle elde edilen özellik vektörü daha yüksek sınıflandırma başarısına ulaşmıştır. GKM yöntemi kullanılan bu çalışmanın, duygu tanımada geleneksel yönteme göre çok daha iyi sonuç verdiği gösterilmiştir (Park ve ark., 2009b).

Özellik çıkarımı ve seçimi duygu tanımada da en önemli adımlardan bir tanesidir. En çok kullanılan özellik çıkarım yöntemleri spektral ve prozodik özellikler olarak iki grupta incelenebilir. Mel Ölçekli Kepstrum Katsayıları (MFCC), Doğrusal Öngörüm Katsayıları (LPC), Dalgacık Dönüşümü (DWT) spektral özellikler arasında sayılabilir. F0 temel frekans değeri ise prozodik özelliklerin başında gelmektedir. Temel Bileşen Analizi (PCA), Özbağımlı Model Parametreleri (AR) ise diğer özellik çıkarım yöntemleri arasında sayılabilir.

Mel Ölçekli Kepstrum Katsayıları (MFCC) yöntemi duygu tanıma uygulamalarında oldukça başarılı sonuçlar vermektedir. Bozkurt ve ark. (2011) yaptıkları çalışmada ağırlıklı MFCC (WMFCC) özelliklerini çıkararak sesten duygu

tanıma işlemi gerçekleştirmişlerdir. FAU Aibo duygu veri tabanını kullanarak yaptıkları çalışmada, iki-sınıflı duygu tanıma oranında en yüksek %70.29; beş-sınıflı duygu tanıma modelinde ise en yüksek %43.59 değerleri elde ederek standart spektral özelliklere göre daha yüksek performans elde etmişlerdir (Bozkurt ve ark., 2011).

MFCC yöntemi uygulanarak başarılı sonuçlar veren çalışmalardan bir tanesini de Albornoz ve ark. (2011) gerçekleştirmiştir. Bu çalışmada 12 tane MFCC özniteliği, Ortalama Logaritma spektrumu, F0 ve Enerji özelliklerinin ortalaması, minimumu, maksimumu gibi özellikler kullanılarak farklı özellik kümeleri elde edilmiş ve bu özellik kümeleri GKM, SMM ve Çok Katmalı Perseptron (ÇKP) yöntemleri ile sınıflandırma işlemine tabi tutulmuştur. EmoDB veri tabanı kullanılarak gerçekleştirilen çalışmada 7 duygu için, 10 çapraz doğrulama yöntemi sonuçlarına göre, GKM yöntemi ile %63.49; ÇKP yöntemi ile %66.83; SMM yöntemi ile %68.57 sınıflandırma başarısı elde edilmiştir. Bu sonuçlara ek olarak önerdikleri hiyerarşik-2 ve hiyerarşik-3 sınıflandırma yöntemleri de sırasıyla %66.99 ve %71.75 başarıya ulaşmıştır (Albornoz ve ark., 2011).

Wu ve ark. (2011), çalışmalarında EmoDB veri tabanının 7 duygusunun tamamını kullanarak Prozodik özellikler, MFCC ve Modüler Spektral Özellikler (MSFs) özellikleri elde ederek DVM sınıflandırıcısı ile sınıflandırma işlemi yapmışlardır. Prozodik özellikler ve MSFs özelliklerinin birleşiminden elde ettikleri yeni özellik kümesiyle %91.6 sınıflandırma başarısı elde etmişlerdir (Wu ve ark., 2011).

Yang ve Lugger (2010) ise çalışmalarında yeni bir harmoni özellik kümesi önermişlerdir. Prozodik özellikler, format özellikleri, sıfır geçiş oranı, ses kalitesi özellikleri elde ederek SB ile sınıflandırma işlemi yapmışlardır. Duyguları ikili, üçlü, altılı olacak şekilde farklı sayıda sınıflandırmaya tabi tutmuşlardır. Kızgın, sıkıntılı, mutlu, üzgün, normal olmak üzere beş duygunun sınıflandırmasından maksimum %73.5 sınıflandırma başarısı ele etmişlerdir (Yang ve Lugger, 2010).

Shahzadi ve ark. (2015) EmoDB veri tabanı kullanarak yedi duygu üzerinde yaptıkları çalışmada doğrusal olmayan dinamik özellikler (NLDs), prozodik ve spektral özellikleri çıkararak cinsiyete bağlı duygu tanıma işlemi gerçekleştirmiştir. İlk olarak altı farklı NLDs özniteliğine 20 farklı tamamlayıcı istatistik hesaplanarak bir özellik kümesi elde edilmiştir. Spektral özellik olarak 12 tane MFCC özniteliği çıkarılmış ve beş farklı prozodik özellik hesaplanarak geniş bir özellik veri kümesi elde edilmiştir. Bu özelliklerin etkinliği Fisher Ayırıcı Oranı (FAO) ve sınıflandırma doğruluğu ile değerlendirilmiştir. Sınıflandırıcı olarak Genetik Algoritmalar kullanılmıştır. Kadın ya

da erkek olarak iki farklı grupta gerçekleştirilen bu çalışmada kadın veri kümesi için %82.72 ve erkek veri kümesi için %85.90 tanıma oranı elde edilmiştir (Shahzadi ve ark., 2015).

Konuşma işaretinin incelenerek kişinin duygu durumunun tanımlanması üzerine yapılan bir diğer çalışmada (Durukal, 2015) özellik olarak MFCC katsayıları alınmış ve farklı duygu sınıfları için tanıma başarısı sonuçları irdelenmiştir. Çalışmada sınıflandırıcı olarak kNN ve DVM kullanılmıştır. Tek başına MFCC özellikleri kullanılan çalışmalarda LPC, prozodik ve kalite özelliklerini kullanan diğer çalışmalarla kıyaslandığında daha başarılı sonuçlar elde edildiği görülmüştür.

Spektral ve prozodik özellikler kullanılarak yapılan bir diğer çalışmayı da Karimi ve Sedaaghi (2016) gerçekleştirmiştir (Karimi ve Sedaaghi, 2016). Bu çalışma EmoDB veri tabanının dört duygusu; Kızgın, Sıkıntılı, Korku, Mutlu üzerinde gerçekleştirilmiştir. Temel frekans, enerji, MFCC, LPC ve algısal doğrusal öngörü (PLP) özelliklere ve bunlara ait maksimum, minimum, ortalama ve varyans değeri hesaplanarak elde edilen özellikler kullanılarak DVM ve Bayes yöntemleriyle sınıflandırılmıştır. Konuşmacı bağımlı olarak gerçekleştirilen ve 10 çapraz doğrulama kullanılan çalışmada DVM ile %93.10 ve Bayes ile %83.73 sınıflandırma başarısı elde edilmiştir.

Kotti ve Paternò (2012) çalışmalarında psikolojiden ilham alan ikili ve kademeli sınıflandırıcı önermişlerdir. Perde frekansı, enerji gibi özelliklerin yanında spektral özelliklerin de çıkarıldığı uygulamada sınıflandırma işlemi kNN ve DVM ile gerçekleştirilmiştir. Cinsiyet bağımlı, konuşmacı bağımsız yapılan çalışmada EmoDB veri tabanı kullanılmıştır. Erkeklere ait verilerde daha yüksek sınıflandırma başarısı elde edildiği çalışmada en yüksek %87.7 sınıflandırma başarısı gerçekleştirmiştir (Kotti ve Paternò, 2012).

Duygu sınıflandırma işlemi yapılan bir diğer çalışma (Arias ve ark., 2014) EmoDB veri tabanı kullanılarak gerçekleştirilmiştir. Arias ve ark. (2014) PCA'nın "Temel Fonksiyonlar" fonksiyonunu kullanarak F0 değerlerinden duygu modülasyonu tespiti yapmışlar. İkili duygu sınıflandırma işlemi sonunda %75.8 tanıma oranı elde etmişlerdir.

Öz bağımlı Model (AR) ile elde edilen özelliklerin de duygu tanıma uygulamalarında yüksek performans gösterdiği gözlemlenmiştir. Örneğin Milton ve Tamil Selvi (2014) yaptıkları çalışmada, MFCC, LPC ve AR parametrelerini kullanarak kNN, GKM ve YSA yöntemleriyle, sınıf-bağımlı çoklu sınıflandırıcı yöntemi

önermişlerdir. Sadece konuşmacı bağımsız, sadece metin bağımsız ve hem konuşmacı hem metin bağımsız olarak gruplandırarak sınıflandırma işlemini gerçekleştirmişlerdir. Tek çıkışlı 10 çapraz doğrulama uygulanarak gerçekleştirilen performanslarda sınıflandırma metriği olarak özgüllük, duyarlılık ve f-ölçümü değerleri temel alınmıştır. Uygulama sonuçlarına bakıldığında farklı özellik kümeleri ve farklı ölçümler temel alarak yapılan sınıflandırmalarda SD değeri AR ve MFCC özellikleri birlikte kullanılarak konuşmacı bağımsız sınıflandırmada %57.8, metin-bağımsız sınıflandırmada %83.7 ve konuşmacı ve metin bağımsız sınıflandırmada ise %80.2 olarak elde edilmiştir (Milton ve Tamil Selvi, 2014).

Zhao ve Zhang (2015) yaptıkları çalışmada Konum Kısıtlı Çekirdek Seyrek Gösterime Dayalı Sınıflandırma (LC-KSRC) adını verdikleri konum kısıtlı çekirdek seyrek gösterime dayalı duygu sınıflandırma metodunu sunmuşlardır. EmoDB ve Polish veri tabanı üzerinde yaptıkları çalışmada diğer metotlara göre daha tatmin edici performans gerçekleştirmişlerdir. EmoDB veri tabanında, Doğrusal Ayrımcılık Analizi (LDC) ile %69.74, kNN ile %70.84, Radyal Taban Fonksiyonu Sinir Ağları (RBFNN) ile %71.22, DVM ile %78.75 ve önerdikleri LC-KSRC yöntemi ile %83.95 doğruluk elde etmişlerdir (Zhao ve Zhang, 2015).

Sun ve ark. (2015) Mel Katsayılarının bir çerçevedeki Mel filtrelerinin komşuluk katsayıları arasındaki ilişkinin göz önünde bulundurulmaması sebebiyle spektrogramda bir çok faydalı özelliklerin kaybedildiğini iddia etmektedir. Bu nedenle spektrogramdaki yerel enerji dağılımını tanımlamak için spektrogramdaki yerel bölgeden Hu Moment değerlerini hesaplamışlar ve bu şekilde enerjinin merkezde yoğunlaşma derecesini incelemişlerdir. Önerdikleri HuWSF yöntemiyle üç farklı veri tabanı kullanılan deneylerden EmoDB ile yapılan uygulamalara bakıldığında konuşmacı bağımsız için, önerilen HuWSF yöntemiyle bulunan sonuç %74.71; HuWSF ile prozodik özelliklerin birleşimiyle elde edilen sonuç ise %81.01'dir. Konuşmacı bağımlı olarak gerçekleştirilen denemelerde sırasıyla %84.72, %87.66, %89.32 sınıflandırma başarısı elde etmişlerdir (Sun ve ark., 2015).

Lee ve ark. (2011) duygu tanımak için hiyerarşik bir hesaplama yapısı öne sürmüştür. Önerilen yapısal haritalamayla, ikili sınıflandırmanın sonucunda elde edileni bir sonraki sınıflandırıcıya giriş olarak vererek çoklu duygu sınıflandırma gerçekleştirmişlerdir. Sınıflandırma yapısı akustik özellikler kullanılarak AIBO ve USC IEMOCAP veri tabanları üzerinde test edilmiştir. Deney sonuçlarına bakıldığında performans kriteri olarak kullanılan ortalama ağırlıksız duyarlılık değerinin AIBO veri

tabanında DVM yöntemi ile %3.37 arttığı; USC IEMOCAP veri tabanında ise %7.44 oranında iyileşme sağlandığı görülmüştür (Lee ve ark., 2011).

Koolagudi ve ark. (2018) Telugu Filmlerinden konuşmacı bağımsız, içerikleri farklı ses verilerinden oluşturdukları veri tabanı üzerinde çalışmışlardır. Bu çalışmada spektral ve prozodik özellikler birlikte kullanılmıştır. Prozodik özelliklerin minimum, maksimum, ortalama ve standart sapması alınmıştır ve MFCC özellikleriyle birlikte kullanılmıştır. Kızgın, korku, mutlu, normal ve üzgün gibi beş temel duygu üzerinde çalışılmıştır. Sınıflandırıcı olarak GKM, k-Ortalama kümeleme ve Vektör Nicelemesi (VQ) ve YSA kullanılmıştır. Farklı özellik kümelerine farklı sınıflandırıcı uygulayarak yapılan çalışmada en iyi sonuçlar GKM ile elde edilmiştir (Koolagudi ve ark., 2018).

Duygu tanımada çokça kullanılan bir diğer özellik çıkarım yöntemi dalgacık dönüşümüdür. Kısa zamanlı Fourier dönüşümüne (STFT) alternatif olarak geliştirilen sürekli dalgacık dönüşümü (CWT) frekansa bağlı geçici çözümleri optimize etmek için geliştirilmiştir. CWT yöntemiyle yapılan çalışmalardan bir tanesi Ray ve Chan (2001) aittir. Çalışmada CWT ile elde edilen özellikler GA ile sınıflandırılmıştır. Bir başka çalışmada ise ayrık dalgacık tabanlı çoklu çözünürlük analizi kullanılarak yüz görüntülerini ve konuşma sinyallerini kullanarak bir duygu tanıma yöntemi önermişlerdir (Ray ve Chan, 2001; Go ve ark., 2003).

Reddy ve Vijayarajan (2020) çalışmalarını EmoDB ve Telugu veri tabanı üzerinde gerçekleştirmişlerdir. Verileri %80 test %20 eğitim olarak ayırmışlardır. MFCC, LPC, DWT ve Mel Frekanslı Ayrık Dalgacık Dönüşümü (MFDWT) yöntemleri ile özellikler elde ettikleri çalışmayı DVM ve kNN yöntemleriyle sınıflandırmışlar ve %76.82 SD elde etmişlerdir. Tek tek duyguların tanıma oranları incelendiğinde EmoDB veri tabanı için normal duygusunu %89.87 ve kızgın duygusunu %91.38 olarak tanımışlardır (Reddy ve Vijayarajan, 2020).

Bu tezde araştırılan konulardan biri de sesten duygu tanıma probleminin etmen tabanlı yaklaşım yöntemiyle çözümüdür. Esas itibariyle Etmenler ve Çoklu Etmen Sistemleri'nin (MAS) makine öğrenmesi sistemlerinde kullanımı ile ilgili çok fazla çalışma bulunmamakta ve araştırılmaya açık bir alan olma özelliğini korumaktadır. Bu alanda yapılan ilk çalışmalardan (van Deemter ve ark., 2008) gömülü konuşma etmenleri (ECAs) arasında diyalog oluşturmak için NECA yaklaşımı sunmaktadır. ECAs, insan benzeri bir sentetik ses ve bilgisayar animasyonu içeren bir yazılımdır. Bu

yazılımda modüller oluşturulurken diyalogların üretilmesinde otonom etmenler kullanılmıştır.

Bir başka çalışmada Montano ve ark. (Montano ve ark., 2008) serbest-metin sorgu cevaplarına uygun etmeni tanımlamayı öğrenmek için bir MAS kullanmışlardır.

Navarro ve ark. (2016) belirli kuralları izleyerek melodi kompozisyonu yapabilen (yani bir diğer adıyla müzik oluşturabilen) uzman bir MAS tabanlı sanal bir organizasyon önermişlerdir (Navarro ve ark., 2016).

Bir başka çalışma da Taha ve ark. (2007) Arapça konuşma tanıma için yeni bir etmen tabanlı tasarım geliştirmişlerdir. Arapça ses tanıma, belirli bir amaç ve yalnızca bu hedefle ilgilenen etmenlerden oluşan bir MAS olarak tanımlanmıştır. YSA sınıflandırıcının geriye yayılım algoritması kullanılarak gerçekleştirilen çalışmada tüm kelimelerde %80.95 başarıya ulaşılmıştır (Taha ve ark., 2007).

Tezde uygulanan bir diğer yöntem Derin Öğrenmedir. Derin Öğrenme, özellik çıkarma ve dönüştürme için birçok doğrusal olmayan işlem birimi katmanını kullanır. Derin Öğrenme kullanılarak geliştirilen mimariler, elle çıkarılan özellikler yerine veriyi en iyi temsil eden hiyerarşik özellik çıkarımını gerçekleştirmektedir.

Derin Öğrenme Algoritmaları, imge ve video işleme (Zia ve ark., 2017), doğal dil işleme (Sulubacak ve Eryigit, 2018), biyomedikal sinyal ve görüntü işleme (Yin ve ark., 2019) ve ses analizi (Sainath ve ark., 2013; Zhang ve ark., 2018b) gibi birçok alanda problem çözümü için kullanılmıştır.

Derin Öğrenme yöntemleriyle sesten duygu tanıma ile ilgili çalışmalardan biri Zhang ve ark.'nın (2018) yaptığı çalışmadır. Bu çalışmada Derin Evrişimsel Sinir Ağı (DCNN) kullanarak, konuşma sinyalinden duygu elde etmedeki boşluğu doldurmak amaçlanmıştır. Çalışmada üç kanallı log – mel - spektrogram görüntüleri DCNN'e giriş olarak verilmiştir. AlexNet DCNN modeli kullanılarak eğitilmiş ve EmoDB, RML, eNTERFACE05 ve BAUM-1 veri tabanlarına uygulanarak umut verici performans elde etmişlerdir (Zhang ve ark., 2018b).

Bir başka çalışmada Özseven (2018) spektrogram tabanlı özellik çıkarma yaklaşımı, Gabor Filtresi (GF), Yönlendirilmiş degradelerin histogramı (HOG), Gri Seviye Eş Oluşum Matrisi (GLCM) ve Dalgacık Ayrışması (WD) yöntemleri ile spektrogramdan özellikler çıkartmıştır. Frekans bilgisi içeren Spektrogram ile akustik özellikler içeren temel frekans, format frekansı ve MFCC yöntemleri ile elde edilen iki farklı özellik grubu kullanılmıştır. Sınıflandırıcı olarak Destek Vektör Makineleri ve Keel programları kullanılarak yapılan çalışmada spektrogramdan elde edilen özellikler

ile akustik özelliklerin birlikte kullanımıyla iyi sonuçlar elde edildiği gözlemlenmiştir (Ozseven, 2018).

Bir başka çalışmada Mirsamadi ve ark. (2017) IEMOCAP duygu veri tabanı üzerinde gerçekleştirdikleri çalışmada Derin Tekrarlayan Sinir Ağı (DRNN) kullanarak kısa süreli çerçeve seviyesindeki akustik özelliklerin öğrenilebileceğini göstermişlerdir. Aynı zamanda sinyal üzerindeki duygu yoğunluğu olan bölgelere odaklanmak için yeni bir strateji önermişlerdir (Mirsamadi ve ark., 2017).

Fayek ve ark. (2017) ileri beslemeli ve Tekrarlayan Sinir Ağı (RNN) mimarilerini ve çeşitlerini kullanarak gerçekleştirdikleri çalışmada IEMOCAP duygu veri tabanını kullanmışlardır. Önerdikleri yöntemde test başarısı %60.89 (SD) olarak elde edilmiştir (Fayek ve ark., 2017).

Harár ve ark., (2017) yaptıkları çalışmada, EmoDB'nin üçlü duygu grubunu (kızgın, normal, üzgün) kullanmışlardır. DSA mimarisi ile gerçekleştirilen çalışmada veriler Eğitim-test olarak iki kısma ayrılmıştır. Her bir veri 20 milisaniyelik, üst üste bindirmesiz bölünmüştür. Test doğruluğu olarak %96.97 başarı elde etmişlerdir (Harar ve ark., 2017).

Verilerin bölümlere ayrılması çok tercih edilen bir yöntem olsa da sabit boyutlu küçük parçalara bölmenin veri bütünlüğünü bozduğunu düşünen araştırmalar da bulunmaktadır. Ma ve ark. (2018) değişken uzunluklu konuşma segmentlerin spektrogramlarını direkt DSA e uygulamışlardır. Çalışma IEMOCAP veri tabanının dört duygusu üzerinde gerçekleştirilmiştir. Değişken uzunlukta verilerin kullanılabilmesi yönüyle diğer geleneksel yöntemlerle karşılaştırıldığında bu çalışma da önem kazanmaktadır (Ma ve ark., 2018).

Bir başka çalışmada (Badshah ve ark., 2019) ise veri bütünlüğünü bozmadan spektrogramları çıkartmışlardır. Bir sonraki aşamada spektrogramları %50 örtüştürerek küçük parçalara bölmüştür. Burada amaç işlem hattındaki sürekliliği sağlamak ve ESA'i eğitmek ve daha ince ayar yapmak için spektrogram sayısını arttırmaktır. Buna rağmen tanıma EmoDB'nin 7 duygusu için %60 ve altında kalmıştır. Bu alanda yapılan çalışmalar özellikle DSA mimarisinin uygulanmasıyla spektrogram görüntülerinden ayırt edici veri elde etmeye yönlendirse de hala değişkenlik-zorluk göstermektedir.

3. MATERYAL

Duygu tanıma problemlerinde kullanılan çok sayıda veri tabanı bulunmaktadır. Bunlardan bazılarının söyleniş diline göre bir listesi Çizelge 3.1'de (Ververidis ve Kotropoulos, 2003) verilmiştir.

Veri tabanlarını içerdiği duygu sayısına, hangi dilde oluşturulduğuna, genel (public) ya da özel (private) oluşuna veya profesyoneller veya aktörler tarafından söylenişine göre farklı şekillerde sınıflandırmak mümkündür (El Ayadi ve ark., 2011).

Profesyonel aktörler tarafından seslendirilen duygu veri tabanlarının başında LDC Duygusal Prozodik Konuşma ve Metin Kaydı Veri tabanı (M. Liberman ve ark., 2002), Berlin duygu veri tabanı (EmoDB) (Burkhardt ve ark., 2005) ve INTERFACE veri tabanları (Martin ve ark., 2006) gelmektedir. LDC veri tabanı 7 profesyonel aktör tarafından ve 15 duygunun İngilizce olarak seslendirildiği 10 farklı söyleyişten oluşan duygu veri tabanıdır. EmoDB ise 10 farklı aktör tarafından 7 duygunun, 10 farklı şekilde Almanca olarak seslendirilmesi ile oluşturulmuştur. INTERFACE veri tabanı ise İngilizce, Slovence, İspanyolca, Fransızca olacak şekilde 4 farklı dilde 8 farklı duyguda profesyonel aktörler tarafından seslendirilmiş verilerden oluşmaktadır.

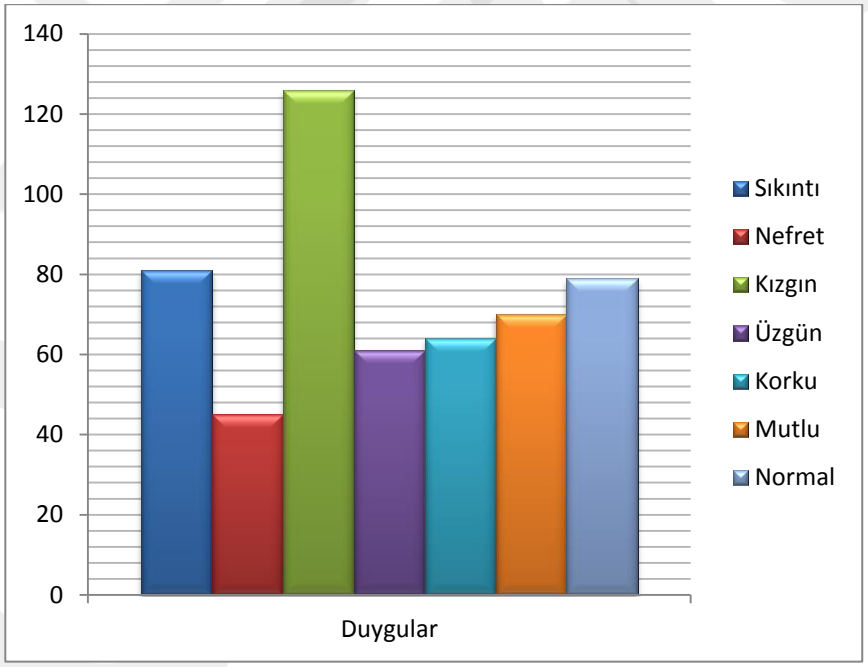
Profesyonel olmayan aktörler tarafından seslendirilen veri tabanları ise Danca duygu veri tabanı (I. Engberg ve A. Hansen, 1996), KISMET (Breazeal ve Aryananda, 2002), ESMBS (Nwe ve ark., 2003), BabyEars (Slaney ve McRoberts, 2003) veri tabanlarıdır. Danca duygu veri tabanı 4 profesyonel olmayan aktör tarafından toplamda 2 kelime, 9 cümle, 2 okuma parçasından oluşan metnin 5 farklı duygu ile seslendirmesinden oluşmaktadır. KISMET, Amerikan İngilizcesi ile 3 Kadın konuşmacı tarafından 5 farklı duyguda seslendirilmiş verilerden oluşmaktadır. ESMBS, Çince dilinde 12 profesyonel olmayan aktör tarafından, 6 farklı duyguda seslendirilen verilerden oluşan bir duygu veri tabanıdır. BabyEars 12 anne-baba (6 Erkek + 6 Kadın) tarafından, 3 duygu türünde oluşturulan kayıtlardan oluşan veri tabanıdır.

Çizelge 3.1. Yaygın kullanılan duygu veri tabanları ve özellikleri

Duygu veri tabanı	**Dil**	**Örneklem Boyutu**	**Seslendiren Kişiler**	**Duygular**
LDC duygusal prozodik konuşma ve metin kaydı veri tabanı (M. Liberman ve ark., 2002)	İngilizce	7 Aktör 15 Duygu 10 Söyleyiş	Profesyonel Aktörler	Nötr, panik, kaygı, sıcak, öfke, soğuk, öfke, umutsuzluk, üzüntü, sevinç, ilgi, can sıkıntısı, utanç, gurur, saygısızlık
Berlin duygu veri tabanı (Burkhardt ve ark., 2005)	Almanca	10 Aktör 7 Duygu 10 Söyleyiş +Bazı ikinci söyleyişler 800 Söyleyiş	Profesyonel Aktörler	Öfke, sevinç, hüzün, korku, tiksinti, can sıkıntısı, tarafsız öfke
INTERFACE veri tabanı (Martin ve ark., 2006)	İngilizce, Slovence, İspanyolca, Fransızca	186 İngilizce, 190 Slovence, 184 İspanyolca, 175 Fransızca Söyleyiş	Profesyonel Aktörler	Öfke, tiksinti, korku, sevinç, sürpriz, üzüntü, yavaş normal, hızlı normal
Danca duygu veri tabanı (I. Engberg ve A. Hansen, 1996)	Danca	4 Aktör 5 Duygu 2 Kelime 9 Cümle 2 Okuma parçası	Profesyonel Olmayan Aktörler	Öfke, sevinç, hüzün, sürpriz, normal
KISMET veri tabanı (Breazeal ve Aryananda, 2002)	Amerikan İngilizcesi	1002 Söyleyiş 3 Kadın konuşmacı 5 Duygu	Profesyonel Olmayan Aktörler	Onaylama, dikkat, yasaklama, yatıştırıcı, tarafsız
ESMBS veri tabanı (Nwe ve ark., 2003)	Çince	720 Söyleyiş, 12 Konuşmacı 6 Duygu	Profesyonel Olmayan Aktörler	Öfke, sevinç, hüzün, tiksinti, korku, sürpriz
BabyEars veri tabanı (Slaney ve McRoberts, 2003)	İngilizce	509 Söyleyiş, 12 Aktör 6 Erkek&6 Kadın, 3 Duygu	Anne Babalar	Onaylama, dikkat, yasaklama

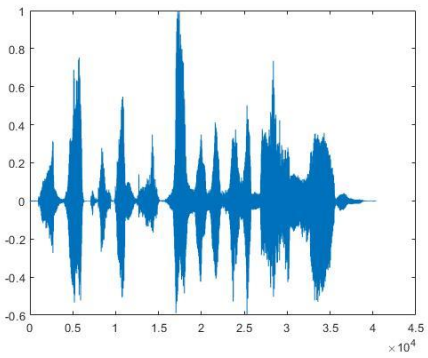
3.1. Berlin Duygu Veri tabanı (EmoDB)

Bu tez çalışmasında Berlin Duygu veri tabanı (EmoDB) (Burkhardt ve ark., 2005) olarak isimlendirin veri tabanı kullanılmıştır. Profesyoneller tarafından seslendirilerek oluşturulan veri tabanları genellikle "özel" olduğundan erişimi pahalı olmaktadır. Buna rağmen EmoDB "genel" bir veri tabanıdır. Genel olması ve bu konuda karşılaştırma yapılabilecek çoklukta çalışmanın bulunması sebebiyle, EmoDB veri tabanı tez çalışmasında tercih edilmiştir. Bu veri tabanı Berlin Teknik Üniversitesi'nde yankısız bir odada oluşturulmuş, 5 bay, 5 bayan olmak üzere toplam 10 farklı aktör tarafından seslendirilmiştir. Beşi uzun, beşi kısa olmak üzere 10 farklı cümle 7 duygu ile seslendirilmiştir. Seslendiren kişilere ait cümleler değişkenlik göstermektedir. Bu duygular şunlardır: Kızgın, Sıkıntılı, Nefret, Korku, Mutlu, Üzgün, Normal. Duyguların veri tabanındaki dağılımı Şekil 3.1'de gösterilmiştir.

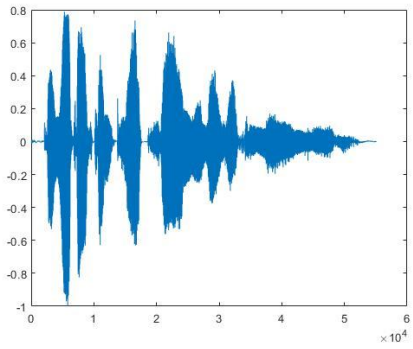


Şekil 3.1. EmoDB veri tabanında duygu dağılımı

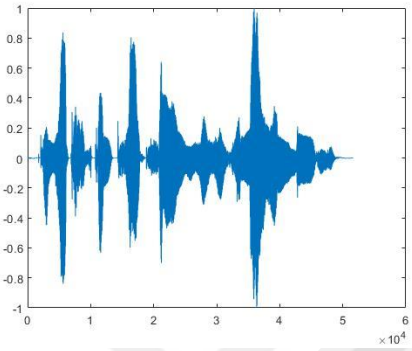
Her bir duyguya ait birer ses kaydının sinyal gösterimi Şekil 3.2'de gösterilmiştir.



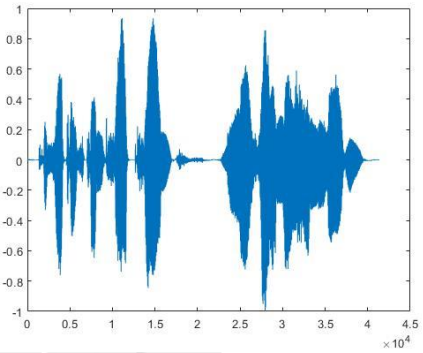
Kızgın



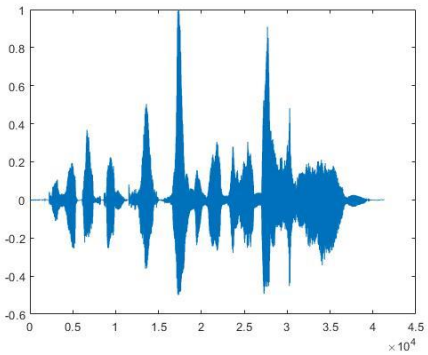
Sıkıntılı



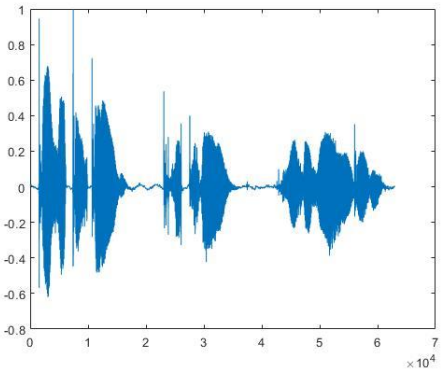
Nefret



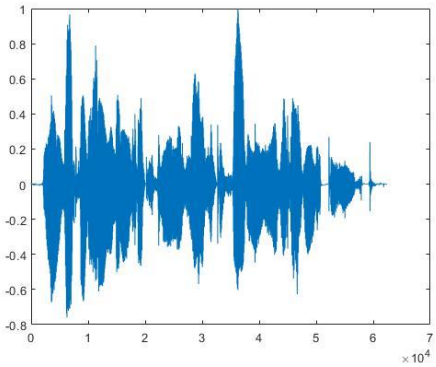
Korku



Mutlu



Üzgün



Normal

Şekil 3.2. Her bir duyguya ait ses sinyal örnekleri

Duyguların veri tabanındaki sayıları şu şekildedir: 81 Sıkıntılı (S), 46 Nefret (N), 126 Kızgın(K), 61 Üzgün(Ü), 64 Korku (Ko), 70 Mutlu (M), 79 Normal (No). Seslendirilen cümleler ise günlük hayatta sıklıkla kullanılan cümlelerdir ve Almanca olarak seslendirilmiştir. Kullanılan cümleler Çizelge 3.2'de gösterilmektedir

EmoDB veri tabanında kayıtlar, 48 kHz örnekleme frekansı ile kaydedilmiş ve daha sonra örneklenme frekansı 16 kHz'e düşürülmüştür. Ses formatı PMC'dir. Veri tabanı toplamda yaklaşık 30 dakikalık veriden oluşmaktadır. Çalışmada kullanılan her bir veri bir segment olarak ifade edilmiştir ve toplam 527 segment kullanılmıştır.

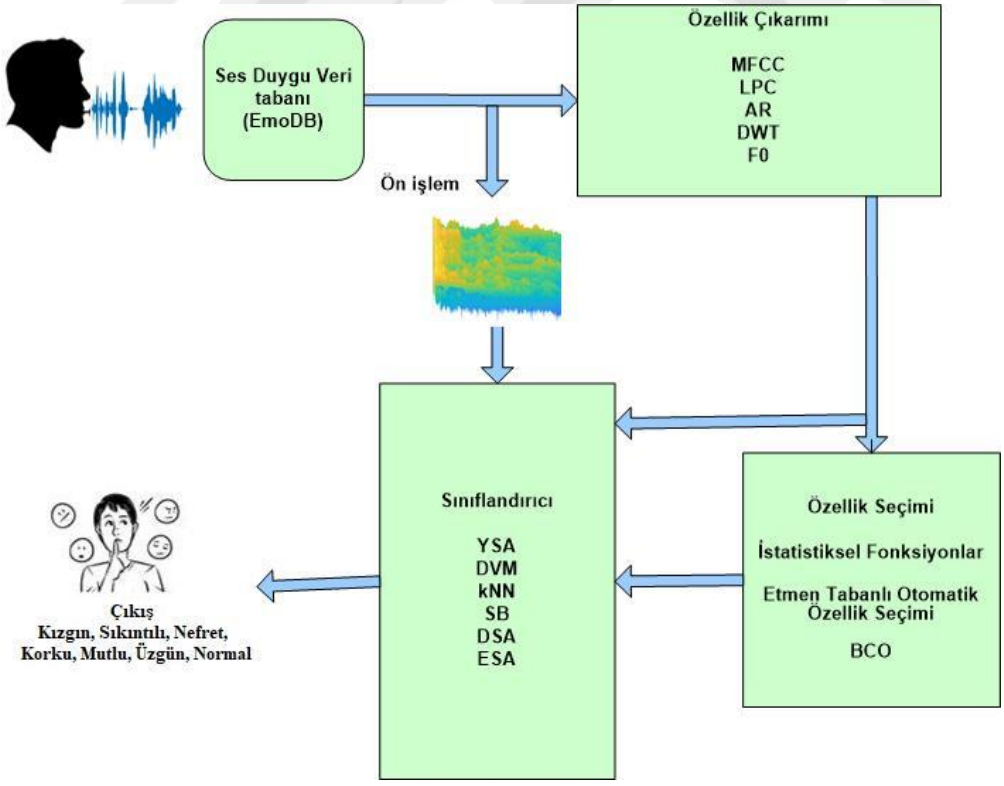
Çizelge 3.2. EmoDB veri tabanında seslendirilen cümleler

Cümle No	**Seslendirilen Cümle (Almanca)**	**Seslendirilen Cümlenin Türkçesi**
a01	Der Lappen liegt auf dem Eisschrank	Kıyafet buzdolabının üstündeydi.
a02	Das will sie am Mittwoch abgeben	Çarşamba günü teslim edecek.
a04	Heute Abend könnte ich es ihm sagen	Bu gece ona anlatabilirdim.
a05	Das schwarze Stück Papier befindet sich da oben neben dem Holzstück	Siyah kâğıt parçası tahta parçasının yanındadır.
a07	In sieben Stunden wird es soweit sein	Yedi saat içinde hazır olacak.
b01	Was sind denn das für Tüten, die da unter dem Tisch stehen?	Masanın altında ne tür çantalar var?
b02	Sie haben es gerade hochgetragen und jetzt gehen siewieder runter.	Sadece üst kata taşıdılar ve şimdi tekrar aşağı iniyorlar.
b03	An den Wochenenden bin ich jetzt immer nach Hause gefahren und habe Agnes besucht.	Hafta sonlarında hep eve giderim ve Agnes'i görürüm.
b09	Ich will das eben wegbringen und dann mit Karl was trinken gehen	Sadece bunu götürüp Karl'la bir şeyler içmek istiyorum.
b10	Die wird auf dem Platz sein, wo wir sie immer hinlegen	Her zaman koyduğumuz yerde olacak.

4. YÖNTEM

Bu tez çalışmasında EmoDB veri tabanı kullanılarak üç uygulama gerçekleştirilmiştir. İlk iki uygulama çıkarılan özelliklerden önerilen yöntemlerle özellik seçimi yapıldıktan sonra YSA, DVM, kNN ve SB ile sınıflandırılmıştır. Üçüncü uygulama ise özellik çıkarımı yapılmadan direk ses verilerinden oluşturulan spektrogram görüntüleri üzerinde ESA uygulanarak ve özniteliklerden oluşturulan veri kümesine DSA uygulayarak gerçekleştirilmiştir.

Ses verilerinden duygu tespitinde izlenen işlem adımları incelendiğinde temel olarak ön işleme, özellik çıkarımı, özellik seçimi ve sınıflandırıcı işlem adımları görülmektedir. Şekil 4.1'de giriş olarak verilen ses sinyalinin duygusunun tespiti için gerekli aşamalar gösterilmiştir. Giriş sinyali ses sinyalidir ve farklı boyutta olabilir. İlk olarak ses sinyalinden özellikler çıkarılır. Çıkarılan bu özellikler direkt kullanılabildiği gibi özellik seçimi yapılarak da kullanılabilir. Özellik seçiminde amaç anlamlı ve ayırt edici özelliklerin seçimidir. En son olarak sınıflandırıcı ile verinin hangi duygu grubuna ait olduğu tespit edilir.



Şekil 4.1. Duygu tanıma işlem adımları

4.1. Çalışmada Kullanılan Özellik Çıkarım Yöntemleri

Sesten duygu tanımada en önemli adımlardan bir tanesi en çok bilgi içeren özelliklerin elde edilmesi aşamasıdır. Özellikler genel olarak spektral ve prozodik özellikler olarak iki grupta incelenebilir. Spektral ve prozodik özellikler Çizelge 4.1'de detaylı olarak gösterilmiştir.

Çizelge 4.1 Spektral ve prozodik özellikler

Spektral Özellikler	Mel Frekansı Kepstrum Katsayıları (MFCC)
	Doğrusal Öngörüm Katsayıları (LPC)
	Ayrık Dalgacık Dönüşümü (DWT)
	Öz bağımlı Model Parametreleri (AR, AR-MA)
Prozodik Özellikler	Temel Periyot (T0),Temel Frekans (F0), Perde Frekansı

4.1.1. Mel Frekansı Kepstrum Katsayıları (MFCC)

Mel Frekansı Kepstrum Katsayıları (MFCC) ses sinyallerinden özellik elde edilirken en çok kullanılan spektral özelliklerin başında gelir. MFCC özellikleri insan kulağını modellemektedir (Becchetti ve Prina-Ricotti, 2004). Bunun sonucu olarak MFCC yöntemi duygu tanıma uygulamalarında oldukça başarılı sonuçlar vermektedir. MFCC katsayılarını elde ederken aşağıdaki işlem adımları kullanılmaktadır (Şekil 4.2):

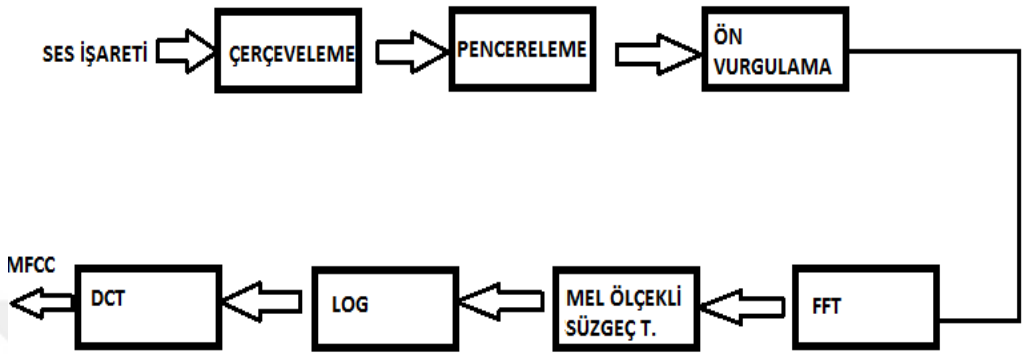
- Çerçeveleme
- Pencereleme
- Ön Vurgulama
- Hızlı Fourier Dönüşümü (FFT)
- Mel Ölçekli Süzgeç Takımı
- Logaritma Alma (LOG)
- Ayrık Kosinüs Dönüşümü (DCT)

Çerçeveleme

Giriş işareti, M örnekten oluşan kısımları örtüşen N uzunluğundaki konuşma parçalarına bölünür $(M<N)$. İlk çerçeve N örnekten oluşurken sonraki çerçeve, ilk

çerçeveden *M* örnek sonra başlar ve böylece *N-M* kadar örnek örtüşür (Becchetti ve Prina-Ricotti, 2004).

Pencereleme



Şekil 4.2. Mel Ölçekli kepstrum katsayılarının işlem adımları

Çerçeveleme işleminden sonraki adım olan pencereleme işleminde amaç sinyalin başındaki ve sonundaki süreksiz kısımları azaltmak, dolayısıyla sinyalin başındaki ve sonundaki bilgi içermeyen bölümleri bastırarak spektral bozulmayı engellemektir. Giriş işaretimizi *x(n)*, pencere fonksiyonunu *w(n)* ve çıkış işareti ise *y(n)* ile ifade edecek olursak, çıkış işaretin ifadesi Denklem 4.1'deki gibi olur (Becchetti ve Prina-Ricotti, 2004):

$$y(n) = x(n)w(n) \tag{4.1}$$

Genellikle pencere fonksiyonu olarak Hamming penceresi kullanılır. Hamming penceresinin matematiksel ifadesi, *n* örnek sayısı olacak şekilde Denklem 4.2.'de verilmektedir (Becchetti ve Prina-Ricotti, 2004).

$$w(n) = 0.54 - 0.46\,cos\left(\frac{2\pi n}{N-1}\right), \quad 1 \leq n \leq N \tag{4.2}$$

Ön Vurgulama

Ön vurgulama işleminde giriş işareti birinci dereceden bir Sonlu Dürtü Yanıtı Filtresi FIR süzgeç girişine uygulanır. Amaç sinyalin yüksek frekans bileşenlerini daha baskın hale getirmektir. Birinci dereceden süzgecin transfer fonksiyonu Denklem 4.3'de verilmiştir:

$$H(z) = 1 - 0.95z^{-1} \tag{4.3}$$

Hızlı Fourier Dönüşümü

N örnekten oluşan konuşma parçasını zaman domeninden, frekans domenine çevirmek için Hızlı Fourier Dönüşümü (FFT) uygulanır. FFT, Ayrık Fourier Dönüşümünü (DFT) hızlandırmak için uygulanan bir algoritmadır. *N*, örnekli bir küme için DFT'nin matematiksel ifadesi Denklem 4.4'de gösterilmiştir (Becchetti ve Prina-Ricotti, 2004).

$$X_n = \sum_{k=0}^{N-1} x_k \, e^{\frac{-2\pi kn}{N}} \quad , \qquad n = 0,1,2, \ldots\ldots, N-1 \tag{4.4}$$

Mel Ölçekli Süzgeç Takımı

Konuşma sinyallerinin frekans ortamındaki içeriklerinin doğrusal ölçekli olmadığı akustik çalışmalar sonucunda ortaya konulmuştur. 1 *kHZ* frekanslı bir sesin, insan kulağının algısal duyma eşiğinin 40 *dB* yukarısı 1000 mel *kHz* olarak tanımlanır. Diğer değerler referans sese göre ayarlanır Mel ölçeği 1 *kHz*'e kadar doğrusal, 1 *kHz*'den sonra ise logaritmik olarak değişen aralıklarla ifade edilen bir ölçektir. Verilen bir *f (Hz)* frekansını mel frekansı ölçeğinde ifade etmek için Denklem 4.5 kullanılır (C. Becchetti; L. P. Ricotti, 2004).

$$mel(f) = 2595 \, log\left(1 + \frac{f}{700}\right) \tag{4.5}$$

Logaritma Alma

Logaritma alma işleminde süzgeç çıkışında elde edilen işaretin logaritması alınmaktadır. *l*. filtrenin logaritmik enerji çıkışı c(*l*) Denklem 4.6'daki gibi hesaplanır (Becchetti ve Prina-Ricotti, 2004).

$$\boldsymbol{c}(l) = log(\frac{1}{A_1}\sum_{k=L_1}^{U_1} F_1[k]X[k]) \tag{4.6}$$

Denklem 4.7'deki A_1 filtrelerin band genişliğine bağlı olarak kullanılan normalizasyon katsayısıdır (Becchetti ve Prina-Ricotti, 2004).

$$A_1 = \sum_{k=L_1}^{U_1} F_k[k] \tag{4.7}$$

Ayrık Kosinüs Dönüşümü (DCT)

MFCC çıkarma işleminin en son adımı olan Ayrık Kosinüs Dönüşümünde (DCT) logaritmik mel ölçeğindeki veriler tekrar zaman domenine dönüştürülür. Sonuç olarak da elde edilen veriler Mel frekansı kepstrum katsayıları (MFCC) olarak adlandırılır. Logaritma alma işleminden sonra elde edilen işareti c_l ile gösterirsek MFCC katsayıları Denklem 4.8'deki gibi hesaplanır (Becchetti ve Prina-Ricotti, 2004).

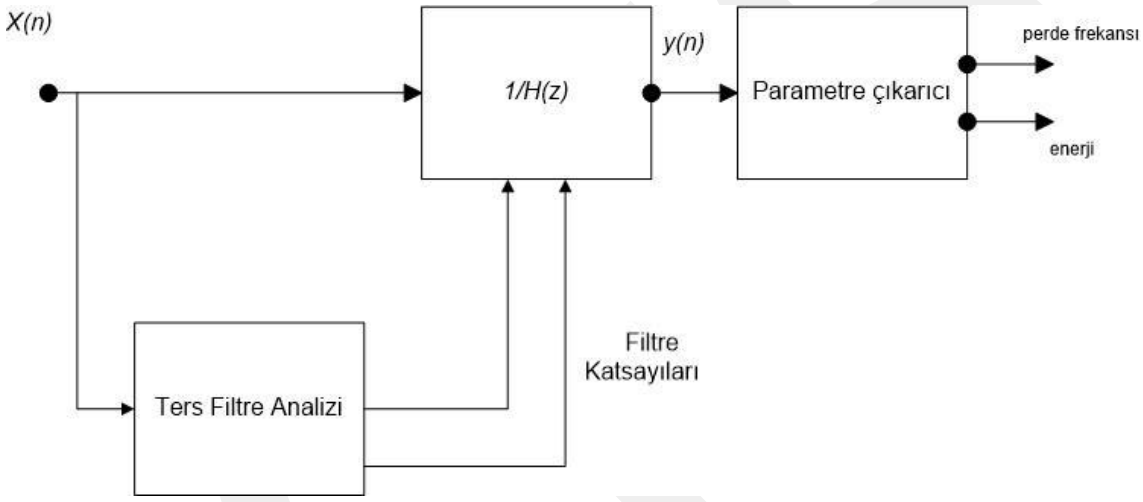
$$mfcc\ (i) = \frac{1}{FS}\sum_{i=1}^{FS} c(l)\ cos\left(i\left(l-\frac{1}{2}\right)\frac{\pi}{FS}\right), \qquad i = 1, \ldots . FS-1 \tag{4.8}$$

4.1.2. Doğrusal Öngörüm Katsayıları (LPC)

LPC ses ve konuşma sinyallerinde filtre tasarımı ve ses kodlama uygulamalarında yaygın olarak kullanılan bir yöntemdir. Bu yöntem zamanın bir fonksiyonu olarak değişir. LPC katsayılarının hesaplaması aynı zamanda bir otokorelasyon yöntemidir.

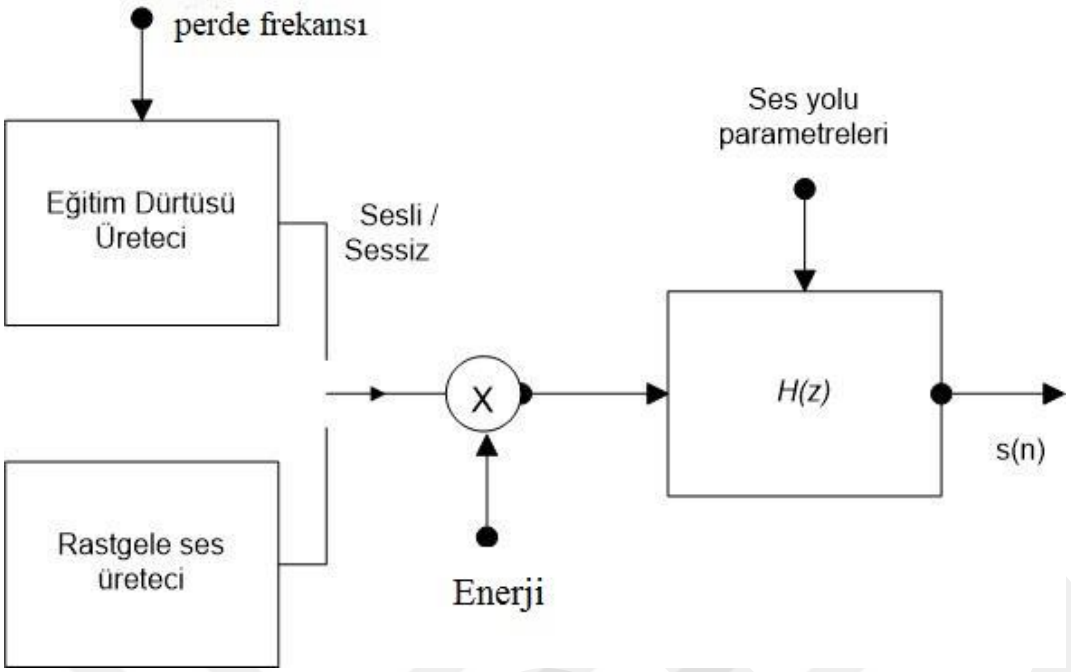
LPC ses sinyalinin bir rastgele jeneratör kaynağı tarafından üretilen dalga darbelerinin bir tüpün sonunda üretildiğini varsayar. Rastgele kaynağından üretilen ses

kıvrımlıdır ve dalga darbeleri ses yüksekliği ve perde frekansı ile karakterize edilir. Ses yolu, format frekansı ile karakterize edilen tüpü oluşturur. Bu tüp tüm-kutuplu bir filtre olarak modellenmiştir. Bu filtre katsayıları doğrusal denklem sistemlerini çözerek hesaplanır ve ses yolunun yanıtını tahmin eden bir tüm-kutuplu filtre H*(z)* oluşturmak için kullanılır. *x(n)* giriş sinyali ve *1/ H(z)* filtrenin tersi ise *y(n)* ses bölümünün etkisinin ortadan kaldırdığı artık sinyaldir. Bu sinyal sadece *x(n)* kaynağı hakkında bilgiler içerir ve kaynağın perde frekansı ve enerji bilgilerini çıkarmak için kullanılır. Ses girişinden LPC kullanılarak çıkarılan konuşma parametreleri işlem adımları Şekil 4.3'de gösterilmektedir. (Hassan, 2012)



Şekil 4.3. LPC uygulanarak parametre çıkarma işlemi

LPC bir ses kodlayıcı olarak kullanıldığında sadece filtre katsayıları, perde frekansı ve enerji bilgileri alıcıya gönderilir. Bu işlemin blok diyagramı da Şekil 4.4'de gösterilmektedir.



Şekil 4.4. LPC parametreleri kullanılarak konuşma ses üretim diyagramı

4.1.3. Temel Periyot (T0), Temel Frekans (F0), Perde Frekansı (Pitch)

Temel frekansın doğru bir şekilde elde edilmesi konuşma analizinde en önemli konulardan biri olmasına rağmen kesin bir tanımı belirlenememiştir (Furui, 2000). Konuşmanın temel frekansının tahmini genellikle perde frekans analizi ile karıştırılır. Perde frekansı ses üretiminde bir parametre olmaktan çok idrak edilen algılanan şeyi ifade eder (McKinney, 1965). Buna rağmen literatürde perde frekansı ifadesi temel frekans ve temel periyot kavramlarını ifade eder.

Temel frekans çıkarım yöntemleri üç ana başlık altında toplanabilir (Furui, 2000).

- Dalgaform işlemi
- Bağıntı (korelasyon) işlemi
- Tayf (spektrum) işlemi

En çok kullanılan temel frekans metotları Çizelge 4.2'de detaylı olarak anlatılmıştır (Itakura, 1975; Furui, 2000).

Temel frekans temel periyot baz alınarak da hesaplanabilir. Temel Periyot üç farklı şekilde hesaplanır (Howard, 1991)

- Uzun-Vade (Long-term) Tanımı, T_0 kesinlikle periyodik olan sinyalin periyot süresidir.
- Kısa-Vade (Short-term) Tanımı, T_0 ardışık uyarımlar arasındaki ortalama geçen süre nedeniyle bir şekilde belirli kısa-vade pencere üzerindeki ortalamanın alınmasıdır.
- Periyoda göre Periyot (Period-by-period) tanımı, T_0 ardışıl periyot işaretleyicileri arasındaki geçen zamandır.

Uzun-Vade tanımı konuşma için uygulanamaz, çünkü konuşma yarı-periyodik bir sinyaldir ve bu tanım sadece sabit sinyaller için uygulanabilir. Kısa-Vade tanım konuşma sinyalinin Kısa-Vade analizi yapıldığı anlamına gelir. Periyoda göre Periyot ise konuşma sinyalinin zaman domeni vasıtasıyla elde edilebilir.

Her durumda Temel frekansın Temel periyot tanımıyla ilişkisi Denklem 4.9'da verilmektedir.

Çizelge 4.2. Temel frekans elde etme metotları

Sınıf	Perde Frekans çıkarım yöntemi	Temel özellikler
DalgaForm İşlemi	Paralel İşleme Yöntemi	Perde frekansı hesabında çoğunlukla, çok türlü basit dalgaformu tepe noktası bulucusunu kullanır.
	Veri Azaltım Yöntemi	Çeşitli mantıksal işlemler üzerinden gereksiz dalgaform verisini çıkararak, sadece perde titreşimlerini bırakır.
	Sıfır-Geçiş Sayısı Yöntemi	Dalgaform sıfır geçiş değeri içerisindeki tekrarlayan örüntülerden yararlanılır.
Bağıntı İşlemiyle	Otokorelasyon Yöntemi	Dalgaformuna otokorelasyon fonksiyonu uygulanır.
	Değiştirilmiş Korelasyon yöntemi	LPC sinyalinin artık sinyali için otokorelasyon fonksiyonundan faydalanılır.
	Basitleştirilmiş Ters Filtre İzleme Algoritması	Konuşma dalgasının indirgenmiş örneklenmesi (down-sampling) sonrasında spektrum düzleşmesi için LPC analizi uygulanır.
	AMDF Yöntemi	Ortalama Büyüklük Türev Fonksiyonu (Avarage magnitude differential function – AMDF)
Spektrum İşlemiyle	Kepstrum Metodu	Spektral örtüşmeleri ayırmak ve düşük güç spektrumunun Fourier Dönüşümünün tersi ile yapıları berraklaştırmak.
	Periyot Histogram Metodu	Harmonik bileşenler için ortak bölenler perde frekansı olarak karar verilir.

$$F_0 = 1/T_0 \tag{4.9}$$

Duygusal durum tespitinde temel frekans ayırt edici bir özellik olarak kullanılmıştır. Farklı ruh halleri için temel frekansa ait özellik bilgisi ele alındığında mutluluk, korku ve utangaçlık için perde frekansı yüksek, kızgınlık, korku ve mutluluk için ortalama F0 diğer ruh hallerine göre daha yüksek, üzüntü ve sıkıntılı için ise en düşük F0 değerine sahiptir (Polat, 2007).

Bu tez çalışmasında Kepstrum Metodu ve Otokorelasyon yöntemi ile Temel Frekans öznitelikleri belirlenmiştir.

4.1.4. Ayrık Dalgacık Dönüşümü (DWT)

Dalgacık dönüşümü sinyali içerdiği farklı frekans bileşenlere ayıran ve her bir bileşeni; kendi ölçek değerine göre bir çözünürlükte inceleyen matematiksel dönüşüm fonksiyonudur (Graps, 1995; Tezel, 2007). Dalgacık dönüşümü, sinyalin farklı frekans ve zaman ölçeklerinde analiz edilebilme imkânı verdiği için çoklu çözünürlük analizi için de uygundur (Verma ve Tiwary, 2014).

Dalgacık dönüşümü STFT'ye bir alternatifi olarak çözünürlük probleminin üstesinden gelmek için ortaya atılmıştır. STFT'de zaman penceresi bütün frekanslarda değişmeyen büyüklüktedir. Sinyallerin çoğunda daha hassas bir hesaplamaya ihtiyaç duyulur. Bu hassasiyet pencerenin büyüklüğüne bağlıdır. Bunun içinde pencere büyüklüğünün sürekli değişebilmesi gerekir. Dalgacık dönüşümünde düşük frekanslı sinyali açmak için geniş bir pencere, sinyaldeki yüksek frekans detaylarını incelemek için sıkıştırılmış bir pencere tercih edilir. Dolayısıyla Dalgacık dönüşümü odağa tam olarak ayarlanarak sinyalin farklı kısımlarını gözetleyebilen bir matematik mikroskobu gibi davranır (Tezel, 2007).

Sinyalin türüne bağlı olarak çeşitli dalgacık dönüşümleri kullanılabilir. Sürekli bir giriş sinyali için (analog), zaman ve ölçek parametreleri de sürekli olabilir ki bu durumda sürekli dalgacık dönüşümü (CWT) kullanılabilirken, ayrık zamanlı sinyaller (dijital) için ise ayrık dalgacık dönüşümü (DWT) kullanılabilir (Rioul ve Vetterli, 1991; Luna ve ark., 2011). CWT'de bütün zaman ekseni boyunca, tüm ölçek değerleri için dalgacık katsayıları hesaplanmaktadır. Bunun sonucunda çok sayıda veri ortaya çıkmaktadır ve işlem yükü artmaktadır. Bu sorunu ortadan kaldırmak için DWT geliştirilmiştir.

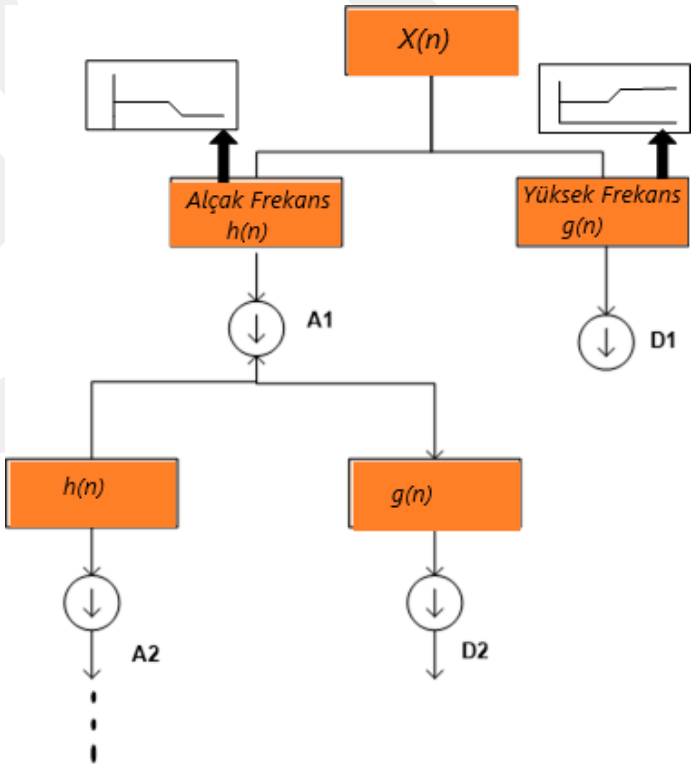
DWT dönüşümünde dalgacıklar mesafeye bağlı olarak ayrık nümerik değerler çevresinde kaydırılır ve genişletilir. Genlik değişiminde genellikle 2 faktörü kullanılır ve matematiksel ifadesi Denklem 4.10'da gösterildiği gibidir (Arı ve ark., 2008):

$$DWT(k,t) = C\psi(2^k t + l) \tag{4.10}$$

Denklem 4.10'da l, kaydırma faktörünü, k Ölçekleme faktörünü ifade eder. ψ: dalgacık fonksiyonu, t: zaman DWT yönteminde sinyal işlemede kullanılan 2 kanal alt band geçiren bir donanım süzgeci kullanılır. Temel olarak kesin aritmetik bir kurala göre sinyal alçak frekans bileşenlerine (A, yaklaşım) ve yüksek frekans bileşenlerine (D, detay) ayrıştırılır (Denklem 4.11) (Arı ve ark., 2008).

$$X(n) = A(n) + D(n) \tag{4.11}$$

Şekil 4.5'de sinyalin (*X(n)*) alçak (A) ve yüksek (D) frekans bileşenlerine ayrılması gösterilmiştir. *X* sinyali 1. dereceden dönüşümünde *A1* ve *D1* bileşenlerine ayrılır veya *A1/D1* biliniyorsa, bu bileşenlerden *X* sinyali ters dönüşüm yöntemi ile elde edilir (Arı ve ark., 2008).



Şekil 4.5. Sinyalin alçak ve yüksek frekans bileşenlerine ayrılması

4.1.5. Öz Bağımlı Model Parametreleri (AR)

Öz bağımlı Model Parametreleri (AR) metodu parametrik metotlardan biridir. Parametrik spektral analiz; güç spektrum kestirimi, maksimum frekans, bant genişliği ya da güç bileşenleri olarak bir parametre dizisiyle ifade edilebilir.

AR modellemede, farklı genlikli önceki örnekler toplanarak verilen bir periyottaki sinyalin genliğinin elde edildiği varsayılır. Bu işleme tahmini hata (beyaz gürültü) eklenir. Buna göre *i*. dereceden AR işlemi Denklem 4.12'de formüle edilmiştir:

$$x(n) = -\sum_{m=1}^{i} a_m x\,(n-m) + e(n) \tag{4.12}$$

Burada *x(n),* örnek dizisini, *a*'lar AR model katsayılarını, *e(n)* hata terimini (beyaz gürültü) ifade eder (Vaseghi, 2000).

Doğrusal denklemlerin çözümüyle AR parametrelerinin tahmini kolay olduğundan AR parametreleri sıklıkla kullanılan model tabanlı metottur. AR parametreleri farklı tahmin metotlarıyla hesaplanabilir. Bunlardan biri Burg metodudur (Übeylı ve Güler, 2004). Burg metodu kullanıldığında *k.* dereceden Burg parametrelerinin tahmini güç kestirimi Denklem 4.13'de gösterildiği şekilde formüle edilebilir (Übeylı ve Güler, 2004).

$$\widehat{K}_{BURG(f)} = \frac{\hat{e}_k}{\left|1 + \sum_{i=1}^{k} \hat{a}_k(i) e^{-j2\pi f i}\right|^{2,}}$$

$$\hat{e}_k = \hat{e}_{f,k} + \hat{e}_{b,k} \quad Toplam\ en\ küçük\ kare\ hatası \tag{4.13}$$

4.2. Özellik Seçimi Amacıyla Kullanılan Algoritmalar

4.2.1. OPT-aiNET Optimizasyon Algoritması

aiNET optimizasyon algoritması kümeleme için geliştirilmiş bir ayrık bağışıklık ağ algoritmasıdır. Opt-aiNET algoritması ise bu algoritmanın biraz daha geliştirilmiş ve optimizasyon problemlerine uyarlanmış halidir (Castro ve Timmis, 2002).

Opt-aiNET muhtemel çözümleri barındıran antikor ağından oluşan bir popülasyonu geliştirir. Popülasyonu klonlama, mutasyon, uygunluk hesabı ve klonal seçimden oluşan aşamalardan geçirerek bu işlemi gerçekleştirir. Opt-aiNET beş ana özellik ile karakterize edilebilir. Bunlar (Timmis ve Edmonds, 2004):

- Popülasyon boyutu dinamik olarak ayarlanabilir.
- Arama uzayında araştırmaya izin verir.
- Çoklu optimumların yerlerini belirler.
- Birden çok optimal çözümü bulabilme kapasitesine sahiptir.
- Durdurma kriteri tanımlanır.

Opt-aiNET algoritmasının optimizasyon versiyonu aşağıda özetlenmiştir. (de Castro ve Timmis, 2002).

1. Rastgele şekilde ağ hücrelerinden oluşan başlangıç popülasyonu oluşturulur.
2. Durdurma şartı sağlanıncaya kadar tekrar et:

2.1. Her bir ağ hücresinin uygunluğunu belirle ve uygunluk vektörünü normalize et.

2.2. Her bir ağ hücresinden N_c tane klon üret.

2.3. Her bir klonu, ebeveyn hücresinin uygunluğu ile ters orantılı olarak Denklem (4.14)'deki formülü kullanarak mutasyona uğrat. Ebeveyn hücresinin aslını sakla.

2.4. Mutasyona uğramış hücrelerin uygunluklarını hesapla.

2.5. Her bir klonu kendi ebeveyni ile kıyasla. Uygunluğu daha iyi olanı seç ve bu şekilde güncellenen popülasyonun ortalama uygunluğunu hesapla.

2.6. Eğer popülasyonun ortalama uygunluğu, bir önceki iterasyondaki popülasyonun ortalama uygunluğundan anlamlı bir şekilde farklı ise adım *(2.1)* ' e git; değilse devam et.

2.7. Ağdaki her bir hücrenin birbirleri ile olan duyarlılıklarını hesapla. Duyarlılığı baskılama eşiğin altında olan hücrelerden birini baskıla. Baskılamadan sonra, belli sayıda ağ hücresini hafıza hücresi olarak sakla.

2.8. Yeni hücre oluştur.

3. Durdurma şartı sağlanmışsa algoritmayı sonlandır değilse adım *(2.1)*' e git.

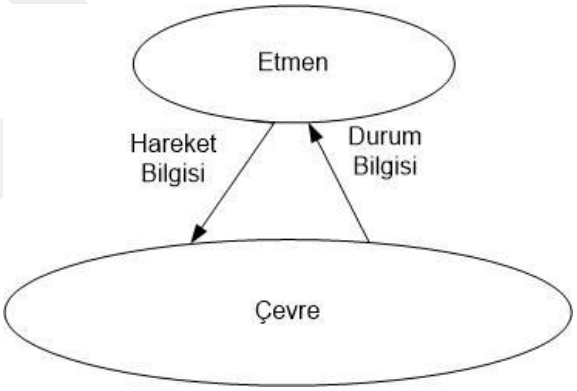
$$c' = c + \alpha * N(0,1)$$
$$\alpha = (1/\beta)\exp(-f^*) \quad (4.14)$$

c	mutasyona uğrayacak klon
c'	Klonun mutasyona uğradıktan sonraki hali
α	Mutasyon miktarı
$N(0,1)$	Ortalaması 0, varyansı 1 olan Gaussiyan Operatörü
β	α Mutasyonundaki değişim oranını kontrol eden parametredir.
f^*	Tüm ağ hücrelerinin [0,1] aralığında normalize edilmiş uygunluk değerini içeren vektör

Bu tez çalışmasında Opt-aiNET optimizasyon algoritması kullanılarak ses verilerinden duygu tanıma için özellik seçimi yapılmıştır. Bu sayede en optimum özelliklerin seçimi amaçlanmıştır.

4.2.2. Önerilen Etmen Tabanlı Modelleme

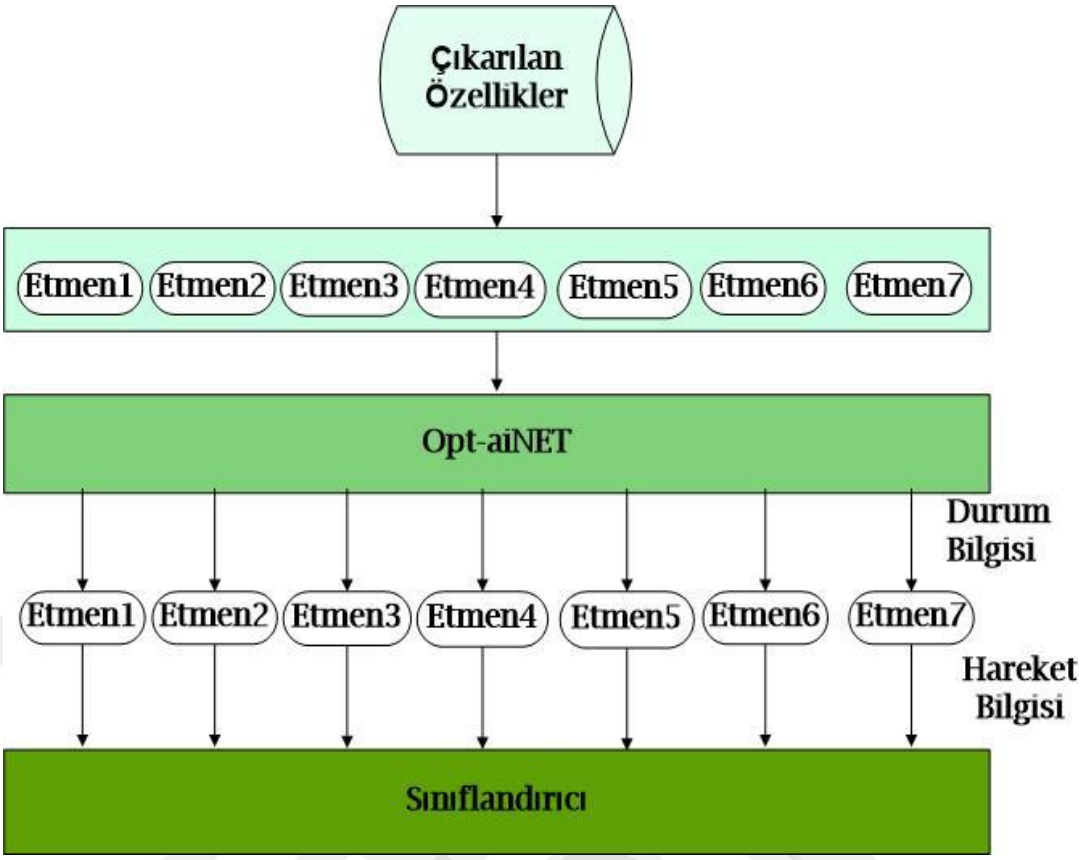
Etmen kesin bir tanımı olmamakla birlikte, belli bir çevre içinde, bir alan içinde hedef, hareket ve durum özellikleri bulunan bir nesne olarak tanımlanabilir (Stone ve Veloso, 2000). Ayrıca belli bir amacı gerçekleştirmek için otomatik olarak hareket edebilen bilgisayar sistemi olarak da ifade edilebilir (Şekil 4.6).



Şekil 4.6. Etmen Yapısı

Otonom bir etmen, içinde bulunduğu ortamın (sanal veya gerçek) algısını alan, ardından durumsal farkındalık oluşturan ve amacı doğrultusunda bu algı bilgisini kullanarak bir sonraki davranışını belirleyen ve belirlenen aksiyonunu ortamda gerçekleştiren bir sistem olarak tanımlanır (Weiss, 2013).

Bu tez çalışmasında geliştirilen etmen tabanlı modelleme ile özellik seçiminde etmen yapısına benzer şekilde, çıkarılan özellik grupları *"etmen",* optimizasyon işlemi ise bir *"çevre"* gibi düşünülmüştür. Her bir etmen için *"hareket bilgisi"* olarak 0 veya 1 değerleri tanımlanır. Hareket bilgisi "0" veya "1 " olmasına göre etmen grubunun temsil ettiği özellik grubu sınıflandırma işlemine tabi tutulur veya tutulmaz. Sınıflandırma başarısı maksimum yapılırken etmen kullanılıyorsa *"durum bilgisi=1"* olarak etmene gönderilmekte etmen de *"hareket bilgisi"* "1" olacak şekilde sınıflandırmaya tabi tutulmaktadır. Aksi takdirde *"durum bilgisi = 0"* ve aynı zamanda etmenin *"hareket bilgisi"* de "0" olduğundan sınıflandırmaya tabi tutulmayacağına karar verilmektedir. Burada amaç sınıflandırma doğruluğunu maksimum yapacak etmen grubunu seçmektir. Bu seçimi yapmak amacıyla kullanılacak optimizasyon işleminde Opt-aiNET algoritması kullanılmıştır (Şekil 4.7). Etmenler *"hareket bilgisi"* ile optimizasyon işlemine alınmaktadır. Opt-aiNET ile oluşturulan popülasyondaki bireylerin yapısına Şekil 4.8.'de bir örnek verilmiştir.



Şekil 4.7. Özelliklerin Etmen Tabanlı modellenmesi

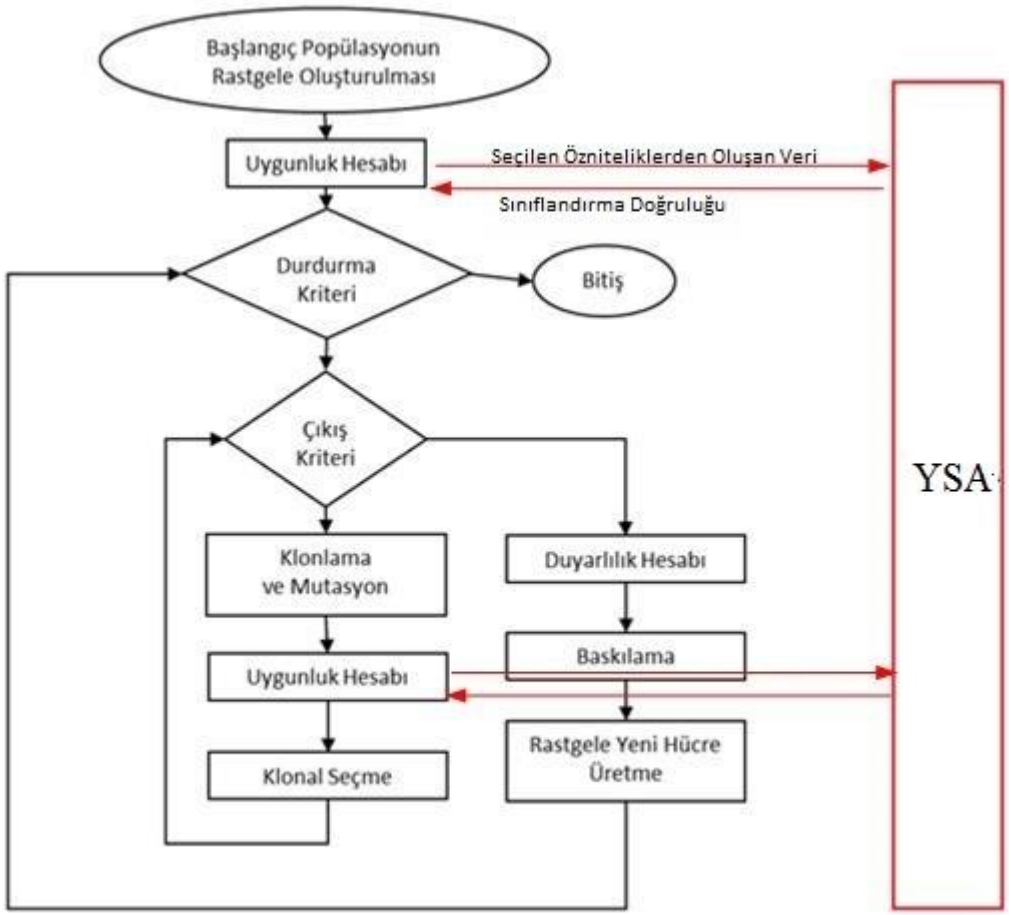
0	0	1	1	0	1	0

Şekil 4.8. : Opt-aiNET birey örneği

Örneğin 7 özellik grubuna sahip özelliklerin olduğunu varsayalım. Bu durumda Şekil 4.8'de görüldüğü gibi her bir birey uzunluğu yedidir. Bu örnekte gösterilen *sıfır* ve *birlerin* anlamı şu şekildedir:

- Birinci, ikinci, beşinci, yedinci özellik grubunu temsil eden etmen durum bilgisi = 0 'dır. Bu grubu temsil eden özellikler sınıflandırmaya tabi tutulmamaktadır.
- Üçüncü, dördüncü, altıncı özellik grubunu temsil eden etmen *durum bilgisi = 1* 'dir. Bu grubu temsil eden özellikler sınıflandırmaya tabi tutulmaktadır.

Uygunluk fonksiyonu hesaplanırken her bir bireyin temsil ettiği özellik grubu yeni bir veri kümesi olarak oluşturularak yeni sınıflandırma işlemine tabi tutulur. Sınıflandırma sonucunda elde edilen sınıflandırma başarısı uygunluk fonksiyonu olarak geri döndürülür. Şekil 4.9'da önerilen bu yöntem detaylı bir şekilde gösterilmiştir.



Şekil 4.9. Önerilen Etmen tabanlı modelleme ile özellik seçimi algoritması

Bu işlem sırasında uygunluk fonksiyonu olarak sınıflandırma doğruluğu alınmıştır. Sınıflandırma doğruluğunu maksimum yapan etmen grubu seçilmiştir. Seçilen etmen grubunun temsil ettiği özellik veri kümesi yeniden YSA'ya verilerek sınıflandırma doğruluğu hesaplanmıştır. Hesaplanan sınıflandırma doğruluğu yeni uygunluk hesabı sonucu olarak döndürülmüştür. Bu işleme Opt-aiNET algoritması sonlandırma kriteri sağlanıncaya kadar devam edilmiştir.

4.2.3. Bulanık C-Ortalama Kümeleme Yöntemi (BCO)

Tez çalışmasında kullanılan bir diğer yöntem kümelemedir. Kümeleme mevcut verilerin belirli özelliklerine göre alt kümelere ayrılması işlemine denir. Kümeleme benzer özellikteki nesnelerin aynı gruba toplanması olarak da ifade edilir. Bulanık c-ortalama algoritması (BCO) en yaygın kullanılan bulanık bölünmeli kümeleme tekniklerinden biridir. Bulanık c-ortalama algoritması ilk olarak Dunn (Dunn, 1973) tarafından 1973 de ortaya atılmış ve 1981 de Bezdek tarafından geliştirilmiştir. Bulanık c-ortalama yöntemi amaç fonksiyonunu minimize etmeye çalışır.
BCO algoritması adımları aşağıda verilişitir:

1. *Başlarken* $U=[u_{ij}]$ *matrisi,* $U^{(0)}$
2. *k. adımda:* $U^{(k)}$ *ile* $C^{(k)}= [c_j]$ *merkez vektörleri hesapla*

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m}.x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \quad (4.15)$$

3. $U^{(k+1)}$ *yı* $U^{k)}$ *olarak yenile*

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \quad (4.16)$$

4. *Eğer* $\|U^{k+1} - U^{k}\| < \varepsilon$ *ise SONLANDIR; değilse adım 2 ye dön*

Algoritma da kullanılan *m* 1 den büyük bir reel sayıdır. u_{ij} ise *j* kümesi içindeki x_i nin üyelik fonksiyonlarıdır. x_i ise *d* boyutla ölçülendirilmiş *i.* veridir. Son olarak c_j ise kümenin *d* - boyutlu merkezidir (Matteucci, 2002).

BCO veriyi, birbiri üstüne binen bulanık alt bölümlere ayırarak kümeleme işlemini gerçekleştirir. Kümelerin üyelik değerleri [0,1] aralığında bir değerdir. Verinin tüm sınıflara ait üyelik değeri toplamda 1 olmalıdır. Veri hangi kümeye ait üyelik değeri daha büyük ise o küme merkezine daha yakındır. Yani bulanık kümelemede amaç hangi kümeye ait olma ihtimalinin daha fazla olduğunu tespit etmektir (Castellano ve Fanelli, 2000; Ceylan ve Özbay, 2007).

Bu tez çalışmasında BCO, özelliklerin seçiminde kullanılmıştır.

4.3. Sınıflandırma Amacıyla Kullanılan Algoritmalar

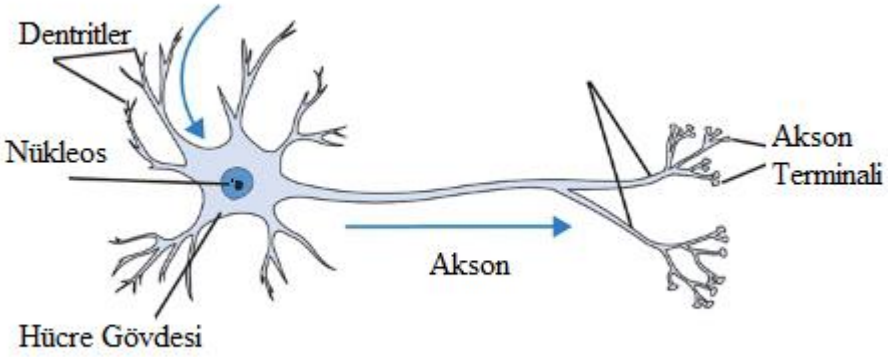
Bu bölümde tez çalışmasında kullanılan sınıflandırıcılar hakkında bilgi verilmiştir. Kullanılan sınıflandırıcılar YSA, DVM, kNN, SB ve Derin Öğrenme Algoritmaları olan ESA ve DSA'dir.

4.3.1. Yapay Sinir Ağları (YSA)

YSA insan beyninden esinlenerek geliştirilmiş, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır (Elmas, 2003). Bir başka deyişle, bilgi akışını aksonlar yardımıyla sağlayan bir grup sinir hücresinin meydana getirdiği sinir sisteminin bir benzeri olarak tanımlanmakla birlikte genellikle yenilenebilir (adaptif) olan basit elemanların yoğun bir şekilde paralel bağlanmasıyla ortaya çıkan ağlar olarak tanımlanmıştır (Kohonen, 1988).

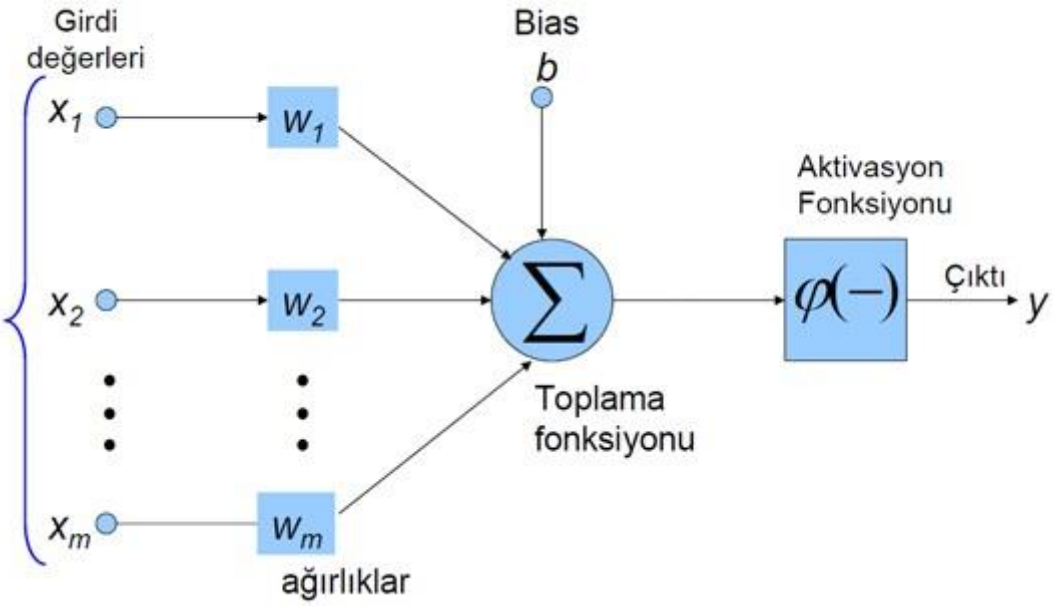
Yapay sinir ağlarının (YSA) adının geçtiği ilk çalışma 1943 yılında McCullogh ve Pitts (Mc Culloch W. ve Pitts W., 1943) ilk olarak yapay sinir tanımını yaparak hücre modelini geliştirdikleri çalışma olarak kabul edilir. Uzunca bir süre yapay sinir ağlarının birçok probleme çözüm bulamayacağı düşünülse de Kohenen (Kohonen, 1982) ve Anderson (Anderson, 1983) tarafından yapılan çalışmalar sonucunda danışmansız öğrenen ağların geliştirilmesiyle çalışmalar yeniden ivme kazanmıştır. Ayrıca 1986 yılında Rumelharlt ve ark. (2004) tarafından çok katmanlı algılayıcı tipi ağlar için "geriye yayılımlı" olarak adlandırılan bir eğitme algoritması geliştirdi. Halen en çok kullanılan öğretim sistemlerinden biri olan bu algoritmanın ortaya atılması YSA alanında bir çığır açmıştır (Rumelhart ve ark., 1986a) (Şen, 2004).

YSA'nin temel birimi işlem elemanı ya da düğüm olarak adlandırılan yapay bir sinirdir. Bir yapay sinir hücresi biyolojik sinir hücresine (Şekil 4.10) oranla daha basit olmasına karşın, biyolojik sinir hücrelerinin dört temel işlevini taklit eder. Şekil 4.11'de yapay bir sinir ağı düğümü gösterilmiştir (Elmas, 2003).



Şekil 4.10. Biyolojik sinir hücresi yapısı

Şekil 4.11'de girişler x_i sembolüyle gösterilmiştir. Bu girişlerin her biri ağırlık *w* ile çarpılır. Basitçe, bu ürünler eşik değeri *bias* ile toplanır ve sonucu oluşturmak için aktivasyon fonksiyonu ile işlem yapılır ve y_i çıkışı alınır (Elmas, 2003).



Şekil 4.11. Yapay nöron modeli

Nöron davranışını belirleyen en önemli faktörlerden biri, nöronun aktivasyon fonksiyonudur. Biyolojik nöronda, toplamda gelen sinyal belli bir eşik değeri aştığında tepki oluşturduğu biliniyor. Yapay nöronlarda da bu işlem bir aktivasyon fonksiyonu

yardımıyla sağlanır. Bu fonksiyon literatürde sıkıştırma veya eşik fonksiyonu olarak da adlandırılır. Çizelge 4.3 en çok kullanılan aktivasyon fonksiyonlarını göstermektedir. İlgilenilen problem türüne ve ağ yapısına göre farklı aktivasyon fonksiyonları kullanılabilir (Hamzaçebi, 2011).

Çizelge 4.3. Aktivasyon fonksiyonları

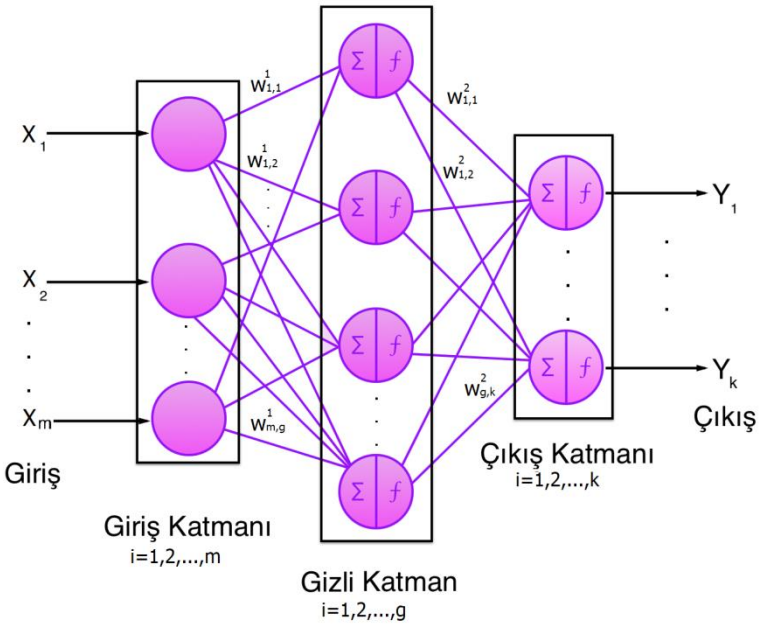
Fonksiyon adı	Fonksiyonun Formülasyonu
Özdeşlik Fonksiyonu	$g(x) = x$
Basamak Fonksiyonu	$g(x) = \begin{cases} b & x \geq c \\ a & x < c \end{cases}$
Rampa Fonksiyonu	$g(x) = \begin{cases} b & x > d \\ a & x < c \\ a + \frac{(x-c)(b-a)}{d-c} & degilse \end{cases}$
Sigmoid Fonksiyou	$g(x) = \frac{1}{1+e^{-x}}$
Hiperbolik Tanjant	$g(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$
Gaus	$g(x) = e^{\frac{(x-c)^2}{r^2}}$

YSA sahip olduğu özelliklerden dolayı alışılagelmiş bilgi işleme yöntemlerinden farklılık göstermektedir. Hatta sahip olduğu bazı özellikler bakımından birçok yönteme göre daha sağlıklı sonuçlar vermektedir. Bu özelliklerden bazıları paralellik, hata toleransı, öğrenilebilirlik ve gerçekleme kolaylığı olarak tanımlanabilir (Şen, 2004).

Yapay sinir ağları bir programcının geleneksel yeteneğini gerektirmeyen, kendi kendine öğrenme düzenekleridir. Bu ağlar öğrenmenin yanı sıra ezberleme ve bilgiler arasında ilişkiler oluşturma yeteneğine de sahiptir (Elmas, 2003).

Bu çalışmada Çok Katmanlı YSA modeli kullanılmıştır. Giriş ve çıkış verisi arasında doğrusal bir ilişki olmadığında, öğrenme işlemini gerçekleştirebilmek için giriş ve çıkış katmanları arasına, bir veya daha fazla sayıda gizli katman yerleştirilirse elde edilen modele çok katmanlı yapay sinir ağı modeli denir. Her bir gizli katmanda bulunan nöron sayısı, modeli oluşturan kişi tarafından belirlenir. Şekil 4.12'de tek gizli katmana sahip, çok katmanlı bir yapay sinir ağı modeli verilmiştir. Amaç ağ girişine uygulanan veriye karşılık gelen çıkışın üretilmesini sağlayan ağırlık bulunmasıdır. Bu işleme ağın eğitilmesi denir. Modeldeki ağırlıkların, çıkış hatasına göre sürekli güncellenerek en uygun değerinin öğrenilmesi için kullanılan eğitim algoritmalarından

bir tanesi Hatanın Geriye Yayılım Algoritmasıdır. Şekil 4.12'de gösterildiği gibi tek bir gizli katmana sahip olan bir YSA ağı için geriye yayılım algoritmasının sözde kodu aşağıda verilmiştir (Mitchell, 1999).



Şekil 4.12. Tek gizli katmana sahip, çok katmanlı bir yapay sinir ağı modeli

Geriye_Yayılım (*X,Y,Og,m,g,k*)

Girişler:

X : p adet veriden oluşan Eğitim Kümesi: $X_j = x_1,x_2,...,x_m \quad , j=1,2,...,p$

Y : p adet hedef çıkış kümesi: $Y_j = y_1,y_2,...,y_k, j=1,2,...,p$

Og: Öğrenme Oranı

m: Giriş katmanındaki nöron sayısı

g: Gizli katmandaki nöron sayısı

k: Çıkış katmanındaki nöron sayısı

Çıkış:

w: öğrenilmiş ağırlık değerleri

1. Ağdaki her bir ağırlığa rastgele küçük bir değer ata.
2. Durdurma kriteri sağlanasıya kadar devam et.

2.1 Eğitim kümesindeki her bir örnek için.

2.1.1 Giriş eğitim verisinin ileri doğru yay.

a. Gizli katmandaki her bir nöronun girişini hesapla: $net^1_{p,j}$

// *f*: Kullanılan aktivasyon fonksiyonu.

b. Gizli katmandaki her bir nöronun çıkışını hesapla: $o^1_{p,j} = f(net^1_{p,j})$

c. Çıkış katmandaki her bir nöronun girişini hesapla: $net^2_{p,k}$

d. Çıkış katmandaki her bir nöronun çıkışını hesapla: $o_{p,k} = f(net^2_{p,k})$

2.1.2 Hatayı hesapla ve geriye doğru yayarak ağırlıkları güncelle:

a. Gizli katman ile çıkış katmanı arasındaki ağırlıkların değişimini hesapla:

$$\Delta w^2_{k,j} = Og * (t_{p,k} - o_{p,k}) * f'(net^2_{p,k}) * o^1_{p,j} \quad (4.17)$$

b. Giriş katman ile gizli katman arasındaki ağırlıkların değişimini hesapla:

$$\Delta w^1_{j,i} = Og * \Sigma_k [(t_{p,k} - o_{p,k}) * f'(net^2_{p,k}) * w^2_{k,j}] * f'(net^1_{p,j}) * x_{p,i} \quad (4.18)$$

c. Ağırlık değişimlerini, mevcut ağırlıklara ekleyerek güncel ağırlık değerlerini hesapla.

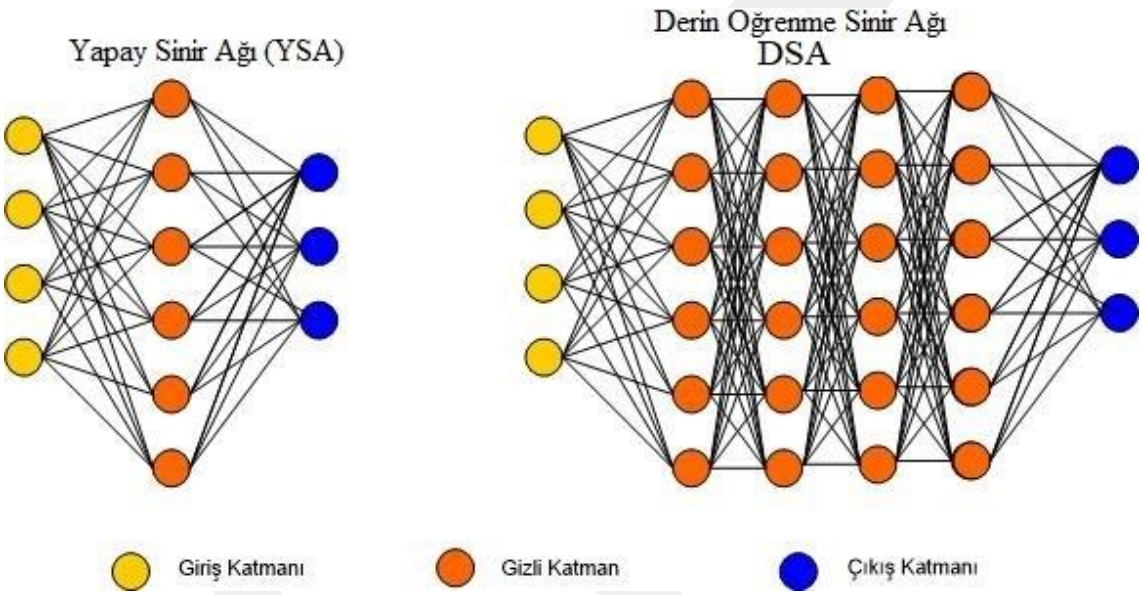
4.3.2. Derin Öğrenme

Derin Öğrenme, doğru kestirim yapabilmek için geliştirilen en popüler yaklaşımlardan biridir. YSA bağlamında ilk kez "Derin Öğrenme" (Deep Learning) ifadesi 2000 yılında Igor Aizenberg ve ark. tarafından tanıtılmıştır (Aizenberg ve ark., 2000).

Bu yaklaşım çoklu soyutlama yapısı ile verinin temsillerini öğrenmek için bir araya getirilmiş çoklu işleme katmanlarından oluşur (Hinton, 2007). Her bir katman, önceki katmandaki çıktıyı girdi olarak alır (Deng ve Yu, 2014).

Derin derken Derin Öğrenmenin bir takım derin bilgiler elde etmesi değil, birbirini takip eden gösterim katmanları ifade edilmektedir. Modeldeki katman sayısı modelin katman derinliğini oluşturmaktadır. Modern Derin Öğrenme modelleri, onlarca hatta yüzlerce birbirini takip eden katmanlar içermektedir. Oysa diğer makine öğrenme algoritmaları, genelde bir veya iki katmandan oluşur (Chollet, 2017).

Derin Öğrenme genel itibariyle yapay sinir ağına çok benzese de en büyük farkı arttırılmış katman sayısı (Şekil 4.13) ve ezberlemeyi önlemek için kullanılan *İletim Sönümü* (*ing. DropOut)* (Srivastava ve ark., 2014) yöntemidir. Bu yöntem sayesinde eğitim yapılırken ağdaki bazı düğümler gelişi güzel kaldırılarak ezberlemenin önüne geçilmesi amaçlanmaktadır. Ayrıca Derin Öğrenme yöntemleri, elle çıkarılan veriyi en iyi temsil eden hiyerarşik özellik çıkarımı için etkin algoritmalar kullanmaktadır (Srivastava ve ark., 2014). Bu yönüyle YSA'dan daha az ön işlem gerektirir.



Şekil 4.13. YSA ve DSA mimarileri

Derin Öğrenme birçok alanda başarıyla uygulanmaktadır. Görüntü ve örüntü tanımaya yönelik ilk model, sıkıca kırpılmış ve oldukça küçük görüntülerdeki tekil nesneleri tanıyabilmekteydi (Rumelhart ve ark., 1986b). Bu tarihten sonra sinir ağlarının işleyebildiği görüntünün büyüklüğü kademeli bir şekilde artarak aynı zamanda fotoğrafta tanınacak nesnenin etrafı kırpılmadan tanıma da artık mümkün olmuştur (Krizhevsky ve ark., 2012).

En büyük nesne tanıma yarışması, her yıl yapılan ImageNet Büyük Ölçekli Görsel Tanıma (*ImageNet Large Scale Visural Recognition Challange, ILSVCR*) yarışmasıdır. Bu yarışmanın 2012'de düzenlenen üçüncü yılında, her modelin ürettiği en olası 5 kategorinin doğru kategoriyi içerip içermediğine bağlı olarak en iyi 5 hatayı %26.1'den yüzde 15.3'e düşüren Krizhevsky ve arkadaşlarının (Krizhevsky ve ark., 2012) birinci olmalarıyla Derin Öğrenmede büyük yükseliş görülmüştür. Konuşma tanımada Derin Öğrenmenin kullanımındaki artış ise daha eskilere 1990'lı yıllara

dayanır. Konuşma tanımadaki hata miktarı ancak 2000'li yıllarda duraksamaya başlamış fakat Derin Öğrenmenin konuşma tanımada kullanılmasıyla (Deng ve ark., 2010; George ve ark., 2010; F. ve ark., 2011) birlikte hata oranlarında büyük düşüşler elde edilmiştir. Hatta bazı hata oranları yarıya indirilmiştir.

Sesten duygu tanıma konusunda Derin Öğrenmenin kullanımı son bir kaç yılda artış göstermiştir. Albornoz ve ark. (2014) Alman duygu veri tabanı üzerinde Boltzmann Makinesi ve Derin İnanç Ağı (ing. *Deep Belief Network)* uygulayarak gerçekleştirdikleri uygulamada %8.76 iyileştirme gözlemlemişlerdir (Albornoz ve ark., 2014).

Bu çalışmada kullanılan Derin Öğrenme Mimarileri şunlardır:

- Derin Sinir Ağları - (*ing. Deep Neural Networks* - DSA)
- Evrişimsel Sinir Ağı - (*ing. Convolutional Neural Network* - ESA)

4.5.2.1 Derin Sinir Ağları (*Deep Neural Networks* - DSA)

DSA mimarisi aynı zamanda "Derin İleri Besleme Ağları" (ing. *Deep Feed-forward Networks*), Çok Katmanlı Algılayıcılar (ing. *Multilayer Perceptrons - ÇKPs*) olarak da bilinir (Bengio, 2009; Şeker ve ark., 2017). İleri besleme ağlarının amacı bir f^* fonksiyonunun yakınsamasını hesaplamaktır. Örneğin bir sınıflandırıcı için x girdisi $y=f^*(x)$ kullanılarak, y kategorisi değerine eşitlenir. İleri besleme ağı, $y=f(x;\ \theta)$ eşlemesini tanımlar ve en iyi fonksiyon yaklaşımını sağlayan θ parametresinin değerini öğrenir (Şeker ve ark., 2017).

Her bir gizli katman bir önceki katmanın çıktılarını girdi olarak alır. Giriş değeri Denklem (4.19) da ifade edildiği gibi, bir önceki katmanın ağırlıklı çıktısı ile bias değerinin toplamıdır. Sonuç doğrusal olmayan bir denklem olan Denklem (4.20)'den geçirilerek elde edilir (Fayek ve ark., 2017).

$$h^{(l)} = y^{(l-1)}W^{(l)} + b^{(l)} \quad (4.19)$$

$$y^{(l)} = \varphi(h^{(l)}) \quad (4.20)$$

Burada l (1,..,L) ye kadar katmanları; h(l) $\in$ R^{n_0} l. katmanın ön aktivasyon fonksiyonunu; y(l−1) $\in$ R^{n_i} (l-1). katmanın çıktısını ve l. katmanın girişini; W(l) $\in$

$R^{n_i x\, n_0}$ l. katmanın öğrenilebilen ağırlık vektörünü; b(l) ∈ R^{n_0} l. katmanın öğrenilebilen bias vektörünü; y(l) ∈ R^{n_0} l. katmanın çıkışını; y(0) giriş; y(L) çıkış ; φ ise doğrusal olmayan aktivasyon fonksiyonunu ifade etmektedir (Fayek ve ark., 2017).

İleri besleme ağları, makine öğrenmesi uygulayıcıları içinde çok önemli bir yer tutar. Çünkü birçok önemli ticari uygulamanın temelini oluştururlar. Örneğin evrişimsel ağlar, görüntülerden nesne tanıma için özelleştirilmiş bir ileri besleme ağı türüdür. Ayrıca ileri besleme ağları çoğu doğal dil uygulamasının gücünü sağlayan yinelemeli ağlar yolunda atılan önemli bir kavramsal adımdır (Şeker ve ark., 2017).

4.5.2.2. Evrişimsel Sinir ağı (ESA / ConvNets)

Evrişimsel Sinir ağı (ing. *Convolutional Neural Networks - ESA*) veya ESA olarak da bilinen Evrişimsel Ağlar (Yann, 1989), ızgara benzeri bilinen bir topolojiye sahip veriyi işlemek için kullanılan özel bir tür sinir ağıdır. Örnekler, düzenli zaman aralıklarında örnekler alan bir boyutlu bir ızgara olarak düşünülebilecek zaman dizisi verisinin ve iki boyutlu bir piksel olarak düşünülebilecek görüntü verisini içerir. Evrişimsel ağlar pratik uygulamalarda muazzam derecede başarılı olmuştur. "Evrişimsel sinir ağı" adı ağın, evrişim (*convolution*) adı verilen bir matematiksel işlemi kullandığını göstermektedir. Evrişim özel bir tür doğrusal işlemdir. Evrişimsel ağlar basitçe, katmanlardan en az birinde genel matris çarpımı yerine evrişim kullanan sinir ağlarıdır (Şeker ve ark., 2017).

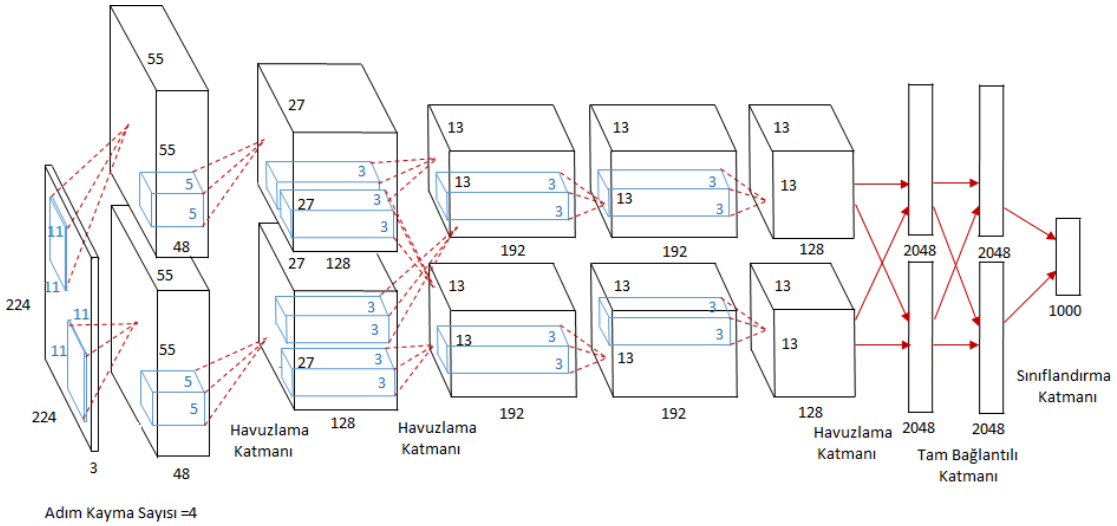
ESA, en az ön işleme ile görsel verileri tanımak için kullanılır. ESA, bir veya daha fazla konvolüsyonel katman, alt örnekleme (ing. subsampling) katmanı ve bunun ardından standart çok katmanlı bir sinir ağı gibi bir veya daha fazla tamamen bağlı katmandan oluşur (LeCun ve ark., 2015; Şeker ve ark., 2017).

Bu mimari kullanılarak; LeNet-5, AlexNet (2012), ZF NET, (2013), GoogleNet/Inception(2014), VGGNet (2014), ResNet(2015), Sınırlı Boltzman Makinelri - (Restricted Boltzmann Machines-RBM), Derin inanç ağları – (ing. Deep Belief Network) modelleri geliştirilmiştir.

AlexNET ESA mimarisi kullanılarak geliştirilen ve yaygın olarak kullanılan bir modeldir. AlexNet mimarisi 5 Evrişim Katmanı ve 3 tam bağlantılı katman (ing. Fully Connected layer) olmak üzere toplam 8 katmandan oluşmaktadır. Doğrusal olmayan

ReLU fonksiyonu evrişim ve tam bağlantılı katmanın tamamının çıkışına uygulanır (Şekil 4.14).

Genel olarak sinir ağlarının temel sorunlarından biri aşırı öğrenmedir (ing. overfitting). Aşırı öğrenmeyi önlemenin başında veri arttırma gelmektedir. Ayrıca AlexNET mimarisinde tam bağlantılı katmandaki aşırı öğrenmeyi önlemek için bir regülasyon metodu olan "iletim sönümü" (ing. dropout) geliştirilmiştir.



Şekil 4.14. AlexNET Mimarisi

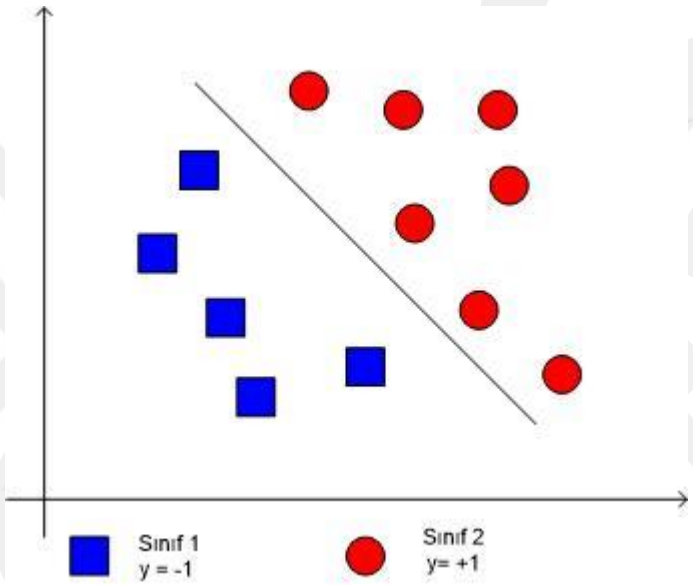
4.3.3. Destek Vektör Makineleri (DVM)

Bu çalışmada Duygu sınıflandırmada kullanılan sınıflandırma yöntemlerinden biri de Destek Vektör Makineleri (DVM) adını taşımaktadır. Bu yöntem sınıflandırmayı doğrusal ya da doğrusal olmayan bir fonksiyon yardımıyla yerine getirir. DVM danışmanlı sınıflandırma yöntemlerinden bir tanesi olup veriyi birbirinden ayırmak için en uygun fonksiyonun tahmin edilmesi esasına dayanır (Özkan ve Erol, 2015).

DVM'nin amacı eğitim verisini iki veya daha fazla sınıfa ayırmak için en iyi sınıflandırıcı fonksiyonu bulmaktır. DVM sinyal işleme, ses tanıma, örüntü tanıma gibi birçok alanda sınıflandırıcı olarak kullanılmaktadır (Davy ve ark., 2002).

DVM yöntemi ile sınıflandırmada iki durum söz konusudur. Bunlar verilerin doğrusal olarak ayrılabilmesi ve verilerin doğrusal olarak ayrılmaması durumudur.

a) Verilerin doğrusal olarak ayrılabilmesi : D veri kümesinin (x_1,y_1), $(x_2,y_2),\ldots,$ (x_n,y_n) biçiminde olduğu varsayılırsa $y \in (-1,1)$ 'dir ve *n* eleman sayısıdır. Amaç bu verileri doğrusal olarak ayırabilmektir. Çok boyutlu uzayda bu doğruların yerini hiper düzlemler alır. İki sınıflı doğrusal ayrılabilen verileri ayıracak birden fazla hiper düzlem olabilir. DVM kendisine en yakın noktalar arası uzaklığı maksimum yapan hiper düzlemi bulmayı amaçlar. Şekil 4.15'da iki boyutlu uzayda doğrusal olarak ayrılabilen veriler gösterilmiştir. Denklem 4.21 bu hiper düzlemin altında kalan noktaları ifade ederken Denklem 4.22 hiper düzlemin üstünde kalan noktaları ifade etmektedir.



Şekil 4.15. İki boyutlu uzayda doğrusal olarak ayrılabilen verilerin görünümü

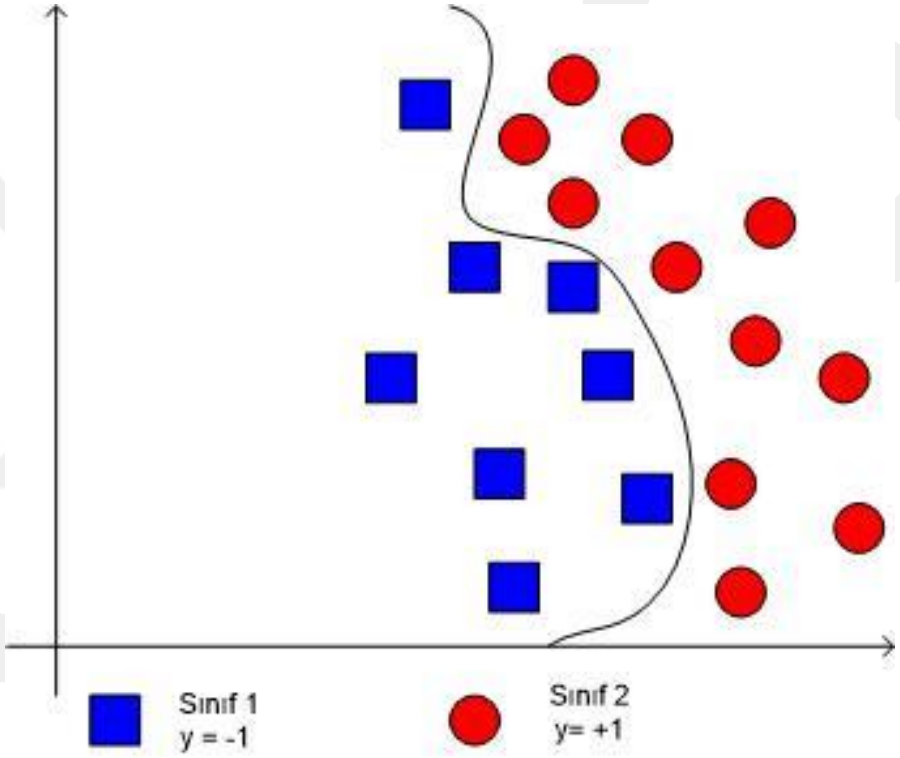
$$w.x_i + b \geq -1 \tag{4.21}$$

$$w.x_i + b \geq +1 \tag{4.22}$$

$$f(x) = sign\left(\sum_{i=1}^{k} \lambda y_i (x.x_i) + b\right) \tag{4.23}$$

Sonuçta doğrusal olarak ayrılabilen iki sınıflı bir problem için karar fonksiyonu Denklem 4.23 şeklinde yazılabilir. Denklemdeki y_i destek vektörler, x_i sınıf etiketini, *x* test gözlemini, λ ve *b* optimizasyon parametrelerini, *w* ağırlık vektörünü (hiper düzlemin normali), *k* ise destek vektör sayısını ifade etmektedir (Kavzoğlu ve Çölkesen, 2010).

b*) Verilerin doğrusal olarak ayrılmaması durumu:* Ses verileri de dâhil birçok verinin doğrusal olarak ayrılması mümkün değildir (Şekil 4.16). Bu tür durumlarda eğitim verilerinden bir kısmının optimum hiper-düzlemin diğer tarafında kalmasından kaynaklanan problem pozitif bir yapay değişkenin *(ξ_i)* tanımlanması ile çözülebilir. Sınırın maksimum hale getirilmesi ve yanlış sınıflandırma hatalarının minimum hale getirilmesi arasındaki denge pozitif değerler alan *C* ile gösterilen bir düzenleme parametresi ile kontrol edilebilir (Cortes ve Vampik, 1995; Kavzoğlu ve Çölkesen, 2010).



Şekil 4.16. İki boyutlu uzayda doğrusal olarak ayrılamayan verilerin görünümü

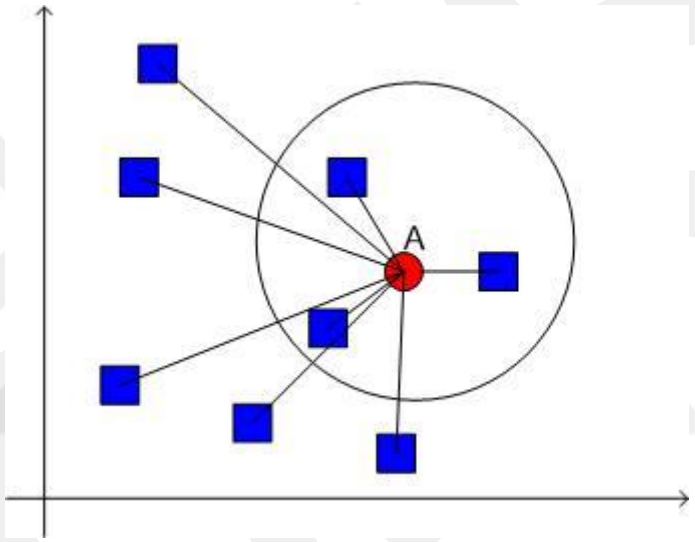
Doğrusal olarak ayrılamayan veriler için karar fonksiyonu yapay değişkende dikkate alınarak elde edilen denklem, Denklem 4.24 ‘de verilmiştir.

$$y_i(w.x_i + b) - 1 + \xi_i \geq 0 \qquad (4.24)$$

4.3.4. K - en Yakın Komşu Algoritması (kNN)

K - en Yakın Komşu algoritması bir örnek tabanlı öğrenme biçimidir, yani yeni gözlemler sınıflanırken hafızada tutulan bir veri kümesi kullanılır. Bu hafızadaki veri kümesinde hedef değişkenin alacağı değerler bellidir (Akküçük, 2011).

kNN örnek kümedeki gözlemlerin her birinin, sonradan belirlenen yeni gözlem değerine uzaklıklarının hesaplanması ve en küçük uzaklığa sahip *k* sayıdaki gözlemin seçilmesi esasına dayanır. Şekil 4.17'de verilen bir *A* noktasına en yakın *k=3* komşunun nasıl belirlendiği gösterilmiştir (Özkan ve Erol, 2015).



Şekil 4.17. Verilen bir A noktasına en yakın k=3 komşunun belirlenmesi

kNN algoritması aşağıdaki gibidir (Akküçük, 2011).

- Öncelikle *k* sayısı belirlenir, bu *k* sayısı yeni gözlem sınıflandırmak için kaç adet en yakın gözlemin kullanılacağını belirler.
- Yeni gözlemde, belirlenen değişkenler üzerinde hesaplanan mesafeye göre en yakın *k* kayıt incelenir. Burada kullanılan mesafe Öklid ya da tercih edilen başka bir mesafe türü olabilir. Denklem 4.25'de x_i sınıf etiketini, x test gözlemini, $a_r(x)$ r. örneğin özniteliği olacak şekilde,

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \tag{4.25}$$

- Basit bir oylama yöntemiyle yeni gözlem gereken sınıfa ayrılır. Örneğin *k* üç olarak seçilmiş ise kayıtlardaki en yakın üç ya da daha fazla kayıt hangi sınıfa ait ise yeni gözlem de o şekilde sınıflandırılır.

4.3.5. Sade Bayes Sınıflandırıcı (SB)

En iyi bilinen Bayes öğrenme metodu, Sade Bayes (SB) öğrenmedir. Örneğin, *X* sınıf üyeliği bilinmeyen veri örneği olsun ve $X= (x_1, x_2, \ldots, x_n)$ nitelik değerlerinden oluşsun. Ele alınan sınıf niteliğinde *m* sınıf oluştuğu varsayılırsa, $C_1, C_2, \ldots, C_m$ sınıf değerleri olsun. *X* niteliği seçildiğinde buna bağlı olarak C_i sınıflarının olasılığı elde etmek amacıyla Denklem 4.26 kullanılır (Özkan ve Erol, 2015).

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{4.26}$$

C burada sınıfı, *P(C|x)* terimi ise verilmiş *x* için *i. C*'nin koşullu olasılığını ifade eder. Bayes teoremi eğer *x* gözlemlenmiş ise, *c* gözleminin ne şekilde güncelleştirilebileceğini ortaya çıkartır. Hesaplamalardaki işlem yükünü azaltmak üzere Denklem (4.26) bağıntısında yer alan $P(X|C_i)$ olasılığı için basitleştirme yoluna gidilebilir. Bunun için, örneğe ait x_i değerinin birbirinden bağımsız olduğu kabul edilerek Denklem 4.27 kullanılabilir (Özkan ve Erol, 2015).

$$P(X|C_i) = \prod_{k=1}^{n} \mathrm{P}(x_k|C_i) \tag{4.27}$$

Bilinmeyen örnek *X*' i sınıflandırmak için Denklem 4.27'de $P(C_i|X)$ tanımında yer alan paydalar birbirine eşit olduğundan sadece pay değerlerini karşılaştırmak yeterlidir. Bu değerler içinden en büyük olanı seçilerek bilinmeyen örneğin bu sınıfa ait olduğu belirlenmiş olur (Denklem 4.28) (Özkan ve Erol, 2015).

$$\arg max_{c_i} \; P(X|C_i)\, P(C_i) \tag{4.28}$$

Sonrasal olasılıkları kullanan yukarıdaki ifade, *en büyük sonrasal sınıflandırma yöntemi* (Maximum A Posteriori Classification = MAP) olarak da bilinir. O halde 4.28

den yararlanılarak, 4.25 ifadesi yerine Bayes sınıflandırıcı olarak Denklem 4.29 kullanılabilir (Özkan ve Erol, 2015).

$$C_{MAP} = \arg max_{c_i} \prod_{k=1}^{n} \mathrm{P}(x_k|C_i) \tag{4.29}$$

4.4. Performans Değerlendirme Kriterleri:

Sınıflandırma Doğruluğu (Doğruluk Oranı) (SD): Model başarımı ölçümünde kullanılan en bilinen ve temel yöntem sınıflandırıcıya ait doğruluk oranıdır. Bazı performans değerlendirilmelerinde Ağırlıksız Doğruluk (Unweighted Accuracy - UA) olarak da kullanılır. Sınıflandırma Doğruluğu formülasyonu Denklem 4.30'da verilmiştir. Burada DP, Doğru sınıflandırılmış Pozitif örnekleri, DN, Doğru sınıflandırılmış Negatif örnekleri; YP Yanlış sınıflandırılmış Pozitif Örnekleri, YN ise Yanlış sınıflandırılmış Negatif örnekleri ifade etmektedir.

$$SD = \frac{\mathrm{DP} + \mathrm{DN}}{\mathrm{DP} + \mathrm{DN} + \mathrm{YP} + \mathrm{YN}} \tag{4.30}$$

Kappa İstatistiği (κ): Kappa katsayısı 0-1 değerleri arasında değişen istatistiksel bir değerdir. Buna göre κ = 1 tam uyumu, κ = 0 ise zayıf uyumu temsil etmektedir. Kappa değeri 0.4'ün üzerinde ise kappa istatistiği kabul edilebilir bir anlaşmayı ifade eder (Landis ve Koch, 1977).

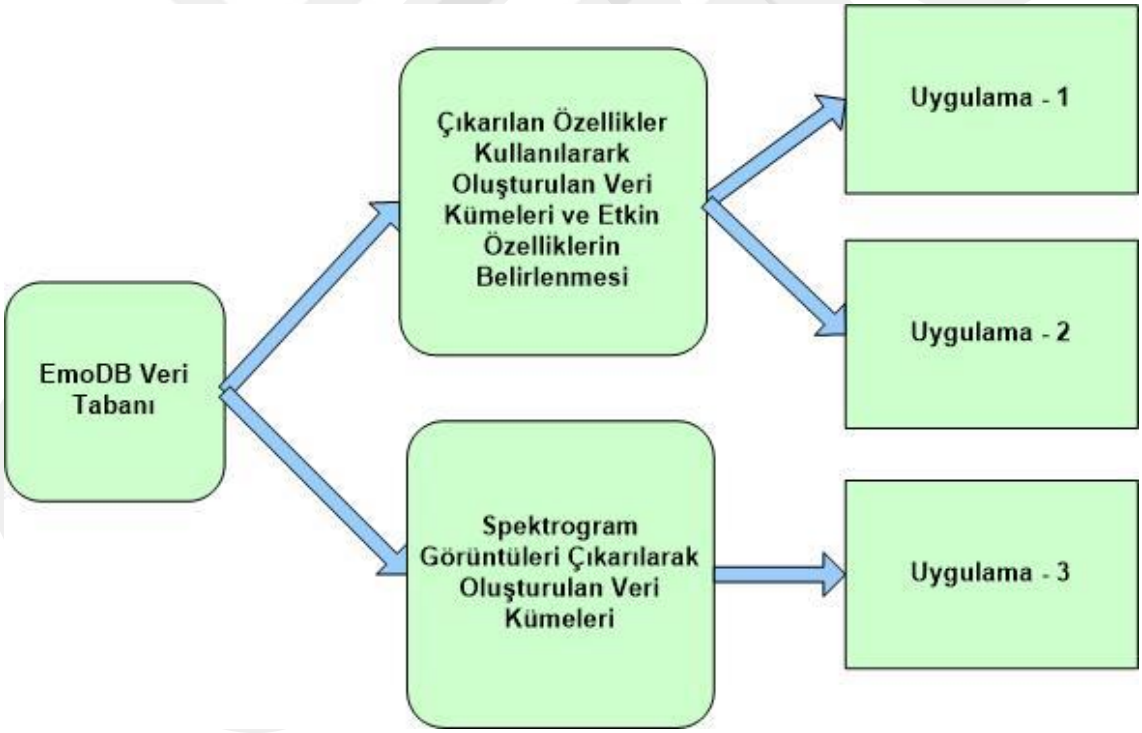
$$P_f = \frac{[(DP + YP)(DP + YN)] + [(YN + DN)(YP + DN)]}{(DP + YP + DN + YN)^2} \tag{4.31}$$

$$\kappa = \frac{\mathrm{SD} - P_f}{1 - P_f} \tag{4.32}$$

Denklem 4.32 Kappa istatistiğini ifade etmektedir. Burada P_0 kabul edilen oranı, P_f kabul edilmesi beklenen oranı ifade etmektedir (Nizam ve Akın, 2014).

5. UYGULAMA

Bu tez çalışmasında ilk olarak EmoDB veri tabanından farklı özellik çıkarım yöntemleri ve farklı sayıda duygu kullanılarak 20 tane veri kümesi oluşturulmuştur. Veri kümeleri herhangi bir özellik seçimi yapılmadan direk sınıflandırıcıya verilmiştir. Elde edilen sonuçlardan en etkin 12 veri kümesi tespit edilmiştir. Bu veri kümeleri kullanılarak 2 uygulama gerçekleştirilmiştir. Bunlardan birincisi "Etmen Tabanlı Otomatik Özellik Seçimi (*ABAfs*)"ve ikincisi "Bulanık C - Ortalama (BCO) ile Duygu Tanıma Yaklaşımı" dır. Daha sonra çalışmada ses sinyallerinden özellik seçimi yapılmadan spektrogram görüntülerinin elde edilmesi ile oluşturulan veri kümeleri DSA ile sınıflandırılmıştır. Bu uygulama ise "Derin Öğrenme Algoritmaları ile Ses Verilerinden Duygu Tanıma" isimli çalışmadır. Şekil 5.1'de uygulama adımları gösterilmiştir.



Şekil 5.1. Uygulama işlem adımları

5.1. Ön Çalışmalar

5.1.1. Uygulama 1 ve Uygulama 2 için Veri Kümelerinin Belirlenmesi ve Ön İşlemler

Uygulama amacı ile "Berlin DB" veri tabanı kullanılmıştır. EmoDB veri tabanında yedi duyguya ait ses verileri mevcuttur. Ayrıca Çizelge 5.1'de çalışmada bu duyguların uygulamalar boyunca hangi kısaltmalar ile temsil edildiği de gösterilmiştir.

Çizelge 5.1. Duygular ve uygulamalarda kullanılan kısaltmaları

Duygu Adı	Duygu Kısaltması
Kızgın	K
Sıkıntılı	S
Nefret	N
Korku	Ko
Mutlu	M
Üzgün	Ü
Normal	No

Çalışmada duygu verileri üzerine spektral ve prozodik özellik seçme yöntemleri uygulanarak özellikler hesaplanmıştır. Çıkarılan özellikler ve bu özelliklerden oluşturulan veri kümeleri Şekil 5.2'de gösterilmektedir.

Uygulamada ilk olarak verilerden MFCC özellikleri çıkartılmıştır. Her bir veriden farklı sayıda Mel Katsayıları elde edilmiştir.

MFCC katsayılarının nasıl çıkarıldığı bir örnekle açıklanmak istenirse örneğin 16 tane Mel Katsayısının çıkarıldığı varsayılsın. Rastgele seçilmiş bir segment için MFCC özelik çıkarma adımlarındaki matris boyutları Şekil 5.3'de gösterilmiştir. Şekilde görüldüğü gibi, segmentin boyutu ilk etapta 24933 iken, Mel Katsayıları çıkarıldıktan sonra boyutu 388X16 olarak elde edilmiştir. Yani 16 tane Mel Katsayısı elde edilmiştir. Son olarak Şekil 5.3'de verilen tamamlayıcı istatistikler her bir Mel Katsayısı için hesaplanarak 7X16'lık yeni sabit boyutlu özellik vektörü elde edilmiştir. Bu şekilde spektral veriden yeni özellikler elde ederek, her bir 7X16 boyutunda MFCC katsayıları elde edilmiştir.

MFCC Özellikleri
Spektral Özellikler

- ***Kullanılan İstatistiksel Hesaplamalar***
- **Maksimum Değeri**
- **Minimum Değeri**
- **Ortalama Değeri**
- **Standart Sapma Değeri**
- **Çarpıklık Değeri**
- **Basıklık Değeri**
- **Medyan Değeri**

Oluşurulan Veri Kümeleri
8 adet MFCC katsayısı - VeriKümesi1
16 adet MFCC katsayısı - VeriKümesi2
32 adet MFCC katsayısı - VeriKümesi3
60 adet MFCC katsayısı - VeriKümesi4

Doğrusal Öngörüm Katsayıları (LPC)
Spektral Özellikler

- **LPC Katsayıları**
-
-

Oluşurulan Veri Kümeleri
16 adet LPC katsayısı - VeriKümesi5
112 adet LPC katsayısı - VeriKümesi6

Temel Periyot (T_0), Temel Frekans (F0), Perde Frekansı
Prosodik Özellikler

- ***Kullanılan İstatistiksel Hesaplamalar***
- **Maksimum Değeri**
- **Minimum Değeri**
- **Ortalama Değeri**
- **Standart Sapma Değeri**
- **Çarpıklık Değeri**
- **Basıklık Değeri**
- **Medyan Değeri**

Oluşurulan Veri Kümeleri
24 adet Temel Frekans katsayısı VeriKümesi7

AR Parametreleri
Spektral Özellikler

- **AR parametreleri**
-

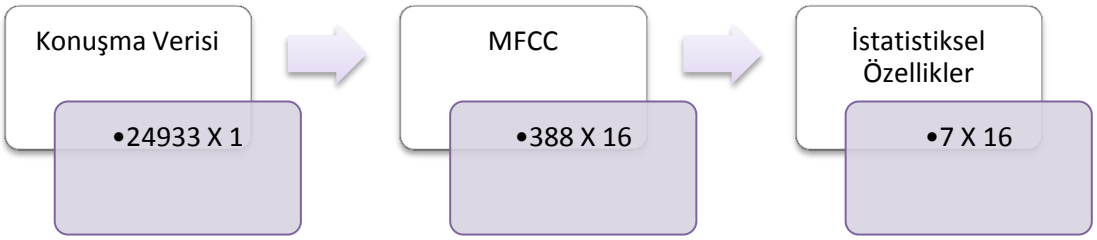
Oluşurulan Veri Kümeleri
32 adet AR parametresi VeriKümesi8

Ayrık Dalgacık Dönüşümü (DWT)
Spektral Özellikler

- **Maksimum Değeri**
- **Maksimum Frekans Değeri**
- **Minimum Değeri**
- **Minimum Frekans Değeri**
- **Ortalama Değeri**
- **Standart Sapma Değeri**
- **Çarpıklık Değeri**
- **Basıklık Değeri**

Oluşurulan Veri Kümeleri
32 adet DWT parametresi VeriKümesi9
56 adet DWT parametresi VeriKümesi10

Şekil 0.2. Çıkarılan özellikler ve oluşturulan veri kümeleri



Şekil 0.3. Rasgele seçilmiş bir segment için MFCC özellik çıkarımı

Her bir segmentin boyutu farklı olduğundan, MFCC'den elde edilen katsayılar da farklı boyuttadır. Bu durumda katsayılar sınıflandırmaya giriş olarak verilirken sorun oluşturacağından ikinci adım olarak her bir katsayı için tamamlayıcı istatistik değerleri kullanılarak boyut eşitleme işlemi yapılmıştır. Şekil 5.2'de gösterilen maksimum, minimum, ortalama, standart sapma, çarpıklık, basıklık ve medyan tanımlayıcı istatistikleri kullanılmıştır.

Oluşturulan matris satır matrisi haline dönüştürülerek sonuçta her bir segment için 7*16=112 tane özellikten oluşan yeni bir veri kümesi elde edilmiştir. EmoDB Veri tabanını erişim tarihi olan 2 Nisan 2014 tarihindeki toplam veri sayısı 527 olduğundan bu çalışmada 527 segment kullanılmıştır.

Şekil 5.3'de de gösterildiği gibi, çıkarılan 8 adet MFCC katsayısından oluşan veri kümesi VeriKümesi1, 16 adet MFCC katsayısından oluşan veri kümesi VeriKümesi2, 32 adet MFCC katsayısından oluşan veri kümesi VeriKümesi3 ve 60 adet MFCC katsayısından oluşan veri kümesi VeriKümesi4 olarak tanımlanmıştır.

Uygulamada ikinci olarak LPC katsayıları elde edilmiştir. LPC katsayıları elde edilirken Matlab'ın[1] "LPC" fonksiyonu kullanılmıştır. Bu fonksiyon En Küçük Kareler formülünden kaynaklanan normal denklemi çözmek için Levinson-Durbin özyinelemesini kullanır. LPC fonksiyonunun bu hesaplama yöntemi otokorelasyon olarak da ifade edilir. LPC kullanılarak 16 tane özellik çıkarılarak oluşturulan veri kümesi VeriKümesi5, 112 tane özellik çıkarılarak oluşturulan veri kümesi VeriKümesi6 olarak tanımlanmıştır.

Uygulamada kullanılan bir diğer özellik F0 olarak da bilinen "Temel Frekans" değeridir. Temel Frekans'ın farklı şekillerde hesaplanabildiği Bölüm 4.1.3'de

[1] https://www.mathworks.com/help/signal/ref/lpc.html

anlatılmıştı. Bu uygulamada F0 değeri "Kompleks Kepstrum (KK)" yöntemi ve "Otokorelasyon" yöntemi ile hesaplanmıştır. F0 değeri toplam 24 tane özellik ile temsil edilmiştir. Bu özellikler Çizelge 5.2'de ifade edilmiştir. Çizelge 5.2'de de görüldüğü üzere öncelikle KK yöntemi ile elde edilen özelliklerden yedi (minimum, maksimum, ortalama, standart, medyan, çarpıklık, basıklık) farklı tamamlayıcı istatistiklerle özellikler hesaplanmıştır. Temel Frekans Maksimum ve Minimum Bölgesi için ise beş (ortalama, standart, medyan, çarpıklık, basıklık) farklı istatistik değerler hesaplanarak özellikler elde edilmiştir. Ayrıca otokorelasyon yöntemi ile yedi (minimum, maksimum, ortalama, standart, medyan, çarpıklık, basıklık) adet özellik elde edilmiştir.

Çizelge 0.2. F0 özellikleri

F0 Özellikleri	KK Yöntemi	Temel Frekans minimum değeri
		Temel Frekans maksimum değeri
		Temel Frekans ortalama değeri
		Temel Frekans standart sapması
		Temel Frekans medyan değeri
		Temel Frekans çarpıklık değeri
		Temel Frekans basıklık değeri
		Temel Frekans minimum bölgesinin ortalama değeri
		Temel Frekans minimum bölgesinin standart sapması
		Temel Frekans minimum bölgesinin medyan değeri
		Temel Frekans minimum bölgesinin çarpıklık değeri
		Temel Frekans minimum bölgesinin basıklık değeri
		Temel Frekans maksimum bölgesinin ortalama değeri
		Temel Frekans maksimum bölgesinin standart sapması
		Temel Frekans maksimum bölgesinin medyan değeri
		Temel Frekans maksimum bölgesinin çarpıklık değeri
		Temel Frekans maksimum bölgesinin basıklık değeri
	Otokorelasyon Yöntemi	Temel Frekans minimum değeri
		Temel Frekans maksimum değeri
		Temel Frekans ortalama değeri
		Temel Frekans standart sapması
		Temel Frekans medyan değeri
		Temel Frekans çarpıklık değeri
		Temel Frekans basıklık değeri

Temel frekans maksimum bölgesi: Maksimum bölge temel frekans ortalamasının üstünde kalan bölge olarak tanımlanmıştır.

$$F0_{max-bolge} = \{F0_i | F0_i \geq F0_{ort}\} \quad (5.1)$$

Temel frekans minimum bölgesi: Maksimum bölgenin tersine Temel Frekans ortalamasının altında kalan bölgedir. Temel Frekansa ait özelliklerden oluşturulan veri kümesi VeriKümesi7 olarak tanımlanmıştır.

$$F0_{min-bolge} = \{F0_i | F0_i \leq F0_{ort}\} \quad (5.2)$$

Uygulamada AR parametreleri Yule-Walker ve Burg methodu olmak üzere iki farklı otoregresif model ile tahmin edilmeye çalışılmıştır. Her iki yöntemle de 16 farklı özellik olacak şekilde toplam 32 adet AR parametresi elde edilmiş ve VeriKümesi8 olarak tanımlanmıştır.

Tez çalışmasında çıkarılan bir diğer özellik seçim yöntemi DWT'dir. Dalgacık dönüşüm parametreleri elde edilirken alçak frekans bilgileri yoksa yüksek frekans bilgileri mi kullanılacağının seçilmesi gerekir. Yapılan denemeler sonucunda ses verilerinde duygu tespit ederken alçak frekansların daha çok bilgi barındırdığı gözlemlenmiştir. Bu nedenle çalışmada alçak frekans bilgileri kullanılmıştır. Özelliklerin çıkarımında ana dalgacığın seçimi için literatürde en çok kullanılan Daubechies 3 (db3) filtresi kullanılmıştır (Sharma ve ark., 2019). DWT katsayılarının boyut azaltımında Şekil 5.3'de belirtilen tamamlayıcı istatistikler hesaplanmıştır. Bu değerler katsayılara ait maksimum değer ve maksimum frekans değeri, minimum değer ve minimum frekans değeri, ortalama değeri, standart sapması, basıklık ve çarpıklık değerleridir.

32 özellik alçak frekans bilgilerine dört kez DWT uygulandıktan sonra, oluşan dördüncü dereceden alçak frekans bilgisi ve birinci, ikinci ve üçüncü dereceden elde edilen yüksek frekans bilgisine ait değerlerin tamamlayıcı istatistik hesaplamaları sonucu elde edilmiştir. Oluşturulan bu veri kümesi VeriKümesi9 olarak tanımlanmıştır.

56 ise alçak frekans bilgilerine yedi kez DWT uygulandıktan sonra yine ilk altı dereceden yüksek frekans bilgileri ile yedinci derece DWT'den elde edilen alçak frekans bilgilerinin tamamlayıcı istatistik hesaplamaları sonucu elde edilmiştir. Oluşturulan bu veri kümesi VeriKümesi10 olarak tanımlanmıştır.

Farklı özelliklerin birlikte kullanımıyla VeriKümesi11 ve VeriKümesi12 oluşturulmuştur. VeriKümesi11, 16 tane MFCC, 16 tane LPC, 16 tane F0, 16 tane AR ve 56 tane DWT parametresi kullanılarak oluşturulmuştur. VeriKümesi12 olarak isimlendirilen diğer veri kümesi ise 16 tane MFCC, 32 tane DWT ve 16 tane AR özelliklerinden oluşmaktadır.

Bu aşamaya kadar oluşturulan VeriKümesi1'den VeriKümesi12'ye kadar bütün veri kümeleri EmoDB veri tabanındaki yedi duygunun tamamı kullanılarak oluşturulmuştur. Bu çalışmada aynı zamanda birbiriyle çokça karıştırılan üçerli duygu grupları da kullanılmıştır. İlk duygu grubu Kızgın-Mutlu-Korku (KMKo) duygularından oluşmaktadır. İkinci duygu grubu ise Sıkıntılı-Normal-Üzgün (SNoÜ) duygularından oluşmaktadır. Bu duygu grupları için, daha önce anlatılan VeriKümesi1, VeriKümesi2, VeriKümesi3 ve VeriKümesi4 kullanılarak oluşturulan veri kümeleri Çizelge 5.3'de gösterilmiştir

Uygulama 1 ve Uygulama 2 için oluşturulan bütün veri kümeleri, veri kümesinde kullanılan özellikler ve veri boyutları toplu olarak Çizelge 5.3'de gösterilmiştir.

Çizelge 0.3. Uygulama 1 ve Uygulama 2 için oluşturulan Veri Kümeleri

Veri Kümesi Tanımı	Veri Kümesinde Kullanılan Özellikler	Veri Boyutu
VeriKümesi1	MFCC özellikleri (8 tane)	527 X 56
VeriKümesi2	MFCC özellikleri (16 tane)	527 X 112
VeriKümesi3	MFCC özellikleri (32 tane)	527 X 224
VeriKümesi4	MFCC özellikleri (60 tane)	527 X 420
VeriKümesi5	LPC özellikleri (16 tane)	527 X 16
VeriKümesi6	LPC özellikleri (112 tane)	527 X 112
VeriKümesi7	Temel frekans (24 tane)	527 X 24
VeriKümesi8	AR parametreleri (32 tane)	527 X 32
VeriKümesi9	DWT parametreleri (32 tane)	527 X 32
VeriKümesi10	DWT parametreleri (56 tane)	527 X 56
VeriKümesi11	MFCC özellikleri (16 tane) LPC özellikleri (16 tane) F0 özellikleri (16 tane) AR özellikleri (16 tane) DWT özellikleri (56 tane)	527 X 216
VeriKümesi12	MFCC özellikleri (16 tane) AR özellikleri (16 tane) DWT özellikleri (32 tane)	527 X 160
VeriKümesi1_KMKo	MFCC özellikleri (16 tane)	260 X 56
VeriKümesi2_KMKo	MFCC özellikleri (16 tane)	260 X 112
VeriKümesi3_KMKo	MFCC özellikleri (32 tane)	260 X 224
VeriKümesi4_KMKo	MFCC özellikleri (60 tane)	260 X 420
VeriKümesi1_SNoÜo	MFCC özellikleri (16 tane)	221 X 56
VeriKümesi2_SNoÜ	MFCC özellikleri (16 tane)	221 X 112
VeriKümesi3_SNoÜ	MFCC özellikleri (32 tane)	221 X 224
VeriKümesi4_SNoÜ	MFCC özellikleri (60 tane)	221 X 420

Bu bölümde ayrıca oluşturulan toplam 20 veri kümesi, önerilen yöntemlerin karşılaştırılabilmesi için, hiçbir özellik seçimine tabi tutulmadan sınıflandırılmıştır. Sınıflandırma işlemi, otomatik sınıflandırma için açık kaynak kodlu ve genel bir araç çubuğu olan Weka (Bouckaert ve ark., 2010) programı ile gerçekleştirilmiştir. Gerçekleştirilen uygulamalar 10 çapraz doğrulama yönteminden gelen sınıflandırma doğruluklarının ortalaması alınarak bulunmuştur. Bunun için öncelikle veri kümesinden rastgele verilerden oluşturulan 10 farklı küme elde edilir. On farklı veri kümesinin dokuzu eğitim, biride test verisi olarak kullanılmıştır. Bu işlem on kez tekrarlanarak her bir kümenin en az bir kez test verisi olarak kullanılması sağlanmıştır. Öğrenme katsayısı ve momentum değerleri denemeler sonucunda en iyi olan değerler seçilerek elde edilmiştir. Sınıflandırıcı ve fonksiyonları aşağıda verilmiştir:

YSA için "MultiLayerPercepteron" Algoritması tercih edilmiştir. Öğrenme katsayısı ve momentum değerlerini tespit etmek için bazı denemeler yapılmıştır. Öğrenme katsayısı ve momentum değerini değiştirerek performanstaki değişim gözlemlenmiştir. On yedi denemeye ait sonuçlar Çizelge 5.4'de gösterilmektedir. Çizelge 5.4'den de görüldüğü üzere en yüksek SD değerin öğrenme katsayısı 0.2 iken momentum değerinin 0.8 olduğu görülmektedir.

Öğrenme katsayısı ve momentum değeri belirlendikten sonra gizli katmandaki nöron sayısı değiştirilerek başarıya etkisi araştırılmıştır.

Çizelge 5.4. Sınıflandırma doğruluğunun öğrenme katsayı ve momentumla değişimi

Öğrenme Katsayısı	Momentum	SD
0.2	0.9	73.81
0.2	0.8	74.38
0.2	0.7	72.86
0.2	0.6	74.19
0.2	0.5	72.48
0.2	0.4	72.10
0.2	0.3	73.05
0.2	0.2	72.65
0.2	0.1	72.29
0.9	0.2	72.67
0.8	0.2	73.81
0.7	0.2	73.81
0.6	0.2	73.43
0.5	0.2	73.43
0.4	0.2	74.37
0.3	0.2	72.48
0.1	0.2	72.10

Gizli katman nöron sayısı için kullanılan kısaltmalar Çizelge 5.5'de gösterilmektedir. Gizli katman nöron sayısı varsayılan olarak '*a*' olarak alınmaktadır. Diğer değerler için bakıldığında hesaplama sonuçları Çizelge 5.6'da verilmiştir. Çizelge 5.6'dan de görüldüğü üzere en yüksek sınıflandırma başarısını veren değer %74.76 ile *'i'* parametresiyle belirtilen özellik sayısı olarak alındığındaki sonuçtur. Sonuç olarak gizli katman sayısı özellik sayısı olarak alınmıştır.

Çizelge 5.5. Gizli katman nöron sayısı için kullanılan kısaltmalarının açıklaması

a	Özellik sayısı + sınıf sayısı /2
i	Özellik sayısı
o	Sınıf sayısı
t	Özellik sayısı + sınıf sayısı

Son olarak deneme sayısının (epoch) performansa etkisi araştırılmış ve elde edilen değerler Çizelge 5.7'de verilmiştir. Çizelge 5.7'den de görüldüğü üzere en yüksek SD değeri 10000 deneme sayısı ile elde edilmiştir.

Çizelge 5.6. YSA için SD'nin gizli katmandaki nöron sayısı ile değişimi

Gizli Katman Nöron Sayısı	SD
a	74.38
i	74.76
o	67.74
t	67.74
1	38.33
2	47.24
3	56.17
4	59.96
5	63.94
6	67.36
8	66.60

Çizelge 5.7. YSA SD'nin deneme sayısı ile değişimi

Deneme Sayısı	SD
500	74.76
1000	74.38
10000	75.14
100000	74.38

Sonuç olarak YSA sınıflandırıcı için kullanılan parametreler Çizelge 5.8'de toplu olarak verilmiştir.

Çizelge 0.8. YSA, DVM, kNN ve SB sınıflandırıcı için kullanılan parametreler

YSA	SVM	kNN	SB
Öğrenme Katsayısı: 0.2	Maliyet c: 0.3	Metrikler: Öklit	debug: Doğru -
Momentum Katsayı: 0.8	Sayısal Hassasiyet: 0.001	Sürekli nitelikler: Normalize edilmiş	Modeli eski biçimde görüntüle: Yanlış
Gizli Katman Sayısı: i	Sınıf olasılıklarını tahmin edin: Hayır	Yok olan bilinmeyen değerler yok sayılıyor: Hayır	Kernel tahmin ediciyi kullan: Yanlış
İterasyon sayısı: 10000	Normaliz edilmiş veri: Evet	Komşu sayısı: 7	Denetimli yaklaşım: Yanlış
		Ağırlıklandırma: Mesafeye göre	

DVM için "SMO" (Sequential Minimal Optimization) Algoritması uygulanmıştır. kNN sınıflandırıcı için " IBk " Algoritması uygulanmıştır. Küme sayısı yapılan denemelerle belirlenmiş ve Çizelge 5.8'de belirtilmiştir. SB sınıflandırıcı için " NaiveBayes" algoritması uygulanmıştır. DVM, kNN, SB sınıflandırıcılar için kullanılan parametreler Çizelge 5.8'de toplu olarak verilmiştir.

Tanımlanan her bir veri kümesinin verilen dört farklı sınıflandırıcı (YSA, DVM, kNN, SB) uygulandıktan sonraki sınıflandırma doğrulukları ise Çizelge 5.9'da verilmiştir.

Çizelge 0.9. Veri Kümeleri için elde edilen sonuçlar

Veri Kümesi	Sınıflandırıcı	SD (%)	Kappa İstatistiği
VeriKümesi1	YSA	**66.22**	0.60
	DVM	65.60	0.60
	kNN	65.08	0.58
	SB	60.91	0.53
VeriKümesi2	YSA	**75.14**	0.71
	DVM	72.11	0.67
	kNN	68.69	0.62
	SB	61.29	0.54
VeriKümesi3	YSA	**76.47**	0.72
	DVM	74.19	0.69
	kNN	65.84	0.59
	SB	35.67	0.24
VeriKümesi4	YSA	**78.94**	0.75
	DVM	73.62	0.68
	kNN	68.69	0.62
	SB	42.50	0.32
VeriKümesi5	YSA	**46.30**	0.36
	DVM	40.61	0.28
	kNN	42.32	0.30
	SB	35.67	0.24
VeriKümesi6	YSA	**46.87**	0.37
	DVM	47.06	0.37
	kNN	28.46	0.16
	SB	31.69	0.21
VeriKümesi7	YSA	37.76	0.26
	DVM	**40.03**	0.26
	kNN	34.72	0.22
	SB	30.55	0.17
VeriKümesi8	YSA	**50.47**	0.41
	DVM	43.45	0.32
	kNN	40.23	0.28
	SB	35.48	0.24
VeriKümesi9	YSA	47.25	0.37
	DVM	**52.75**	0.43
	kNN	52.56	0.43
	SB	46.68	0.36
VeriKümesi10	YSA	54.08	0.45
	DVM	**60.53**	0.53
	kNN	50.10	0.41
	SB	49.15	0.40

Veri Kümesi	Sınıflandırıcı	SD (%)	Kappa İstatistiği
VeriKümesi11	YSA	**78.74**	0.75
	DVM	78.37	0.74
	kNN	69.64	0.64
	SB	62.43	0.55
VeriKümesi12	YSA	**74.38**	0.70
	DVM	73.06	0.68
	kNN	70.02	0.64
	SB	63.58	0.57
VeriKümesi1_KMKo	YSA	**74.23**	0.60
	DVM	72.31	0.55
	kNN	70.77	0.52
	NB	70.77	0.52
VeriKümesi2_KMKo	YSA	**74.66**	0.67
	DVM	76.54	0.62
	kNN	76.15	0.62
	SB	69.62	0.51
VeriKümesi3_KMKo	YSA	**81.92**	0.71
	DVM	79.23	0.67
	kNN	73.08	0.56
	SB	66.54	0.81
VeriKümesi4_KMKo	YSA	**84.61**	0.76
	DVM	78.85	0.66
	kNN	73.84	0.58
	SB	70	0.52
VeriKümesi1_SNoÜ	YSA	66.97	0.50
	DVM	**73.30**	0.60
	kNN	72.39	0.58
	SB	69.23	0.53
VeriKümesi2_SNoÜ	YSA	**76.47**	0.64
	DVM	71.49	0.60
	kNN	74.21	0.61
	SB	65.91	0.48
VeriKümesi3_SNoÜ	YSA	**79.64**	0.69
	DVM	76.92	0.65
	kNN	71.49	0.57
	SB	67.42	0.51
VeriKümesi4_SNoÜ	YSA	**78.28**	0,67
	DVM	76.02	0.64
	kNN	75.11	0.63
	SB	69.23	0.53

VeriKümesi1 MFCC özelliklerinden elde edilen 8 özellik çıkarılarak oluşturulmuştur. VeriKümesi1 için en yüksek SD değeri DVM sınıflandırıcı ile %66.60 olarak elde edilmiştir.

MFCC verilerinden 16 özellik çıkarılarak oluşturulan veri kümesi VeriKümesi2'dir. VeriKümesi2 için en yüksek SD değeri %75.14 ile YSA sınıflandırıcısı ile gerçekleştirilmiştir.

VeriKümesi3 MFCC özelliklerinden 32 özellik çıkarılarak oluşturulan bir diğer veri kümesidir. Bu veri kümesi için %76.47 ile en yüksek SD değeri YSA yöntemi ile elde edilmiştir.

MFCC özellik çıkarım yöntemiyle elde edilen verilerden elde edilen son veri kümesi VeriKümesi4'tür. En yüksek SD YSA yöntemi ile %78.94 olarak gerçekleştirilmiştir.

VeriKümesi5 16 tane LPC özellikleri ile oluşturulan bir veri kümesidir. Çizelge 5.9'dan da görüldüğü üzere LPC özellikleri tek başına duyguları sınıflandırmakta yeterli olmamaktadır. Kappa değerleri de 0.4'ün altında olduğundan kabul edilebilir değildir. En yüksek SD değeri ise YSA yöntemi ile %46.30 olarak gerçekleştirilmiştir.

VeriKümesi6 LPC'den 112 tane özellik elde edilerek oluşturulan diğer bir veri kümesidir. Bu veri kümesi LPC özelliklerden elde edilen özellik sayısını arttırarak SD değerinin artıp artmayacağını analiz etmek için oluşturulmuştur. Ama Çizelge 5.9'da belirtildiği gibi en yüksek SD değeri %46.87 olmuştur. Bu da göstermektedir ki VeriKümesi5'te elde dilen en yüksek SD değerinden (%46.30) anlamlı bir farklılık içermemektedir. Bu nedenle LPC özellikleri ile oluşturulan veri kümeleri daha fazla arttırılmamıştır. Ayrıca Kappa değerleri de 0.4'ün altında olduğundan kabul edilebilir değildir.

VeriKümesi7 temel frekansa ait 17 tane özellik çıkarılarak oluşturulmuştur. Elde edilen en yüksek SD değeri %40.03 ile DVM sınıflandırıcı ile gerçekleştirilmiştir. SD değerinden de görüldüğü üzere temel frekansa ait özellikler duygu tanımada tek başına yeterli gelmemektedir. Kappa değerleri bu veri kümesinde de 0.4'ün altındadır.

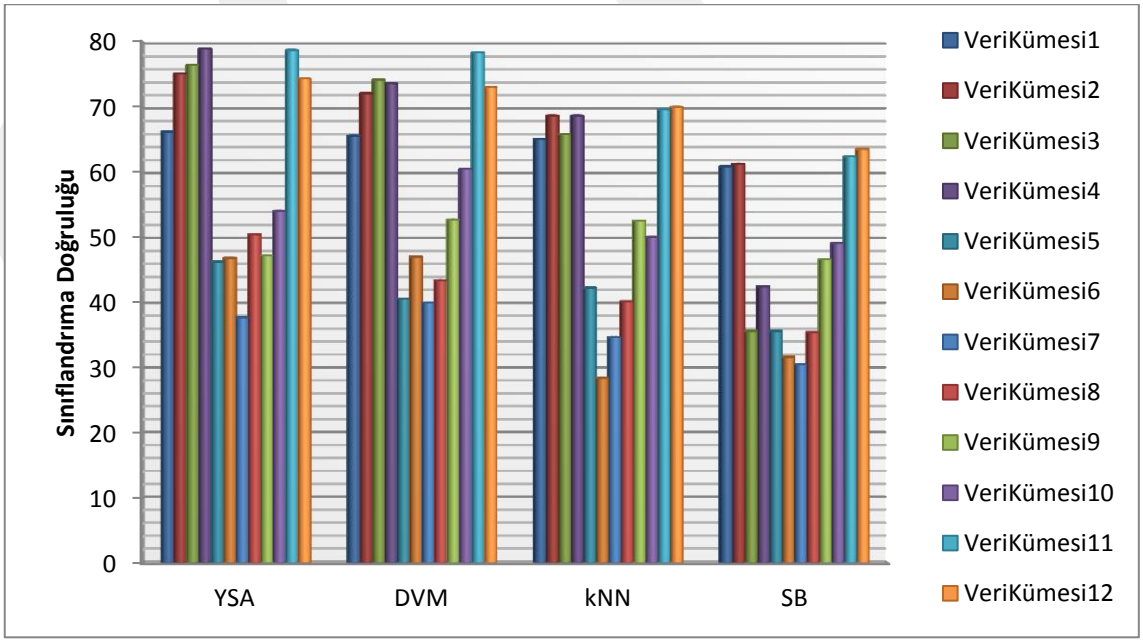
VeriKümesi8 AR parametreleri kullanılarak 32 özellik kullanılarak oluşturulmuştur. Bu veri kümesi ile elde edilen en yüksek SD değeri %50.47 ile YSA sınıflandırıcı ile gerçekleştirilmiştir. SD değerinden de görüldüğü üzere tek başına AR parametreleri ile elde edilen özellikler ile gerçekleştirilen sınıflandırma işlemi duygu tanıma için yeterli gelmemektedir.

VeriKümesi9 DWT özellik çıkarım yöntemiyle oluşturulan 32 tane dalgacık katsayısı kullanılarak oluşturulan veri kümesidir. Gerçekleştirilen en yüksek SD değeri DVM sınıflandırıcı kullanılarak %52.75 bulunmuştur.

VeriKümesi10 DWT özellik çıkarım yöntemiyle oluşturulan 56 tane dalgacık katsayısı kullanılarak oluşturulan veri kümesidir. Gerçekleştirilen en yüksek SD değeri DVM sınıflandırıcı kullanılarak %60.53'tür. DWT katsayılarının arttırılarak gerçekleştirilen bu veri kümesinde 32 özniteliğe göre SD değerinde %7.78'lik bir yükselme meydana geldiği görülmektedir.

VeriKümesi10 ve VeriKümesi11 için veri tabanı analiz sonuçları incelendiğinde VeriKümesi11 için en yüksek SD değerinin %78.74, VeriKümesi12 için ise %74.38 olduğu görülmektedir. Şekil 5.4'de toplu olarak verilen sonuçlardan da anlaşılacağı gibi çıkarılan özelliklerin üçlü ve dörtlü gruplar halinde oluşturulan özellik kümeleri her bir özniteliğin kendi başına kullanımı ile kıyaslandığında elde edilen sonucun sadece MFCC özniteliği ile elde edilen %78.94 'ün altında kaldığı görülmektedir.

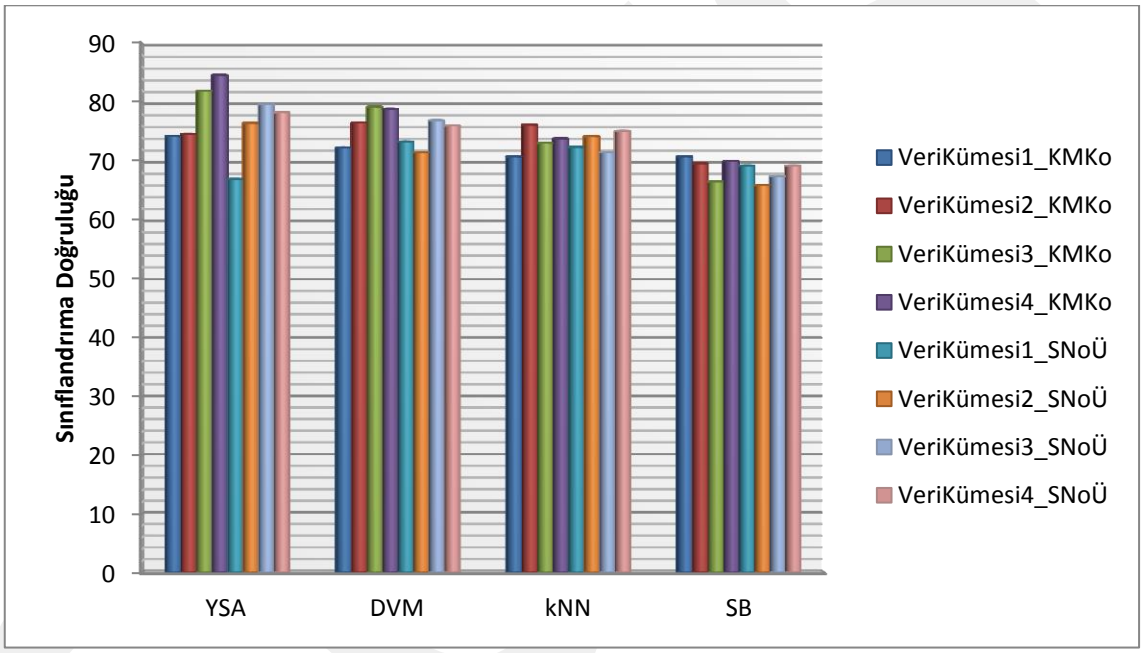
VeriKümesi1'den VeriKümes12'ye kadar EmoDB veri tabanındaki yedi duygudan oluşturulan özellik kümelerinin tamamına ait sonuçlar Şekil 5.4'te toplu olarak verilmiştir. Şekil 5.4'ten de görüldüğü üzere en yüksek SD değeri VeriKümesi4'ün YSA yöntemi ile gerçekleştirilmesinden elde edilen %78.94'tür.



Şekil 5.4. Yedi duygudan oluşturulan veri kümelerine ait analiz sonuçları

Üçerli duygu gruplarına bakıldığında VeriKümesi4_KMKo için en yüksek SD değeri %84.61 ile YSA yöntemi ile hesaplanmıştır. VeriKümesi1_KMKo için SD değeri %74.23, VeriKümesi2_KMKo için SD değeri %76.54 DVM ile, VeriKümesi3_KMKo için SD 81.92 YSA ile, ise YSA ile hesaplanmıştır. VeriKümesi1_SNoÜ için ise en yüksek %73.30 ile DVM ile, VeriKümesi2_SNoÜ %76,46 YSA ile, VeriKümesi3_SNoÜ %79,64 YSA ile, VeriKümesi4_SNoÜ ise %78.28 YSA ile gerçekleştirilmiştir.

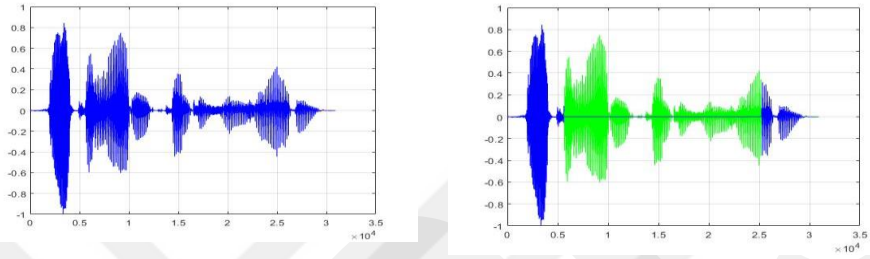
Üçerli duygu grupları kullanılarak hesaplanan SD değerleri de Şekil 5.5'de toplu olarak verilmiştir.



Şekil 5.5. Üçlü duygudan oluşturulan veri kümelerine ait analiz sonuçları

5.1.2. Uygulama 3 için Veri Kümelerinin Belirlenmesi ve Ön İşlemler

Uygulama 3'de Derin Öğrenme algoritmalarıyla duygu tanıma işlemi gerçekleştirmiştir. AlexNET modelinde, ağa vermek için spektrogram görüntüleri kullanılmıştır. Spektrogram, zaman içindeki frekans içeriğine işaret eden bir sinyalin zaman-frekans ayrışmasıdır. Ayrıca spektrogram duygu tanıma için yararlı olan para-lingual bilgileri de bünyesinde taşır (Ma ve ark., 2018).



(a) Sinyalin orijinal görüntüsü (b) Sinyalin kullanılan kısmı

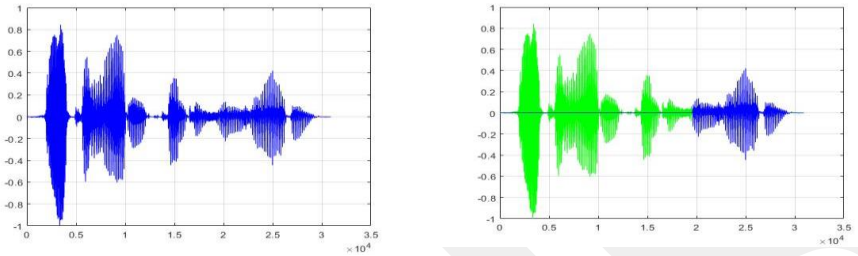
Şekil 5.6. Uygulama 3 için birinci yöntem ile sinyal ön hazırlık işlemi

Spektrogram oluşturulurken veri iki farklı şekilde hazırlanmıştır. Birinci yöntem veri tabanındaki en küçük uzunluklu segmentin örnek sayısı (19608) tespit edilmiştir. Daha sonra mevcut verileri eşit boyutlu hale getirmek için aşağıdaki işlemler gerçekleştirilmiştir:

- Her bir segmentin örnek sayısı bulunmuş ve orta noktası belirlenmiştir.
- Orta nokta merkez olacak şekilde, önceki 9804 örnek ve sonraki 9804 örnek alınarak toplam uzunluğu 19608 olan yeni segment oluşturulmuştur.

Şekil 5.6'da veri hazırlama ön işleminin nasıl yapıldığı gösterilmiştir. Mavi ile gösterilen Şekil 5.6.(a) sinyalin orijinal ilk halidir. Şekil 5.6.(b)'de yeşil ile uygulama için seçilen segmenttir. Spektrogram çıkarılırken, her bir veri parçacığı Hamming pencereleme ile 128 uzunluğunda alt bölümlere ayrılarak üst üste binme örnek sayısı 120 olacak şekilde pencerelenmiştir. Daha sonra Ayrık Fourier Dönüşümü hesaplanmıştır. Bu şekilde hazırlanan *"jpeg"* formatındaki resimler oluşturulmuştur.

İkinci yöntemde sabit boyutlu segmentleri oluştururken, segmentin başından itibaren 19608 örnek alınmıştır. Şekil 5.7.(a) sinyalin orijinal hali iken, Şekil 5.7.(b) ile gösterilen yeşil renkli sinyal seçilen segmenti ifade etmektedir. Seçilen her bir segmentin spektrogram görüntüleri alınmıştır. Elde edilen resim verilerinden yeni veri kümeleri oluşturulmuştur.



(a)Sinyalin orijinal görüntüsü (b)Sinyalin kullanılan kısmı

Şekil 5.7. Uygulama 3 için ikinci yöntem ile sinyal ön hazırlık işlemi

DSA mimarisine vermek için hazırlanan özellikler MFCC özellikleridir.

Tez çalışmasındaki üçüncü uygulamada EmoDB veri tabanında farklı özellikler çıkarılarak toplam beş farklı veri kümesi oluşturulmuştur. Bunlar: Çizelge 5.10'de verilmiştir.

Çizelge 5.10. Uygulama 3 ESA için oluşturulan veri kümeleri

Veri kümesi Adı	Açıklama
VeriKümesi_3	Yedi duyguya ait spektrogram görüntüleri
VeriKümesi_KMKo_3_1	KMKo duygularında birinci yöntem ile elde edilen spektrogram görüntüleri
VeriKümesi_SNoÜ_3_1	SnoÜ duygularından birinci yöntem ile elde edilen spektrogram görüntüleri
VeriKümesi_KMKo_3_2	KMKo duygularından ikinci yöntem ile elde edilen spektrogram görüntüleri
VeriKümesi_SnoÜ_3_2	SnoÜ duygularından ikinci yöntem ile elde edilen spektrogram görüntüleri

Üçüncü uygulamanın ikinci uygulaması olan DSA yöntemi ile sınıflandırma için ise Çizelge 5.11'deki veri kümeleri oluşturulmuştur.

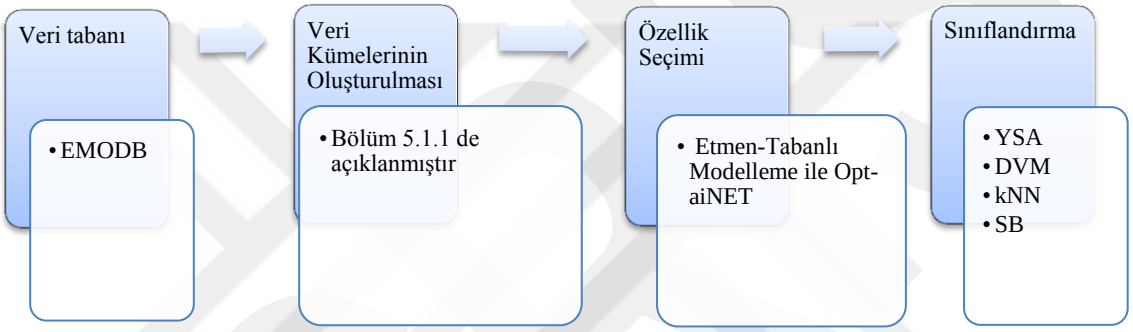
Çizelge 5.11. Uygulama 3 DSA için oluşturulan veri kümeleri

Veri kümesi Adı	Açıklama	Veri Boyutu
VeriKümesi1_3	MFCC özellikleri (8 tane)	527 X 56
VeriKümesi2_3	MFCC özellikleri (16 tane)	527 X 112
VeriKümesi3_3	MFCC özellikleri (32 tane)	527 X 224
VeriKümesi4_3	MFCC özellikleri (60 tane)	527 X 420
VeriKümesi1_KMKo_3	MFCC özellikleri (8 tane)	260 X 56
VeriKümesi2_ KMKo_3	MFCC özellikleri (16 tane)	260 X 112
VeriKümesi3_KMKo_3	MFCC özellikleri (32 tane)	260 X 224
VeriKümesi4_ KMKo_3	MFCC özellikleri (60 tane)	260 X 420
VeriKümesi1_SnoÜ_3	MFCC özellikleri (8 tane)	221 X 56
VeriKümesi2_SnoÜ_3	MFCC özellikleri (16 tane)	221 X 112
VeriKümesi3_SnoÜ_3	MFCC özellikleri (32 tane)	221 X 224
VeriKümesi3_SnoÜ_3	MFCC özellikleri (60 tane)	221 X 420

5.2. Ana Çalışmalar ve Sonuçları

5.2.1. Uygulama 1: Etmen Tabanlı Otomatik Özellik Seçimi (*Agent Based Automatic Feature Selection - ABAfs*) Yaklaşımı

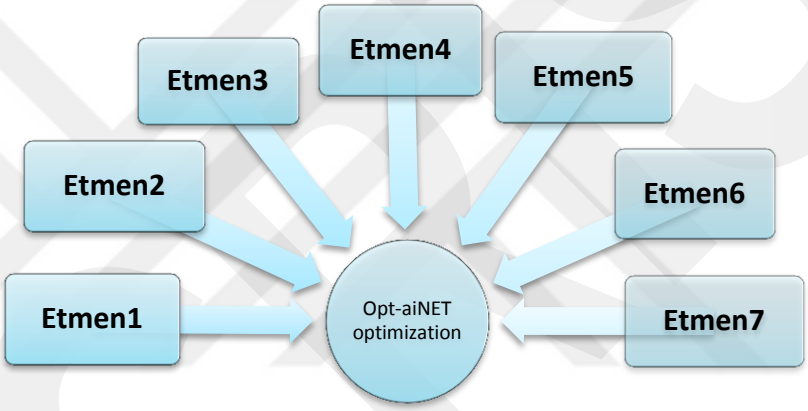
Bu uygulamada EmoDB veri tabanı kullanılarak elde edilen özellikler etmen tabanlı modellenmiş ve daha sonra Opt-aiNET optimizasyon algoritması ile özellik seçimi yapılarak farklı sınıflandırıcılarla duygu tanıma işlemi gerçekleştirilmiştir. Şekil 5.8'de önerilen ABAfs yönteminin işlem adımları gösterilmiştir.



Şekil 5.8. Uygulama 1 işlem adımları

Uygulama 2 EmoDB veri tabanındaki 16 kHz de örneklenmiş toplam 527 segment üzerinde cinsiyet ve cümle bağımsız bir şekilde gerçekleştirilmiştir. İlk olarak duygu verileri üzerine spektral ve prozodik özellik çıkarma yöntemleri uygulanarak özellikler hesaplanmıştır. Spektral özelliklerden MFCC, LPC, AR ve DWT parametreleri, prozodik özeliklerden ise Temel Frekans özellikleri çıkarılmıştır. MFCC için 8, 16, 32 ve 60 adet katsayıdan oluşan özellikler elde edilmiştir. 16 ve 112 adet LPC katsayısı, 24 tane Temel Frekans, 32 tane AR ve 32 ve 56 adet DWT parametresi çıkarılmıştır. Oluşturulan veri kümeleri Bölüm 5.1.1'de verilmiştir. Belirtilen veri kümeleri kullanılarak etmen tabanlı modellenmiş ve bu özelliklere Opt-aiNET optimizasyon algoritması uygulanarak özellik seçimi yapılmıştır. Çalışmada Weka (Bouckaert ve ark., 2010) programı kullanılmıştır. Sınıflandırıcı olarak YSA, DVM, kNN ve SB algoritmaları kullanılmıştır. Uygulamalar 10 çapraz doğrulama ile gerçekleştirilmiş. Sınıflandırıcılarda kullanılan algoritmalar ve parametreler detaylı olarak Bölüm 5.1.1'de verilmiştir.

Otonom bir etmenin, içinde bulunduğu ortamı algılayıp ardından durumsal farkındalık oluşturup ve amacı doğrultusunda bu algı bilgisini kullanarak bir sonraki davranışını belirlemesinden hareketle, çıkarılan özellik grupları etmen, optimizasyon işlemi ise bir çevre (*environment*) gibi düşünülmüş, başarıyı maksimum yapacak şekilde etmenin hareket (*action*) değeri belirlenmiştir. Sınıflandırma başarısı maksimum yapılırken etmenin temsil ettiği özellik kullanılıyorsa durum bilgisi (*state*) "1" olarak etmene gönderilmekte, etmen de hareket değeri "1" olacak şekilde sınıflandırmaya tabi tutulmaktadır. Aksi takdirde durum bilgisi "0" ve aynı zamanda etmenin hareket değeri de "0" olduğundan sınıflandırmaya tabi tutulmayacağına karar verilmektedir. Önerilen ABAfs yönteminin detayları tez çalışmasında Bölüm 4.2.2'de detaylı olarak verilmiştir.



Şekil 5.9. MFCC özellikleri temsil eden etmen modeli

Etmen tabanlı modellemenin MFCC özelliklerine uygulanması Şekil 5.9'da gösterilmektedir. Mel Katsayılarından elde edilen tamamlayıcı istatistik değerlerin her bir grubu bir etmen gibi düşünülmüş, yedi farklı tamamlayıcı istatistik özellik olduğu için de yedi etmen kullanılmıştır. Her bir katsayı grubuna ait maksimum değerleri Etmen1, minimum değeri Etmen2, ortalama değeri Etmen3, Standart Sapma değeri Etmen4, Medyan değeri Etmen5 Çarpıklık değeri Etmen6 ve Basıklık değeri Etmen7 ve olarak tanımlanmıştır.

Uygulama 1 Bölüm 5.1.1'de oluşturulan ve etkinliği gösterilen 12 tane veri kümesi üzerinde uygulanmıştır. Bu veri kümelerinin dört tanesi yedi duyguya ait segmentler içerir. Bunlar VeriKümesi1, VeriKümesi2, VeriKümesi3 VeriKümesi4'tür. Üçerli duygu grupları için oluşturulan veri kümeleri ise VeriKümesi1_KMKo,

VeriKümesi2_KMKo, VeriKümesi3_KMKo, VeriKümesi1_SNoÜ, VeriKümesi3_SNoÜ, VeriKümesi4_SNoÜ'dir.

Belirtilen veri kümelerinden önerilen ABAfs ile özellik seçimi yapılmıştır. Yapılan denemeler sonucunda yedi duygu içeren veri kümeleri için seçilen etmen grubu Etmen1, Etmen3, Etmen4 ve Etmen5 ile modellenen özellik gruplarıdır. Üçerli duygu grupları içeren veri kümelerinin ABAfs yöntemi ile özellik seçimi yapıldığında KMKo etmen grubu için seçilen etmenler Etmen3, Etmen5, Etmen6 ve Etmen7; SNoÜ etmen grubu için seçilen etmenler ise Etmen3 ve Etmen5'dir. Seçilen etmenler Çizelge 5.12'de gösterilmiştir.

Özellik seçimi sonrası oluşan veri kümeleri isimlerinin sonuna "_1" eklenerek birinci uygulamaya ait seçilen özellikler olduğu ifade edilmiştir.

Çizelge 5.12. Uygulama 1 için seçilen özellikler

Etmen	Yedi duygu için seçilen etmenler	Üçerli duygu grupları için etmenler	
		KMKo	SNoÜ
Etmen1	✓		
Etmen2			
Etmen3	✓	✓	✓
Etmen4	✓		
Etmen5	✓	✓	✓
Etmen6		✓	
Etmen7		✓	

Çizelge 5.13. Uygulama 1 için sonuçlar

Veri Kümesi	Sınıflandırıcı	SD (%)	Kappa İstatistiği
VeriKümesi1_1	YSA	**66.22**	0.60
	DVM	65.84	0.60
	kNN	57.31	0.49
	SB	59.58	0.52
VeriKümesi2_1	YSA	**77.99**	0.66
	DVM	72.86	0.67
	kNN	70.40	0.65
	SB	62.24	0.55
VeriKümesi3_1	YSA	**77.41**	0.65
	DVM	74.76	0.70
	kNN	70.78	0.65
	SB	36.43	0.25
VeriKümesi4_1	YSA	**80.28**	0.71
	DVM	76.85	0.72
	kNN	72.11	067
	SB	63.75	0.57
VeriKümesi1_KMKo_1	YSA	**72.30**	0.56
	DVM	70.76	0.52
	kNN	69.61	0.50
	SB	65.38	0.46
VeriKümesi2_KMKo_1	YSA	**78.08**	0.64
	DVM	77.31	0.64
	kNN	76.53	0.61
	SB	76.53	0.61
VeriKümesi3_KMKo_1	YSA	76.15	0.62
	DVM	**78.46**	0.65
	kNN	73.08	0.56
	SB	64.23	0.42
VeriKümesi4_KMKo_1	YSA	**77.31**	0.64
	DVM	76.92	0.63
	kNN	76.53	0.61
	SB	65.38	0.45
VeriKümesi1_SNoÜ_1	YSA	65.15	0.47
	DVM	59.72	0.39
	kNN	64.25	0.45
	SB	**66.52**	0.50
VeriKümesi2_SNoÜ_1	YSA	**80.00**	0.69
	DVM	62.00	0.53
	kNN	71.04	0.56
	SB	68.32	0.52
VeriKümesi3_SNoÜ_1	YSA	**77.82**	0.66
	DVM	68.78	0.52
	kNN	71.04	0.56
	SB	66.06	0.49
VeriKümesi4_SNoÜ_1	YSA	**79.64**	0.69
	DVM	76.15	0.62
	kNN	77.30	0.63
	SB	66.92	0.47

Çizelge 5.13 incelendiğinde en fazla artışın YSA yöntemi ile gerçekleştirildiği görülmektedir. VeriKümesi1 YSA ile sınıflandırıldığında %66.22 olan SD değerinin aynı kaldığı görülmektedir. VeriKümesi2'nin YSA yöntemi ile sınıflandırıldığında SD değeri %75.14 iken VeriKümesi2_1'ün SD değeri ise %77.99 olarak elde edilmiştir. VeriKümesi3 için SD değeri %76.47 iken, VeriKümesi3_1 için SD değeri %77.41 olarak gerçekleştirilmiştir. VeriKümesi4 %78.94 SD değerine sahip iken ABAfs ile özellik seçimi yapıldıktan sonra elde edilen VeriKümesi4_1 SD değeri %80.09'a yükseltilmiştir.

Çizelge 5.14'de VeriKümesi1, VeriKümesi2, VeriKümesi3 ve VeriKümesi4 için elde edilen konfüzyon matrisleri verilmiştir. Çizelge 5.15'de ise VeriKümesi1_1, VeriKümesi2_1, VeriKümesi3_1 ve VeriKümesi4_1 için hesaplanan konfüzyon matrisleri verilmiştir. Önerilen yöntem ile hiçbir özelik seçimi yapılmadan önce hesaplanan SD değerlerinin karşılaştırılması amacıyla yine en iyi performansın elde edildiği YSA yönteminin konfüzyon matrisleri karşılaştırılmıştır.

Çizelge 5.13 ve Çizelge 5.14 incelendiğinde VeriKümesi1'de 99 tane kızgın duygusu doğru tanınırken VeriKümesi1_1'de 104 olarak tanınmıştır. 36 olarak tanınan korku duygu 41, 35 olan mutlu duygusu 37 olarak tanınmıştır. Sıkıntı, nefret duygularının tanınma oranlarında ise düşüş meydana gelmiştir.

VeriKümesi2'nin konfüzyon matrislerine bakıldığında 53 tane sıkıntılı duygusu tanınırken VeriKümesi2_1'de 58 tane tanınmıştır. Nefret duygusu özellik seçimi yapılmadan önce %71.74 oranında tanınırken ABAfs sonrasında %82.61 oranında tanınmıştır. Korku duygusu 106 tanesi tanınırken seçim sonrası 109 tanesi tanınmıştır. Bu şekilde incelendiğinde bütün duyguların tanınma oranlarının arttığı görülmektedir.

VeriKümesi3'de YSA ile 52 tane üzgün duygusu doğru sınıflandırılırken VeriKümesi3_1'de bu sayının 55'e çıktığı görülmektedir. Mutlu duygusu tanıma oranı %51.14 iken özellik seçimi sonrası %62.85'e çıkarılmıştır. Nefret, kızgın, korku ve normal duygularında ise herhangi bir değişiklik gözlemlenmemiştir.

VeriKümesi4'te YSA ile sınıflandırılan sıkıntılı duygusunun 59 tanesini doğru sınıflandırırken VeriKümesi4_1'de 60'a çıktığı görülmüştür. Mutlu duygusunun tanıma oranında herhangi bir değişiklik olmamıştır. Kızgın duygusunun doğru sınıflandırılan veri sayısı 110'dan 113'e, üzüntü duygusunun doğru sınıflandırılan duygu sayısı 52'den 56'e çıkarılmıştır. Korku duygusunun tanınma oranı %79.68 iken %70.31'e düşmüştür. Son olarak normal duygusunda tanıma oranı %79.75 iken %82.28'e yükseldiği görülmüştür.

Çizelge 5.14. VeriKümesi1, VeriKümesi2, VeriKümesi3 ve VeriKümesi4 için konfüzyon matrisleri (YSA)

(a) VeriKümesi1 için konfüzyon matrisi

	S	N	K	Ü	Ko	M	No
S	49	1	1	5	6	1	18
N	1	33	2	2	3	2	3
K	1	1	99	0	4	21	0
Ü	6	0	0	49	3	1	3
Ko	4	5	7	4	36	5	3
M	1	3	20	1	8	35	2
No	19	0	1	3	4	4	48

(b) VeriKümesi2 için konfüzyon matrisi

	S	N	K	Ü	Ko	M	No
S	53	2	0	5	3	2	16
N	0	33	2	1	3	4	3
K	0	3	106	0	0	16	1
Ü	1	1	0	55	1	0	3
Ko	3	3	3	3	44	5	3
M	3	3	15	1	3	45	0
No	16	0	0	1	1	1	60

(c) VeriKümesi3 için konfüzyon matrisi

	S	N	K	Ü	Ko	M	No
S	61	2	0	6	1	0	11
N	0	36	2	2	3	2	1
K	0	2	106	0	1	16	1
Ü	6	0	0	52	1	0	2
Ko	2	3	1	5	48	2	3
M	1	4	20	0	4	40	1
No	14	0	0	1	4	0	60

(d) VeriKümesi4 için konfüzyon matrisi

	S	N	K	Ü	Ko	M	No
S	59	2	0	5	2	2	11
N	0	35	2	3	2	3	1
K	0	0	110	0	0	15	1
Ü	3	1	0	52	2	0	3
Ko	0	4	1	3	51	1	4
M	3	1	15	0	4	46	1
No	14	0	0	0	2	0	63

Çizelge 5.15. VeriKümesi1_1, VeriKümesi2_1, VeriKümesi3_1 ve VeriKümesi4_1 için konfüzyon matrisleri (YSA)

(a) VeriKümesi1_1 için konfüzyon matrisi

	S	N	K	Ü	Ko	M	No
S	38	3	1	6	5	2	26
N	1	28	4	2	3	6	2
K	0	1	104	0	4	16	1
Ü	3	2	0	50	3	0	3
Ko	3	4	6	3	41	4	3
M	2	4	20	1	5	37	1
No	20	0	0	3	4	1	51

(b) VeriKümesi2_1 için konfüzyon matrisi

	S	N	K	Ü	Ko	M	No
S	58	1	0	3	1	1	17
N	1	38	2	1	2	1	1
K	0	0	109	0	1	15	1
Ü	3	0	0	57	1	0	0
Ko	2	5	1	3	45	6	2
M	1	2	14	1	5	46	1
No	18	1	0	0	2	0	58

(c) VeriKümesi3_1 için konfüzyon matrisi

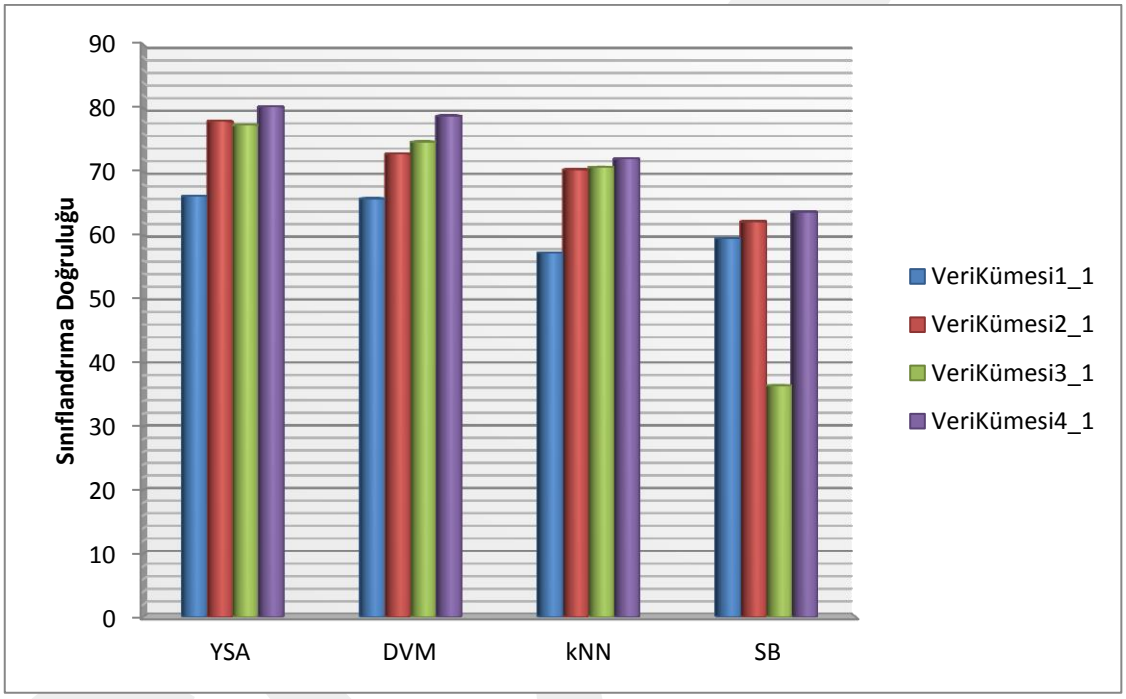
	S	N	K	Ü	Ko	M	No
59	59	1	0	3	0	2	16
0	0	36	2	1	3	3	1
0	0	2	106	0	3	14	1
2	2	0	0	55	1	0	3
1	1	3	1	3	48	3	5
2	2	2	16	0	5	44	1
15	15	1	0	0	1	2	60

(d) VeriKümesi4_1 için konfüzyon matrisi

	S	N	K	Ü	Ko	M	No
S	60	1	0	4	2	1	13
N	0	37	2	1	2	3	1
K	0	0	113	0	1	11	1
Ü	2	1	0	56	1	0	1
Ko	2	2	1	5	45	6	3
M	1	3	15	0	4	46	1
No	10	0	1	2	1	0	65

Çizelge 5.13 ve Çizelge 5.15 incelendiğinde en çok birbiriyle karıştırılan duyguların *Kızgın - Mutlu* ve *Sıkıntılı – Normal* duygularının olduğu görülmektedir. Bu da çalışmada oluşturduğumuz ve en çok birbiriyle karıştırılan üçerli duygu gruplarını desteklemektedir.

Uygulama 1'de EmodB veri tabanındaki yedi duygu için elde edilen sonuçlar Şekil 5.10'da gösterilmiştir. Görüldüğü gibi en yüksek başarı VeriKümesi4_1 ile YSA %80.28 olarak gerçekleştirilmiştir.



Şekil 5.10 Uygulama 1 için yedi duygu sonuç grafiği

Uygulama 1 için üçerli duygu gruplarından oluşan veri kümeleri incelenirse ilk olarak KMKo için VeriKümesi1_KMKo YSA ile sınıflandırıldığında SD değeri %74.23 iken VeriKümesi1_KMKo_1'in SD değeri ise %72.30'a düştüğü görülmüştür. Diğer veri kümelerinde SD oranlarının düştüğü gözlemlenmiştir. VeriKümesi3_KMKo ve VeriKümesi4_KMKo içinde SD değerinde düşüş devam etmiştir. Verikümesi2_KMKo ise %75.77 olan SD %78.08'e yükseldiği görülmüştür. İkinci duygu grubunda ise VeriKümesi1_SNoÜ veri kümesinde %66.97 olan SD değerinin VeriKümesi1_SNoÜ_1'de %65.14'e düştüğü görülmüştür. Yine VeriKümesi3_SNoÜ %79.64 olan başarının %77.82'ye düştüğü görülmüştür. VeriKümesi2_SNoÜ SD değeri %24.66'dan %80'e yükselmiştir. Görüldüğü gibi üçerli duygu gruplarında Uygulama 1

sonuçları değişkenlik göstermektedir. 16 MFCC özelikleriyle oluşturulan VeriKümesi2 için SD değeri artarken diğer veri kümelerinde düştüğü görülmüştür.

Üçerli duygu gruplarda en yüksek SD değeri Verikümesi2_KMKo_1 ve VeriKümesi2_SNoÜ_1 ile YSA yöntemi ile elde edilmişti. Bu veri kümelerine ait konfüzyon matrisleri ve ABAfs ile özellik seçimi yapılmadan önceki konfüzyon matrisleri Çizelge 15'de verilmiştir.

Çizelge 5.16. VeriKümesi2_ KMKo ve VeriKümesi2_ KMKo_1 ve VeriKümesi2_SNoÜ ve VeriKümesi2_SNoÜ_1 için konfüzyon matrisleri

(a) VeriKümesi2_ KMKo

	K	M	Ko
K	104	21	1
M	17	43	10
Ko	5	9	50

(b) VeriKümesi2_ KMKo _1

	K	M	Ko
K	106	19	1
M	16	46	8
Ko	3	10	51

(c) VeriKümesi2_SNoÜ

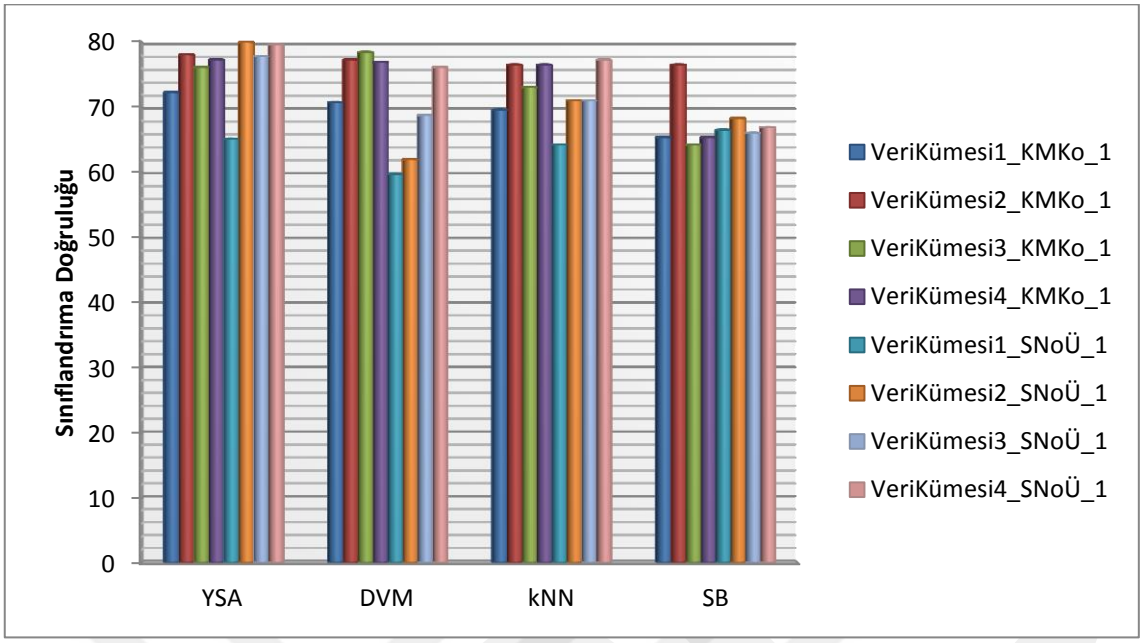
	S	No	Ü
S	52	23	6
No	22	56	1
Ü	3	1	57

(d) VeriKümesi2_SNoÜ_1

	S	No	Ü
S	56	19	6
N	12	66	1
Ü	3	3	55

Çizelge 5.16 incelendiğinde VeriKümesi2_KMKo sınıflandırıldığında 104 tane kızgın duygusu doğru sınıflandırılırken VeriKümesi2_KMKo_1 veri kümesi kullanıldığında 106 tane segment doğru sınıflandırılmıştır ki bu da kızgın duygusunun %82.54 olan sınıflandırma başarısının %84.43'a yükselmesi anlamına gelmektedir. Mutlu duygu verisinde tanınan segment sayısının 43'den 46'ya çıktığı gözlemlenmiştir. Korku duygu verisi ABAfs ile seçim yapılmadan önce 50 tanesi doğru sınıflandırılırken; sonrasında 51 segment doğru sınıflandırılmıştır.

Çizelge 5.16 incelendiğinde VeriKümesi2_SNoÜ sınıflandırıldığında 52 tane sıkıntılı duygusu doğru sınıflandırılırken VeriKümesi2_SNoÜ_1 veri kümesi kullanıldığında 56 tane segment doğru sınıflandırılmıştır ki bu da sıkıntılı duygusunun %64.20 olan sınıflandırma başarısının %69.14 e yükselmesi anlamına gelmektedir. Normal duygu verisinde tanınan segment sayısı 56'dan 66'a çıktığı gözlemlenmiştir. Üzgün duygu verisi ABAfs ile seçim yapılmadan önce 57 tanesi doğru sınıflandırılırken; sonrasında 55 segment doğru sınıflandırılmıştır.



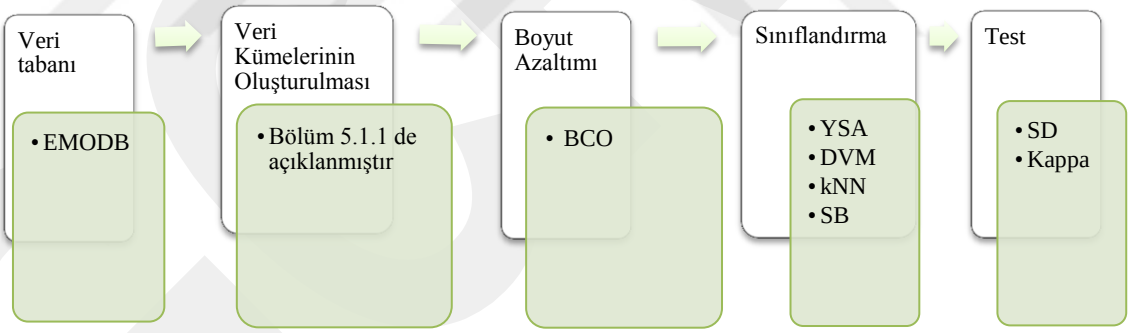
Şekil 5.11. Uygulama 1 için üçerli duygu sonuç grafiği

Uygulama 1'de EmodB veri tabanındaki üçerli duygu grupları için elde edilen sonuçlar Şekil 5.11'da gösterilmiştir. En yüksek SD değeri VeriKümesi2_SNoÜ ile 80 olarak elde edilmiştir.

5.2.2. Uygulama 2: Bulanık C-Ortalama Algoritması (BCO) ile Duygu tanıma Yaklaşımı

Tez çalışmasında gerçekleştirilen ikinci uygulama Bulanık C-Ortalama Algoritması (BCO) kullanılarak ses sinyallerinden duygu tanıma işlemidir. Özellik olarak konuşma sinyallerinden MFCC yöntemi ile elde edilen spektral özellikler kullanılmıştır. Çıkarılan özelliklerin tanımlayıcı istatistikleri (minimum, maksimum, ortalama, standart sapma, medyan, Çarpıklık (skewness), Basıklık (kurtosis)) hesaplanarak, yedi yeni özellik oluşturulmuştur. Bu özellikler kullanılarak oluşturulan veri kümelerine (Bölüm 5.1.1 de ayrıntıları anlatılmıştır.) Bulanık C-ortalama (FCM) algoritması uygulanmış ve her bir duygu için ayrı ayrı küme merkezleri tespit edilerek sınıflandırıcıya giriş olarak verilmiştir. Sınıflandırıcı olarak YSA, DVM, kNN ve SB sınıflandırıcıları kullanılmıştır. Çalışma kişi ve cinsiyet bağımsız bir şekilde gerçekleştirilmiştir.

Gerçekleştirilen işlem adımları Şekil 5.12'te gösterilmiştir. Özellik çıkarımı MFCC yöntemi ile yapılmıştır. Bölüm 5.1.1'de oluşturulan VeriKümesi2 kullanılmıştır. Elde edilen veri kümeleri üzerinde BCO algoritması kullanılarak boyut azaltımı yapılmıştır.



Şekil 0.12. Uygulama 2 işlem adımları

Yedi duygu ve üçerli duygu grupları için oluşturulan veri kümeleri Bölüm 5.1.1'de anlatılmıştır. Her bir veri kümesine BCO yöntemi uygulandıktan sonra oluşturulan yeni veri kümesi çalışmada veri kümelerinin sonuna"_2" eklenerek tanımlanmıştır. BCO ile her bir duygu için onar adet küme merkez hesaplanarak, her bir duygunun bu küme merkezleri ile temsil edilmesi sağlanmıştır. Bu şekilde normalde bir kümeleme algoritması olan BCO, her bir duyguya ait nitelikli özelliklerin tespiti için kullanılmıştır. Küme merkezleri o veriyi en iyi temsil eden aday noktalarıdır. Bu mantıktan yola çıkarak sınıflandırıcı algoritmaların girişine, çıkarılan özelliğin kendisini

değil onu temsil eden küme merkezleri verilmiş ve algoritmalar eğitilmiştir. Küme merkezleri ile eğitilen algoritmaların başarısını test etmek için, EmoDB veri tabanında yer alan her bir cümleden MFCC ile çıkarılan özelliklere ait tanımlayıcı istatistik değerleri giriş olarak sınıflandırıcılara verilmiş ve sonuçlar Çizelge 5.17'de gösterilmiştir.

Uygulama 2 için analiz sonuçları incelendiğinde VeriKümesi1_2 için DVM ile %81.78, VeriKümesi2_2 için DVM ile 92.98, VeriKümesi3_2 için DVM ile %77.79, VeriKümesi4_2 için DVM ile %84.06 olarak hesaplanmıştır.

Uygulama 2 için yedi duygudan oluşturulan veri kümeleri sınıflandırıldığında en yüksek SD değeri VeriKümesi2_2 için %92.98 ile DVM sınıflandırıcı ile elde edilmiştir. VeriKümesi2_2 için elde edilen sonuçların konfüzyon matrisi Çizelge5.18'de verilmiştir.

Çizelge 5.18'deki matrisler incelendiğinde Çizelge 5.13.(b) (VeriKümesi2) ile kıyaslandığında herhangi bir özellik seçimi yapılmadan önce 53 tane sıkıntı duygusu tanınırken, BCO yöntemi ile boyut azaltımından sonra 73 tanesi doğru olarak tanınmıştır. Yine normal duygusunun doğru sınıflandırılma sayısı 33'den 45'e, kızgın duygusu 106'dan 11'e, üzgün duygusu 55'den 69'a, korku duygusu 44'den 60'a 45'den 64'e, normal duygusu ise 60'dan 72'ye çıkarılmıştır. Çizelge 5.18 incelendiğinde en çok birbiriyle karıştırılan duyguların mutlu ile kızgın ve sıkıntılı ile normal olduğu görülmektedir.

Çizelge 5.17. Uygulama 2 için sonuçlar

Veri Kümesi	Sınıflandırıcı	SD (%)	Kappa İstatistiği
VeriKümesi1_2	YSA	81.40	0.78
	DVM	**81.78**	0.78
	kNN	64.14	0.60
	SB	79.51	0.76
VeriKümesi2_2	YSA	91.65	0.90
	DVM	**92.98**	0.92
	kNN	55.98	0.48
	SB	82.35	0.80
VeriKümesi3_2	YSA	77.04	0.74
	DVM	**77.79**	0.73
	kNN	58.82	0.50
	SB	70.78	0.66
VeriKümesi4_2	YSA	78.56	0.82
	DVM	**84.06**	0.81
	kNN	66.79	0.60
	SB	83.49	0.81
VeriKümesi1_KMKo_2	YSA	**82.30**	0.72
	DVM	81.15	0.70
	kNN	72.30	0.53
	SB	80.00	0.69
VeriKümesi2_KMKo_2	YSA	84.61	0.76
	DVM	**85.00**	0.76
	kNN	74.61	0.57
	SB	81.54	0.71
VeriKümesi3_KMKo_2	YSA	**81.53**	0.71
	DVM	81.15	0.71
	kNN	65.38	0.43
	SB	67.69	0.53
VeriKümesi4_KMKo_2	YSA	85.38	0.77
	DVM	**85.38**	0.78
	kNN	73.84	0.56
	SB	**85.38**	0.78
VeriKümesi1_SNoÜ_2	YSA	**86.42**	0.79
	DVM	85.52	0.78
	kNN	66.06	0.48
	SB	78.73	0.68
VeriKümesi2_SNoÜ_2	YSA	**89.14**	0.84
	DVM	**89.14**	0.83
	kNN	59.72	0.39
	SB	87.78	0.81
VeriKümesi3_SNoÜ_1	YSA	**85.07**	0.78
	DVM	**85.06**	0.77
	kNN	53.39	0.28
	SB	83.25	0.75
VeriKümesi4_SNoÜ_2	YSA	89.14	0.84
	DVM	**90.04**	0.85
	kNN	66.06	0.48
	SB	89.14	0.84

Çizelge 5.18. VeriKümesi2_2 için konfüzyon matrisleri

(a) VeriKümesi2_2 YSA sınıflandırıcı

	S	N	K	Ü	Ko	M	No
S	73	0	0	0	2	0	6
N	0	44	0	0	2	0	0
K	0	2	111	0	0	13	0
Ü	2	0	0	59	0	0	0
Ko	0	0	0	1	60	2	1
M	2	1	3	0	0	64	0
No	7	0	0	0	0	0	72

(b) VeriKümesi2_2 DVM sınıflandırıcı

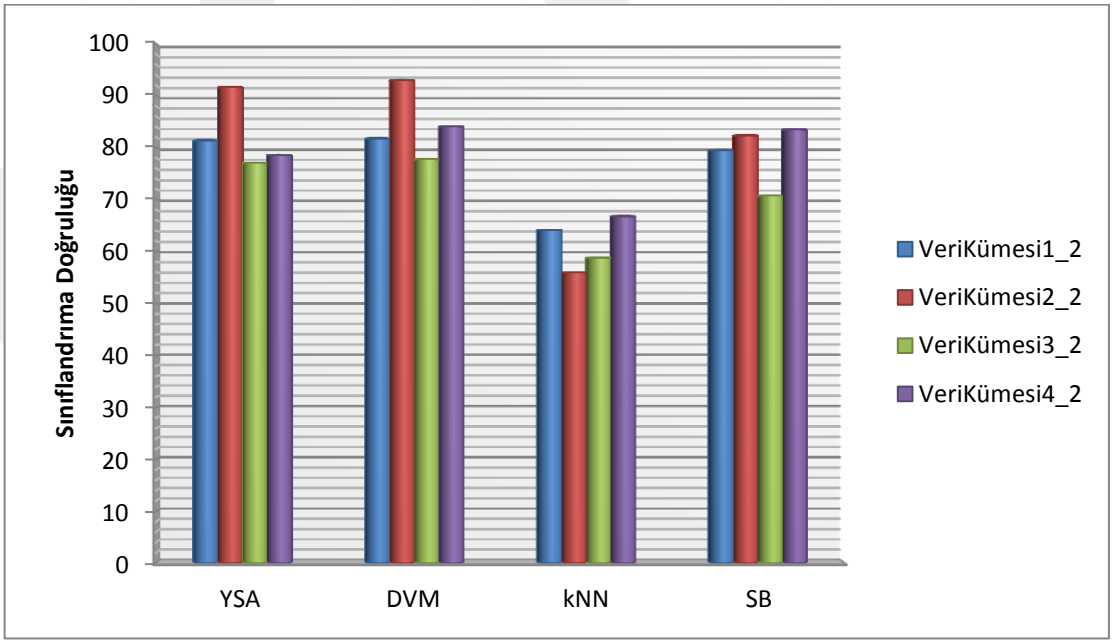
	S	N	K	Ü	Ko	M	No
S	72	0	0	0	3	0	6
N	0	45	0	0	1	0	0
K	0	3	115	0	0	8	0
Ü	1	0	0	60	0	0	0
Ko	0	0	1	0	61	1	1
M	0	1	5	0	1	62	1
No	4	0	0	0	0	0	75

(c) VeriKümesi2_2 kNN sınıflandırıcı

	S	N	K	Ü	Ko	M	No
S	34	6	1	15	3	0	22
N	0	34	0	0	2	6	4
K	0	4	98	0	0	24	0
Ü	0	0	0	55	0	0	6
Ko	1	5	13	3	41	0	1
M	1	15	46	0	0	8	0
No	35	4	0	0	15	0	25

(d) VeriKümesi2_2 SB sınıflandırıcı

	S	N	K	Ü	Ko	M	No
S	76	0	0	0	2	0	3
N	0	44	0	0	2	0	0
K	0	2	114	0	2	8	0
Ü	2	0	0	59	0	0	0
Ko	1	0	0	1	53	9	0
M	2	1	8	0	13	46	0
No	37	0	0	0	0	0	42



Şekil 5.13. Uygulama 2 için yedi duygu sonuç grafiği

Uygulama 2 için yedi duyguya ait sonuçlar toplu olarak Şekil 5.13'de verilmiştir.

Üçerli duygu gruplarını inceldiğinde KMKo duygu grubunda VeriKümesi1_KMKo_2 için YSA ile %82.30, VeriKümesi2_KMKo_2 için YSA ile %85.00, VeriKümesi3_KMKo_2 için YSA ile %81.53, VeriKümesi4_KMKo_2 için DVM ile %85.38 SD değeri elde edilmiştir.

SNoÜ duygu grubunda ise VeriKümesi1_SNoÜ_2 için DVM ile %86.42, VeriKümesi2_ SNoÜ _2 için YSA ve SVM ile %89.14, VeriKümesi3_SNoÜ_2 için YSA ve SVM ile %85.07, VeriKümesi4_KMKo_2 için DVM ile %90.04 SD değeri elde edilmiştir.

Uygulama 3 için en yüksek başarının %90.04 ile VeriKümesi4_SNoÜ_2 kullanılarak YSA ile gerçekleştirildiği görülmektedir. Üçerli duygu gruplarına ait konfüzyon matrisi Çizelge 5.19'da, sonuç grafiği ise Şekil 5.14'de verilmiştir.

Çizelge 5.19. VeriKümesi4_SNoÜ_2 için konfüzyon matrisleri

(a) VeriKümesi4_SNoÜ_2 YSA sınıflandırıcı

	K	M	Ko
K	72	8	1
M	11	67	1
Ko	3	0	58

(b) VeriKümesi4_SNoÜ_2 YSA sınıflandırıcı

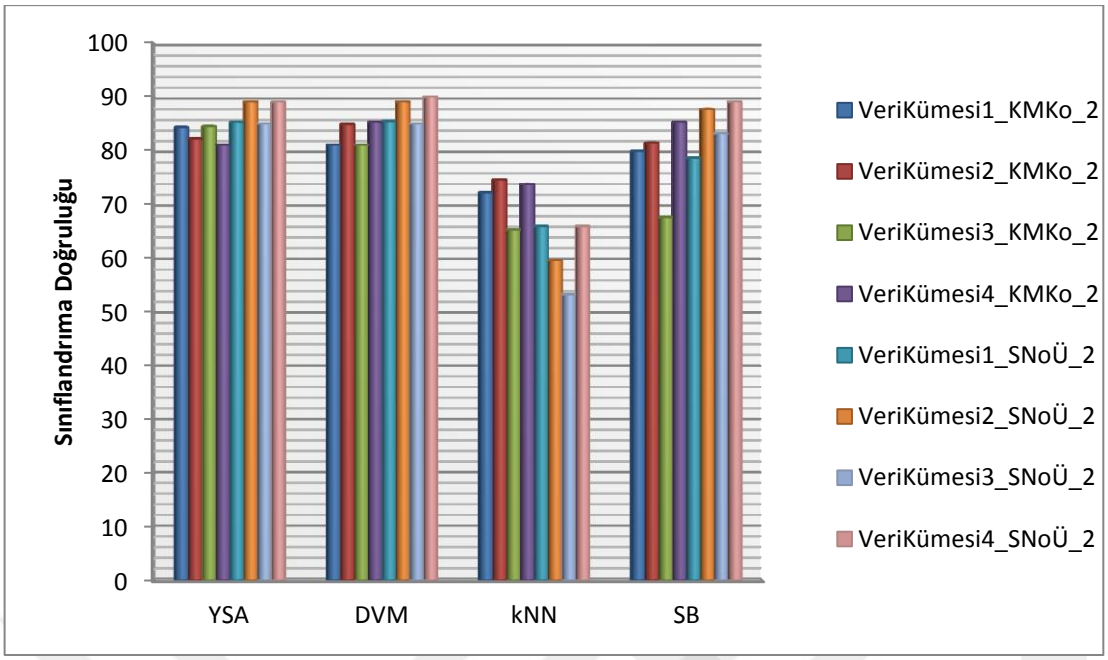
	K	M	Ko
K	72	8	1
M	11	68	0
Ko	2	0	59

(c) VeriKümesi4_SNoÜ_2 YSA sınıflandırıcı

	S	No	Ü
S	60	21	0
No	39	40	0
Ü	8	7	46

(d) VeriKümesi4_SNoÜ_2 YSA sınıflandırıcı

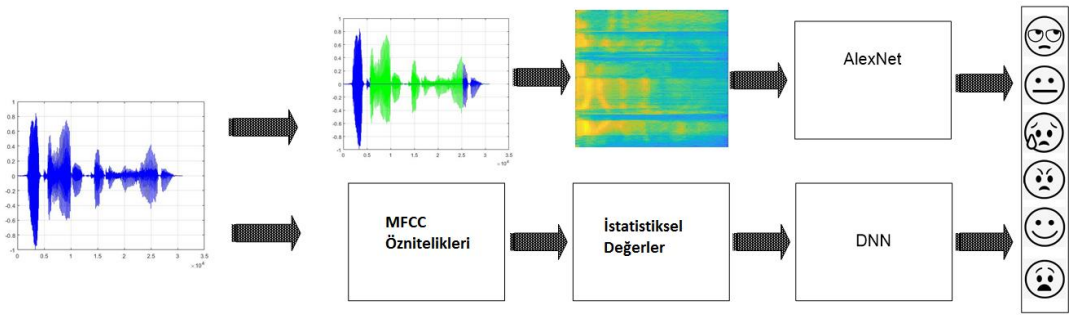
	S	No	Ü
S	72	8	1
N	11	67	1
Ü	3	0	58



Şekil 5.14. Uygulama 2 için üçlü duygu sonuç grafiği

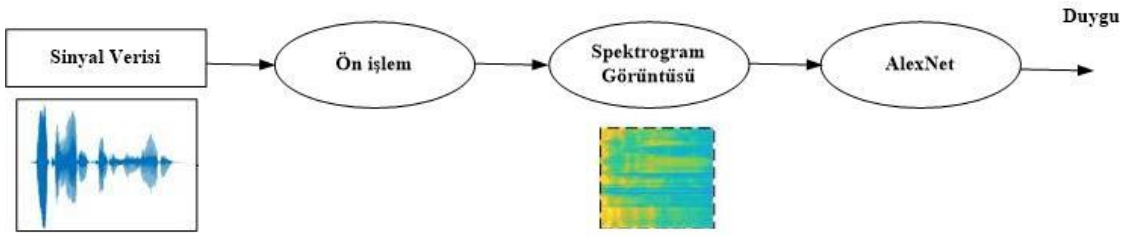
5.2.3. Uygulama 3: Derin Öğrenme Algoritmaları ile Ses Verilerinden Duygu Tanıma

Tez çalışmasındaki 3. uygulamada Derin Öğrenme Algoritmaları ile duygu tanıma gerçekleştirildi. Uygulama 1 ve Uygulama 2'de duygu tanıma işlemi yapılmadan önce özellik çıkarımı ve özellik seçimi işlemleri gerçekleştirilmişti. Oysaki makine öğrenmesinin bir alt bölümü olarak değerlendirilebilecek derin ağlarda ise herhangi bir özellik çıkarımı yapılmadan ham veriler kullanılabilmektedir. Böylece öncelikli olarak özellik çıkarma maliyeti ve zaman kaybı azalmaktadır. Bu sebeple, bu uygulamada ses verilerinden duygu tanıma problemine Derin Öğrenme Algoritmaları ile geliştirilmiş bir yaklaşım sunulmuştur. Çalışma EmoDB duygu verilerinin tüm duygularına ve üçlü duygu gruplarına uygulanmıştır. İlk olarak ön işlemlerden geçirilen sinyal bilgisinden elde edilen spektrogram görüntüleri AlexNET kullanılarak sınıflandırılmıştır. İkinci olarak klasik sesten duygu tanıma yöntemleriyle kıyaslamak adına MFCC özellikleri DSA'ya verilmiştir. Uygulama analiz sonuçları karşılaştırılmıştır. Gerçekleştirilen Derin Öğrenme ile ses verilerinden duygu tanıma uygulaması işlem adımları Şekil 5.15'da gösterilmektedir.



Şekil 5.15. Uygulama 3 işlem adımları

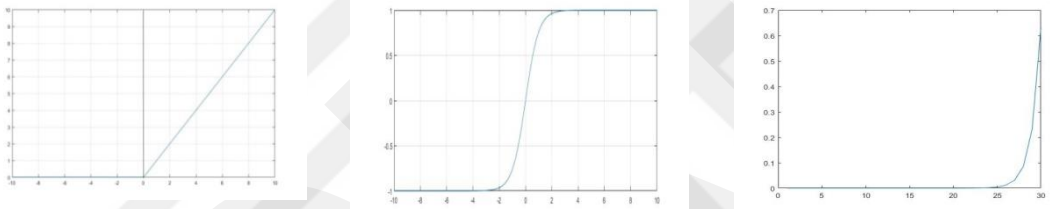
Uygulamada kullanılan veri kümeleri Bölüm 5.1.2'de anlatılmıştır. AlexNet'e giriş olarak verilen spektrogram görüntülerinin Bölüm 5.1.2'de nasıl elde edildiği anlatılmıştır. Şekil 5.16'de çalışmanın işlem adımları gösterilmiştir.



Şekil 5.16. Konuşma sinyallerinin spektrogram görüntüleri kullanarak AlexNet ile sınıflandırılması

Spekrogram görüntülerine Şekil 5.18 de her bir duyguya ait bir tane spektorgram örneği örnek verilmiştir.

AlexNET'de aktivasyon fonksiyonu olarak "*Rectified Linear Unit"* (ReLU) fonksiyonu kullanılmıştır. Şekil 5.17 (a) da ReLU fonksiyonunun grafiği gösterilmiştir.

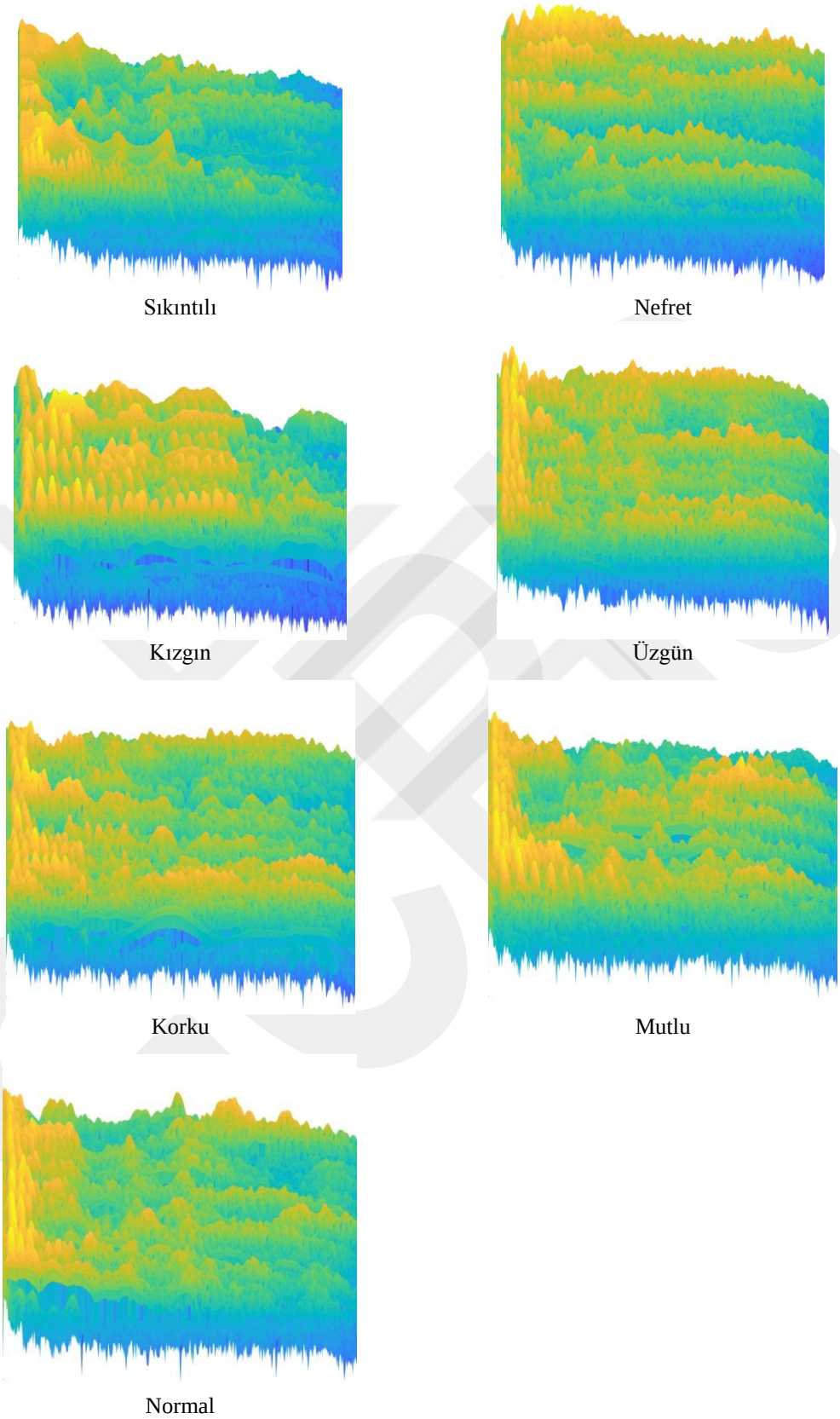


a) ReLU
$F(x) = max(0, x)$

b) Tanh
F(x)= tanh(x)

c) Softmax
$$F(x_i) = \frac{Exp(x_i)}{\sum_{j=0}^{k} Exp(x_j)}$$
$i = 0,1,2, \dots, k$

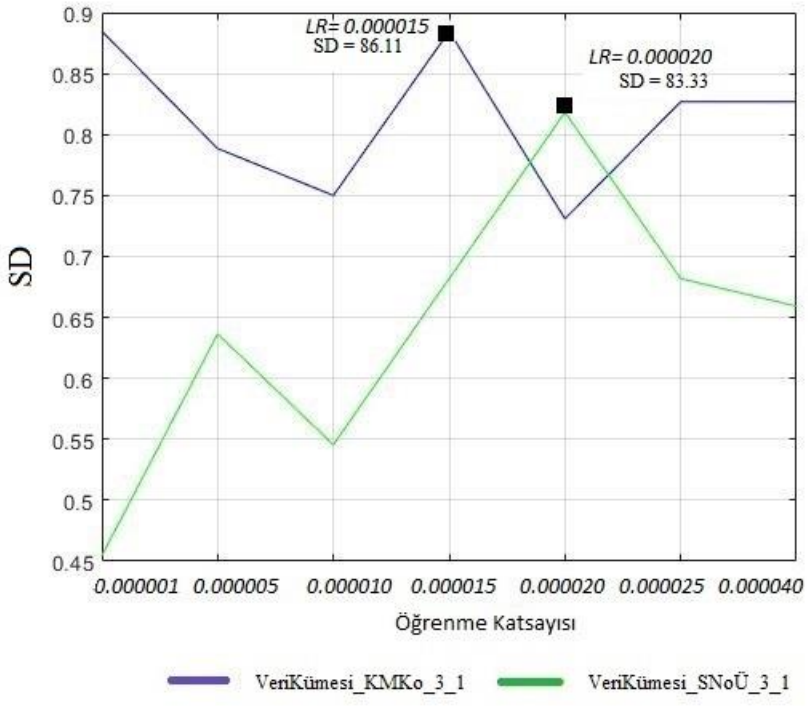
Şekil 5.17. ReLU, Tanh and Softmax fonksiyonları



Şekil 5.18. Spektrogram görüntülerine örnekler

KMKo duygu grubu ve SNoÜ duygu grubu veri kümeleri için kullanılan konvolüsyon filtre boyutu 3X3, aktivasyon fonksiyonu "RELU" ve iletim sönümü *(ing. dropout)* faktörü ise 0.000075 olarak alınmıştır. Optimizer olarak ADAM optimizasyon algoritması kullanılmıştır.

Öğrenme katsayısı 0.000001 değerinden başlatılarak arttırılıp doğruluk değeri incelenmiştir Bir örnek olarak VeriKümesi_KMKo_3_1 ve VeriKümesi_SNoÜ_3_1 için en yüksek SD değeri ve öğrenme katsayısı değişim grafiği Şekil 5.19'da gösterilmiştir. VeriKümesi_KMKo_3_1 için en yüksek SD 0.000015 öğrenme katsayısı ile %86.11 olarak elde edilmiştir. VeriKümesi_SNoÜ_3_1 için en yüksek 0.00002 öğrenme katsayısı ile %83.33 olarak elde edilmiştir. Her iki grupta da ADAM optimizer kullanılmıştır.



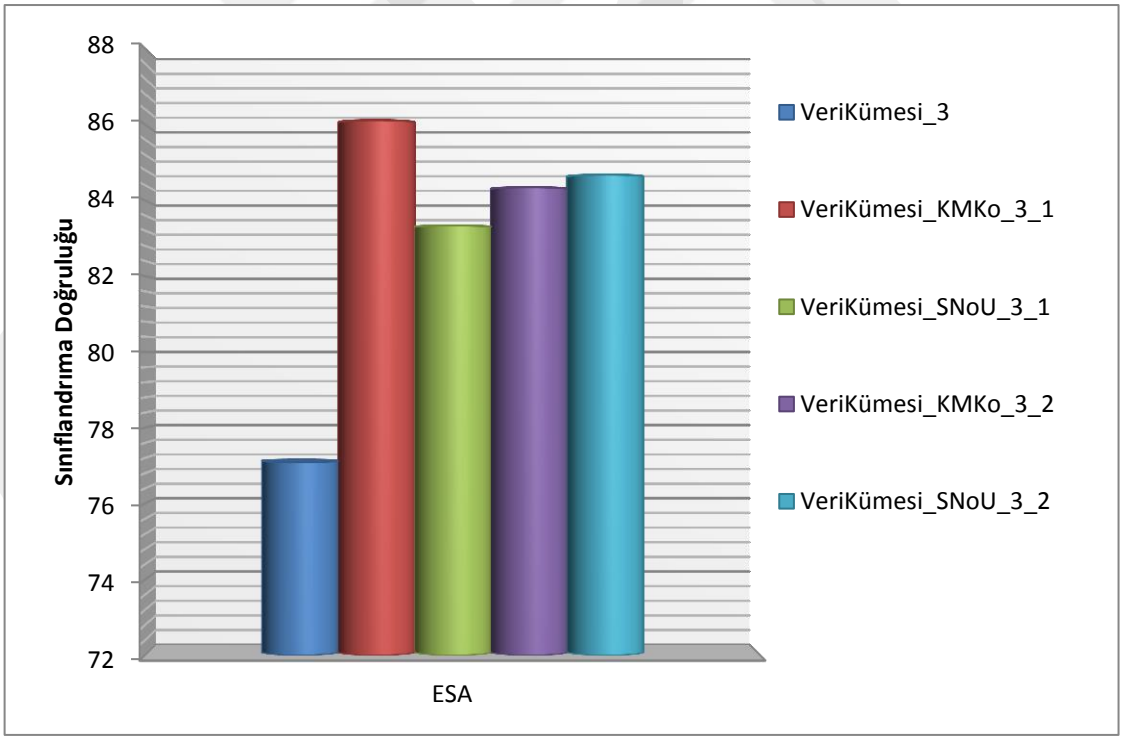
Şekil 5.19. VeriKümesi_KMKo_3_2 ve VeriKümesi_SNoÜ_3_2 için öğrenme katsayısı - SD değişimi

Çalışmanın SD sonuç değerleri Çizelge 5.20'de verilmiştir. Çizelge 5.20'den de görüldüğü üzere VeriKümesi_3 için %77.1, VeriKümesi_KMKo_3_1 için %86.11, VeriKümesi_SNoÜ_3_1 için %83.33, VeriKümesi_KMKo_3_2 için hesaplanan SD değeri %84.34 ve VeriKümesi_SNoÜ_3_2 için SD değeri ise %84.67'dir.

Tüm duygu gruplarının ESA ile elde edilen sonuç grafiği Şekil 5.20'de gösterilmiştir. Spektrogram görüntülerin ESA algoritması ile sınıflandırılması sonucu en yüksek sonuç %86.11 ile VeriKümesi_KMKo_3_1 ile gerçekleştirilmiştir.

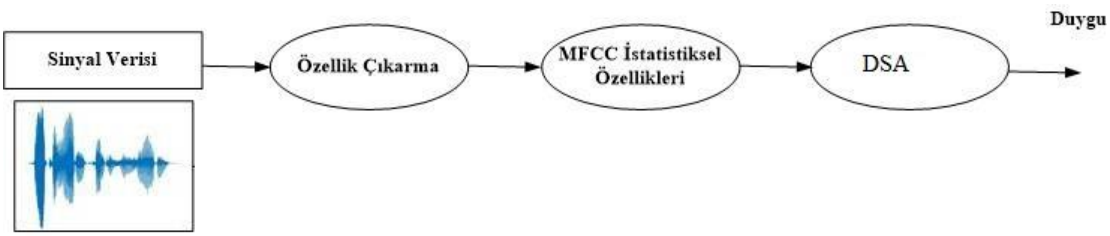
Çizelge 5.20. Uygulama 3 ESA analiz sonuçları

Veri Kümesi	SD (%)	Kappa İstatistiği
VeriKümesi_3	77.1	0.72
VeriKümesi_KMKo_3_1	86.11	0.82
VeriKümesi_SNoÜ_3_1	83.33	0.72
VeriKümesi_KMKo_3_2	84.34	0.68
VeriKümesi_SNoÜ_3_2	84.67	0.76



Şekil 5.20. Uygulama 3 ESA sonuç grafiği

Uygulama 3'de gerçekleştirilen ikinci çalışma DSA yöntemi Bölüm 5.1.2'de gösterilen veri kümeleri üzerine uygulanmıştır. Şekil 5.21'da işlem adımları gösterilmiştir.



Şekil 5.21. MFCC özelliklerinin DSA ile sınıflandırılması

DSA ile gerçekleştirilen ikinci uygulamada ReLU, Tanh ve Softmax (Şekil 5.17) aktivasyon fonksiyonları kullanılmıştır. Optimizer olarak ADAM kullanılmıştır. KMKo veri kümeleri için birinci katmanda 64, ikinci katmanda 64, üçüncü katmanda 64 ve dördüncü katmanda 3 nörondan oluşan dört katmanlı bir ağ yapısı oluşturulmuştur. Birinci, ikinci ve üçüncü katmanda nöron sayısı 6, dördüncü katmanda nöron sayısı 3 olarak alınmıştır.

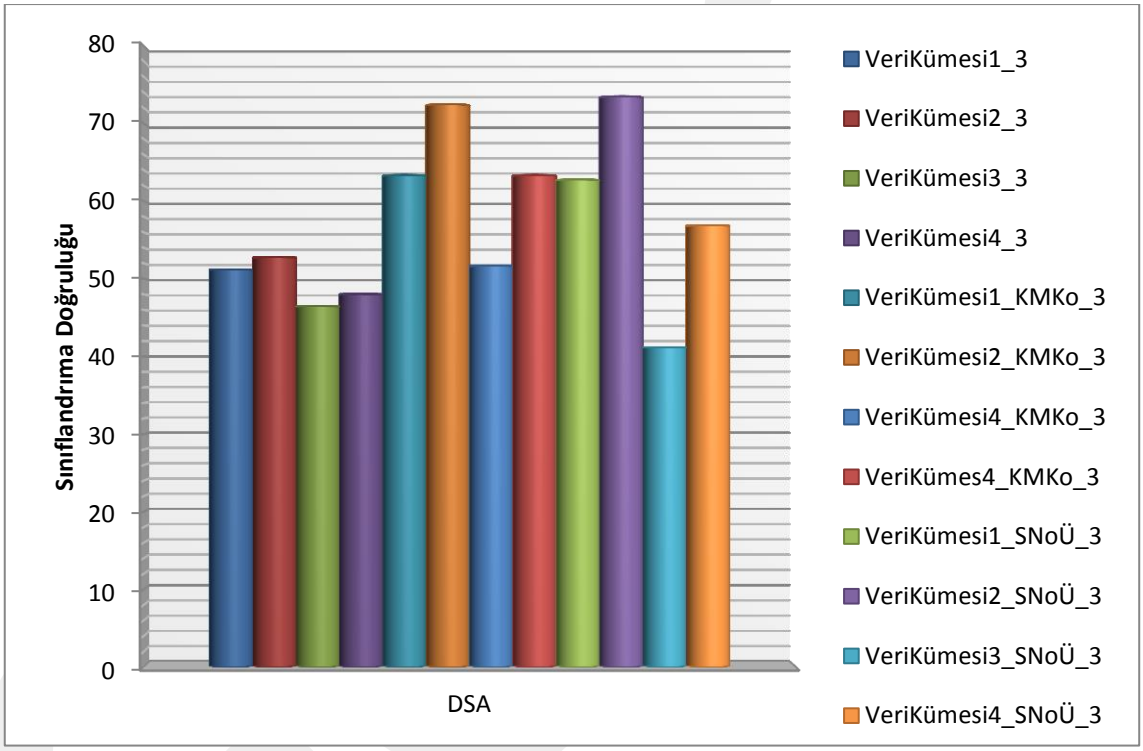
Çizelge 5.21. Uygulama 3 DSA sonuçları

Veri Kümesi	SD (%)	Kappa İstatistiği
VeriKümesi1_3	51.18	0.42
VeriKümesi2_3	52.76	0.43
VeriKümesi3_3	46.46	0.34
VeriKümesi4_3	48.03	0.36
VeriKümesi1_KMKo_3	63.33	0.42
VeriKümesi2_ KMKo_3	75.55	0.61
VeriKümesi3_KMKo_3	51.66	0.17
VeriKümesi4_ KMKo_3	63.33	0.32
VeriKümesi1_SnoÜ_3	62.75	0.44
VeriKümesi2_SnoÜ_3	**83.19**	0.73
VeriKümesi3_SnoÜ_3	41.18	0.01
VeriKümesi4_SnoÜ_3	56.86	0.15

SNoÜ veri kümeleri için 1. katmanda 128, 2. Katmanda 128, 3. Katmanda 64 ve 4.katmanda 3 tane nörondan oluşan 4 katmanlı bir ağ yapısı oluşturulmuştur.

DSA ile gerçekleştirilen çalışmada elde edilen sonuçlar Çizelge 5.21'de gösterilmiştir. En yüksek SD değeri VeriKümesi2_SNoÜ_3 veri kümesi için %83.19 olarak hesaplanmıştır.

Uygulama 3 DSA ile çalıştırılan veri kümelerine ait sonuç grafiği Şekil 5.22'de gösterilmiştir.



Şekil 5.22. Uygulama 3 DSA sonuçları

6. TARTIŞMA

Bu tez çalışmasında ilk olarak kullanılacak veri kümleri ve ön işlemler ön çalışma olarak verilmiştir. İlk olarak MFCC'den farklı boyutta özellikler çıkarılmıştır. Mel katsayısı 8, 16, 32 ve 60 boyutları için sırasıyla SD değerleri %66.60 - %75.14 - %76.47 - %78.94 olduğu gözlemlenmiştir. Dolayısıyla MFCC boyutu arttıkça SD değeri artmıştır. LPC özelliklerinin 16 ve 112 boyutlu veri kümeleri incelendiğinde kayda değer bir artış olmadığı gözlemlenmiştir. F0 özneliktekileri ve AR parametrelerinden elde edilen SD değeri tanıma performansı %50.08'in altında gerçekleştirilmiştir. DWT parametrelerinden oluşan veri kümelerine bakıldığında ise en yüksek %60.53 SD değeri bulunduğu görülmektedir. LPC, F0, AR ve DWT özelliklerinden oluşan veri kümelerinin kappa değerine bakıldığında 0.4'ün altında kaldığı görülmektedir. Bu nedenle kabul edilebilir bir doğruluk vermemektedir. Dolayısıyla bundan sonraki çalışmalar MFCC özelliklerinden oluşturulan yedi duygulu ve üçerli duygu gruplarına ait veri kümeleri üzerinde gerçekleştirilmiştir.

Bu tez çalışmasında 3 uygulama gerçekleştirilmiştir. Belirlenen etkin veri kümeleri üzerinde uygulanmıştır. Uygulama 1 sonucunda görüldüğü gibi en yüksek başarı VeriKümesi4_1 için YSA ile %80.28 ve VeriKümesi2_SNoÜ_1 için YSA ile %80 olarak gerçekleştirilmiştir.

Uygulama 1'de ABAfs özellik seçim yöntemi yedi duygudan oluşan veri kümelerinin başarısında VeriKümesi1_2 (%75.14'ten %77.99'a), VeriKümesi3_1 (%76.47'den %77.41'e), VeriKümesi4_1 (%78.94'ten %80.28'e) SD değerlerinde artış sağlanmıştır. Fakat üçerli duygu kümelerinde 16 tane MFCC öznitelikleriyle oluşturulan Verikümesi2_KMKo (%76.54'ten %78.08'e), VeriKümesi2_SNoÜ (%76.47'den %80'e) ve VeriKümesi4_SNoÜ (%78.28'den %79.64'e) dışında diğer veri kümelerde düşük SD değeri gerçekleştirilmiştir. SD değerinin yükseldiği yedi duyguda seçilen etmenler Etmen 1, Etmen 3, Etmen 4 ve Etmen 5. Bu etmenlerin temsil ettiği özelliklere bakıldığında maksimum, ortalama, standart sapma ve medyan tamamlayıcı istatistik değerlerinin etkin olduğu görülmektedir.

Uygulama 2'de BCO algoritması ile boyut indirgeme yapılmıştır. Uygulanan bütün veri kümelerinde SD değerinde artış meydana gelmiştir. Yedi duyguda en yüksek VeriKümesi2_2 için SVM ile %92.98'e çıkarılmıştır. KMKo duygu grubunda en yüksek VeriKümesi4_KMKo_2 için %YSA ve SVM ile %85.38 olarak elde edilmiştir. SNoÜ duygu grubunda ise VeriKümesi4_SNoÜ_2 için %90.04 olarak

gerçekleştirilmiştir. Bu SD değeri literatürdeki diğer çalışmalarla karşılaştırıldığında (Çizelge 6.1) yedi duygu ve cinsiyet ve konuşmacı bağımsız olarak gerçekleştirilen en iyi tanıma oranıdır.

Uygulama 3'de ilk olarak spektrogram görüntüleri üzerine ESA uygulanarak sınıflandırma işlemi yapılmıştır. AlexNET mimarisi kullanılarak gerçekleştirilen bu uygulamada en yüksek SD VeriKümesi_KMKo_3_1 veri kümesinde elde edilmiştir. Yüksek frekanslı duygu grubu olan KMKo duygu grubunda, elle çıkarılan özelliklerden elde edilen VeriKümesi2_KMKo_1 veri kümesinden elde edilen %78.08 olan başarı, spektrogram görüntüleri kullanılarak oluşturulan VeriKümesi_KMKo_3_1 ile %86.11'e yükseltilmiştir. Daha düşük frekanslı duygular içeren SNoÜ duygu grubunda ise en iyi sonuç yine spektrogram görüntülerinin kullanıldığı ve AlexNET ile sınıflandırılan VeriKümesi_SNoÜ_3_2 duygu veri kümesiyle elde edilmiştir. MFCC özellikleri kullanılarak DSA ile gerçekleştirilen VeriKümesi2_SNoÜ_3 ile %83.19 olarak gerçekleştirilmiştir.

Çalışma sonuçları göstermektedir ki yedi duygu grubunda en iyi SD değeri BCO yöntemi ile gerçekleştirilen uygulama ile elde edilmiştir. Üçlü duygu gruplarında ise BCO yöntemi sınıflandırmada oldukça başarılı sonuçlar vermiştir. Spektrogram görüntülerinden AlexNET kullanılarak gerçekleştirilen çalışmanın, elle çıkarılan özelliklere DSA uygulanarak gerçekleştirilen sonuçlardan daha iyi sonuçlar verdiği gözlemlenmiştir. Bu durum spektrogram görüntülerinin duygusal ifadeleri sınıflandırırken birbiriyle en çok karıştırılan duyguları ayırmakta etkili olduğunu göstermektedir.

Tüm çalışmalar bakıldığında YSA ve DVM yöntemleriyle yapılan sınıflandırmaların daha başarılı olduğu kNN ve SB yöntemlerinin yetersiz kaldığı gözlemlenmiştir.

Çizelge 6.1. EmoDB ile yapılan literatür sonuçları

Yayın	Kullanılan Duygular	Kullanılan Özellik Çıkarım Yöntemleri	Konuşmacı – Metin – Cinsiyet Bağımlı/Bağımsız	Kullanılan Sınıflandırıcı	En yüksek SD (%)
A. Milton, S. Tamil Selvi (Milton ve Tamil Selvi, 2014)	Kızgın, sıkıntılı, nefret, korku, mutlu, üzgün, normal	AR MFCC	Konuşmacı bağımsız, Metin bağımsız, Konuşmacı ve metin bağımsız	KNN, DVM, YSA, GKM	79.4
S. Wu ve ark. (Wu ve ark., 2011)	Kızgın, sıkıntılı, nefret, korku, mutlu, üzgün, normal	MFCC, Prozodik Özellikler, MSFs	Konuşmacı bağımlı Metin Bağımsız Cinsiyet Bağımsız	DVM	91.6
B. Yang , M. Lugger (Yang ve Lugger, 2010)	Kızgın, sıkıntılı, mutlu, üzgün, normal	Prozodik özellikler, Format Özellikleri, Sıfır geçiş oranı, Ses kalitesi	Konuşmacı Bağımsız Metin Bağımsız Cinsiyet Bağımsız	Bayes sınıflandırıcı	73.5
Albornoz ve ark. (Albornoz ve ark., 2011)	Kızgın, sıkıntılı, nefret, korku, mutlu, üzgün, normal	MFCC, Ortalama Logaritma spekrturmu, F0, Enerji özellikleri	Konuşmacı Bağımsız Metin Bağımsız Cinsiyet Bağımsız	GKM SMM ÇKP Hiyerarşik-2 Hiyerarşik-3	71.75
Shahzadi ve ark. (Shahzadi ve ark., 2015)	Kızgın, sıkıntılı, nefret, korku, mutlu, üzgün, normal	Dogrusal olmayan dinamik özellikler, Prozodik ve spektral özellikler	Konuşmacı Bağımsız Metin Bağımsız Cinsiyet Bağımlı	Genetik algoritmalar	Kadın: 82.72 Erkek: 85.90
Karimi ve Sedaaghi (Karimi ve Sedaaghi, 2016)	Kızgın, Sıkıntılı, Korku, Mutlu	Temel frekans, enerji, MFCC, LPC, PLP	Konuşmacı Bağımlı Metin Bağımsız Cinsiyet Bağımsız	DVM Bayes sınıflandırıcı	93.10
Harár ve ark., (Harar ve ark., 2017)	Kızgın, normal, üzgün	_	Konuşmacı Bağımlı Metin Bağımsız Cinsiyet Bağımsız	DSA	96.97
Badshah ve ark. (Badshah ve ark., 2019)	Kızgın, sıkıntılı, nefret, korku, mutlu, üzgün, normal	_	Konuşmacı Bağımlı Metin Bağımsız Cinsiyet Bağımsız	ESA	60
Reddy ve Vijayarajan (Reddy ve Vijayarajan, 2020)		MFCC, LPC, DWT (MFDWT)	SVM, ANN ve kNN		76.82
Bu tez çalışması	Kızgın, sıkıntı, nefret, korku, mutlu, üzgün, normal	MFCC	Konuşmacı Bağımsız Metin Bağımsız Cinsiyet Bağımsız	YSA DVM	91.65 92.98
Bu tez çalışması	Kızgın, mutlu, Korku	MFCC	Konuşmacı Bağımsız Metin Bağımsız Cinsiyet Bağımsız	YSA DVM	85.38 84.61
Bu tez çalışması	Sıkıntılı, Normal, Üzgün	MFCC	Konuşmacı Bağımsız Metin Bağımsız Cinsiyet Bağımsız	YSA DVM	89.14 90.04

7. SONUÇLAR

Bu çalışma kapsamında yedi duygu sınıfını barındıran, on kişinin seslendirdiği EmoDB duygu veri tabanı kullanılmıştır. Bu veriler kullanılarak üç farklı uygulama gerçekleştirilmiştir.

Tez çalışmasında ön çalışma olarak sesten duygu tanıma için gerekli özellik çıkarımı ve sınıflandırma öncesi gerekli parametrelerin ve ön işlemlerin uygulaması gerçekleştirilmiştir. İlk olarak EmoDB veri tabanı kullanılarak özellik çıkarımı yapılmıştır. Bu işlem farklı boyutta olan ses sinyallerini sınıflandırıcıya hazır hale getirmek ve boyut azaltmak için gereklidir. Tez çalışmasında 5 farklı yöntem ile özellik çıkarımı gerçekleştirilmiştir. MFCC yöntemi ile 8, 16, 32 ve 60 tane katsayı, LPC yöntemi ile 16 ve 112 katsayı, Temel frekansa ait 24 özellik, 32 tane AR parametresi ve 32 ve 56 boyutlu DWT özellikleri çıkarılmıştır. Yedili duygu grubu için ve üçerli duygu grubu için toplamda yirmi tane özellik kümesi oluşturulmuştur. Her bir özellik kümesi YSA, DVM, kNN ve SB yöntemleri ile sınıflandırılmıştır. Sınıflandırma sonucunda etkin veri kümeleri seçilmiş ve Uygulama 1, Uygulama 2'de kullanılmak üzere belirlenmiştir. Ayrıca Uygulama 3 için ön gerekli ön işlemler ve spektrogram görüntüleri oluşturulmuştur. Sınıflandırıcılarda kullanılan parametreler ön çalışmada belirtilmiştir.

Tez çalışmasının birinci uygulaması önerilen etmen tabanlı otomatik özellik seçimi (ABAfs) üzerine kurulmuştur. Her bir özellik grubu bir etmen olarak modellenmiştir. Opt-aiNET optimizasyon yöntemi ile modellenen etmenler üzerinden otomatik bir özellik seçim algoritması geliştirilmiştir. Önerilen ABAfs yöntemi ve Opt-aiNET özellik seçimin birlikte kullanımı ile elle denenerek tek tek hesaplanması gereken birçok hesaplama otomatik olarak yapılarak zaman kaybı giderilmiş ve en iyi SD değerini veren özelliklerin seçimi gerçekleştirilmiştir. ABAfs yöntemi ile özelliklerin seçimi bir optimizasyon problemi haline getirilmiş ve "*hangi özellikler seçilirse daha yüksek tanıma gerçekleşir"* sorusu bir maksimizasyon problemi olarak kurulmuştur. Geliştirilen yöntem MFCC veri kümeleri üzerinde çalıştırılmış ve literatürde yaygın olarak kullanılan YSA yöntemi ile sınıflandırılmıştır. Yedi duygu için gerçekleştirilen en yüksek SD değeri %80.28 iken üçerli duygu grupları için en yüksek SNoÜ duygu grubu için %80.00'dir.

Çalışmadaki ikinci uygulamada BCO kümeleme algoritması kullanılarak boyut azaltma işlemi yapılmış ve 16 boyutlu MFCC özelliklerden oluşan veri kümesi kullanılmıştır. BCO yöntemi ile her bir duygu için küme merkezleri tespit edilmiş ve bu küme merkezleri YSA, DVM, kNN ve SB yöntemleri ile sınıflandırılmıştır. En yüksek SD değeri %92.98 ile DVM sınıflandırıcılarıyla hesaplanmıştır. Gerçekleştirilen yüksek tanıma başarısı ile literatüre önemli bir katkı sağlayacağı düşünülmektedir. Ayrıca üçerli duygu gruplarından KMKo için %85.38 ve SNoÜ duygu grubu için %90.04 SD değeri elde edilmiştir.

Üçüncü ve son uygulamada klasik özellik çıkarım yöntemlerinin aksine ham veri üzerinde çalışılmıştır. Ham sinyal verilerinden spektrogram görüntüleri oluşturulmuş ve AlexNET mimarisi ile sınıflandırma işlemi gerçekleştirilmiştir. Üçlü duygu grupları üzerinde geliştirilen çalışmada en yüksek SD değeri %86.11 ile KMNo duygu grubunda elde edilmiştir. Spektrogram görüntülerinden oluşturulan veri kümeleri ile klasik özelliklerden oluşturulan veri kümelerinin SD değerleri karşılaştırılmış ve ses verilerinin resim olarak gösterimi diyebileceğimiz spektrogram görüntülerinin duygu tespitinde önemi vurgulanmıştır. Derin Öğrenme Mimarileri kullanılarak gerçekleştirilen bu çalışmanın duygu tanıma probleminde ayırt edici performans gerçekleştirdiği gösterilmiştir.

Bu tez çalışmasında bütün duygular birlikte ve birbiriyle en çok karıştırılan duygular üçerli gruplar halinde kullanılmıştır. Ayrıca çalışma cinsiyet ve kişi bağımsız bir şekilde gerçekleştirilmiştir. Literatürde kişi bağımlı ya da cinsiyet bağımlı çok fazla uygulama bulunduğundan gerçek dünya problemlerine daha gerçekçi sonuçlar sunabilmek amacıyla bu çalışmada herhangi bir ayrım yapılmadan tüm veri tabanındaki duygular uygulamalarda kullanılmıştır. Sesten duygu tanıma için çıkarılabilecek özellikler ve hangi özelliklerin daha iyi sınıflandırma başarısı gerçekleştirdiği detaylı olarak açıklanmıştır. Çalışma boyunca literatürdeki bu alanda yapılmış bütün çalışmalar incelenmeye çalışılmıştır. Hem önerilen yöntemler hem de işlenilen veriler açısından bu tez çalışmasının literatüre önemli katkı sağlayacağına inanılmaktadır.

Duygu veri tabanının en büyük sorunlarında biri veri sayısıdır. Veri sayısının azlığı derin öğrenme mimarisinde büyük bir dezavantajdır. Fakat söz konusu duygu olduğunda veri tabanı oluşturmanın zorluğu da göz ardı edilmemelidir. Duyguların algılanma şekli kişiden kişiye bile değişkenlik gösterirken sesten duygu tanıma çalışmasında yüzde yüz başarı elde etmek olası değildir.

KAYNAKLAR

Aizenberg, I., Aizenberg, N., Butakov, C. ve Farberov, E., 2000, Image recognition on the neural network based on multi-valued neurons, *15th International Conference on Pattern Recognition, Vol 2, Proceedings*, 989-992.

Akküçük, U., 2011, Veri Madencilği - Kümeleme ve Sınıflandırma Algoritamaları, *İstanbul*, Yalın Yayıncılık, p. 133.

Albornoz, E. M., Milone, D. H. ve Rufiner, H. L., 2011, Spoken emotion recognition using hierarchical classifiers, *Computer Speech and Language*, 25 (3), 556-570.

Albornoz, E. M., Sanchez-Gutierrez, M., Martinez-Licona, F., Rufiner, H. L. ve Goddard, J., 2014, Spoken Emotion Recognition Using Deep Learning, *Progress in Pattern Recognition Image Analysis, Computer Vision, and Applications, Ciarp 2014*, 8827, 104-111.

Anderson, J. A., 1983, Cognitive and Psychological Computation with Neural Models, *Ieee Transactions on Systems Man and Cybernetics*, 13 (5), 799-815.

Arı, N., Özen, Ş. ve Çolak, Ö. H., 2008, DALGACIK TEORİSİ (WAVELET), *Ankara*, Palme Yayıncılık, p.

Arias, J. P., Busso, C. ve Yoma, N. B., 2014, Shape-based modeling of the fundamental frequency contour for emotion detection in speech, *Computer Speech & Language*, 28 (1), 278-294.

Badshah, A. M., Rahim, N., Ullah, N., Ahmad, J., Muhammad, K., Lee, M. Y., Kwon, S. ve Baik, S. W., 2019, Deep features-based speech emotion recognition for smart affective services, *Multimedia Tools and Applications*, 78 (5), 5571-5589.

Başaran, İ. E., 1981, Ses Frekans Tekniği Temel ders Kitabı, Milli Eğitim Basımevi, p.

Becchetti, C. ve Prina-Ricotti, L., 2004, Speech Recognition;theory an C++ Implementation, John Wiley &Sons, p.

Bengio, Y., 2009, Learning Deep Architectures for AI, *Found. Trends Mach. Learn.*, 2 (1), 1-127.

Bhatti, M. W., Yongjin, W. ve Ling, G., 2004, A neural network approach for human emotion recognition in speech, *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)*, II-181.

Bozkurt, E., Erzin, E., Erdem, Ç. E. ve Erdem, A. T., 2011, Formant position based weighted spectral features for emotion recognition, *Speech Communication*, 53 (9-10), 1186-1197.

Breazeal, C. ve Aryananda, L., 2002, Recognition of Affective Communicative Intent in Robot-Directed Speech, *Autonomous Robots*, 12 (1), 83-104.

Burkhardt, F., Paeschke, A., Rolfes, M. ve Sendlmeier, W., 2005, A Database of German Emotional Speech, *Proc. Interspeech*,.

Burkhardt, F., Polzehl, T., Stegmann, J., Metze, F. ve Huber, R., 2009, Detecting Real Life Anger, *2009 Ieee International Conference on Acoustics, Speech, and Signal Processing, Vols 1- 8, Proceedings*, 4761-4764.

C. Becchetti; L. P. Ricotti, 2004, Speech Recognition: Theory and C++ Implementation,, p.

Castellano, G. ve Fanelli, A. M., 2000, Fuzzy inference and rule extraction using a neural network, *Neural Network World Journal*, 3, 361-371.

Castro, L. N. d. ve Timmis, J., 2002, An artificial immune network for multimodal function optimization, *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, 699-704 vol.691.

Ceylan, R. ve Özbay, Y., 2007, Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network, *Expert Systems with Applications*, 33 (2), 286-295.

Chollet, F., 2017, Deep Learning with Python, Buzdağı Yayınevi p. 361.

Cortes, C. ve Vampik, V., 1995, Support vector machine, p.

Davy, M., Gretton, A., Doucet, A. ve Rayner, P. J., 2002, Optimized support vector machines for nonstationary signal classification, *IEEE Signal Processing Letters*, 9 (12), 442-445.

de Castro, L. N. ve Timmis, J., 2002, An artificial immune network for multimodal function optimization, *Cec'02: Proceedings of the 2002 Congress on Evolutionary Computation, Vols 1 and 2*, 699-704.

Demirci, M. D., 2005, BİLGİSAYAR DESTEKLİ SES TANIMA SİSTEMİ TASARIMI, *İSTANBUL ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ*.

Deng, J., Berg, A. C., Kai Li ve Fei-Fei, L., 2010, What Does Classifying More Than 10,000 Image

Categories Tell Us? 11th European Conference on Computer Vision: Part V, ECCV'10. Berlin, Heidelberg, Springer-Verlag**:** 71-84.

Deng, L. ve Yu, D., 2014, Deep Learning: Methods and Applications, *Foundations and Trends® in Signal Processing*, 7 (3–4), 197-387.

Dunn, J. C., 1973, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters., *J Cybern*, 3, 32-57.

Durukal, M., 2015, Performance Analysis of MFCC Features On Emotion Recognition from Speech, *International Journal of Scientific and Technological Research*, Vol 1 (No.5).

El Ayadi, M., Kamel, M. S. ve Karray, F., 2011, Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern Recognition*, 44 (3), 572-587.

Elmas, Ç., 2003, Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama), *Ankara*, Seçkin Yayıncılık, p.

F., S., G., L. ve D., Y., 2011, Conversational Speech Transcription Using Context-Dependent Deep Neural Networks. In Interspeech 2011**:** 437-440.

Fayek, H. M., Lech, M. ve Cavedon, L., 2017, Evaluating deep learning architectures for Speech Emotion Recognition, *Neural Networks*, 92, 60-68.

France, D. J., Shiavi, R. G., Silverman, S., Silverman, M. ve Wilkes, D. M., 2000, Acoustical properties of speech as indicators of depression and suicidal risk, *Ieee Transactions on Biomedical Engineering*, 47 (7), 829-837.

Furui, S., 2000, Digital Speech Processing, Synthesis, and Recognition - p.

George, D., Ranzato, M. A., Mohamed, A.-r. ve Hinton, G., 2010, Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine. In NIPS'2010.

Gilke, M., Kachare, P. H., Kothalikar, R., Rodrigues, V. P. ve Pednekar, M., MFCC-based Vocal Emotion Recognition Using ANN.

Go, H. J., Kwak, K. C., Lee, D. J. ve Chun, M. G., 2003, Emotion recognition from the facial image and speech signal, *Sice 2003 Annual Conference, Vols 1-3*, 2890-2895.

Graps, A., 1995, An Introduction to Wavelets, *IEEE Comput. Sci. Eng.*, 2 (2), 50–61.

Hamzaçebi, C., 2011, Yapay Sinir Ağları - Tahmin Amaçlı Kullanımı MATLAB ve NEUROSOLITIONS Uygulamalı, *BURSA*, Ekin Basım Yayın Dağıtım p. 133.

Hanilçi, C., 2007, Konuşmacı Tanıma Yöntemlerinin Karşılaştırmalı Analizi, *Uludağ Üniversitesi*, Bursa, 72.

Harar, P., Burget, R. ve Dutta, M. K., 2017, Speech Emotion Recognition with Deep Learning, *2017 4th International Conference on Signal Processing and Integrated Networks (Spin)*, 137-140.

Hassan, A., 2012, On Automatic Emotion Classification Using Acoustic Features, *UNIVERSITY OF SOUTHAMPTON*, 2012, 169.

Hinton, G. E., 2007, Learning multiple layers of representation, *Trends in Cognitive Sciences*, 11 (10), 428-434.

Howard, I., 1991, Speech fundamental period estimation using pattern classification., London.

I. Engberg ve A. Hansen, 1996, Documentation of the Danish Emotional Speech Database, *database des /http://cpk.auc.dk/tb/speech/Emotions/S*.

Itakura, F., 1975, Minimum Prediction Residual Principle Applied to Speech Recognition *IEEE TRANSACTIONS ON ACOUSTICS, SPICIPCH, AND SIGNAL PROCESSING* (ASSP-23), 67-72.

Karimi, S. ve Sedaaghi, M. H., 2016, How to categorize emotional speech signals with respect to the speaker's degree of emotional intensity, *Turkish Journal of Electrical Engineering & Computer Sciences*, 24, 1306-1324.

Kavzoğlu, T. ve Çölkesen, İ., 2010, Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi *Harita Dergisi*, 144.

Kohonen, T., 1982, the self-organized formation of topography correct feature maps *biological cybernetics*, 43, 59-69.

Kohonen, T., 1988, An Introduction to Neural Computing, *Neural Networks*, 1 (1), 3-16.

Koolagudi, S. G., Murthy, Y. V. S. ve Bhaskar, S. P., 2018, Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition, *International Journal of Speech Technology*, 21 (1), 167-183.

Kotti, M. ve Paternò, F., 2012, Speaker-independent emotion recognition exploiting a psychologically-inspired binary cascade classification schema, *International Journal of Speech Technology*, 15 (2), 131-150.

Krizhevsky, A., Sutskever, I. ve Hinton, G., 2012, ImageNet Classification with deep Convoutional Neural Network In NIPS'2012.

Landis, J. R. ve Koch, G. G., 1977, The Measurement of Observer Agreement for Categorical Data, *Biometrics*, 33 (1), 159-174.

LeCun, Y., Bengio, Y. ve Hinton, G., 2015, Deep learning, *Nature*, 521 (7553), 436-444.

Lee, C. C., Mower, E., Busso, C., Lee, S. ve Narayanan, S., 2011, Emotion recognition using a hierarchical binary decision tree approach, *Speech Communication*, 53 (9-10), 1162-1171.

Lulia Lefter, L. M. R., DAvid A. van Leeuwen,Pascal Wiggers, 2011, Automatic Stress Detection in Emergency (telephone) calls, *International J. Intelligent Defence Support Systems*, 4 (2), 148-168.

Luna, A. E. V., Jaramillo-Núñez, A. t., Sanchez-Lucero, D., Ortiz-Lima, C. M., Aguilar-Soto, J. G., Flores-Gil, A. ve May-Alarcon, M., 2011, De-Noising Audio Signals Using MATLAB Wavelets Toolbox.

M. Liberman, K. Davis, M. Grossman, N. Martey ve Bell, J., 2002, Emotional prosody speech and transcripts *University of Pennsylvania Linguistic Data Consortium*, *https://www.ldc.upenn.edu/*, *Ocak 2016*

Ma, X., Wu, Z., Jia, J., Xu, M., Meng, H. ve Cai, L., 2018, Emotion Recognition from Variable-Length Speech Segments Using Deep Learning on Spectrograms Interspeech 2018. Hyderabad.

Martin, O., Kotsia, I., Macq, B. ve Pitas, I., 2006, The eNTERFACE'05 Audio-Visual Emotion Database. Proceedings of the 22nd International Conference on Data Engineering Workshops, IEEE Computer Society**:** 8.

Matteucci, M., 2002, http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html:

Mc Culloch W., S. ve Pitts W., A., 1943, a logical calculus of the ideas immantent in nerveus activity, *Bull. Matt. Biophysics*, 5, 115-133.

McKinney, N. P., 1965, Laryngeal freguency analaysis for lingusitic research. , *University of michigan, Communication Sciences Lab.* .

Milton, A. ve Tamil Selvi, S., 2014, Class-specific multiple classifiers scheme to recognize emotions from speech signals, *Computer Speech & Language*, 28 (3), 727-742.

Mirsamadi, S., Barsoum, E. ve Zhang, C., 2017, Automatic Speech Emotion Recognition Using Recurrent Neural Networks with Local Attention, *2017 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp)*, 2227-2231.

Mitchell, T. M., 1999, Machine learning and data mining, *Communications of the Acm*, 42 (11), 30-36.

Montano, B. R., Yoon, V., Drummey, K. ve Liebowitz, J., 2008, Agent learning in the multi-agent contracting system [MACS], *Decision Support Systems*, 45 (1), 140-149.

Nasukawa, T. ve Yi, J., 2003, Sentiment analysis: capturing favorability using natural language processing. Proceedings of the 2nd international conference on Knowledge capture. Sanibel Island, FL, USA, ACM**:** 70-77.

Navarro, M., Corchado, J. M. ve Demazeau, Y., 2016, MUSIC-MAS: Modeling a harmonic composition system with virtual organizations to assist novice composers, *Expert Systems with Applications*, 57, 345-355.

Nizam, H. ve Akın, S. S., 2014, Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması, *XIX. Türkiye'de İnternet Konferansı*.

Nwe, T. L., Foo, S. W. ve De Silva, L. C., 2003, Speech emotion recognition using hidden Markov models, *Speech Communication*, 41 (4), 603-623.

Ozseven, T., 2018, The Acoustic Cues of Fear: Investigation of Acoustic Parameters of Speech Containing Fear, *Archives of Acoustics*, 43 (2), 245-251.

Özkan, Y. ve Erol, Ç. S., 2015, Bioenformatik DNA Mikrodizi Veri Madenciliği, *İstanbul*, Papatya Yayıncılık, p. 432.

Park, J., Kim, J. ve Oh, Y., 2009a, Feature vector classification based speech emotion recognition for service robots, *IEEE Transactions on Consumer Electronics*, 55 (3), 1590-1596.

Park, J. S., Kim, J. H. ve Oh, Y. H., 2009b, Feature Vector Classification based Speech Emotion Recognition for Service Robots, *Ieee Transactions on Consumer Electronics*, 55 (3), 1590-1596.

Polat, G., 2007, Duygu Sınıflandırma problemlerine öznitelik seçim Algoritmaları Kullanarak Sonof Aırt edebişlme Kabiliyetinin İyileştirimesi, Master, *Niğde üniversitesi*, 90.

Ray, S. ve Chan, A., 2001, Automatic feature extraction from wavelet coefficients using genetic algorithms, *Neural Networks for Signal Processing Xi*, 233-241.

Reddy, A. P. ve Vijayarajan, V., 2020, Audio compression with multi-algorithm fusion and its impact in speech emotion recognition, *International Journal of Speech Technology*.

Rioul, O. ve Vetterli, M., 1991, Wavelets and signal processing, *Ieee Signal Processing Magazine*, 8 (4), 14-38.

Rumelhart, D. E., Hinton, G. E. ve Williams, R. J., 1986a, Learning Representations by Back-Propagating Errors, *Nature*, 323 (6088), 533-536.

Rumelhart, D. E., McClelland, J. L. ve Group, P. R., 1986b, Paralel Distributed Processing: Exploration in the Microstructure of Cognition, *Cambridge* MIT Press, p.

Sainath, T. N., Kingsbury, B., Sindhwani, V., Arisoy, E. ve Ramabhadran, B., 2013, Low-rank matrix factorization for Deep Neural Network training with high-dimensional output targets, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 6655-6659.

Schuller, B., Rigoll, G. ve Lang, M., 2004, Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - Belief Network architecture, *2004 Ieee International Conference on Acoustics, Speech, and Signal Processing, Vol I, Proceedings*, 577-580.

Schuller, B. W., Müller, R., Lang, M. K. ve Rigoll, G., 2005, Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles, *19th Annual Conference of the International Speech Communication Association (Interspeech 2018), Vols 1-6*.

Shahzadi, A., Ahmadyfard, A., Harimi, A. ve Yaghmaie, K., 2015, Speech emotion recognition using nonlinear dynamics features, *Turkish Journal of Electrical Engineering & Computer Sciences*, 23, 2056-2073.

Sharma, V., Prakash, N. R. ve Kalra, P., 2019, Audio-video emotional response mapping based upon Electrodermal Activity, *Biomed. Signal Process. Control.*, 47, 324-333.

Slaney, M. ve McRoberts, G., 2003, BabyEars: A recognition system for affective vocalizations, *Speech Communication*, 39 (3-4), 367-384.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. ve Salakhutdinov, R., 2014, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, 15, 1929-1958.

Stone, P. ve Veloso, M., 2000, multi-agent systems a survey from a machine learning perspective, *To Appear in Autonomous Robotics*, 8 (3).

Sulubacak, U. ve Eryigit, G., 2018, Implementing universal dependency, morphology, and multiword expression annotation standards for Turkish language processing, *Turkish Journal of Electrical Engineering and Computer Sciences*, 26 (3), 1662-1672.

Sun, L., Zou, B., Fu, S., Chen, J. ve Wang, F., 2019, Speech emotion recognition based on DNN-decision tree SVM model, *Speech Communication*, 115, 29-37.

Sun, Y., Wen, G. ve Wang, J., 2015, Weighted spectral features based on local Hu moments for speech emotion recognition, *Biomedical Signal Processing and Control*, 18, 80-90.

Şeker, A., Diri, B. ve Balık, H. H., 2017, Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme, *Gazi Mühendislik Bilimleri Dergisi*, 3 (3), 47-64.

Şen, Z., 2004, Yapay Sinir Ağları İllkeleri, *İstanbul*, Bilge Kültür Sanat, p.

Taha, M., Helmy, T. ve Alez, R. A., 2007, Multi-agent Based Arabic Speech Recognition, 433-436.

Tezel, G., 2007, Classification of Biomedical Signals via a New Artificial Neural Network with Adaptive Activation Function, *Selcuk University* Konya, 80.

Timmis, J. ve Edmonds, C., 2004, A Comment on Opt-AiNET: An Immune Network Algorithm for Optimisation, Berlin, Heidelberg, 308-317.

Turk, O., Şayli, Ö., Ozsoy, S. A. ve Arslan, L. M., 2004, Türkçe'de Ünlülerin Formant Analizi, *18. Ulusal Dilbilim Kurultayı*, Ankara Üniversitesi.

Übeylı, E. D. ve Güler, I., 2004, Spectral analysis of internal carotid arterial Doppler signals using FFT, AR, MA, and ARMA methods, *Computers in biology and medicine*, 34 (4), 293-306.

van Deemter, K., Krenn, B., Piwek, P., Klesen, M., Schröder, M. ve Baumann, S., 2008, Fully generated scripted dialogue for embodied agents, *Artificial Intelligence*, 172 (10), 1219-1244.

Vaseghi, S. V., 2000, Advanced Digital Signal Processing and Noise Reduction, *2000 John Wiley & Sons Ltd*, p.

Verma, G. K. ve Tiwary, U. S., 2014, Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals, *Neuroimage*, 102, 162-172.

Ververidis, D. ve Kotropoulos, C., 2003, A state of the art review on emotional speech databases, *Proceedings of 1st Richmedia Conference*, 109-119.

Villerreal, M. R., 2017, Respiratory system complete tr.svg, https://commons.wikimedia.org/wiki/File:Respiratory_system_complete_tr.svg:

Weiss, G., 2013, MULTIAGENT SYSTEMS, MIT Press, p.

Wu, S. Q., Falk, T. H. ve Chan, W. Y., 2011, Automatic speech emotion recognition using modulation spectral features, *Speech Communication*, 53 (5), 768-785.

Yang, B. ve Lugger, M., 2010, Emotion recognition from speech signals using new harmony features, *Signal Processing*, 90 (5), 1415-1423.

Yann, L., 1989, Generalization and Network Design Strategies, *University of Toronto*.

Yin, W. F., Yang, X. Z., Li, L., Zhang, L., Kitsuwan, N., Shinkuma, R. ve Oki, E., 2019, Self-adjustable domain adaptation in personalized ECG monitoring integrated with IR-UWB radar, *Biomedical Signal Processing and Control*, 47, 75-87.

Zhang, S., Zhang, S., Huang, T. ve Gao, W., 2018a, Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching, *Ieee Transactions on Multimedia*, 20 (6), 1576-1590.

Zhang, S. Q., Zhang, S. L., Huang, T. J. ve Gao, W., 2018b, Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching, *Ieee Transactions on Multimedia*, 20 (6), 1576-1590.

Zhao, X. ve Zhang, S., 2015, Spoken emotion recognition via locality-constrained kernel sparse representation, *Neural Computing and Applications*, 26 (3), 735-744.

Zia, S., Yüksel, B., Yüret, D. ve Yemez, Y., 2017, RGB-D Object Recognition Using Deep Convolutional Neural Networks, *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 887-894.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	**:**	Semiye DEMİRCAN
Uyruğu	**:**	T.C.
Doğum Yeri ve Tarihi	**:**	Çumra / KONYA
Telefon	**:**	0506 6119430
Faks	**:**	
E-Posta	**:**	sdemircan@ktun.edu.tr

EĞİTİM

Derece		**Adı**	**İlçe**	**İl**	**Bitirme Yılı**
Lise	:	**Atatürk Kız Lisesi**	**Selçuklu**	**KONYA**	**1998**
Üniversite	:	**Selçuk Üniversitesi**	**Selçuklu**	**KONYA**	**2002**
Yüksek Lisans	:	**Selçuk Üniversitesi**	**Selçuklu**	**KONYA**	**2006**
		Selçuk Üniversitesi	**Selçuklu**	**KONYA**	**2009**
Doktora	:	**Konya Teknik Üniv.**	**Selçuklu**	**KONYA**	

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görevi**
2002-2018	**Selçuk Ünv. Müh. Mim. Fakültesi**	**Araş. Gör.**
2018-	**Konya Teknik Ünv. Müh. Ve Doğa Bil. Fakültesi**	**Araş. Gör.**

UZMANLIK ALANI

Makine Öğrenme Algoritmaları
Ses Sinyali İsleme
Duygu Tanıma

YABANCI DİLLER

İngilizce

YAYINLAR*

MAKALELER

1. S. DEMİRCAN & H. KAHRAMANLI ÖRNEK, *Comparison of the Effects of Mel Coefficients and Spectrogram Images via Deep Learning in Emotion Classification*, 2020, Traitement du Signal, Vol. 37, No. 1,

pp. 51-57. https://doi.org/10.18280/ts.370107 **(Science Citation Index Expanded) (Doktora tezinden yapılmıştır)**

2. S. DEMIRCAN & H. KAHRAMANLI ÖRNEK, *Application of ABM to Spectral Features for Emotion Recognition,* Mehran University Research Journal of Engineering and Technology, 2018, 0254-7821, 37, 4, 452-462**. (Emerging Sources Citation Index) (Doktora tezinden yapılmıştır)**

3. S. DEMIRCAN & H. KAHRAMANLI ÖRNEK, Application of fuzzy C-means clustering algorithm to spectral features for emotion classification from speech, Neural Computing and Applications, 2018, 0941-0643, 29, 8, 59-66. **(Science Citation Index Expanded) (Doktora tezinden yapılmıştır)**

4. S. DEMIRCAN, M. AYDIN & S. S. DURDURAN, *Finding optimum route of electrical energy transmission line using multi-criteria with Q-learning*, Expert Systems With Applications, 2011, 0957-4174, 38, 4, 3477-3482. **(Science Citation Index Expanded) (Yüksek Lisans tezinden yapılmıştır)**

BİLDİRİLER

1. S. DEMIRCAN & H. KAHRAMANLI ÖRNEK, (2017). *Application of Opt AiNET Algorithm to Spectral Features for Feature Selection*. International Conference on Engineering Technologies (Tam Metin Bildiri/Sözlü Sunum). **(Doktora tezinden yapılmıştır)**

2. S. DEMIRCAN & H. KAHRAMANLI ÖRNEK, (2017) *Etmen Tabanlı Modelleme ile Duygu Tanıma -Emotion Recognition via Agent-Based Modelling*.2017 25th Signal Processing and Communications Applications Conference (SIU), 2017, pages: 1 - 4, DOI: 10.1109/SIU.2017.7960513**(Doktora tezinden yapılmıştır)**

3. S. DEMIRCAN & H. KAHRAMANLI ÖRNEK, (2016), *An Agent Based Approach for Emotion Recognition Problem, International Conference on Latest Trends in Computer Science and Engineering* (ICLTCSE 2016) Hamburg, Germany, September 4 – 5, 2016 **(Doktora tezinden yapılmıştır)**

4. S. DEMIRCAN & H. KAHRAMANLI ÖRNEK, (2014*) Emotion recognition from Speech signal: Feature extraction and classification*, International Conference and Exhibition on Electronic, Computer and Automation Technologies (ICEECAT'14), May 9-11,2014 Konya, Turkey **(Doktora tezinden yapılmıştır)**

5. S. DEMIRCAN & H. KAHRAMANLI ÖRNEK, (2013). *Feature extraction from speech verifor emotion recognition*. 4th International Conference on Software and Computing Technology, (ICSCT), 2(1), 28-30, KONYA. **(Doktora tezinden yapılmıştır)**

6. S. DEMIRCAN, M. AYDIN & S. S. DURDURAN, *Route Optimization With Q-learning*, Sözlü Sunum, 8th Wseas International Conference On Applied Computer Science (acs 08) Location: Venice, Italy, 21 Kasım 2008, 23 Kasım 2008, 1790-5109, 416. **(Yüksek Lisans tezinden yapılmıştır)**

7. S. DEMİRCAN, A. ARSLAN (2010). *Artificial Immune Based Multi Agent Systems A Pursuit Domain Application*. The 1st International Symposium on Computing in Science Engineering (ISCSE)(8), 532-536. (Tam Metin Bildiri/Sözlü Sunum)(Yayın No:385146) **(Yüksek Lisans tezinden yapılmıştır)**