

T.C.
BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TRAFİĞE İLİŞKİN SESLERİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE
SINIFLANDIRILMASI

Mariem Mine CHEIKH MOHAMED FADEL

Akıllı Ulaşım Sistemleri ve Teknolojileri Anabilim Dalı

Akıllı Ulaşım Sistemleri ve Teknolojileri Programı

TEMMUZ 2022

T.C.
BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TRAFİĞE İLİŞKİN SESLERİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE
SINIFLANDIRILMASI

Mariem Mine CHEIKH MOHAMED FADEL

DANIŞMAN
Dr. Öğr. Üyesi Zeynep ÖZER

Akıllı Ulaşım Sistemleri ve Teknolojileri Anabilim Dalı

Akıllı Ulaşım Sistemleri ve Teknolojileri Programı

TEMMUZ 2022

ONAY

Mariem Mine CHEIKH MOHAMED FADEL tarafından hazırlanan “**Trafiğe İlişkin Seslerin Makine Öğrenmesi Yöntemleri ile Sınıflandırılması**” adlı tez çalışması 22/07/2022 tarihinde yapılan sınavla aşağıdaki jüri tarafından oybirliği/oyçokluğu ile Bandırma Onyedli Eylül Üniversitesi Fen Bilimleri Enstitüsü Akıllı Ulaşım Sistemleri ve Teknolojileri Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

Dr. Öğr. Üyesi Zeynep ÖZER

(Danışman)

Prof. Dr. Mehmet TEKTAŞ

(Üye)

Doç. Dr. Gültekin BASMACI

(Üye)

İmza

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu’nun .../.../...gün ve .../... sayılı kararı ile onaylanmıştır.

Doç. Dr. Abdullah YEŞİL

Enstitü Müdürü

BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI
ETİK BEYANI

Bandırma Onyedi Eylül Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre hazırlamış olduğum “**Trafiğe İlişkin Seslerin Makine Öğrenmesi Yöntemleri ile Sınıflandırılması**” adlı tezimin özgün bir çalışma olduğunu, tez hazırlanırken tüm aşamalarda bilimsel etik ilkelerine uygun davrandığımı, tez kapsamında sunulan tüm verileri bilimsel etik ilkelerine uygun elde ettiğimi, tezde faydalandığım tüm eserlere atıf yaptığımı ve kaynaklar kısmında bu eserleri gösterdiğimi beyan ederim. 22/07/2022

Mariem Mine CHEIKH MOHAMED FADEL

İmza

ÖNSÖZ

Kara yollarındaki araçlardan kaynaklanan gürültü problemi dünya genelinde her geçen gün artış göstermektedir. İstikrarlı kentleşme eğilimi ve nüfus artışı gibi faktörlerde bu problemi daha da zorlu hale getirmektedir. Buna bağlı olarak şehirlerdeki ses ortamlarını izlemeye yönelik çalışmalarda artış göstermiştir. Bu çalışmada trafikte yaygın olarak karşılaşılan 4 farklı araç türüne ait seslerin sınıflandırılması gerçekleştirilmektedir. Kullanılan yöntem kara yollarındaki trafikten kaynaklanan gürültünün değerlendirilmesi için kullanılabileceği gibi trafik ölçümlerinde kullanılan mevcut sensörlere alternatif ucuz bir yöntem olarak da değerlendirilebilir. Bu doğrultuda spektrogram ve periodogram olmak üzere iki farklı 2 boyutlu temsil yönteminin sınıflandırma performansına olan etkisi değerlendirilmiştir.

Bu tez çalışması sırasında, tez konusunun belirlenmesinden başlayarak son aşamaya kadar her konuda benden yardımlarını esirgemeyen danışman hocam Sayın Dr. Öğr. Üyesi Zeynep ÖZER'e, değerli katkılarıyla çalışmama destek veren Sayın Dr. Öğr. Üyesi İlyas ÖZER'e ve Akıllı Ulaşım Sistemleri ve Teknolojileri Programı Anabilim Dalı Başkanı Sayın Prof. Dr. Mehmet TEKTAŞ'a değerli desteği için şükranlarımı sunarım.

ÖZET

YÜKSEK LİSANS TEZİ

TRAFİĞE İLİŞKİN SESLERİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE SINIFLANDIRILMASI

Mariem Mine CHEIKH MOHAMED FADEL

Bandırma Onyeddi Eylül Üniversitesi

Fen Bilimleri Enstitüsü

Akıllı Ulaşım Sistemleri ve Teknolojileri Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Zeynep ÖZER

Temmuz 2022, 69 sayfa

Dünya üzerinde yaşanan nüfus artışı ve kentleşme gibi unsurlara bağlı olarak şehirlerdeki gürültü seviyeleri de rahatsız edici ve hatta insan sağlığına zarar verecek şekilde artmaktadır. Karayollarındaki taşıtlardan kaynaklanan sesler özellikle kentsel alanlar için önemli bir gürültü kaynağıdır. Trafik sıkışıklıklarının artması da karayollarındaki gürültüyü artıran önemli bir etmendir. Bu doğrultuda Akıllı Ulaşım Sistemleri, yerel trafik durumlarına uyum sağlayarak trafik sıkışıklıklarını azaltmak ve optimize etmek için etkili karşı önlemler almaktadır. Geçtiğimiz on yılda, trafik ışığı döngülerini kontrol etme, trafik kazası algılama, lojistik izleme ve diğer akıllı şehir uygulamaları gibi uygulama senaryoları için trafik izleme için çeşitli otomatik yöntemler geliştirilmiştir.

Trafik izleme sistemleri, kamera sensörlerinden, manyetik döngü sensörleri, radyo dalgalarına (Radar) ve ışık dalgalarına (Lidar) dayalı ölçüm sistemlerine kadar trafik akışını ölçmek için çeşitli sensör modalitelerini kullanmaktadır. Bu tür sistemler, geniş alanları kapsayacak şekilde dağıtılmış sensör ağları olarak kurulabilirken, kurulum ve bakım maliyetleri

genellikle yüksektir. Akustik Trafik İzleme(ATİ) modeli ise , müdahaleci olmayan trafik ölçümleri için daha ucuz bir alternatif sağlar ve bu tez çalışmasının odaklandığı nokta olmuştur.

Gürültü kirliliğinin azaltılması amacıyla kentlerdeki ses ortamlarını izlemeye yönelik teknolojilerin gelişimi son yıllarda hız kazanmıştır. Sinyal işleme ve makine öğrenimi tekniklerini kullanarak gerçekleştirilen otomatik ses tanıma bu alanda son yıllarda dikkat çekici bir gelişme kaydetmiştir. Bu tarz sistemlerde, ses olayları, sinyal işleme ve makine öğrenimi yöntemleri kullanarak otomatik olarak tanımlanır. Bu sistemlerinin kabiliyetleri karayollarındaki akustik olayların kaynaklarını belirlemek için kullanılabilir. Öte yandan bu görev çeşitli zorlukları da bünyesinde barındırmaktadır. Bir yoldaki işitilebilir sesler, araçlar geçerken oluşan motorlar, egzozlar, tekerlekler ve hava türbülansı gibi çeşitli ses kaynaklarının birleşimidir. Aynı tipteki ses kaynaklarının içsel değişkenliği ve farklı konumlar ve durumlar için akustik ortamın etkisi göz önüne alındığında, her sınıfın akustik özellikleri büyük çeşitlilik gösterebilir. Bu da ses kaynağının türünü sınıflandırmayı zorlaştırır.

ATİ görevi için kullanılacak el işi özelliklerin belirlenmesi, uzmanlık gerektiren zor ve zaman alıcı bir süreçtir. Son yıllarda bu sorunun üstesinden gelmek için CNN'ler kullanarak 2B zaman-frekans gösterimlerinden elde edilen özniteliklerin kullanımı giderek artmıştır. Burada en uygun 2B gösterimi belirlemek çok önemlidir. Spektrogram ve Mel-spektrogram tabanlı özellikler geçmiş çalışmalarda yaygın olarak kullanılmıştır. Spektrogramlar eşit bant genişliği ve eşit aralıklı frekans bileşenleri sunduğundan, enerjinin düşük frekanslı bileşenlerde yoğunlaştığı görevler için ideal değildir. Mel spektrogramının performansı genellikle benzer görevler için biyolojik olarak esinlenilmiş modellerin performansından daha düşüktür. Ayrıca Mel ölçekli özniteliklerdeki düzgünlük kısıtı, öznitelikler arasındaki varyansı artırarak sınıflandırma performansını olumsuz yönde etkiler.

Bu çalışma, yukarıda bahsedilen problemleri çözmek için logaritmik ölçekte oranı haritası özelliklerinin kullanılması önerilmiştir. Bir işitsel model olan oran haritası, frekans bileşenlerinin oluşturulması sırasında insan işitsel sisteminin frekans seçim özelliğini modellemek için bir gamaton filtresi kullanır. Bu filtre kullanılarak üst frekans bölgesinde geniş frekans bileşenleri, alt frekans bölgesinde dar frekans bileşenleri elde edilir. Böylece yüksek frekanslarda bilgi tamamen kaybolmadan düşük frekans aralığında daha iyi frekans çözünürlüğü sağlanarak daha fazla spektral bilgi elde edilir. Ek olarak, oran haritası yumuşatma özellikleri sağlar. Bunlar özellikler arasındaki farkı azaltarak sınıf ayrılabilirliğini artırır. Ayrıca önerilen model spektral enerjinin yoğun olduğu bölgelere odaklanmak amacıyla bir eşik fonksiyonunun kullanımını da içermektedir. Önerilen yöntemde öncelikle oran haritası gösterimleri

oluřturulur. Bu gsterimlerin dinamik aralıęı daha sonra logaritma fonksiyonu kullanılarak sıkıřtırılır. Bu iřlemden sonra eřik fonksiyonu uygulanır. Ayrıca nerilen modelde 4 farklı IHC ynteminin sınıflandırma performansına olan etkisi kıyaslanmıřtır. Bunlara ilave olarak periodogram ve spektrogram zellikleri kullanılarak da testler gerekleřtirilmiřtir. Kamuya aık bir veri seti zerinde Joergen IHC ynteminin kullanıldıęı model hem apraz doęrulama testlerinde hem geleneksel test ynteminde dięer modellere kıyasla belirgin bir iyileřme saęlamıřtır.

Anahtar kelimeler: Akustik Trafik İzleme, Derin ęrenme, Oran Haritası zellikleri



ABSTRACT

M.Sc. THESIS

CLASSIFICATION OF TRAFFIC SOUNDS BY MACHINE LEARNING METHODS

Mariem Mine CHEIKH MOHAMED FADEL

Bandırma Onyedi Eylül University

Graduate School of Natural and Applied Sciences

Department of Intelligent Transportation Systems and Technologies

Supervisor: Asst. Prof. Dr. Zeynep ÖZER

July 2022, 69 pages

Depending on factors such as population growth and urbanization in the world, noise levels in cities are increasing in a way that is disturbing and even harmful to human health. Noises from vehicles on highways are an important source of noise, especially for urban areas. The increase in traffic jams is also an important factor that increases the noise on the highways. In this direction, Intelligent Transportation Systems take effective countermeasures to reduce and optimize traffic jams by adapting to local traffic situations. In the past decade, several automated methods for traffic monitoring have been developed for application scenarios such as controlling traffic light loops, traffic accident detection, logistics monitoring, and other smart city applications.

Traffic monitoring systems use a variety of sensor modalities to measure traffic flow, from camera sensors, magnetic loop sensors, to measurement systems based on radio waves (Radar) and light waves (Lidar). While such systems can be deployed as distributed sensor networks to cover large areas, installation and maintenance costs are often high. The Acoustic Traffic Monitoring (ATM) model, on the other hand, provides a cheaper alternative for non-intrusive traffic measurement and has been the focus of this thesis.

The development of technologies for monitoring sound environments in cities in order to reduce noise pollution has accelerated in recent years. Automatic voice recognition using signal processing and machine learning techniques has made remarkable progress in this field in recent years. In such systems, audio events are automatically identified using signal processing and machine learning methods. The capabilities of these systems can be used to identify sources of acoustic events on highways. On the other hand, this task also includes various difficulties. Audible noises on a road are a combination of various sound sources, such as engines, exhausts, wheels and air turbulence, created as vehicles pass by. The acoustic characteristics of each class can vary greatly, given the inherent variability of sound sources of the same type and the influence of the acoustic environment for different locations and situations. This makes it difficult to classify the type of sound source.

Determining the handcrafted features to be used for the ATM task is a difficult and time-consuming process that requires expertise. In recent years, the use of features obtained from 2D time-frequency representations using CNNs has increased to overcome this problem. It is very important to determine the optimal 2D representation here. Spectrogram and Mel-spectrogram-based features have been widely used in past studies. Because spectrograms offer equal bandwidth and equally spaced frequency components, they are not ideal for tasks where energy is concentrated in low-frequency components. The performance of the Mel spectrogram is generally inferior to that of biologically inspired models for similar tasks. In addition, the smoothness constraint on Mel-scale features increases the variance among the features and negatively affects the classification performance.

In this study, it is proposed to use ratio map features in logarithmic scale to solve the above-mentioned problems. An audio model, rate mapping uses a gammaton filter to model the frequency selection feature of the human auditory system during the generation of frequency components. By using this filter, wide frequency components in the upper frequency region and narrow frequency components in the lower frequency region are obtained. Thus, more spectral information is obtained by providing better frequency resolution in the low frequency range, without completely losing the information at high frequencies. Additionally, the ratio map provides smoothing features. These increase class separability by reducing the difference between features. In addition, the proposed model includes the use of a threshold function to focus on regions where spectral energy is concentrated. In the proposed method, firstly ratio map representations are created. The dynamic range of these representations is then compressed using the logarithm function. After this process, the threshold function is applied. In addition, the effects of 4 different

IHC methods on the classification performance were compared in the proposed model. In addition to these, tests were carried out using the periodogram and spectrogram features. The model using the Joergen IHC method on a publicly available dataset provided a significant improvement in both cross-validation tests and conventional testing compared to other models.

Keywords: Acoustic Traffic Monitoring, Deep Learning, Rate Map Features



İÇİNDEKİLER

	Sayfa
ÖNSÖZ	I
ÖZET	I
ABSTRACT	IV
İÇİNDEKİLER.....	VII
ŞEKİL LİSTESİ	X
1. GİRİŞ.....	1
2. LİTERATÜR TARAMASI.....	5
3. AKUSTİK ÖZELLİKLER	15
3.1 AKUSTİK OLAY VE ARKA PLAN GÜRÜLTÜSÜ	15
3.2 SES ÖZELLİKLERİ	16
3.3 ZAMAN ALANINDAKİ ÖZELLİKLER	16
3.4 FREKANS ALANINDAKİ ÖZELLİKLER	17
3.5 DİĞER ÖZELLİKLER.....	19
3.6 ÖN İŞLEME.....	19
3.6.1 SİNYAL İŞLEME	19
3.6.2 GÜRÜLTÜ GİDERME.....	20
3.7 ÖZELLİK ÇIKARMA VE SEÇİMİ.....	20
4. DERİN SİNİR AĞLARI	22
4.1 DERİN SİNİR AĞLARI TÜRLERİ	23
4.1.1 Evrişimli Sinir Ağları	23
4.1.2 Tekrarlayan Sinir Ağları	24
4.1.3 Evrişimli Tekrarlayan Sinir Ağları.....	24
4.2 MAKİNE ÖĞRENME	25
4.2.1 Sınıflandırma Teknikleri	29

4.2.2	Karar Ağacı.....	30
4.2.3	Destek vektör makineleri	31
4.2.4	Naive Bayes Sınıflandırıcı.....	32
4.2.5	Tembel Sınıflandırıcı.....	33
4.2.6	Yapay Sinir Ağları.....	33
4.2.7	Diğer Teknikler.....	34
4.3	MAKİNE ÖĞRENME İLKELERİ.....	34
4.3.1	Denetimli ve Denetimsiz Öğrenme.....	35
4.4	EĞİTİM VE TEST VERİLERİ.....	35
4.5	AKUSTİKLER MAKİNE ÖĞRENME.....	36
4.6	MAKİNE ÖĞRENİMİ KAPSAMI	38
4.6.1	Denetimli Öğrenme.....	39
4.6.2	Denetimsiz Öğrenme	40
4.6.3	Güçlendirme Makine Öğrenimi	41
4.6.4	Destek Vektör Makinesi (SVM)	41
4.6.5	Karar Ağaçları.....	41
4.7	REGRESYON ALGORİTMALARI.....	42
5.	VERİ SETİ VE YÖNTEM.....	43
5.1	AKUSTİK ÖZELLİKLER	43
5.1.1	PERİODOGRAM VE SPEKTROGRAM ÖZELLİKLERİ	43
5.1.2	ORAN HARİTASI TEMSİLLERİ	44
5.1.2.1	Ön Vurgulama Süreci.....	44
5.1.2.2	Koklear Filtreleme	44
5.1.2.3	Nöromekanik Transdüksiyon	46
5.1.2.4	Oran Haritası.....	48
5.2	DERİN ÖĞRENME MODELİ.....	48
5.3	PERFORMANS DEĞERLENDİRME METRİKLERİ.....	49

5.4	VERİ SETİ	50
6.	BULGULAR VE TARTIŞMA.....	52
6.1	SPEKTROGRAM VE PERİODOGRAM TESTLERİ.....	52
6.2	ORAN HARİTASI TESTLERİ.....	55
6.2.1	ÇAPRAZ DOĞRULAMA TESTLERİ	56
6.2.2	DİĞER MODELLER İLE KİYASLAMA	58
7.	SONUÇ	61
8.	KAYNAKÇA.....	63



ŞEKİL LİSTESİ

<u>Sekil</u>	<u>Sayfa</u>
Şekil 1-1: Gürültü kaynağının türüne göre insan gürültü rahatsızlığının dağılımı	2
Şekil 4-1: Ağırlıklarla birbirine bağlanan dört düğüm, basit bir sinir ağı oluşturur	22
Şekil 4-2: Bir gizli katmana sahip yapay bir ileri beslemeli sinir ağı	23
Şekil 4-3: Basit bir CRNN	25
Şekil 5-1: Koklea (solda) ve (sağda) korti (büyütülmüş) görüntüleri	46
Şekil 5-2: CNN Mimarisi	48
Şekil 6-1: Karışıklık matrisleri	54
Şekil 6-2: Makro ortalama değerler	55
Şekil 6-3: Tüm sınıflar için kıyaslama metriklerinin ortalaması	58

TABLO LİSTESİ

<u>Tablo</u>	<u>Sayfa</u>
Tablo 5-1: Veri setinin dağılımı	50
Tablo 5-2: Veri setinin nihai halindeki sınıflar ve olay sayıları	51
Tablo 6-1: Sınıf bazlı sonuçlar	53
Tablo 6-2: IHC metodunun sınıf bazlı değerlendirme sonuçları	56
Tablo 6-3: Geleneksel test yöntemiyle elde edilen sonuçlar	60
Tablo 6-4: Literatürdeki F-Skor Değerleri	60
Tablo 6-5: VGGNet Karışıklık matrisi	60

1. GİRİŞ

Dünya çapında yaşanan nüfus artışı, istikrarlı kentleşme eğilimi ve insanların kırsal alanlardan daha büyük şehirlere taşınması gibi faktörler şehirlerdeki nüfusu her geçen gün artırmaktadır. Buna bağlı olarak şehirlerdeki gürültü seviyeleri de rahatsız edici ve hatta insan sağlığına zarar verecek şekilde artmaktadır. Sonuç olarak gürültü problemi, şehir sakinlerinin en büyük şikayetleri arasında yer almaktadır (Hubel et al., 2022). Karayollarındaki taşıtlardan kaynaklanan gürültüler kentsel alanlardaki gürültünün önemli bir bölümünü oluşturur. Ayrıca kentleşmenin ve buna bağlı olarak nüfus yoğunluğunun artması araçlarla yapılan seyahat sürelerini, gürültü ve hava kirliliği seviyeleri de artırır.

Akıllı ulaşım sistemleri (AUS), yerel trafik durumlarına uyum sağlayarak trafik sıkışıklıklarını azaltmak ve optimize etmek için etkili karşı önlemlerdir. Geçtiğimiz on yılda, trafik ışığı döngülerini kontrol etme, trafik kazası algılama, lojistik izleme ve diğer akıllı şehir uygulamaları gibi uygulama senaryoları için trafik izleme için çeşitli otomatik yöntemler geliştirilmiştir.

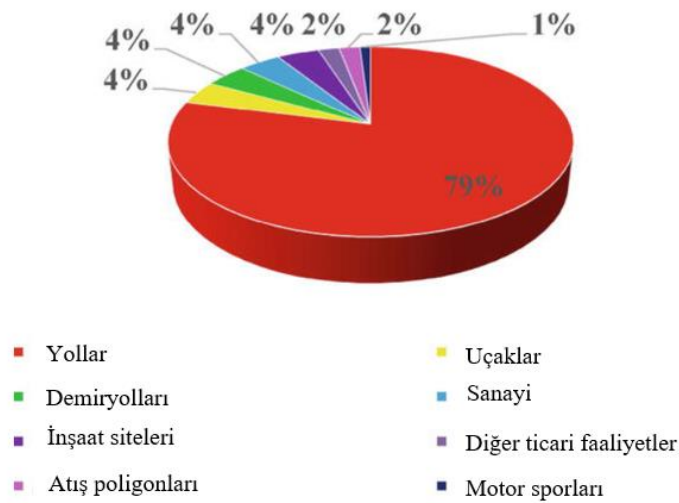
Trafik izleme sistemleri, görsel nesne algılama ve izleme için kamera sensörlerinden, geçen araçları saymak için manyetik döngü sensörlerinden radyo dalgalarına (Radar) ve ışık dalgalarına (Lidar) dayalı ölçüm sistemlerine kadar trafik akışını ölçmek için çeşitli sensör modalitelerini kullanır. Bu tür sistemler, geniş alanları kapsayacak şekilde dağıtılmış sensör ağları olarak kurulabilirken, kurulum ve bakım maliyetleri genellikle yüksektir. Akustik trafik izleme (ATİ), müdahaleci olmayan trafik ölçümleri için daha ucuz bir alternatif sağlar ve bu tez çalışmasının odak noktasıdır.

Hali hazırda gürültü kirliliği özellikle kentsel alanlarda önemli bir sorundur ve ayrıca trafik gürültüsü şehirlerdeki en önemli gürültü kaynağıdır. İçten yanmalı motorlara sahip çok sayıda otomobil ve diğer karayolu taşıtları, karayolu trafik gürültüsünü önde gelen bir gürültü kirliliği kaynağı haline getirmektedir. Günümüzde yerini yavaş yavaş almaya başlayan elektrikli ve hibrit otomobiller, motorları genellikle sessiz olduğu için kentsel alanlarda daha düşük gürültü düzeyine neden olmaktadır (Zakir et al., 2010). Ancak, yalnızca içten yanmalı motorların yokluğu, kentsel alanlarda gürültü seviyelerini düşürmenin tek ölçüsü olamaz.

Avrupa Çevre Ajansı (EAA) tarafından sağlanan verilere göre, “41 milyon kişinin şehir içi trafik gürültüsü nedeniyle 55 dB Lden'nin üzerinde maruz kaldığı bildirilmektedir” ve Avrupa'da yaklaşık 90 milyon olduğu tahmin edilmektedir (Brambilla et al., 2020). Yol trafik gürültüsü en yaygın olanıdır. Gürültü azaltma eylemlerine rağmen kentsel alanlarda gürültü kaynağı, mekan ve

zaman içinde geniş yayılım uluslararası ve ulusal düzeyde politikalar ve mevzuatlar tarafından uygulanmaktadır. Sonuç olarak, insanların karayoluyla üretilen hem doğrudan hem de dolaylı olumsuz sağlık etkilerine karşı farkındalığı trafik gürültüsü artıyor. Dünya Sağlık Örgütü (WHO), “Avrupa'nın batı kesiminde trafik kaynaklı gürültüden her yıl en az bir milyon sağlıklı yaşam yılı kaybedildiğini” tahmin etmektedir (World Health Organization, 2011). Literatürde “çoğunlukla karayolu trafiği ile ilgili uyku bozukluğu ve rahatsızlığı” olduğuna dair kanıtlar vardır. Gürültü, çevresel gürültünün ana yükünü oluşturmaktadır”.

Günümüzde gürültü kirliliği, insan sağlığı ve iş verimliliği üzerindeki kanıtlanmış önemli etkisi nedeniyle çeşitli çalışmaların ve araştırmaların odak noktasıdır. Araştırmalar, kentsel alanlardaki trafik gürültüsünün, özellikle insan ve mal taşımacılığının artması nedeniyle yüzyılın başından beri muazzam bir şekilde arttığını göstermektedir. Kentsel alanlarda en büyük gürültü kaynağının, tüm ortak gürültü kaynaklarının %80'ini oluşturan trafik kaynaklı gürültü olduğu sonucuna varılabilir (Aktürk vd., 2003). Karayolu trafiğinden kaynaklanan trafik gürültüsü, kentsel alanlarda en yaygın gürültü türüdür ve bu nedenle ciddi bir sorun teşkil etmektedir., gürültü kaynağının türüne göre insan gürültüsü rahatsızlığının dağılımını göstermektedir.



Şekil 1-1: Gürültü kaynağının türüne göre insan gürültü rahatsızlığının dağılımı.

Gürültü kirliliği, genel yaşam kalitesini etkileyen ciddi bir sorundur. Bu sorun, özellikle trafik tarafından önemli miktarda gürültü kirliliğinin üretildiği kentsel alanlarda fark edilir (Wang et al., 2022). Uzun süreli gürültüye maruz kalmanın ciddi sonuçları göz önünde bulundurularak, gürültü seviyelerini azaltmak için en azından bazı önlemlerin uygulanması gerekmektedir. Bugün bu sorunun nasıl çözüleceğine dair birçok girişim ve plan bulunmaktadır; ancak bu tez, doğrudan

Makine öğrenimi ile bağlantılı önlemlere odaklanmıştır. Bu çalışma, makine öğrenmesi tabanlı trafik gürültüsü sınıflandırma yöntemlerini araştırmaktadır.

Son yıllarda gürültü kirliliğinin azaltılması amacıyla kentlerdeki ses ortamlarını izlemeye yönelik teknolojilerin gelişimi hızlandı(Ozer vd., 2018; Lecun vd., 1998). Sinyal işleme ve makine öğrenimi tekniklerini kullanan otomatik ses tanıma (OST), bu alanda son yıllarda dikkat çekici bir şekilde gelişen bir teknoloji olarak öne çıkmaktadır. Bir OST sisteminde, ses olayları, sinyal işleme ve makine öğrenimi yöntemleri kullanılarak otomatik olarak tanımlanır. OST sistemlerinin bu kabiliyetleri karayollarındaki akustik olayların kaynaklarını belirlemek için kullanılabilir. Öte yandan bu görev çeşitli zorlukları da bünyesinde barındırmaktadır. Bir yoldaki işitilebilir sesler, araçlar geçerken oluşan motorlar, egzozlar, tekerlekler ve hava türbülansı gibi çeşitli ses kaynaklarının birleşimidir (Hubel vd., 1968). Aynı tipteki ses kaynaklarının (örneğin arabalar) içsel değişkenliği ve farklı konumlar ve durumlar için akustik ortamın (örneğin yankılanma, mesafe) etkisi göz önüne alındığında, her sınıfın akustik özellikleri büyük çeşitlilik gösterebilir. Bu da ses kaynağının türünü sınıflandırmayı zorlaştırır.

Ses tanıma aslında sinyal ön işleme, özellik çıkarma ve sınıflandırma olmak üzere 3 temel aşamadan oluşan bir örüntü tanıma problemidir. Sinyal ön işleme aşamasında, giriş sinyali genellikle daha küçük segmentlere bölünür ve özellik çıkarımına hazır hale getirmek için hamming penceresi gibi bir pencere işlevi uygulanır. Özellik çıkarma, karmaşık verilerin boyutunu küçülterek bir özellik vektörü olarak temsilidir. Son olarak elde edilen özellikler bir sınıflandırma algoritması kullanılarak sınıflandırılır.

Ses tanıma görevlerinde prozodik özellikler (Rao vd., 2013), ses kalitesi (Valstar vd., 2016), Teager enerji operatörü (TEO) (Bandela vd., 2019) ve spektral özellikler gibi birçok farklı özellik kullanılabilir. Bununla birlikte son yıllarda geleneksel el ile seçilmiş özelliklerin yanı sıra 2 boyutlu zaman-frekans gösterimleri de yaygın olarak kullanılmaya başlanmıştır. Bu gösterimler üzerinden CNN ile özellik çıkarımı ve sınıflandırması gerçekleştirilmektedir. Böylelikle el ile seçilmiş özelliklerinin belirlenmesindeki uzmanlık ve zaman gerektiren süreçler önemli ölçüde azaltılmaktadır. Fakat 2 boyutlu zaman-frekans gösterimleri kendi içerisinde çeşitlilik göstermektedir ve bunlardan hangisinin en uygun özellik olduğuna dair genel bir kanı bulunmamaktadır.

İnsan işitme sistemi çevresel sesleri ayırt etme konusunda oldukça başarılıdır. Dolayısıyla insan işitme sistemini modelleyen yaklaşımlar bir ATİ görevi için başarılı olabilir. Bu çalışmada (Dau et al., 1998) inner hair-cell modeli tabanlı rate-map özelliklerini kullanan bir ATİ sistemi önerilmiştir. İnsan kokleasındaki mekanik transdüksiyon sürecinin bir sonucu olarak, işitsel sinir

ateşlemesi meydana gelir ve rate-map özellikleri, bu ateşleme işlemi hızının spektro-zamansal temsilini kodlar. Oran haritası özellikleri dört temel aşamada oluşturulur (Brown vd., 1994). İlk olarak, ses sinyaline ön vurgulama işlemi uygulanır. Daha sonra Koklear filtreleme gammatone filtresi (GTF) kullanılarak gerçekleştirilir. Bir sonraki aşamada, baziler membran segmentinin çıkışındaki sinyal, yarım dalga doğrultulmuş ve 1 kHz'de düşük geçişli filtrelenmiştir. Bu aşama, baziler membranın mekanik salınımlarının iç tüy hücrelerinde reseptör potansiyellerine dönüşümünü kabaca simüle eder. Son olarak, Hamming penceresi ile meydana gelen ani artışlar entegre edilerek oran haritası özellikleri elde edilir.

GTF, üst frekans bölgesinde geniş frekans bileşenleri ve alt frekans bölgesinde dar frekans bileşenleri sunar (Sharan vd., 2019). Alt frekans aralığında sağlanan daha kesin frekans çözünürlüğü, üst frekans aralığında önemli spektral bilgileri kaybetmeden daha fazla spektral bilgi elde etmede faydalıdır (Sharan vd., 2019; Sharan vd., 2015). Bu, OST görevi gibi düşük frekans aralığında yüksek enerji seviyelerine sahip uygulamalar için spektrogram ve Mel-spektrogram gibi 2 boyutlu zaman-frekans gösterimlerine göre bir avantaj sağlar (Sharan vd., 2019; Sharan vd., 2020). Bu çalışmada ilk olarak WAV dosyaları normalleştirmeden önceki durumlarına ölçeklendi. Daha sonra yukarıda bahsedildiği gibi oran haritası özellikleri elde edilmiştir. Bu doğrusal hız haritalarının dinamik aralığı log fonksiyonu kullanılarak sıkıştırılmış ve log-güç hız haritaları elde edilmiştir. Bu adımı takiben, hız haritalarının güçlü olduğu bölgelere odaklanmak için doygun olmayan doğrusal olmayan bir fonksiyon uygulanmıştır. Bu işlem spektral tepelere odaklanmak için gerçekleştirilir. Literatürdeki çalışmalar, spektral tepelerin spektral vadilere kıyasla sağlamlık açısından göreceli önemini ortaya koymaktadır (Josifovski, 2002). Spektral vadilerdeki bilgiler, daha önce gürültü ile dolduruldukları için sağlam olmayacaktır. Elde edilen hız haritası özellikleri, sabit artışlarla gri tonlamalı olarak normalleştirildi. Bu gri tonlamalı görüntü, KYM (kırmızı, yeşil, mavi) bileşenlerine nicelleştirildi. Bu görüntüler, OST görevi için bir CNN kullanılarak sınıflandırıldı. Evrişimli sinir ağları, son yıllarda görüntü ile ilgili birçok uygulamada oldukça başarılı bir şekilde kullanılmıştır. Bununla birlikte, sabit boyutlu girdilere olan ihtiyaç nedeniyle CNN'ler, konuşma sinyallerinin doğasına aykırı olarak önemli bir eksikliğe sahiptir. Bu nedenle tüm monokrom görüntüleri aynı boyuta ayarladıktan sonra CNN'ye uyguladık. Önerilen yaklaşım, kamuya açık olarak yayınlanmış bir veri seti üzerinde test edildi ve oldukça başarılı sonuçlar elde edildi.

2. LİTERATÜR TARAMASI

Akıllı ulaşım sistemi, modüler bir yaklaşımı benimseyen ancak birbiriyle uyumlu çalışan birçok akıllı süreçten oluşur. Araç trafiği tıkanıklığını anlamak, etkili hareketlilik ve yüksek kaliteli trafik yönetimi ve güvenlik sistemleri için bir anahtardır. Ortaya çıkan trafik tıkanıklığı, ağlardaki trafik akışları üzerinde ters etkiye sahiptir. Daha ampirik bir yaklaşım, yoldaki tıkanıklığın ani bir arıza nedeniyle olduğunu söyler. Araç hızı ani bir şekilde azalır ve başlangıçta bir otoyol trafik yolunda araç yoğunluğu artar. Sonraki araştırma pini, trafik ağlarının çok karmaşık uzaysal zamansal davranışına işaret etmektedir. Trafik arızasının ampirik özelliklerini ve sonuçta ortaya çıkan tıkanıklığı açıklamak için bir trafik modeline ihtiyaç vardır. Bunu açıklamak için çok sayıda model ve teori geliştirilmiştir. Bu tezin çalışmasının temel amacı, akustik özellikleri kullanarak derin öğrenme modelleriyle trafikteki farklı türdeki araçları sınıflandırmaktır. Böylelikle ATİ alanındaki çalışmalara bir katkı sunmaktır.

Nüfusun hızla artmasıyla birlikte toplum kentleşme ve hızlı sanayileşme yönünde ilerlemektedir. Bu nedenle, trafik sıkışıklığının azaltılmasının ABD Ulaştırma Bakanlığı'nın en önemli direktiflerinden biri olması şaşırtıcı değildir. Yol tıkanıklığı olaylarını tespit etmek ve takip etmek için, birkaç araştırmacı tarafından derin öğrenmeye dayalı bazı yaklaşımlar uygulanmaktadır. Bununla birlikte, bunu yapmanın ilk ve en önemli adımı, trafiğin nasıl aktığını ve çeşitli koşullar altında belirli yol ağlarındaki tıkanıklığın kısa ve uzun vadeli etkilerinin ne olacağını anlamak olacaktır. Trafik izleme daha büyük bir ölçekte gerçekleştiğinden, incelenen tüm yol ağları için trafik parametreleri hakkında gerçek zamanlı geri bildirim sağlayacak, ölçeklenebilir, merkezi olmayan bir sisteme sahip olmak eşit derecede önemlidir. Yaklaşımlarının ana katkıları, farklı yol ağları arasında ölçeklendirmenin kolay olması ve trafik akışlarının doğasını tahmin etmek için tamamen merkezi olmayan hale getirilebilmesiydi. 90'ların başında ilk kez, trafik akışını modellemek için bir yapay sinir ağı kullanıldı ve o zaman bu yaklaşım kullanılarak tıkanıklık tahmin edilebileceği ortaya kondu. Yazarlar çalışmalarında sırasıyla 1 giriş katmanı, 1 çıkış katmanı ve 1 gizli katmandan oluşan bir ağ tanıtmışlardır. Bu tasarım, trafik akışlarının doğasını tahmin etmede ve doğru seyahat sürelerini tahmin etmede daha yüksek doğruluk elde ettiğini kanıtladı. Ancak, temel yapısı, özellikle büyük bir veri kümesi içeren diğer trafik uygulamaları için yetersizdi. Derin öğrenmenin ortaya çıkmasıyla birlikte, temel tek katmanlı sinir ağları, trafik akışı ve tıkanıklık parametrelerini incelemek için tipik olarak kullanılmamaktadır. Modelleri önce uzaysal-zamansal özellikleri öğrenmek için Evrişimsel Sinir Ağı'nı kullanır. Daha

sonra, tekrarlayan bir sinir ağı, tam diziyi kategorize etmeye yardımcı olan ilk adım modellerinin çıktısı kullanılarak eğitilir.

Bazı yeni araştırmalar, tıkanıklığı tespit etmek için tahmini trafik hızlarını kullanmayı araştırmıştır. Bir yol kesiminde tahmin edilen hız, gerçek zamanlı hızdan belirli bir farkla düşük olduğunda, trafik sıkışıklığı olarak işaretlenir. Modellerini kullanarak ve uzamsal-zamansal trafik bilgilerini entegre ederek, şehrin çevre yollarındaki tıkanıklık kaynaklarını tespit edebildiler. Genel olarak, çoğu tıkanıklık algılama sistemi, döngü dedektörlerinden veya mikrodalga radar sensörlerinden gerçek zamanlı trafik verileri alır. Bu veri akışları daha sonra hız-hacim ve doluluk ilişkilerine dayalı tıkanıklığın başlangıcını algılayan ve büyüdükçe ve dağıldıkça onları kesintisiz olarak izleyen algoritmalar aracılığıyla kabul edilir.

Video veya görüntü tabanlı tıkanıklık algılama çerçeveleri, trafik akışı verilerinin toplanması için kullanılan geleneksel sensörlerin sınırlamaları nedeniyle de popülerlik kazanmıştır. Derinlemesine öğrenilmiş bir sınıflandırıcı sistemin trafik sıkışıklığını yaklaşık %95 doğrulukla tespit edebildiğini göstermişlerdir. Çalışmada bunun için, araçları tanımak için kümeleme, arka plan çıkarma ve bölütleme gibi görüntü işleme tekniklerini kullandılar ve son olarak, farklı kavşak konumlarındaki çeşitli kalibre edilmiş kameralar için yaklaşık kuyruk uzunlukları kullandılar. Benzer bir yaklaşım Hao ve arkadaşları tarafından kullanılmıştır. Yazarlar en yüksek benzerlik eşleştirme yöntemini kullandılar ve gerçek zamanlı görüntülerdeki ve ayardaki şerit bölgesindeki sapmaları düzeltmek için statik bir arka plan alanı penceresine odaklandılar.

Araç algılama ve sayma, herhangi bir trafik gözetim sisteminin önemli bir bileşenidir. GPU'ların daha yüksek işlem hızlarıyla, araçları algılamak ve sayıları gerçek zamanlı olarak tablolamak için sağlam, derin öğrenme tabanlı modeller eğitilebilir. Son yıllarda çeşitli geleneksel ve geleneksel olmayan araç algılama ve analiz yöntemleri önerilmiştir. Yazar, farklı trafik senaryolarında hareket eden araçların sayısını tahmin etmek ve tablolamak için sanal döngü tabanlı bir yaklaşım kullandı. Araçları tespit etmek ve sınıflandırmak için ResCeption adlı bir sinir ağı kullandılar. Bu ağ, artık öğrenmeyi, tek bir bakışta araba sayısını algılayabilen ve sayabilen Başlangıç tarzı katmanlarla bütünleştirir. Yaklaşımları, lokalizasyon veya yoğunluk tahmini ile yapılan sayımlara kıyasla doğru araç sayımları elde etmede üstündür.

Algılama örneği sayaçları, her arabayı özel olarak yerleştirir ve ardından yerleştirmeyi sayar. Ancak bu, yerleştirme oluşturmak için tüm görüntünün piksel piksel incelenmesini gerektirecek bir soruna sahip olabilir. Benzer şekilde, dedektör örtüşen nesneleri birleştirebileceğinden tıkanmalar başka bir engel oluşturabilir. Yoğunluk tahmin edicileri,

sayılabilir araçlar için bir yoğunluk tahmini oluşturmaya ve daha sonra onları bu yoğunluk üzerinde özümsemeye çalışarak içgüdüsel bir şekilde çalışır. Genellikle çok fazla eğitim verisi örneğine ihtiyaç duymazlar, ancak yoğunlukla eğitildikleri sahnede sınırlı kalırlar.

Araç algılama, tanıma ve farklı trafik sahnelerindeki araçların takibi alanlarına büyük ilgi vardır. Yaklaşımları, araç tipini sınıflandırmak için araçların geometrik özelliklerini kullanan hareketli nesne bölütleme tekniğini kullanarak hareketli araçları durağan olanlardan ayırdı. Yaklaşımları gerçek zamanlı trafik senaryolarında uygulanabilir ve aksi takdirde tıkanma durumunda tespit edilemeyen araçları tespit edecek kadar sağlamdır. Yazarlar yaklaşımlarında, her seferinde tüm aracı izlemek yerine araçların özelliklerini izleyen bir araç algılama sistemi kullandılar. Böylece kısmi tıkanma sorununu ortadan kaldırmaktadır ve daha sağlamdır. Sistem, araçları hem gündüz hem de gece koşullarında yüksek hassasiyetle algılayabilir. Yaklaşımlarında, araç algılama modelini eğitmek ve trafik videolarından araçları yakalamak için bir Faster-RCNN kullandılar. Önerilen modellerinin doğruluğu, farklı araç sınıfları için %95'in üzerindeydi. Aynı şekilde, hafiften yoğun trafik durumu için trafik akış istatistiği de %90'ın üzerindeydi.

Görüntü izleme bileşenleri, trafik bilgilerini araçların tespit edilen konumlarından ayırt edebilir ve gerçek zamanlı olarak araç takibi yapabilir. Önerilen yaklaşım, çeşitli çevre ve aydınlatma koşulları altında trafiği etkin bir şekilde izleme yeteneğine sahiptir. Önerilen gerçek zamanlı görüş sistemi VISATRAM, araç sayısını sayabilir, hızlarını tahmin edebilir ve ayrıca 3 boyutlu ölçümler yoluyla bunları sınıflandırabilir (Zhu vd., 2000). Araç tespiti ve sınıflandırması için Evrişimli Sinir Ağları (CNN) üzerine yakın zamanda yapılan birkaç çalışma, diğer algoritmalara göre daha yüksek performans kapasitesi üretmiştir.

Ayrıca trafik koşullarını tahmin etmek için de birçok çalışma yapılmıştır. Bunlardan biri, Vlahogianni ve diğerlerine (2014) göre trafik özelliklerinin modellenmesidir, birkaç saniye veya birkaç saate ait mevcut ve geçmiş trafik bilgilerine dayanan tahminle ilgilidir. Kısa vadeli trafik akışı tahmini, trafik yönetiminin dinamik ve yüksek verimli kontrolünü geliştirmek için hayati önem taşır. Yolcu bilgileri ve trafik yönetimi için temel girdiler, kısa vadeli trafik akışı tahmini ile sağlanır (Yang, vd., 2006).

Araç yoğunluğu olarak da bilinen kare alan başına araç sayısı, kısa vadeli trafik akışını tahmin etmedeki değerli doğruluğu nedeniyle iyi bir gösterge olma eğilimindedir. Kısa vadeli trafik akışı tahmin yaklaşımlarının farklı çalışmaları, araç yoğunluğunu belirlemek ve böylece trafik akışını tahmin etmek için gözetim görüntüleri gibi çeşitli değişken özellikler veya parametreler kullanmıştır. Araştırmacıların önceki çalışması, araç sınıflarını saymak ve

sınıflandırmak için Mask R-CNN'yi kullanılmıştır (Piedad vd., 2019). Bu, araştırmacılara araç sınıflandırma koleksiyonlarında yardımcı olmuştur.

Literatürde kısa dönemli trafik akışı tahmini için hertz veya ses frekansı olarak ölçülen ses yoğunluğu kullanılması görece olarak çok daha yeni bir konudur. Bu, ses frekansını kısa dönemli trafik akışını tahmin etmek için yeni bir özellik haline getirir. Trafik gürültüsü veya sesinin araç sayısı ile ilişkisini araştıran ancak kısa dönemli trafik akışını tahmin etmek için kullanmayan benzer çalışmalar vardır. Halim ve Abdullah (2014) tarafından yapılan bir araştırma, yollardaki yoğun trafik akışının, düşük trafik akışına kıyasla daha yüksek gürültü kaydettiğini göstermektedir. Yol kenarlarında kaydedilen ses frekansı, araç yoğunluğu ile doğru orantılı bir ilişki gösterir. Bununla birlikte, bu ilişki yalnızca, özellikle bir yerleşim bölgesinde üç tekerlekli bisikletlerin ürettiği gürültüye odaklanmaktadır. Diğer araç türleri tarafından verilen ses verileri de dahil olmak üzere açık alanlarda ses ölçümü yapıldığında daha doğru sonuçlar elde edildi (Vergel vd., 2004).

Kısa dönemli trafik akışının tahmini, tahmin ve karar vermeyi destekleyen yararlı bilgilere sahip olmak için verileri incelemek, temizlemek, dönüştürmek ve modellemek için verilerin analizini gerektirir. Yapay zekanın temel alt alanlarından biri, verimli ve hızlı algoritmalar ve veriye dayalı modeller geliştirerek doğru sonuçlar ve analizler üreten makine öğrenmesidir. Bu makale, trafiğe ilişkin seslerin makine öğrenmesi yöntemleri ile sınıflandırılmasını esas almıştır.

Ses, bir yandan sürücü üzerinde sakinleştirici bir etki sağlayabilir ve sürüş deneyimini iyileştirebilirken, diğer yandan seslerin insan üzerindeki fizyolojik ve psikolojik etkilerinden dolayı araç içi gürültü trafik kazası olasılığını artırabilir.

Geçtiğimiz yirmi yıl, araç içi gürültüsünün ses kalitesi üzerine artan bir araştırma grubuna tanık oldu. Böylece müşteriler, motor sesi, uyarı sesleri, kapı sesleri, araç ses sistemi vb. araçlarla ilgili seslere karşı oldukça hassas hale geldi. Araç ses özellikleri, müşteri araç modeli tercihini etkileyen en önemli faktörlerden biridir. Araç sesi karakterizasyonundaki önceki çalışmalar, sessizlikten sesin hoşluğuna kadar çeşitli yönleri içerir, ancak öncelikle yeni araçların tasarımı ve üretimi için belirlenen gereksinimlere uygulanmıştır. Ancak, rahatsız edici gürültülerin ortadan kaldırılması ile dinleyicinin belirli bir marka ve model otomobilin ses kalitesine ilişkin beklentileri arasında karmaşık bir ödünleşim vardır.

Sürücülerin güvenliği, sürücü davranışlarını tahmin etme anlayışını genişleten sürücü modelleme çalışmalarıyla on yıllar boyunca iyileştirildi. Çoğunlukla otonom sistemlerde uygulanan görsel algı ve etkileşimli sistemlerde kullanılan sürücü modellemedeki dikkate değer gelişmelere rağmen, bir vibro-akustik sistemin bileşenleri olarak yolcular, sürücü ve araç

düşünüldüğünde önemli bir boşluk bulunmaktadır. Bu çalışma, ses tanıma ve psikoakustik ses değerlendirme alanlarından elde edilen bulgulara dayanmaktadır. Sürücünün duygusal durumunun sürüş güvenliğini önemli ölçüde etkilediği kabul edilmiştir.

Araçta gürültü işleme ile ilgili hemen hemen tüm çalışmalar, bir aracın tasarım süresi boyunca ayarlamalar için gürültü algılama veya azaltma üzerine odaklanmıştır. Ses kalitesi değerlendirme mühendisliğinde çeşitli psikoakustik indeksler tanıtılmış ve çeşitli sürüş senaryolarında değerlendirilmiştir. Nor ve arkadaşları (2008), örneğin, araç konfor indeksini değerlendirmek için hem öznel hem de nesnel testler incelenmiştir. Ses yüksekliği, keskinlik, pürüzlülük ve dalgalanma kuvveti ölçümlerinin, insan denek çalışmalarıyla ilişkili olduğunu buldular. Deneyleri ayrıca akustik konforun her bir ölçümden ne kadar etkilendiğini gösterdi. Nor ve diğerlerinin çalışmasının bir başka ilginç sonucu, yol pürüzlülüğü (örneğin, otoyol, düzgün şehir, çift vagon, kaldırım ve banliyö) ile konfor endeksi arasındaki ilişkinin incelenmesidir.

Duan ve arkadaşları (2015) birden fazla çalışma koşulunda araç iç gürültüsünü incelemiştir: rölanti, sabit hız, hızlanma ve frenleme. Giriş özellikleri olarak dört psikoakustik parametre, ses yüksekliği, keskinlik, artikülasyon indeksi ve A ağırlıklı ses basınç seviyeleri (SPL'ler) ve ağı eğitmek için bir etiket olarak öznel rahatsızlık kullandılar. Rölanti, sabit hız, hızlanma ve frenleme koşullarında araçtan alınan 36 iç gürültü örneğinden oluşan bir veri seti oluşturdular. Ölçüm yöntemi GB/T 18697 standardını takip eder. Sonuçlar %95,57'nin üzerinde bir doğruluk göstermektedir.

Geçmiş yıllarda çeşitli psikoakustik indeksler önerilmiş olsa da, sürüş sırasında bu bilgi tek başına rahatsız edici ses kaynağının belirlenmesi için kullanışlı değildir. Basit bir yaklaşım, ses olaylarını ayırmak ve her birinin rahatsızlığını hesaplamaktır. Ne yazık ki, genel ses olaylarının tespiti ve segmentasyonu hala iyi çözülmüş bir problem değildir.

Mel-frekans cepstrum (MFCC), Spektrogramlar ve el ile seçilmiş özellikler gibi özellik çıkarma ve SVM ve GMM gibi genel sınıflandırıcılardan oluşan geleneksel ardışık düzen, konuşma tanımadan sahne sınıflandırmaya ve olay algılamaya kadar ses tanıma sorunlarının üstesinden gelmek için yaygın olarak kullanıldı. Bununla birlikte, son birkaç yılda, transfer öğrenme teknikleri ve çapraz modlu öğrenme derin öğrenme yöntemindeki gelişmeler sayesinde, çoğu ses tanıma problemi için en gelişmiş sonuçlara ulaşılmaktadır. Ses tanımada zorlu bir görev, tek tek seslerin bileşik seslerde sınıflandırılmasıdır. Bu görev, polifonik ses olayı algılama olarak adlandırılır ve birden çok örtüşen ses olayını algılamaktan oluşur. Transfer öğrenme, çapraz modlu öğrenme ve derin öğrenme yöntemleri dahil olmak üzere çeşitli yaklaşımlar önerilmiştir.

Bununla birlikte, ses örneklerinden öğrenmenin özel bir sorunu, büyük açıklamalı verilerin eksikliğidir. Mevcut ses veri kümelerinin çoğu ortam, olay veya karışık seslerden oluştuğundan, bu eksiklik özellikle araç içinden gelen sesler için vurgulanır. Bir aracın içindeki ses, birbiriyle ilişkisiz ve dinamik birçok faktörden etkilenmenin yanı sıra, insanları (yani sürücüleri) farklı şekilde etkiler. Bu nedenle, veri eksikliğine ek olarak, zorlu bir görev, belirli bir sürücü için araç içinde hoş bir ortam sağlamak için çevreden ve ayrıca sürücü tercihlerinden nasıl uyum sağlanacağını öğrenmektir.

Sağlam sınıflandırma görevlerinde etiketlenmiş verilerin bu kıtlığı nedeniyle, bazı derin öğrenme yaklaşımları bazı transfer öğrenme teknikleri üzerine kuruludur. Han ve arkadaşları (2016), insan açıklamasını en aza indirmek için aktif öğrenme ve kendi kendine eğitim yaklaşımını birleştirerek etiketli verilerin yokluğunu ele alıyor. Metodolojileri, FindSounds veri kümesi 1'de insan etiketli örneklerde %52.2'lik bir azalma sağlayabildi. Bu veri kümesinde dokuz kategori ve 1 ila 10 saniye arasında değişen sürelerle 16.930'dan fazla ses örneği bulunmaktadır.

Örneğin SoundNet veri seti, ham ses dalga biçimlerinin bir temsilini öğrenmek için etiketli videolardan kareler kullanarak doğal sesin geniş ölçekli ve zengin bir temsilini sağlar. Eğitim aşamasında görüntü kullanılmasına rağmen, SoundNet yalnızca test aşamasında ses gerektirir. Bu ağlardan gömme özelliklerini kullanmanın AudioSet'teki veriler (akustik olaylar etiketli 1 milyon 10 saniyeden fazla alıntıdan oluşan bir veri kümesi) ham özelliklerden daha iyi performans gösterdiğini buldular. Sonuç olarak, son teknoloji görüntü ağlarının, ses sınıflandırmasında, basit tam bağlantılı ağlara veya AlexNet gibi daha eski görüntü sınıflandırma mimarilerine göre daha iyi sonuçlar sağladığı sonucuna varıldı. Çoklu evrişim katmanları ve maksimum havuzlama ile spektrogramları, olay sınıflandırması için çıktı katmanlarını ikili hale getiren tekrarlayan bir sinir ağını (RNN) beslediler. Deneyleri, önerilen sinir ağının geleneksel Gauss Karışım Modelleri-Gizli Markov Modelleri (GMM-HMM) sınıflandırıcılarından ve evrişimli sinir ağı veya RNN kullanan yöntemlerden daha iyi performans gösterdiğini gösterdi.

Denetimli yaklaşımlardaki etkileyici sonuçlara rağmen, gerçek dünya senaryolarına tam entegrasyonu çeşitli zorluklarla karşı karşıyadır. Her şeyden önce, etiket kümesinin yetersiz tanımlandığı veya çok büyük olduğu bazı durumlarda oluşturulması mümkün olmayan, açıklamalı büyük bir veri kümesinin gerekliliğidir.

Hareket halindeyken araç tanımlama, trafik ve hız yönetimi, sınıflandırılmış araç sayısı, trafik sinyali zaman optimizasyonu, boşluk/yol mesafesi ölçümü ve askeri amaçlar için bir ön koşuldur. Hareket eden araçlar çevreyi farklı şekillerde etkiler. Araç ısı, ses ve manyetik alan yayar. Farklı türde sinyallere dayalı olarak araç tanımlamayı araştıran birçok yaklaşım vardır.

Görüntü işleme teknikleri, araçları gerçek zamanlı trafik yönetimi altında sınıflandırmak ve Akıllı Ulaşım Sistemleri (AUS) için kullanılır. Endüktif döngü tabanlı sistemler, araç sayımlarının belirlenmesinde yaygın olarak kullanılmaktadır. Araç tanımlama için umut verici yaklaşımlardan bir tanesi akustik sinyallere dayanan yaklaşımdır. Hareket eden araçlar karakteristik sesler çıkarır. Bu sesler hareketli parçalardan, sürtünmelerden, rüzgarlardan, emisyonlardan, lastiklerden vb. oluşur. Aynı çalışma koşullarına sahip benzer araçların birbirine benzer sesleri çıkardığını varsayılırsa; daha sonra bu sesler araçları sınıflandırmak için kullanılabilir. Ses sinyallerine dayalı araç sınıflandırması, araştırmacılar tarafından son yıllarda ilgi odağı haline gelmektedir. Araç tipini ve mesafesini sınıflandırmak için olasılıksal sinir ağı (PNN) modelleri geliştirilmiştir (Paulraj vd., 2013). Deneysel sonuçlar, araç sesleri ile uğraşırken TESPAP (Time Encoded Signal Processing and Recognition) yönteminin etkinliğini göstermektedir. Özellik setinin etkinliği genellikle sinyali ne kadar iyi temsil ettikleriyle ölçülür; bununla birlikte, özellikler de birkaç koşulu karşılamalıdır: sayıları küçük olmalıdır (15'ten az); ayrıca pratik uygulamayı sağlamak için hesaplanması kolay olmalıdır. Araç sınıflandırmasında kullanılan bir dizi özellik çıkarma yöntemi vardır. Bununla birlikte, hiçbir özellik çıkarma yöntemi tüm gereksinimlerimizi karşılamaz ve kriterimize göre en iyi özelliği seçmenin sistematik bir yolu yoktur. Yöntemler, ayrılabilirlikleri ve doğru sınıflandırma oranları bağlamında karşılaştırıldı. Birincisi spektrum dağılımına, ikincisi ise dalgacık paket dönüşümüne dayanmaktadır. K-en yakın komşu algoritması (KNN) ve destek vektör makinesi (SVM) gibi farklı sınıflandırıcıların performansını değerlendirmişlerdir. Araç ses verileri için, ayrık spektrum tabanlı bir öznetelik çıkarma yönteminin, dalgacık paket dönüşümü yönteminden daha iyi performans gösterdiği bulunmuştur.

Ses sınıflandırma ile ilgili ilk çalışmalar, konuşma tanıma uygulamalarına dayanmaktadır. Konuşma tanıma uygulamalarında kullanılan sıfır geçiş hızı, çerçeve güç enerjisi, perde ve benzeri özellikler, KNN ve karar ağaçları gibi basit sınıflandırıcılar kullanılarak kategorize edilmiştir. Ancak son yıllarda veri miktarındaki yüksek artış oranları ve daha fazla sınıf tanımlama ihtiyacı, Ses sınıflandırma görevlerine özgü yeni tekniklerin geliştirilmesini zorunlu kılmıştır. Yukarıdaki bölümlerde de belirtildiği gibi ATİ görevi için kullanılabilecek birçok ses özelliği ve sınıflandırma metodu bulunmaktadır. Ses sınıflandırma görevlerinde özelliklerin manuel olarak el ile seçilmesi yaygın olarak kullanılan yöntemlerden bir tanesidir. Fakat bu süreç hem zaman alıcıdır hem de uzmanlık gerektirir. Bu nedenle son yıllarda spektrogram ve mel-spektrogram gibi 2 boyutlu zaman-frekans alanı temsilleri üzerinden CNN kullanılarak otomatik özellik çıkarılması ve sınıflandırılması yaygın olarak kullanılmaya başlanmıştır.

Ses sınıflandırma görevlerinde kullanılan birçok özellik gürültüsüz koşullarda oldukça başarılı olurken, ortam gürültüsünün artmasıyla sınıflandırma performansı belirgin şekilde düşmektedir. Öte yandan spektrogram gibi 2 boyutlu görsel temsiller ise gürültülü koşullar altında da diğer özelliklere nazaran oldukça iyi sonuçlar vermektedir. Bu özelliklerin gürültüye karşı dayanıklı olmasının temel unsuru ses sinyalinin spektrogramının seyrek olması gerçeğiyle açıklanır. Bu nedenle saçılan gürültü yoğunluğu genellikle spektrogramın dinamik aralığının alt bölgelerinde yer alırken, en güçlü ses bileşenleri daha yüksek olan bölgelerde yer almaktadır (Dennis vd., 2010). Bu durumda da ortam gürültüsünün artmasıyla birlikte spektrogramın tamamında bir bozulma gerçekleşmemiş olur.

Esasında 2 boyutlu zaman-frekans temsilleri, bir sinyalin frekans spektrumunun görsel temsili olduğundan, görüntü uygulamalarında kullanılan yöntemler, bu temsiller üzerinden özellik çıkarma ve sınıflandırmada kullanılabilirler. Ayrıca ses sinyalleri genellikle gürültü kaynaklarına kıyasla daha seyrek ve ayırt edilebilir modeller zayıf mutlak yerelliğe sahiptir. Bu, tanınabilir bir ses olayının farklı zamanlarda ve biraz farklı frekans aralıklarında görülebileceği anlamına gelir. Öte yandan, CNN, görüntü üzerindeki örüntü pozisyonuna (Krizhevsky vd., 2012) duyarsızdır, bu nedenle 2 boyutlu zaman-frekans temsilleri sınıflandırmak için uygundur.

CNN yöntemi görüntü sınıflandırmasında yaygın olarak kullanılır ; ancak buradaki kritik konu, ses sinyallerinin görüntü olarak en iyi nasıl temsil edileceğidir Genel olarak, bu sorunu çözmek için iki temel yaklaşım uygulanır (Sharan vd., 2019). İlkinde, ses sinyalleri spektrogram görüntülerine dönüştürülür (Sajjad vd., 2020; Papakostas vd., 2017; Bertero vd., 2019; Sharan vd., 2019). . İkincisinde Mel-spektrogram görüntü temsilleri Mel-filtre bankaları (Jiang vd., 2019; Luengovd., 2010; Abdel-Hamid vd., 2014) kullanılarak elde edilir. Spektrogram temsilleri eşit bant genişliği ve eşit aralıklı frekans bileşenleri sunar Bu durum, çoğu spektral enerjinin daha düşük frekans aralığında olduğu görevler için ideal değildir (Sharan vd., 2019). Mel filtresinin kullanımı bir dereceye kadar fayda sağlar. Bununla birlikte, birkaç çalışma, performansının biyolojik olarak esinlenilmiş modellerden daha düşük olduğunu ortaya koymuştur (Sharan vd., 2019; 2020). Ek olarak, Mel ölçekli filtre bankalarında bir düzgünlük sorunu vardır (Josifovski, 2002). Öznitelik vektörlerinin yüksek varyansı, sınıf ayrılabilirliğini olumsuz etkiler ve ses sınıflandırma performansını azaltır (Alam vd., 2013). Ayrıca aynı tipteki ses kaynaklarının içsel farklılıkları ve değişik konumlar ve durumlar için akustik ortamın yankı ve mesafe gibi etkisi göz önüne alındığında, her sınıfın akustik özellikleri büyük çeşitlilik gösterebilir. Bu durumda ATİ görevi de daha zor hale gelir. Bununla birlikte, insan işitme sistemi, seslerin temel frekansını ve yoğunluğunu algılamada çok başarılıdır. Bu yetenek, algılanan bilginin zamansal dinamiklerinin,

fark edilebilir duygusal alanlara odaklanarak izlenmesine olanak tanır (Peng vd., 2020). Bu nedenle, insan işitsel sistemini modelleyen yaklaşımlar, ATİ görevi için etkili olabilir (Peng vd., 2020).

Yukarıda verilen nedenlerle, bu çalışma esas olarak insan işitsel sisteminden esinlenen işitsel ateşleme oran haritası özelliklerine odaklanmaktadır (Brown vd., 1994). İnsan kokleasındaki mekanik transdüksiyon sürecinin bir sonucu olarak, işitsel sinir ateşlemesi meydana gelir ve oran haritası özellikleri, bu ateşleme işlemi hızının spektro-zamansal temsilini kodlar. Oran haritası özellikleri dört temel aşamada oluşturulur (Brown vd., 1994). İlk olarak, ses sinyaline ön vurgulama işlemi uygulanır. Koklear filtreleme daha sonra gammaton filtresi (GTF) kullanılarak gerçekleştirilir. Bir sonraki aşamada, nöromekanik transdüksiyon işlemi ile hesaplanan işitsel sinir ateşleme hızında ani bir artış olasılığı bulunur. Son olarak, Hamming penceresi ile meydana gelen ani artışlar entegre edilerek oran haritası özellikleri elde edilir.

GTF, üst frekans bölgesinde geniş frekans bileşenleri ve alt frekans bölgesinde dar frekans bileşenleri sunar (Sharan vd., 2019). Alt frekans aralığında sağlanan daha kesin frekans çözünürlüğü, üst frekans aralığında önemli spektral bilgileri kaybetmeden daha fazla spektral bilginin çıkarılmasında faydalıdır (Sharan vd., 2019; Sharan vd., 2015). Bu, ATİ görevi (Sharan vd., 2019; Sharan vd., 2020) gibi düşük frekans aralığında yüksek enerji seviyelerine sahip uygulamalar için spektrogram ve Mel-spektrogram yöntemlerine göre bir avantaj sağlar. Oran haritası, genellikle hesaplamalı işitsel sahne analizi ve konuşmacı tanıma görevlerinde kullanılan işitsel bir modeldir (Josifovski, 2002). İşitsel tümevarım, kullanıldığı çalışmalar için birincil motivasyondur (Warren vd., 1994). İşitsel deneyler, bir sesin algılanmasına karşı bir kanıt yoksa, boşlukları doldurabileceğini göstermektedir (Josifovski, 2002). Deneyler, belirli bir sesin olasılığına izin vermek için yeterli enerji varsa, insanların seslerdeki boşluklardan habersiz olduklarını göstermiştir (Warren vd., 1994). Buna "fonem restorasyonu" etkisi denir. Oran haritası, her bir frekans kanalının ortalama enerjisini ayrı ayrı yansıttığı için bu görevler için kullanışlı bir özelliktir (May, 2012). Mel ölçekli filtre bankalarından farklı olarak bir oran haritası, özellikler üzerinde yumuşatma sağlar (Josifovski, 2002), böylece özellikler arasındaki varyansı azaltır. Özellik vektörünün varyansındaki azalma, sınıf ayrılabilirliğini iyileştirir ve ATİ performansını olumlu yönde etkiler (Alam vd., 2013). Ayrıca oran haritaları, Mel ölçeğinde frekans filtreli enerji temsilleri grubuna girer (Trowitzsch, 2012). Bu grup altındaki çalışmalar, diğer özelliklere kıyasla ses olayı algılama görevlerinde önemli başarı elde etmiştir (Trowitzsch, 2012; Mesaros vd., 2017; Mesaros vd., 2019).

Oran haritaları elde edilirken kritik unsurlardan bir tanesi nöromekanik transdüksiyon işleminin nasıl modelleneceğidir. Koklea, ses basıncını baziler membranın titreşim hareketlerine dönüştürür ve daha sonra bu titreşimlerden sinirsel uyarılar oluşturulur (Russo vd., 2019). Koklea'yı karmaşık seslerin frekans bileşenlerini analiz eden bir sistem olarak düşünmek mümkündür çünkü baziler membran boyunca her pozisyonun belirli bir frekansa karşılık geldiği bir frekans analizörü işlevi gerçekleştirir (Russo vd., 2019). Koklea içerisinde yer alan iç tüy hücreleri (inner hair cell - IHC) mekanik transdüksiyon yoluyla nöral uyarılar üreten ana sensörlerdir. Bu çalışmada 4 farklı IHC metodunun ATİ görevi üzerindeki sınıflandırma performansına olan etkileri değerlendirilmektedir. Ayrıca sonuçlar spektrogram ve periodogram olmak üzere iki farklı görsel temsil metodu ile de kıyaslanmaktadır.

Çalışmada kamuya açık bir veri seti üzerinde karayollarındaki trafiğe ilişkin 5 farklı ses olayı sınıflandırılmıştır. Bu ses sınıfları araba, kamyon, otobüs, motosiklet ve araç sesinin bulunmadığı durumlardan oluşmaktadır. Sınıflandırma aşamasında hem geleneksel bölme işlemiyle (holdout test) hem de 5-parça çapraz doğrulama ile testler gerçekleştirilmiş ve sonuçlar kıyaslanmıştır. Hilbert ve Joergensen IHC metodlarının kullanıldığı yöntemlerin diğer modellere kıyasla daha iyi sonuç verdiği gözlemlenmiştir.

3. AKUSTİK ÖZELLİKLER

Sesler, zamanla değişen sinyallerdir ve tüm anlamları bu tür zaman değişkenliği ile ilgilidir. Ses analizi teknikleri, sinyaller için karşılaştırma, modifikasyon ve yeniden sentezi anlama görevlerini kolaylaştırmak amacıyla zamanla değişen ayırt edici özelliklerin en azından bazılarını kavramak için geliştirilmiştir. Ses analizinin uygulama alanları, video işleme, konuşma işleme, biyoakustik analiz gibi akustik ortamların birçok yönünü kapsamaktadır.

Bir ses frekansı veya duyulabilir frekans, frekansı ortalama bir insan tarafından duyulabilen periyodik bir titreşim olarak tanımlanır. Perdeyi en çok belirleyen ve hertz (Hz) cinsinden ölçülen sesin özelliğidir.

Genel olarak standart duyulabilir frekans aralığı 20 ila 20.000 Hz'dir, ancak bireylerin duyduğu frekans aralığı çevresel faktörlerden ve yaştan büyük ölçüde etkilenir. Titreşimin genliğinin yeterince büyük olduğu varsayıldığında, 20 Hz'nin altındaki frekanslar genellikle duyulmak yerine hissedilir. 20.000 Hz üzerindeki frekanslar bazen gençler tarafından hissedilebilir. Yüksek frekanslara uzun süre maruz kalınırsa işitme kaybına sebep olabilir. İşitilebilir aralığın üzerindeki ve altındaki frekanslar sırasıyla "ultrasonik" ve "infrasonik" olarak adlandırılır.

En önemli ses analiz teknikleri Kısa Zamanlı Fourier Dönüşümü (STFT) ve Doğrusal Öngörülü Kodlamadır (LPC). STFT, zaman alan sinyalinin dilimleri üzerinde gerçekleştirilir ve STFT'nin işlevi, sinyalleri zaman alanından frekans alanına dönüştürmektir (Griffin ve Lim, 1984). Esasında spektrogram temsilleri de Hızlı Fourier Dönüşümü ile elde edilen büyüklük spektrogramlarından oluşan bir yığından meydana gelmektedir ve son dönemde bu özellikler ses olaylarının sınıflandırılmasında yaygın olarak kullanılmaya başlanmıştır. LPC analizi ise sentetik konuşma ve konuşma sinyali iletişimine ulaşmak için verimli ve etkili bir araçtır.

3.1 AKUSTİK OLAY VE ARKA PLAN GÜRÜLTÜSÜ

Bir ses akışındaki zaman damgaları akustik olay anlamına gelir. Bazı olaylar ilgi çekiciyken bazıları ise değildir. Daha spesifik olarak, çevresel akustik çalışmalarda ilgi çağrıları akustik olaylar olarak adlandırılır. Oysa ilgi dışı olan olayların tümüne arka plan gürültüsü denir. Sonuç olarak, arka plan gürültüsünün tanımı belirsizdir. Çalışmamızda, şiddetli yağmur sesleri,

yaprakların hışırtısı veya uzaktaki trafik gibi bir zaman aralığı boyunca yalnızca sürekli olarak devam eden sesleri arka plan gürültüsü olarak varsayılmıştır.

Bazı uygulamalardaki arka plan sesleri, diğer uygulamalarda ilgi çeken sinyallerdir. Örneğin, kuş cıvıltısı sınıflandırmasının amacı ile kullanılan bir ses kaydında, ağustosböcekleri gibi diğer türlerin sesleri arka plan gürültüsü olarak kabul edecek belirli bir türe ait çağrılar bulmaktır. Bu nedenle gürültünün tanımı uygulama alanına bağlıdır. Genel olarak, arka plan gürültüsü doğal ve yapay gürültü olarak ikiye ayrılır. Doğal gürültü, yaprakların hışırtısı, rüzgar ve yağmur sesi olabilir. Yapay gürültü insan faaliyetlerinden gelir: konuşma, uzaktaki trafik ve inşaat sesi vb.

3.2 SES ÖZELLİKLERİ

İstatistiksel özellikler, ortalama temel frekans, maksimum temel frekans, minimum temel frekans, sıfır geçiş oranı, kısa süreli enerji ve sinyal bant genişliğini içerir. İstatistiksel olmayan özellikler arasında doğrusal tahmin katsayıları, mel-frekans cepstral katsayıları, spektral akı, bant enerji oranı gibi özellikler bulunmaktadır.

3.3 ZAMAN ALANINDAKİ ÖZELLİKLER

Zaman alanında, konuşma ve müzik olaylarını sınıflandırmak için dalga biçimindeki kısa süreli enerji ve ortalama sıfır geçiş oranı gibi özellikler ölçülebilmektedir. Bunun temel nedeni konuşma ve müzik gibi farklı akustik olaylar farklı spektral dağılıma ve zamansal değişim modellerine sahip olmalarıdır.

- Sıfır geçiş oranı (ZCR): farklı ses sinyallerinin karakterize edilmesinde faydalı olduğu kanıtlanmıştır. Birçok konuşma/müzik sınıflandırma algoritmasında kullanılır. Sıfır geçişi, ardışık örneklerin genliği pozitiften negatife veya tersi değiştiğinde meydana gelir. ZCR, sinyalin kısa zaman aralığında ortalama kaç kez işaret değiştirdiğini gösterir.

- Kısa süreli enerji (STE): bir sinyal çerçevesindeki toplam enerji olarak tanımlanır (Pohjalainen, 2007).

İstatistiksel özellikler doğrudan dalga biçiminden türetilmediğinden, birçok araştırmacı frekansla ilgili daha fazla özellik elde etmek için (istatistiksel olmayan) spektrogramlara yönelmektedir. Ayrıca doğrudan akustik sinyalin dalga biçiminden elde edilen özellikler çevresel gürültüler karşısında çok ciddi bir şekilde bozulmaya uğrayabilmektedirler.

3.4 FREKANS ALANINDAKİ ÖZELLİKLER

MFCC'ler, ses sınıflandırması için en yaygın kullanılan özelliklerden biridir. Buradaki fikir, öncelikle büyüklük spektrumuna benzeyen Mel-frekans katsayılarını (MFC'ler) hesaplamaktır (büyüklük spektrumu, farklı frekanslarda bir sinyal çerçevesi sırasında sesin yoğunluğunu temsil eder), ancak Hertz (Briggs) yerine Mels birimlerinde hesaplamaktır (Briggs vd., 2009; Chu vd., 2009). Bu nedenle, MFCC özellikleri, genel spektrumun şekline göre modellenir, bu da onu tek ses kaynaklarının (konuşma) modellenmesi için daha uygun hale getirir. Bununla birlikte, çevresel sesler tipik olarak, kuşların cıvıltısı, yağmur damlaları gibi MFCC'lerin etkili bir şekilde kodlayamadığı dar spektral tepe noktaları ile karakterize edilen koşullar dahil olmak üzere çok çeşitli sesleri içerir (Chu vd., 2008). MFCC için filtre bankaları, insan işitsel sistemine dayanmaktadır. Konuşma ve müzik gibi yapılandırılmış sesler için özellikle iyi çalıştığı gösterilmiştir, ancak performansları gürültü varlığında düşmektedir.

Yaygın olarak kullanılan diğer bir özellik ise lineer tahmin cepstral katsayılarıdır (LPCC'ler). Doğrusal tahminin ardındaki temel fikir, mevcut örneğin, ani değişikliklere karşı daha sağlam bir özellik sağlayacak olan önceki örneklerin doğrusal bir kombinasyonu olarak tahmin edilebilir veya tahmin edilebilir olmasıdır.

Aşağıdaki paragrafta diğer popüler spektral özelliklerden bazılarını açıklanmaktadır:

- Bant Enerji Oranı: belirli bir frekans bandındaki enerjinin toplam enerjiye oranıdır.
- Spektral akı (SF): birbirini izleyen iki çerçeve arasındaki spektral genlik farkını ölçmek için kullanılır (Zeppelzauer vd., 2009).
- Spektral roll-off: frekans yanıtı büyüklüğünün kümülatif değerinin toplam büyüklüğün belirli bir yüzdesine ulaştığı frekans değerinin miktarını belirler. Yaygın olarak kullanılan bir eşik %85'tir (Pfeiffer ve Vincent, 2001).

İnsan işitsel algısının daha fazla yönünü hesaba katan ve perde, ses yüksekliği ve parlaklık gibi algısal özellikler olarak da adlandırılan özellikler de vardır.

Birçok araştırmacı, farklı türde öznitelikler, çeşitli sınıflandırıcılar ve küçük öznitelik setleri kullanarak çevresel seslerin sınıflandırılmasının ve tanınmasının doğruluğunu geliştirmek için çaba sarf etmiştir.

Chu ve arkadaşlarının (2009) tezinde, yeni bir dizi zaman-frekans özelliği kullandılar ve bir ses sensörünü çevreleyen bir sahneyi veya bağlamı anlamak için çevresel sesleri tanıma görevini düşündüler. Ortam seslerini analiz etmek için küçük bir zaman-frekans özelliği seti

seçmek için Eşleştirme Takip algoritmasını (MP) kullanan yeni bir özellik çıkarma yöntemi önerilmiştir. 14 tür çevresel sesi sınıflandırmak için bir Gauss karışım modeli (GMM) sınıflandırıcısını benimseyen bu çalışmanın sonuçlarında, MFCC ve MP özelliklerini ayrı ayrı kullanmanın zayıf bir doğruluk oranı verdiğini bulunmuştur. MFCC ve MP özelliklerinin birleştirilmesiyle elde edilen ortalama doğruluk oranı, on dört sınıfı ayırt etmede %83,9'dur. Ayrıca MP ve MFCC özelliklerinin kombinasyonunun yüksek bir doğruluk oranına ulaşıldığı da gösterilmiştir. Destek vektör makinesini çevresel ses sınıflandırma sistemi için bir sınıflandırıcı olarak kullanılmıştır. %92 doğruluk oranına ulaşılmıştır.

Mitrovic ve arkadaşları (2009), çevresel sesler için optimal bir özellik setinin kompozisyonu için temel bileşen analizini kullanmış ve özellik kombinasyonlarının kalitesini değerlendirmek için geri alma deneyleri yürütmüştür. Alma sonuçları, istatistiksel veri analizinin, çevresel ses tanımda özellik seçimi için faydalı ipuçları verdiğini göstermektedir.

Genel olarak perde, sesin algısal bir özelliğidir ve algısı, insanın farklı sesleri duymasında ve anlamasında önemli bir rol oynar. Akustik bir ortamda, insan dinleyiciler birkaç gerçek zamanlı sesin perdesini tanıyabilir ve bir karışımındaki bir sesi akustik olarak ayırmak için perdeyi verimli bir şekilde kullanabilir (Bregman ve McAdams, 1994). Ancak, sokak gürültüsü, yağmur, çığlık veya silah sesi gibi gürültüye benzer konuşma olmayan ses sinyallerinin sabit bir perdesi değil, bir dizi değeri vardır.

Daha önce yapılan çalışmalarda, ses sinyallerini karakterize etmek için bazı özelliklerin bir kombinasyonunu veya hatta daha büyük bir özellik setini kullanılmaktadır. Ancak daha fazla özellik eklemek her zaman yardımcı olmamıştır. Özellik boyutu arttıkça, veri noktaları daha seyrek hale gelir ve sınıflandırma sonuçlarını olumsuz etkileyebilecek potansiyel olarak alakasız özellikler vardır. Colonna ve arkadaşları (2012), sınıflandırma için yüksek öznelite kümesi boyutunun kullanılmasının, ses sınıflandırma problemlerinde her zaman iyi performans üretmediğini göstermiştir. Hesaplama maliyetini ve çalışma süresini azaltmak ve kabul edilebilir bir sınıflandırma oranı elde etmek için daha küçük bir özellik seti elde etmek için basit bir özellik seçim algoritması kullanılmıştır.

3.5 DİĞER ÖZELLİKLER

Son zamanlardaki bazı indeksler biyoçeşitlilik için kullanılmış, ancak özellikle yağmur sınıflandırması üzerinde değerlendirilmemiştir. Aşağıdaki endeksler ana özelliklerimiz olarak kullanılmış ve aşağıda kısaca açıklanmıştır.

Akustik entropi (H), bir kayıt içindeki akustik enerjinin zaman veya frekans bantları yoluyla dağılımının bir ölçüsüdür (Sueur vd., 2008). Yazarlar, bireysel tür tanıyıcıları oluşturmanın zorluğunu kabul eder ve bu nedenle, ses çıkaran türlerin sayısının bir yerellik içindeki ses verilerinin akustik heterojenliği ile olumlu bir şekilde ilişkili olduğu basit varsayımını yaparak, biyolojik çeşitliliğin dolaylı ölçümlerine yönelir. Akustik entropinin akustik heterojenite ile ilişkili olduğu sonucuna varmışlardır. Sonuçları, bireysel olarak izole edilmiş çeşitli kuş çağrılarını birleştirerek elde edilen yapay olarak oluşturulmuş kayıtlara dayanmaktadır.

3.6 ÖN İŞLEME

Ön işleme, sınıflandırma sürecinde önemli bir adımdır. Amacı, sinyal işleme ve gürültü giderme dahil olmak üzere çeşitli işlemler yoluyla tanınacak kalıplar oluşturmak ve geliştirmektir. Toplanan veriler her zaman iyi kalitede olmadığından, ön işleme sonraki işlemlerin sonuçlarını iyileştirmek için gerekli bir adımdır.

3.6.1 SİNYAL İŞLEME

Uygun bir temsilin oluşturulması için ses sinyallerinin işlenmesi gerekir. Genel olarak, Kısa Zamanlı Fourier Dönüşümü (STFT), Hızlı Fourier Dönüşümü (FFT), dalgacık dönüşümü (WT) ve Doğrusal Öngörülü Kodlama (LPC), sinyal temsili oluşturmanın üç ana yoludur. Dalga biçimi, zaman alanındaki sinyallerin temel bir biçimidir. Ses sinyallerinin daha kullanışlı görsel bilgilerini keşfetmek için Fourier dönüşümü genellikle zaman alanlı sinyalleri frekans alanı sinyallerine dönüştürmek için kullanılır. Özellikle Fourier dönüşümü ile spektrumlar oluşturulmakta ve STFT ile spektrogramlar üretilmektedir (Griffin ve Lim, 1984).

3.6.2 GÜRÜLTÜ GİDERME

Çevrenin ses kayıtları bağlamında “gürültü”nün birkaç anlamı olabileceğini belirtmek önemlidir. Gürültü, mühendislerin anladığı gibi sadece elektronik veya mikrofon gürültüsü anlamına gelmemektedir. Rüzgar ve yağmur, çevresel akustik kayıtlarda sıklıkla bulunur ve genellikle gürültü olarak kabul edilebilir.

Gürültünün ortamın kayıtlarına katkısı tipik olarak artan frekansla azalır. Ancak, standart bir gürültü modeli varsaymıyoruz. Bunun yerine, her bir dakikalık kaydın spektrogramındaki frekans kutularının her biri için modal gürültü gücünü bağımsız olarak tahmin etmektedir. Lamel ve arkadaşları (1981) nedeniyle aynı uyarlanabilir seviye eşitleme algoritmasının değiştirilmiş bir versiyonunu kullanılmaktadır. Uyarlanabilir seviye eşitleme, sürekli arka plan akustik aktivitesini kaldırma ve bu seviyeyi sıfır genliğe ayarlama etkisine sahiptir. Böylece, birden fazla frekans kutusuna yayılan bir akustik olayın tespiti için tek bir mutlak eşik tanımlamak mümkün hale gelir. Bu değiştirilmiş versiyonun, spektrogram değerlerinin desibele dönüştürülüp dönüştürülmediğine bakılmaksızın uygulanabilir. Her frekans kutusu için bir eşik yoğunluk değeri hesapladıktan sonra, bunu o kutudaki her bir değerden çıkarılır. Çalışmamızda, daha sonraki ACI ve akustik entropi (spektral ve zamansal entropiler) hesaplaması için uygun değerleri korumak için spektrogramlar desibele dönüştürülmemiştir. Gürültü azaltmayı ses kaydında değil, iki boyutlu spektrogramda yapılmıştır.

3.7 ÖZELLİK ÇIKARMA VE SEÇİMİ

Cowling ve Sitte'ye (2003) göre öznelik çıkarımı iki geniş türe ayrılabilir: durağan (frekans tabanlı) öznelik çıkarma ve durağan olmayan (zaman-frekans tabanlı) öznelik çıkarma. Durağan özellik çıkarımı, tüm sinyalde bulunan frekansları detaylandıran genel bir sonuç üretir. Durağan özellik çıkarımı ile bu frekansların sinyalde nerede oluştuğuna dair bir ayırım yapılmaz. Buna karşılık, durağan olmayan özellik çıkarımı, sinyali ayrık zaman birimlerine böler. Bu, frekansın sinyalin belirli bir alanında meydana geliyor olarak tanımlanmasına izin vererek sinyalin anlaşılmasına yardımcı olur.

Durağan olmayan özellik çıkarımı (Vapnik, 1999) şunları içerir:

- Kısa zamanlı Fourier dönüşümü (STFT)
- Hızlı (ayrık) dalgacık dönüşümü (FWT)
- Sürekli dalgacık dönüşümü (CWT)

- Wigner-Ville dağılımı (WVD)

Durağan öznitelik çıkarma, konuşma dışı sesler için uygun sekiz popüler tekniği (aşağıda listelenmiştir) içerir (Lamel vd., 1981):

- Frekans çıkarma (FE)
- Cepstral katsayıları (CC'ler)
- Mel-frekans cepstral katsayıları (MFCC'ler)
- Doğrusal tahmine dayalı kodlama (LPC)
- Doğrusal tahmin cepstral katsayıları (LPCC'ler)
- Mel-frekans LPC katsayıları (MFLPCC'ler)
- Kabuk frekans cepstral katsayıları (BFCC'ler)
- Kabuk frekans LPC katsayıları (BFLPC'ler)

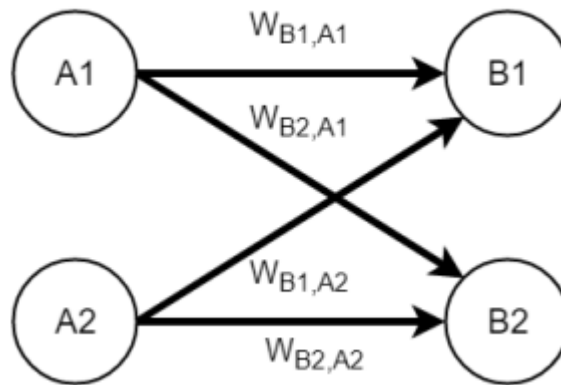
Özellik seçimi, özellik çıkarma modülünün tasarımında ele alınması gereken önemli bir konudur. Özellik vektörü gösterimine hangi özelliklerin dahil edileceğine karar vermeyi ifade eder. Çoğu zaman, özellikler, alan bilgisi ve deneylerin bir kombinasyonu kullanılarak seçilir (Pohjalainen, 2007).

Makine öğrenimi ve istatistikte, özellik seçimi, bir sınıflandırma algoritmasındaki en önemli görevlerden biridir. Yanlış sınıflandırma hatasını arttırmadan düşük bir hesaplama yüküne izin verir. Amaç, eğitilmekte olan sınıflandırma algoritmaları için girdi modelini temsil eden verimli ve küçük bir akustik özellik vektörü elde etmektir.

4. DERİN SİNİR AĞLARI

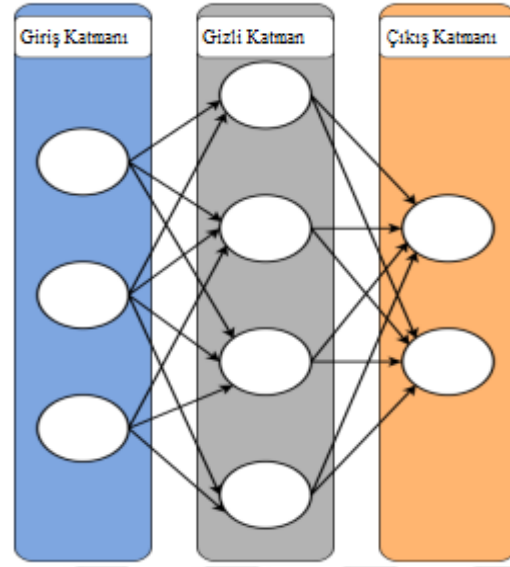
Derin sinir ağları, biyolojik beyni oluşturan sinir ağlarından ilham alan bir tür yapay sinir ağlarıdır (YSA). Bu ağlar, impulsları alıp birbirlerine gönderebilen nöronlar veya düğümler olarak adlandırılan birbirine bağlanarak oluşturulur. Derin sinir ağlarında, bunlar genellikle farklı katmanlardan gelen düğümlerin bağlandığı, ancak bir katman içindeki nöronlar arasında bağlantı olmadığı katmanlar halinde oluşturulur. YSA'lardan farklı olarak, derin sinir ağları her zaman daha soyut ve doğrusal olmayan bir bilgi işlemeyi sağlayan birden fazla nöron katmanından oluşur. Şekil 1'de oklarla gösterilen ağırlıklarla birbirine bağlanmış dört düğüm görülmektedir. Bu, nöronların sinapslar yoluyla birbirine bağlandığı biyolojik bir sisteme benzer.

Bu şekilde ağ, daha küçük (düşük seviyeli) özellikleri bir araya getirerek özellik hiyerarşilerini öğrenebilir. Örneğin, bir derin sinir ağlarının bir yüzü tanıması için önce görüntüyü çizgilere böler, ardından bu bilgiyi alarak yüz özellikleri (gözler, burun, ağız vb.) haline gelen şekiller (daireler, üçgenler, kareler vb.) oluşturur ve son olarak bu özellikleri bir yüze yerleştirir. Bir derin sinir ağı ne kadar derinse, o kadar soyut özellikleri tanıyabileceği gösterilmiştir. Ancak daha derin ağlar da problemlere tabidir, biri gradyan iniş problemidir. Sık görülen bir diğer sorun da fazla takmadır; bu, ağın, verilerin test setinde doğru tahminler için optimize etmeye çalışması nedeniyle ortaya çıkan bir problemdir. Yanlış eğitilirse, derin sinir ağları, verilerin çok özel özelliklerini öğrenir ve eğitim verileri tarafından açıkça kapsanmayan örnekleri genelleştiremez ve doğru şekilde sınıflandıramaz.



Şekil 4-1: Ağırlıklarla birbirine bağlanan dört düğüm, basit bir sinir ağı oluşturur.

Birçok modern ağ gibi, bu çalışmada ele alınan derin sinir ağları ileri beslemeli ağlardır, bu, nöronların çıktısının ağda yalnızca bir yönde hareket ettiği anlamına gelir. Resimlerde bu, uygun olduğunda ağdaki bilginin akışını veya yönünü gösteren oklarla temsil edilir



Şekil 4-2: Bir gizli katmana sahip yapay bir ileri beslemeli sinir ağı.

Oldukça basitleştirilmiş bir örnek kullanarak, YSA terminolojisini yüksek sesle kitap okuyan bir insana uygulanabilir. Giriş katmanı, gözlerden beyne gönderilen sinyal olacaktır ve çıkış, yüksek sesle söylenen kelimeler olacaktır. Gizli katmanlar, aradaki tüm mantık, daha spesifik olarak kağıt üzerinde görülen formların ve çizgilerin kelimelere dönüştürülmesi ve kelimelerin insanı konuşturan kaslarımıza gönderilen dürtülere çevrilmesidir.

4.1 DERİN SİNİR AĞLARI TÜRLERİ

Bu bölümde açıklanan DSA, en çok çok katmanlı algılayıcı (ÇKA) olarak bilinen sinir ağı türüne benzemektedir. Yani, her katmanın komşularına tamamen bağlı olduğu yığılmış algılayıcı katmanları bulunmaktadır. Bununla birlikte, belirli görevler için tasarlanmış başka ağ türleri de vardır. Aşağıda, bu projede kullanılan daha özel ağların bir açıklaması bulunmaktadır.

4.1.1 Evrişimli Sinir Ağları

Evrişimli sinir ağları (CNN'ler), katmanlardan en az birinin, bilgileri bir sonraki katmana geçirmeden önce girdi verileri üzerinde evrişimler (çapraz korelasyon veya nokta çarpımı) kullandığı bir DNN biçimidir. Bu konvolüsyonlar, her filtrenin verilerin bazı özelliklerini vurguladığı özellik çıkarma filtreleri olarak yaygın olarak kullanılır. Örneğin, görüntü tanımda bu, bir tür kenar veya eğri algılama olabilir.

CNN'lerde özel havuzlama katmanları da yaygındır, bunlar bir nöron bölgesinin çıktısını birleştirir ve bir ayırıklaştırma katmanı görevi görür. Bu, girişin boyutsallığını ve uzamsal boyutunu azaltır, bu da ağda daha az sayıda hesaplama yol açar. Orijinal görüntünün bu şekilde küçülmesi, ağın başlangıçta bulamayacak kadar dağınık olan daha büyük yapıları bulmasına da olanak tanır.

Ana uygulama alanı görüntü tanıma gibi görünse de, konuşma tanıma DNN'lerinde evrişimli katmanların uygulamaları da vardır. Bu durumda, CNN genellikle bir frekans spektrumu üzerinde çalışır ve verileri sınıflandırabilmesi için özellik haritalarını çıkarmak için ona konvolüsyonlar uygular.

4.1.2 Tekrarlayan Sinir Ağları

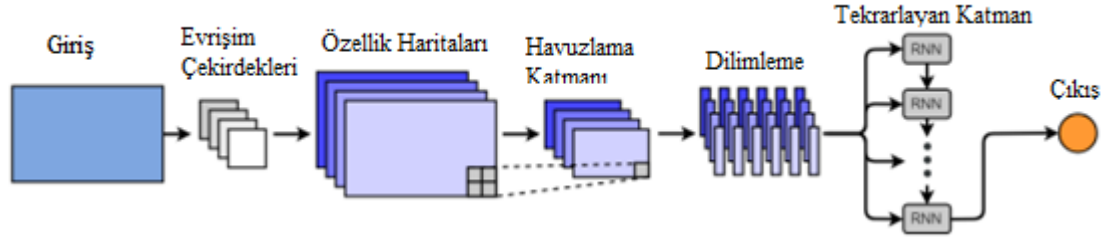
Bu tezdeki problem cümlesi durumunda, ses dosyasının sınıflandırma açısından daha ilginç olan bazı kısımları olabilir. Örneğin, bir motorun ateşleme sesi, motorun markası ve hatta arabanın modeli hakkında çok fazla bilgi verebilir, ancak kaydedilen sesin çoğunluğunun motorun çalıştığı bir ses dosyasını incelerken bu özellik kaybolabilir. Bu nedenle, bir tür zamansal perspektife sahip olan ve sesin nihai sınıflandırması düşünüldüğünde birden çok zaman adımından gelen verileri kullanabilen bir DNN tercih edilir. Bu yeteneklere sahip bir DNN, tekrarlayan sinir ağıdır (RNN). RNN'ler, ses tanımada başarılı olduğu kanıtlanmış ve konuşma tanıma sistemlerinde yaygın olarak kullanılan belirli bir DNN türüdür. Bir RNN'nin en az bir geri besleme bağlantısı vardır, bu tipik olarak katmandan katmana ölçektedir. Yani bir katmandaki tüm nöronların kendisiyle bir geri besleme bağlantısı vardır.

Bir zaman adımından nöronun çıktısı kaydedilerek ve bir sonraki adımda normal girdiye eklenerek bir geri besleme bağlantısı kurulur. Bu, ağın önceki çıktıları "hatırlamasına" neden olur ve birden çok zaman adımında devam edebilir, bu da geçmiş çıktıların mevcut sınıflandırmayı etkileyebileceği bir sistemle sonuçlanır.

4.1.3 Evrişimli Tekrarlayan Sinir Ağları

Adından da anlaşılacağı gibi, bir evrişimli tekrarlayan sinir ağı (CRNN), CNN ve RNN mimarilerinin bir karışımıdır. Genellikle bu ağ, öznitelik çıkarma için CNN'nin ve verilerin

zamansal yönlerini yakalamak için RNN'nin güçlü yanlarını kullanır. Bu şekilde, CRNN genellikle bir CNN'ye benzer şekilde oluşturulur, ancak son yoğun katman(lar) bir RNN ile değiştirilir.



Şekil 4-3: Basit bir CRNN.

Ağ akışında adım adım ilerleyen CRNN'nin ilk kısmı bir CNN gibi çalışır. Ancak evrişim ve havuzlama adımları yapıldığında havuzlanmış özellik haritaları bölümlere ayrılır. Bu bölümler, en ilginç veri dizilerini (genellikle zaman eksenini) içeren eksen boyunca ayrılır. Segmentler daha sonra RNN'ye gönderilir. Son olarak, son RNN nöronu, sonucunu örneğin sınıfını belirleyen çıktı katmanına gönderir.

2017 yılında Choi ve arkadaşları müzik parçalarındaki özellikleri (tür, ruh hali, enstrüman vb.) sınıflandırmak için bir CNN ve RNN melezi kullanmıştır. Özellik çıkarma için bir CNN ve özelliklerin zamansal yorumu için bir RNN kullanarak, müziği sınıflandırma görevi verildiğinde CNN'lerden daha iyi performans gösteren bir CRNN yarattılar.

4.2 MAKİNE ÖĞRENME

Yapay zeka (AI), son zamanlarda en çok konuşulan kelimelerden biri haline geldi. Yapay zeka, makinelerin akıllı olarak kabul edilen görevlerini üstlenmelerini sağlayan kontrol algoritmalarını ve sistemleri tanımlamak için kullanılan daha geniş bir terimdir. Makine öğrenimi (ML), yapay zekanın makinelerle uygulanmasıdır, çünkü ilgili verilere makinelerle erişim, onların kendileri için öğrenmelerini sağlayacaktır. Makine öğrenmenin çeşitli şekillerde farklılık gösteren farklı türleri olabilir, ancak ortak olan bir şey, hepsinin veriler üzerinde çalışmasıdır. Bu nedenle, veri alaka düzeyi ve erişilebilirlik, herhangi bir makine öğrenme performansındaki ana anahtarlardır. Makine öğrenme, AI'nın alt kümesi olarak da adlandırılır. Makine öğrenimini mevcut teknolojinin en son hali olarak düşünmek yanlış olmaz.

Yapay Zeka (AI), amacı gerçekleştirdiği görevlerde akıllı davranış sergileyen bir sistem oluşturmak olan bir bilgisayar bilimi alanıdır. Bir sistem, kendisine atanan süreçle ilgili bir görevi

herhangi bir insan müdahalesi olmadan ve yüksek doğrulukla gerçekleştirmeyi öğrendiğinde akıllı olduğu söylenebilir. Makine Öğrenimi (ML), bir bilgisayarın öğrenmesini sağlamak için algoritmaların ve tekniklerin geliştirilmesi, anlaşılması ve değerlendirilmesi ile ilgilenen bir AI alt alanıdır. ML istatistik, finans, insan psikolojisi ve beyin modelleme gibi diğer disiplinlerle bağlantılıdır. Birçok ML algoritması, model oluşturmak için veri analizini kullandığından, istatistik bu alanda önemli bir rol oynar.

Bir bilgisayarın ilgilenmek üzere atandığı bir işlem veya görev, bilgi veya görev alanı (veya yalnızca etki alanı) olarak adlandırılabilir. Alan tarafından üretilen veya alandan elde edilen bilgiler, bilgi tabanını oluşturur. Bilgi tabanı, Boolean, sayısal ve ayrık değerler, ilişkisel değişmezler ve bunların kombinasyonları kullanılarak çeşitli şekillerde temsil edilebilir. Bilgi tabanı genellikle girdi-çıkı çiftleri şeklinde temsil edilir, burada girdi tarafından temsil edilen bilgi etki alanı tarafından verilir ve etki alanı tarafından üretilen sonuç çıktıdır. Bilgi tabanından gelen bilgiler, etki alanının veri oluşturma sürecini (yani, belirli bir girdi için çıktı sınıflandırmasını) tasvir etmek için kullanılabilir. Veri üretme süreci bilgisi, etki alanının çalışmasının iç kısımlarını tanımlamaz, ancak yeni girdileri buna göre sınıflandırmak için kullanılabilir.

Bilgi tabanı büyüdükçe veya karmaşıklaştıkça, veri üretme süreci (alan) hakkında yeni ilişkiler çıkarmak insanlar için zorlaşıyor. ML algoritmaları, alanı doğru ve verimli bir şekilde temsil eden hesaplama modelleri oluşturmak için alandan ve bilgi tabanından öğrenmeye çalışır. Oluşturulan model, alanın veri oluşturma sürecini yakalar ve bu modelin kullanımıyla algoritma, alandan daha önce gözlemlenmemiş örnekleri eşleştirebilir.

Oluşturulan modeller, kullanılan ML algoritmasına göre farklı biçimler alabilir. Model formlarından bazıları karar listeleri, çıkarım ağları, kavram hiyerarşileri, durum geçiş ağları ve arama-kontrol kurallarıdır. Çeşitli ML algoritmalarının kavramları ve işleyişi farklıdır ancak ortak amaçları temsil ettikleri alandan öğrenmektir.

ML algoritmaları, etki alanının bir modelini oluşturmak için bir veri kümesine ihtiyaç duyar. Veri kümesi, etki alanındaki örneklerin bir koleksiyonudur. Her örnek, etki alanından o örneğin özelliklerini tanımlayan bir dizi öznitelikten oluşur. Bir öznitelik, öznitelik türüne göre ayrık veya sürekli olabilen bir dizi değer alır. Ayrık (veya nominal) nitelikler farklı değerler alırken (örneğin, renk = kahverengi, hava durumu = güneşli), sürekli (veya sayısal) nitelikler sayısal değerler alır (örneğin, mesafe = 3,5 metre, yağmur = 40 mm).

Her örnek, bir dizi girdi özniteliği ve bir çıktı özniteliğinden oluşur. Girdi öznitelikleri, öğrenme algoritmasına verilen bilgilerdir ve çıktı özniteliği, aktivitenin bu bilgi üzerindeki geri bildirimini içerir. Çıktı özniteliğinin değerinin, girdi özniteliklerinin değerlerine bağlı olduğu

varsayılır. Öznitelik, kendisine atanan değerle birlikte, bir örneği bir özellik vektörü yapan bir özelliği tanımlar. Bir algoritma tarafından oluşturulan model, örnekteki girdi özniteliklerini çıktı özniteliğinin bir değerine eşleyen bir işlev olarak görülebilir.

Basit bir gözle bakıldığında büyük miktarda veri rastgele görünebilir, ancak daha yakından incelendiğinde, içinde örüntüler ve ilişkiler bulunabilir. Ayrıca verileri oluşturan mekanizma hakkında bir fikir verebilir. Ayrıca verilen verilerden ilişkilerin çıkarılması işlemi olarak da adlandırılır. Genel olarak veri madenciliği, bir model öğrenmenin verimliliği konusunun, öğrenmenin etkinliği ile birlikte ele alınması bakımından makine öğrenmesinden farklıdır. Veri madenciliği problemlerinde, veri üretme sürecini alan olarak ve alan tarafından üretilen verilere bilgi tabanı olarak bakılabilir. Böylece, ML algoritmaları, kendisine verilen veri kümesine dayalı olarak veri üretme sürecini tanımlayan bir modeli öğrenmek için kullanılabilir. Modeli oluşturmak için algoritmaya verilen verilere, bilgisayar bu verilerden öğrenmek için eğitildiğinden eğitim verileri denir ve oluşturulan model, öğrenme sürecinin sonucudur. Bu model artık daha önce görülmemiş örnekleri tahmin etmek veya sınıflandırmak için kullanılabilir. Modeli değerlendirmek için kullanılan yeni örneklerle test seti denir. Bir modelin doğruluğu, test kümesindeki hedef özniteliğin tahmin edilen ve gerçek değeri arasındaki farktan tahmin edilebilir.

Bir algoritmanın, girdi verileriyle etkileşime dayalı olarak bir sorunu modellemesinin farklı yolları vardır. Bir algoritmanın benimseyebileceği öğrenme stillerini veya öğrenme prosedürlerini dikkate almak, makine öğrenimi ve veri madenciliğinde yaygındır. Bu öğrenme stilleri şu şekilde tanımlanır:

Denetimli öğrenme: Girilen verilere eğitim verileri denir ve bilinen etikete sahiptir. Tahmin yapılması gereken bir eğitim süreci ile bir model hazırlanır ve bu tahminler yanlış olduğunda düzeltilir. Eğitim süreci, model eğitim verilerinde istenen doğruluk düzeyine ulaşana kadar devam eder. Örnek problemler sınıflandırma ve regresyondur.

Denetimsiz öğrenme: Girdi verileri etiketlenmemiştir ve bilinen bir sonucu yoktur. Girdi verilerinde bulunan yapılar çıkarılarak bir model hazırlanır. Örnek problemler birliktelik kuralı öğrenme ve kümelemedir.

Yarı denetimli öğrenme: Girdi verileri, etiketli ve etiketsiz örneklerin bir karışımıdır. İstenen bir tahmin problemi var ama modelin tahminlerde bulunmanın yanı sıra verileri organize edecek yapıları da öğrenmesi gerekiyor. Örnek problemler sınıflandırma ve regresyondur.

Makine öğrenimi, belirli bir veri kümesinden belirli görevleri bulmak ve gerçekleştirmek için kullanılan teknikler kümesidir. Mevcut olan belirli bir alandan gelen veriler, daha sonra sınıflandıran ve geleceği tahmin eden sınıflandırıcının öğrenme sürecine ana itici güç olarak

hizmet eder. Algoritma daha sonra kurallarını ve işlevlerini ya kendisi (denetimsiz) ya da kullanıcıların seçtiği özelliklerden (denetimli) oluşturur. Bir algoritmayı öğrenmenin ve eğitmenin bu aşamasına eğitim aşaması denir. Bundan sonra öğrenilen algoritma örneği, eğitim ve test aşamalarında sırasıyla doğrulama ve test veri kümesine karşı test edilir. Doğruluk ve performanslar daha sonra bir simülasyon ortamında standart kıyaslama algoritmalarına göre sınıflandırılır. İnsanların çok fazla çaba harcamasına gerek olmadığı için bu yaklaşım, geleneksel programlama yaklaşımlarından oldukça daha uygundur. Gittikçe daha fazla veri mevcut hale geldikçe, örnek verilerden daha fazla genellemeler çıkarılabilir. Bununla birlikte, makine öğrenimi uygulamasının başarılı olması için, ders kitaplarında bulunması çok zor olan 'siyah sanatı' iyi kavramak gerekir.

Makine öğrenimi, veri analizi sürecini otomatikleştiren bir analitik çözüm biçimidir. Makine öğrenimi, yinelemeli olarak verilerdeki kalıpları arayan ve sonunda bilgisayarı açıkça programlamadan farkında olmasını sağlayabilecek bazı yararlı gizli bilgileri öğrenen algoritmalarından oluşur. Günümüzün makine öğrenimi, geçmişin makine öğreniminden çok değişti. Günümüzde, bir gerçeklik haline gelen büyük verilere karmaşık matematiksel işlemler uygulama olasılığı nedeniyle, Makine öğrenimi algoritmaları daha hızlı ve daha hızlı tasarlanmaktadır.

Günümüzde görüşmelerde çok popülerlik kazanan yaygın olarak uygulanan uygulamalardan bazıları şunlardır: Google ve tesla'nın kendi kendine giden arabaları. Makine öğrenimi algoritmalarının kapasitesi nedeniyle hepsi pratik olarak mümkündür.

Müşterinin satın alma gücü, zevki ve geçmiş sipariş geçmişine göre ML tekniklerine dayalı ilgi alanı ürün tavsiye sistemleri. Satın alma geçmişi, bu öneri sistemlerinin ruhunu yönlendirir. Müşterilere sunulan ürün seti, k ortalama kümeleme tekniği kullanılarak zaten atanmış olan satın alma güçlerine veya dinamik fiyat grubu aralığına göre seçilir. Son olarak, potansiyel müşteriler tarafından ürün satın alımlarını tahmin etmek için test verilerine dayalı bir ikili doğrusal lojistik regresyon eğitilmiş sınıflandırıcı çağrılır.

Bir veri madenciliği ve dilsel kurallar toplama yaklaşımı olan twitter tweet'lerinde Duygu Analizi, birinin başkalarına karşı tuttuğu yanıt türlerini tahmin edebilen bir öğrenme algoritması biçimidir.

Çevrimiçi Kredi işlem satıcıları, önceki vakalara dayalı olarak istenmeyen posta vakalarını tespit etmek ve tahmin etmek için şu anda makine öğrenimi tekniklerini kullanıyor. Bu, bu kredi satıcılarının hizmetlerini ve müşterilerinin memnuniyet düzeylerini iyileştirmeye yardımcı olur.

E-posta spam filtreleme, gelen e-postalardan gelen spam filtreleme sınıflandırmasının genel olarak anlaşılması için bugün bile sunulan klasik makine öğrenimi örneğinin bir parçasıdır.

Bu, her seferinde bir e-postayı kontrol etmek için gereken insan çabasını azaltmaya yardımcı olur ve bilgisayar korsanlığına yatkın bilgisayarların bir dereceye kadar güvenli olmasını sağlayabilir. Basit filtreleme teknikleri, temel karar ağacı tabanlı yaklaşımı kullanırken, karmaşık anti-virüsler geleneksel-hibrit algoritma kombinasyonel yaklaşımını kullanabilir.

Kablosuz yerel ağ (WLAN) çağının başlangıcından beri, genel ağ güvenliği ve izinsiz giriş ile ilgili endişeler uzun süredir tartışılmaktadır. Şimdi, farklı ağ izinsiz giriş türlerinin olası nedenlerini tespit etmek için en son makine öğrenimi yaklaşımları yapılmaktadır. İzinsiz giriş tespiti, öğrenme sınıflandırıcısının bunu tahmin etmesi için sağlanan geçmiş izinsiz giriş verileri kadar yüksek doğruluktadır. r2I, u2r, ağ araştırmaları ve DoS Saldırılarının tespiti, farklı belirgin özellik tabanlı eğitilmiş sınıflandırıcıların varlığını gerektirir. Saldırı verisi eğitim özellikleri, izinsiz giriş saldırılarının türüne ve düzeyine bağlı olarak, ana bilgisayar tabanlı özellikler veya ağ tabanlı olabilir. Son zamanlarda genel olarak makine öğrenimine artan ilgi, sınıflandırma, veri madenciliği ve tahmin fenomeninin daha popüler hale gelmesinden kaynaklanmaktadır. Kurumsal gelişmeler, büyük kuruluşlar tarafından üretilen büyük veriler, bilgi işlem güçlerindeki teknolojik gelişmeler, giderek daha gelişmiş CPU/GPU mimarileri ve hepsinden öte, giderek daha ucuz hale gelen toplu veri depolamaları gibi şeyler, araştırmacının ilgisini çekmiştir. Bu, araştırmacıları toplanan verileri kullanmaya ve ondan düşünceli içgörüler çıkarmaya daha yakın hale getirir.

4.2.1 Sınıflandırma Teknikleri

Belirli bir örneği bir dizi ayrı kategoride sınıflandıran algoritmalara sınıflandırma algoritmaları denir. Bu algoritmalar, belirli bir girdiyi bir dizi ayrı çıktı değerinden birine sınıflandıran bir model veya bir dizi kural oluşturmak için bir eğitim seti üzerinde çalışır. Çoğu sınıflandırma algoritması, herhangi bir biçimde, ayrık veya sürekli girdi alabilir, ancak sınıflandırma algoritmalarından bazıları, tüm girdilerin de ayrık olmasını gerektirir. Çıktı her zaman ayrı bir değer biçimindedir. Karar Ağaçları sınıflandırıcıları; kural tabanlı sınıflandırıcılar, destek vektör makineleri ve Naives Bayes sınıflandırıcıları, sınıflandırma algoritmalarının örnekleridir.

Karar ağacı, verilerin kendisinden çıkarılan sınıflardaki verileri sınıflandırmak için kullanılır. Dinamik Karar ağaçları, gerçek zamanlı olarak beslenen verilere göre yeni sınıflar veya dallar oluşturan daha karmaşık uygulamalar olabilir. Geleneksel bir ulaşım analizi çözümü, uygulamada karar ağaçlarını bilmek için daha uygun bir yaklaşım olabilir. Ulaştırma hizmeti sağlayıcıları söz konusu olduğunda, gerçek zamanlı karar desteği, Eksik verilerin işlenmesi ve

insan bilinçli karar verme yetkileri gibi temel sorunlara çözümler oluşturmaları gerekir. Ulaştırma planlamasının kalitesini artırmak için çevrimdışı planlama çözümleri sağlamak için verilerden kuralları okumak için makine öğrenme teknikleri kullanılır. Operatörler tarafından gerçek zamanlı planlama sistemlerinde zaman tüketimi büyük bir sorundur. Yuvayı gelen istek için tahsis etmek için geleneksel optimal çözümlerin, o yuvadan yararlanma maliyeti açısından daha büyük bir hata oranına sahip olduğu bulunmuştur. Böylece sistem tarafından müşterilere anında karar ve geri bildirim sağlanmaktadır.

Sınıflandırma örnekleri, bir ses kaydında bulunan kuş türlerinin tanınmasını içerir (Duan vd., 2011), bir ses parçasını kategorize eder.

Konuşma ve konuşmacı tanımada birçok sınıflandırma tekniği kullanılmıştır (Cowling ve Sitte, 2003):

- Dinamik zaman atlama (DTW).
- Gizli Markov modelleri (HMM).
- Kuantizasyon vektörünü (LVQ) öğrenme.
- Yapay sinir ağları (YSA).
- K-En yakın komşu (kNN).
- Gauss karışım modelleri (GMM).
- Naives Bayes (NB)
- Karar Ağacı (DT).
- Destek vektör makineleri (SVM).

Araştırma çalışmalarının çoğu, Karar Ağaçları, Yapay Sinir Ağları, Gizli Markov Modelleri ve Destek Vektör Makineleri gibi denetimli makine öğrenme algoritmalarının (Rokach ve Maimon, 2005), yüksek doğruluklarından dolayı ses sınıflandırması ve segmentasyonu için en iyi seçimdir. Araştırmamızda farklı sınıflandırma algoritmaları kullanılmış, bu algoritmaların karşılaştırması yapılmıştır. Araştırmamızda farklı sınıflandırma algoritmaları kullanılmış, bu algoritmaların karşılaştırması yapılmıştır. Bu tezde kullanılan sınıflandırma algoritmalarını aşağıdaki alt bölümlerde detaylı olarak açıklanmaktadır.

4.2.2 Karar Ağacı

Karar Ağacı, örnek uzayının özyinelemeli bir bölümü olarak ifade edilen bir sınıflandırıcıdır (Rokach ve Maimon, 2005). Karar Ağacı, değerleri hiyerarşik bir biçimde ayıran

bir dizi kuralı temsil eden bir dallanmadır. Bu temsil, kolayca anlaşılan bir dizi IFTHEN kuralına çevrilebilir.

Genellikle ağacın üç tür düğümü vardır:

- Gelen kenarı olmayan ve sıfır veya daha fazla giden kenarı olmayan bir kök düğüm.
- Her biri tam olarak bir gelen kenarı ve iki veya daha fazla giden kenarı olan dahili düğümler.
- Her biri tam olarak bir gelen kenarı olan ve giden kenarı olmayan yaprak veya uç düğümler.

Bir Karar Ağacında, her yaprak düğüme bir sınıf etiketi atanır. Kök ve diğer dahili düğümleri içeren terminal olmayan düğümler, farklı özelliklere sahip örnekleri ayırmak için öznitelik test koşulları içerir. Örnekler, yol boyunca yapılan testlerin sonucuna göre ağacın kök düğümünden bir yaprak düğüme kadar gezinerek sınıflandırılır.

Ağaç yapısında, iki nokta üst üste, belirli bir yaprağa atanmış sınıf etiketlerini ve ardından, algoritmanın eksik değerleri işlemek için kesirli örnekleri kullanma biçimi nedeniyle ondalık sayı olarak ifade edilen, o yaprağa ulaşan örneklerin sayısını tanıtır.

Bu Karar Ağacının yalnızca sayısal öznitelikleri içerdiğine dikkat edin. Bu sınıflandırıcı göz önüne alındığında, akustik kaydın şiddetli yağmur içerip içermediğini (ağaçta sıralayarak) tahmin edilebilir. Her düğüm, test ettiği nitelik ile etiketlenir ve dalları, karşılık gelen değerlerle etiketlenir.

Doğal olarak karar vericiler, daha kapsamlı olarak kabul edilebilecekleri için daha az karmaşık Karar Ağacını tercih ederler (Rokach ve Maimon, 2005).

4.2.3 Destek vektör makineleri

Destek vektör makineleri temelde ikili sınıflandırıcılardır, ancak herhangi bir sayıda sınıf, ikili DVM sınıflandırıcıları birleştirilerek yerleştirilebilir. DVM sınıflandırma ilkesi, ilk olarak doğrusal olarak ayrılabilir sınıflar, yani bir karar sınırı olarak doğrusal bir hiperdüzlem kullanılarak mükemmel bir şekilde ayrılabilen iki sınıf dikkate alınarak tanımlanabilir. DVM eğitimi, herhangi bir karar sınırı ile hiper düzlemin her iki tarafındaki en yakın gözlem arasındaki marjı maksimize etme fikrine dayanır, yani amaç, en yakın sınıf temsilcisi noktalarından karar sınırına olan mesafeyi maksimize etmektir. Bu temsilcilere destek vektörleri denir. Maksimum marjlı hiperdüzlem tasarlamının optimizasyon problemi, Lagrange çarpanları kullanılarak çözülebilir.

Sınıfların doğrusal olmayan bir karar sınırıyla bile ayrılabilir olmadığı genel durumda, doğrusal olmayan SVM sınıflandırıcı, özellik vektörlerini, eğitim kümesinin doğrusal olarak ayrılmasının mümkün olduğu daha yüksek boyutlu bir uzaya etkin bir şekilde eşler. Marj daha sonra eğitim prosedürü sırasında daha yüksek boyutlu uzayda maksimize edilir. DVM eğitiminde marjın maksimize edilmesi, daha önce görülmemiş verilerle sunulduğunda sınıflandırıcının genelleştirme performansının iyileştirilmesini amaçlar.

4.2.4 Naive Bayes Sınıflandırıcı

Naive Bayes sınıflandırıcısı istatistiksel bir sınıflandırıcıdır. Belirli bir örneğin belirli bir sınıfa ait olma olasılığı gibi sınıf üyeliği olasılıklarını tahmin edebilir (Langley, 1994). Bayes sınıflandırması Bayes teoremine dayanmaktadır. Sınıflandırma algoritmalarını karşılaştıran çalışmalar, Karar Ağacı ve sinir ağı sınıflandırıcıları ile performans açısından karşılaştırılabilir olan Naive Bayes Sınıflandırıcı olarak bilinen basit bir Bayes sınıflandırıcı bulmuştur. Naive Bayes algoritması, çıktı özniteliği tarafından alınan olası bir değer verilen her bir girdi özniteliğinin koşullu olasılıklarını öğrenerek olasılıklı bir model oluşturur. Bu model daha sonra bize bir dizi girdi verildiğinde bir çıktı değerini tahmin etmek için kullanılır. Bu, verilen örnekteki öznitelik değerleri birlikte görüldüğünde olası bir çıktı değeri görmenin koşullu olasılığına Bayes teoremi uygulanarak yapılır. Algoritmayı tanımlamadan önce Bayes teoremini tanımlanmaktadır.

Naive Bayes algoritması, Bayes yaklaşımı kullanılarak kendisine verilen yeni bir örneği sınıflandırmak için bir dizi eğitim örneği kullanır. Bir örnek için, girdi nitelikleri verilen her bir çıktı sınıfını gözlemleme olasılığını bulmak için Bayes teoremi uygulanır ve örneğe en yüksek olasılığa sahip sınıf atanır. Kullanılan olasılık değerleri, eğitim setinde görülen öznitelik değerlerinin sayılarından elde edilir.

Naive Bayes algoritması, örnekteki tüm özniteliklerin ayrık olmasını gerektirir. Sürekli değerli nitelikler, kullanılmadan önce ayrıklaştırılmalıdır. Bir özniteliğin olasılık değerlerinin hesaplanmasında zorluklara yol açabileceğinden, bir öznitelik için eksik değerlere izin verilmez. Eksik değerlerle başa çıkmak için yaygın bir yaklaşım, onları o öznitelik için varsayılan bir değerle değiştirmektir. Bayes sınıflandırıcıları da büyük veri kümelerine uygulandığında yüksek doğruluk ve hız sergilemiştir (Hall vd., 2009).

4.2.5 Tembel Sınıflandırıcı

Tembel öğrenciler eğitim örneklerini depolar ve sınıflandırma zamanına kadar gerçek bir iş yapmazlar. IB1, verilen test örneğine Öklid mesafesinde en yakın eğitim örneğini bulan ve bu eğitim örneğiyle aynı sınıfı tahmin eden temel örnek tabanlı bir öğrencidir. Birkaç örnek en yakın olarak nitelendirilirse, bulunan ilki kullanılır.

Tembel IBk, sabit komşuluk k ile Örnek tabanlı öğrenen anlamına gelir (Cufoglu vd., 2008). IBk, etiketlenmemiş öznitelik vektörüne, öznitelik uzayında kendisine en yakın olan eğitim vektörünün etiketinin atanmasından oluşan bir k-en yakın komşu yöntemidir (Deng vd., 2006). kNN'de, daha önce görülmemiş bir x örneğinin sınıfını belirlemek için bir eğitim seti T kullanılır. İlk önce T'deki ortalama ve maksimum değerleri belirlenir ve benzer şekilde görünmeyen örnek x için de yapılır. Daha sonra, x'e en yakın T'deki k elemanını belirlemek için öznitelik uzayında uygun bir uzaklık ölçüsü kullanılır. Bu k en yakın komşunun çoğu benzer değerler içeriyorsa, x buna göre sınıflandırılır.

Bu sınıflandırma şeması, doğrusal olmayan karar sınırlarını açıkça tanımlar ve böylece performansı iyileştirir. Ayrıca, özellik dağılımı, örnek set T'de kullanılan veri noktalarının sayısının daha hızlı işleme için önemli ölçüde azaltılabileceğini önermektedir; sadece karar sınırına yakın olan örnekler gerçekten gereklidir.

4.2.6 Yapay Sinir Ağları

Bir sinir ağı (NN), en basit insan beynini temsil eden ve onun karar verme gücünü taklit eden popüler bir sınıflandırıcı biçimidir. Farklı katmanlar şeklinde birbirine bağlanan algılayıcıların kombinasyonlarından oluşan yapay sinir ağları (YSA). Sinir ağları, gerçek dünyada öğrenilmesi ve taklit edilmesi başka şekillerde zor olan karmaşık girdi ve çıktı korelasyonlarını temsil eder. Her bir algılayıcıya atanan ağırlıklar, ağırlık değerleri ve aktivasyon fonksiyonları şeklinde bilgi akışını kontrol eden güçlendirme zinciri görevi görür. Optik Elektrik devresi tanıma, YSA'nın birçok uygulamasından biridir. Elektrik devre şemasındaki sembolleri anlamak için, farklı sayıda YSA katmanı, aktivasyon fonksiyonları, eğitim verilerindeki olası benzersiz varyasyon tekniklerinin çeşitliliğini öğrenmek için kullanılır. Test verileri daha sonra, elle çizilmiş elektrik devresi bileşenlerinin gerçek tahmini ve sınıflandırma gücüne dayalı olarak nihai model performansını ölçmek için kullanılır.

4.2.7 Diğer Teknikler

En yaygın olarak kullanılan diğer iyi kurulmuş ML sınıflandırma ve tahmin teknikleri, rastgele ormanları (RF) içerir. Derin veri yorumlama, araştırma ve anlama için değişken seçimi kullanılabilir. RF, onları önemlerine göre özel bir şekilde sıralayarak korelasyon etkisini azalttığı için farklı değişken seçim teknikleri altında daha iyi performans göstermiştir.

İlişkilendirmeler ve sıra bulma, diğer araçlarla birlikte verilerin sınıflandırılmasına yardımcı olur. Gelen bir verinin mevcut veriye göre sınıflandırılabilmesi için veri kayıtlarının ontoloji tabanlı olay grafiklerine dönüştürülmesi gerekmektedir. Bu grafikler, zaman içindeki olay dizilerinin görsel temsilleridir. Olaylar açısından bu haritalama tekniği, olay şeklinde çizilen toplu kayıtlar arasındaki veri çakışmalarının çözülmesine yardımcı olacaktır. Benzer bir şekilde, hibrit yapay sinir ağı (YSA) - destek vektör makinesi (SVM), sürekli artan insan nüfusu ile bina enerji tüketimini tahmin etmek için kullanılmaya başlandı. En yakın komşu haritalama, K-ortalama kümeleme, kendini tanıyan haritalar, yerel optimal arama teknikleri (genetik algoritmalar), beklenti maksimizasyonu, Bayes ağları, temel bileşen analizi (PCA), çekirdek yoğunluğu tahmini, tekil değer ayrıştırması, gauss karışım modelleri , sıralı kapsama kuralı binaları, sınıflandırma, tahminler yapma ve verilerden faydalı içgörüler elde etme açısından en uygun çözümleri bulmak için genellikle tek başına veya bir dizi veri kümesindeki diğer algoritmalarla birlikte kullanılan gelişmiş makine öğrenimi algoritmalarından bazılarıdır.

4.3 MAKİNE ÖĞRENME İLKELERİ

Makine öğrenme veri odaklıdır ve verilerde geleneksel yöntemlere göre potansiyel olarak daha karmaşık modelleri modelleyebilir. Verileri modellemek ve tahmin etmek için klasik sinyal işleme teknikleri, kanıtlanabilir performans garantilerine dayanmaktadır. Bu yöntemler, Gaussian bağımsız ve özdeş olarak dağıtılmış (iid) değişkenler ve ikinci dereceden istatistikler (kovaryans) gibi basitleştirici varsayımları kullanır. Bununla birlikte, makine öğrenme yöntemleri ve özellikle son zamanlarda DL yöntemleri, geleneksel yöntemlere kıyasla bir dizi görevde gelişmiş performans göstermiştir. Ancak, makine öğrenme modellerinin artan esnekliği bazı zorluklarla birlikte gelir.

Makine öğrenme modellerinin karmaşıklığı ve eğitim algoritmaları, performanslarının garanti edilmesini zorlaştırır ve model yorumlamasını engelleyebilir. Ayrıca, makine öğrenimi modelleri önemli miktarda eğitim verisi gerektirebilir, ancak makine öğrenimi tekniklerinden

yararlanmak için "çok büyük" miktarlarda eğitim verisi gerekmediğini belirtelim. Bedava öğle yemeği teoremi nedeniyle, performansı bir görev için maksimize edilen modeller, diğerlerinde muhtemelen daha kötü performans gösterecektir. Yalnızca belirli bir görev için yüksek performans isteniyorsa ve yeterli eğitim verisi varsa, makine öğrenmenin faydaları bu sorunlardan daha ağır basabilir.

4.3.1 Denetimli ve Denetimsiz Öğrenme

Makine öğrenimi yöntemleri genellikle denetimli veya denetimsiz öğrenme görevleri olarak kategorize edilebilir. Denetimli öğrenmede görev, etiketlenmiş girdi ve çıktı çiftleri verilen girdilerden çıktıları tahmine dayalı bir eşleme öğrenmektir. Denetimli öğrenme en yaygın olarak kullanılan makine öğrenme kategorisidir ve lineer regresyon (ridge regresyon olarak da adlandırılır) ve en yakın komşu sınıflandırıcılar gibi tanınmış yöntemlerin yanı sıra daha karmaşık destek vektör makinesi (SVM) ve sinir ağı (NN) modelleri içerir. Denetimsiz öğrenmede, hiçbir etiket verilmez ve görev, veriler içindeki ilginç veya faydalı yapıyı keşfetmektir. Bu, veri görselleştirme, keşifsel veri analizi, anormallik tespiti ve özellik öğrenmeyi içeren birçok faydalı uygulamaya sahiptir. PCA, K-means, ve Gauss karışım modelleri (GMM'ler) gibi denetimsiz yöntemler onlarca yıldır kullanılmaktadır. Daha yeni yöntemler, t-SNE, sözlük öğrenme, ve derin temsilleri (örneğin, otomatik kodlayıcılar) içerir. Önemli bir nokta, denetimsiz yöntemlerin sonuçlarının, örneğin gizli faktörlerin keşfi veya veri görselleştirme gibi doğrudan kullanılabilmesi veya denetimli öğrenme performansını iyileştirmek için özelliklerin dönüştürülmüş sürümlerini sağladıkları denetimli öğrenme çerçevesinin bir parçası olmasıdır.

4.4 EĞİTİM VE TEST VERİLERİ

Makine öğrenmenin merkezinde, öğrenilen modellerin gözlemlenen veriler kadar gözlemlenmemiş veriler üzerinde de iyi performans göstermesi gerekliliği vardır. Modelin görünmeyen verileri iyi tahmin etme yeteneğine genelleme denir. Önce ilgili terminoloji, ardından bir makine öğrenme modelinin genelleştirilmesinin nasıl değerlendirilebileceğini açıklanmıştır.

Genellikle karmaşıklık terimi, veri ilişkilerinin veya makine öğrenme görevinin karmaşıklık düzeyini belirtmek için kullanılır. Belirli bir makine öğrenme modelinin, belirli bir karmaşıklıkta veri ilişkilerini (örneğin, özellikler ve etiketler arasındaki) iyi tahmin etme yeteneği, kapasitedir. Bu terimler kesin olarak tanımlanmamıştır, ancak bu kavramları

matematiksel olarak resmileştirmek için çaba gösterilmiştir. Örneğin, Vapnik-Chervonenkis (VC) boyutu, ikili sınıflandırıcılar durumunda model kapasitesini ölçmek için bir araç sağlar. Veri karmaşıklığı, özellikler arasında yararlı ilişkilerin bulunduğu boyutların sayısı olarak yorumlanabilir. Daha yüksek karmaşıklık, daha yüksek boyutlu ilişkiler anlamına gelir. Makine öğrenme modelinin kapasitesinin, eğitim verilerinin miktarı ile sınırlanabilir.

Genel olarak makine öğrenme modelleri, kapasiteleri sağlanan verilerin ve görevin karmaşıklığına uygun olduğunda en iyi performansı gösterir. Uyumsuz model-veri/görev karmaşıklıkları için iki durum ortaya çıkabilir. Düşük karmaşıklıkta bir görev için yüksek kapasiteli bir model kullanılırsa, model fazla sığacak veya eğitim setinin gürültüsünü veya kendine özgü özelliklerini öğrenecektir. Ters senaryoda, yüksek karmaşıklıkta bir görev üzerinde eğitilmiş düşük kapasiteli bir model, örneğin, verileri eksik bırakma veya temel fizik hakkında yeterli ayrıntıyı öğrenmeme eğiliminde olacaktır. Hem fazla uydurma hem de eksik uydurma, makine öğrenme modeli genellemesini bozabilir. ML modelinin model parametrelerine göre eğitim ve test gözlemleri üzerindeki davranışı, uygun model karmaşıklığını belirlemek için kullanılabilir. Daha sonra bunun nasıl yapılabileceğini tartışacağız. Eksik ve fazla uydurmanın, makine öğrenme modelinin önyargısı ve varyansı kullanılarak ölçülebilir. Önyargı, tahmini hedeflerimizin ortalaması y^* ile gerçek ortalama arasındaki farktır ve varyans, tahmin edilen ortalama değer civarında tahmini hedeflerin beklenen sapma karesidir.

Makine öğrenme modellerinin görünmeyen gözlemler üzerindeki performansını tahmin etmek ve böylece genellemelerini değerlendirmek için, tam eğitim setinden alınan bir dizi test verisi, model eğitiminden çıkarılabilir ve mevcut parametrelere verilen genellemeyi tahmin etmek için kullanılabilir.

4.5 AKUSTİKLER MAKİNE ÖĞRENME

Akustik veriler, insan konuşmasının makine yorumu ve hayvan seslendirmeleri, okyanus kaynağı lokalizasyonu ve okyanustaki jeofizik yapıların görüntülenmesi dahil olmak üzere çok geniş bir alanda bilimsel ve mühendislik anlayışları sağlar. Tüm bu alanlarda, veri analizi, veri bozulması, eksik veya seyrek ölçümler, yankılanma ve büyük veri hacimleri gibi bir dizi zorluk nedeniyle karmaşıktır. Örneğin, tek bir olayın veya ifadenin çoklu akustik varışları, kaynak lokalizasyonunu ve konuşma yorumlamasını makineler için zor bir görev haline getirir. Akustik tomografi ve biyoakustik gibi birçok durumda, büyük miktarda veri toplanabilir. Akustik özellikleri ve olayları manuel olarak tanımlamak için gereken insan çabası miktarı, veri

kümelerinin boyutu arttıkça hızla sınırlayıcı hale gelir. Ayrıca, verilerde insan bilişi tarafından kolayca tanınmayan örüntüler bulunabilir.

Makine öğrenimi teknikleri bilgisayarla görme, görüntü işleme, konuşma işleme ve (jeo)fizik bilimi dahil olmak üzere birçok alanda otomatikleştirilmiş veri işleme ve örüntü tanıma yeteneklerinde geniş ilerlemeler sağlamıştır. Akustikte makine öğrenme hızla Yukarıda belirtilen akustik zorluklara birçok zorlayıcı çözüm ile gelişen alan, makine öğrenme tabanlı tekniklerin akustik alanındaki potansiyel etkisi ve son zamanlarda gördükleri ilgi bu incelemeyi motive ediyor.

Geniş anlamda makine öğrenme, verilerdeki kalıpları otomatik olarak algılamak ve kullanmak için bir teknikler ailesidir. Makine öğrenmede modeller, örneğin, bir hayvanın türü veya akustik dizilerden alınan kayıtlara dayalı olarak konumları gibi ölçülen niteliklere dayalı olarak veri etiketlerini tahmin etmek için kullanılır. Bu ölçümler ve etiketleri genellikle belirsizdir; bu nedenle, istatistiksel yöntemler sıklıkla işin içindedir. Bu şekilde makine öğrenme, makinelerin bilgi kazanması veya "öğrenmesi" için bir araç sağlar. Makine öğrenme yöntemleri genellikle iki ana kategoriye ayrılır: denetimli ve denetimsiz öğrenme. Bu incelemede tartışılmamasına rağmen, pekiştirmeli öğrenme adı verilen üçüncü bir kategori de vardır. Denetimli öğrenmede amaç, etiketli girdi ve çıktı çiftleri verilen girdilerden çıktıları tahmine dayalı bir eşleme öğrenmektir. Etiketler, sırasıyla sınıflandırma ve regresyon için kategorik veya gerçek değerli skaler olabilir. Denetimsiz öğrenmede, hiçbir etiket verilmez ve görev, veriler içindeki ilginç veya faydalı yapıyı keşfetmektir. Denetimsiz öğrenmeye bir örnek, kümeleme analizidir (örneğin, K-ortalamlar). Denetimli ve denetimsiz modlar da birleştirilebilir. Yani, etiketler yalnızca kısmi veya bağlamsal bilgi verdiğinde yarı ve zayıf denetimli öğrenme yöntemleri kullanılabilir.

Akustik araştırmaları geleneksel olarak yüksek seviyeli fiziksel modeller geliştirmeye ve bu modelleri çevrenin ve ortamdaki nesnelerin özelliklerini çıkarmak için kullanmaya odaklanmıştır. Fiziksel prensip tabanlı modellerin karmaşıklığı, Fizik bilimlerinde daha fazla veri mevcut oldukça, gelişmiş akustik modelleri makine öğrenme ile daha iyi birleştirebileceğimiz beklenmektedir.

Makine öğrenmede, eldeki ML görevi için verilerde faydalı modeller sağlayan verilerin temsil modellerini, temsilleri tasarlamak için belirli alan bilgisini kullanmak yerine doğrudan verilerden öğrenmek tercih edilir. ML fiziksel modeller üzerine inşa edilebilir. Makine öğrenimindeki temsiller, verilerin belirli öznelilikleri olan girdi özellikleridir. Özellikler, insan konuşmasının spektral özelliklerini veya fiziksel bir ortamın morfolojik özelliklerini içerir. Bir makine öğrenme ardışık düzenine yönelik özellik girdileri, bir sinyalin (verilerin) ham ölçümleri

veya örneğin klasik temel bileşenler analizi (PCA) yaklaşımıyla elde edilen verilerin dönüşümleri olabilir. Gauss karışım modelleri (GMM'ler) dahil olmak üzere daha esnek temsiller, beklenti-maksimizasyon (EM) kullanılarak elde edilir. Makine öğreniminin temel kavramları hiçbir şekilde yeni değildir. Örneğin, temel bir sınıflandırma modeli olan lineer diskriminant analizi (LDA), 1930'ların başlarında geliştirilmiştir. K-means kümeleme algoritması ve modern sinir ağlarının (NN'ler) öncüsü olan perceptron algoritması, daha sonraki yıllarda geliştirildi. Algılayıcı algoritması yayınlandıktan kısa bir süre sonra, NN'lere olan ilgi, geri yayılım algoritmasının geliştirildiği 1980'lere kadar azalmıştır.

Akustikte makine öğrenme, son yıllarda önemli ilerleme kaydetmiştir. ML tabanlı yöntemler, geleneksel sinyal işleme yöntemlerine göre üstün performans sağlayabilir. Bununla birlikte, ML tabanlı yöntemlerin açık bir sınırlaması, veriye dayalı olmaları ve bu nedenle test ve eğitim için büyük miktarda veri gerektirmeleridir. Geleneksel yöntemler ayrıca birçok ML modelinden daha yorumlanabilir olma avantajına sahiptir. Özellikle derin öğrenmede, makine öğrenimi modelleri "kara kutular" olarak kabul edilebilir; yani, makine öğrenimi sisteminin girdileri ve çıktıları arasındaki araya giren işlemlerin mutlaka fiziksel olarak sezgisel olmadığı anlamına gelir. Ayrıca, serbest öğle yemeği yok teoremi nedeniyle, bir görev için optimize edilmiş modeller, diğerlerinde muhtemelen daha kötü performans gösterecektir. Bu derlemenin amacı, bu zorluklara rağmen ML'nin akustikte önemli bir potansiyele sahip olduğunu göstermektir.

4.6 MAKİNE ÖĞRENİMİ KAPSAMI

Dijital devrim, insanların karşılaştığı yeni sorunları hızlandırdı. Hızlı teknoloji gelişimi, elektrikli cihazlar ve çeşitli teknolojiler ile insan etkileşimi, elektronik kayıtlar ve web'deki insanların kayda değer gelişimi, her saniye devasa bir veri ölçüsü oluşturmak için el ele tutuşuyor. Önceki yirmi yıldan bu yana, dernekler, kolejler, bilim adamları ve akademisyenler, bu bilgileri örneğin tespitler, tanıma, araştırma, ayırt edici kanıt parçaları, tavsiyeler vb. için çeşitli amaçlarla kullanmak için yeni modeller ve yenilikler geliştirmeye çalışıyorlar. Pratikte bugünlerde tüm işletmeler, Tıp, mühendislik, finans, imalat ve benzeri alanlarda kullanılan çalışma prosedürlerinin ve yöntemlerinin iyileştirilmesi ve kesinliği için AI kullanıyor.

Makine Öğrenimi (ML) terimi, 1959'da Arthur Samuel tarafından kullanılan ve şimdiye kadar devasa farklı araştırmacılar tarafından yaratılan ve tamamlanan yarasadan hemen sonra. ML, Bilgisayar uygulamalarına kapasite sunmak için istatistik stratejilerini kullanan ve daha önceki yönergeler ve karakterize edilmiş programlar olmadan dinamik olarak verilerden kazanmalarını

sağlayan bir Yapay Zeka (AI) parçasıdır. ML, veriler ve durumlar arasındaki ilişkileri ve örnekleri değerlendirerek ve algılayarak, gizlenmiş verileri bilinir hale getirmeyi yapıyor. Bu amaçlar için ML, bilgisayar algoritmalarını kullanır, çoğunlukla, verilerden elde etmek için modeller üzerinde çalışılır, ayrıca hazırlanan benzer tür bilgiler üzerinde tahminlerde bulunur. Geleneksel uygulamalar, yazılım mühendisleri tarafından karakterize edilen kısıtlı ve sınırlı yönergelerde çalışmaktadır. Algoritmaların ilerlemesi ve kendi kendine öğrenme özellikleri, bilgisayar görüntüleri veya e-posta eleme ve benzeri bilgi odaklı uygulamalar oluşturmak geleneksel programlama stratejileriyle neredeyse olanaksız olduğundan, geleneksel uygulamaları fethetmelerini etkiler. Bu algoritmalar daha iyi kararlar almamıza ve güvenilirlik getirmemize yardımcı olur. ML'nin aşağıdakilere göre 3 temel türü vardır:

- Denetimli
- Denetimsiz
- Güçlendirme

4.6.1 Denetimli Öğrenme

Çoğu durumda, akılcı yapay zeka sorunları için bile denetimli öğrenme kullanılır. Denetimli öğrenmede biri bilgi girdileri, diğeri çıktılar için olmak üzere iki faktör vardır.

Algoritmalar, bir haritalama çalışması aracılığıyla bu faktörler arasındaki haritalamada ustalaşmaya çalışmaktadır. Algoritmalar bilgi kaynakları ile hazırlanır ve sonucu düşünülerek erişilebilir çıktılara dönüştürülür.

Algoritmaların çıktıyı tahmin ettiği noktada, uygun yanıtın doğru mu yanlış mı yoksa çıktıya yakın mı yoksa çıktıdan çok uzakta mı olduğunu kontrol edecektir. Bu prosedürler, algoritmaların performanslarını öğrenmelerine ve geliştirmelerine yardımcı olur. Öğrenme prosedürü, değerli bir doğruluk düzeyine geldiğinde sona erer. Yeni verinin geldiği noktada, geçmiş öğrenmelere ve bilgi girdi ve çıktıları arasındaki tahmini haritalamaya bağlı olarak çıktıyı öngörmeye çalışır.

Denetimli öğrenme algoritmaları, son bölümde incelenen regresyon ve sınıflandırmada toplanmıştır. Bu algoritmaların, veriler etiketlendiğinde çalışır durumda olduğu gibi olduğuna dair birkaç endişe vardır. Verileri gezegende başka bir petrol olarak kullanmaktan dolayı, veriler artık ücretsiz değil ve öğrenme için sosyal olaylar bilgisi maliyetli olabilir. Sağduyu kısmı için, bir ton denetimli algoritma test edildi. Her birinin kendine has özellikleri ve eksiklikleri vardı. Makine öğreniminde temel fikir olarak, her tür görev için en iyi sonucu veren böyle bir algoritma yoktur.

Buna göre algoritma seçmek, ML'de çalışırken herkesin düşünmesi gereken önemli bir konudur. En çok kullanılan yönetilen algoritmalar aşağıdaki gibidir:

- Vektör makineleri desteklemek
- Naif bayanlar
- Lojistik regresyon
- Doğrusal regresyon
- Doğrusal diskriminant analizi
- Karar ağaçları
- k-en yakın komşu
- Nöral ağlar

4.6.2 Denetimsiz Öğrenme

Denetimsiz algoritmalar, denetimli gibi değil, doğru cevaplara sahip değiller. Olduğu gibi, herhangi bir çıktı değişkeni ve aksaklıkları gidermek için bir rehber veya eğitmen yoktur. Algoritmalar bilgi özelliklerini anlamaya çalışıyor. Sadece girdi faktörlerine sahip olarak çıktıyı tahmin etmek için veri setinde örtülü ve gizli örnekler ararlar. Öngörü kapasitelerini öğrenmek ve geliştirmek için kullanacakları bir isim yoktur. Denetimsiz öğrenmeler, kümeleme ve ilişkilendirme problemlerinde toplanır.

Kümeleme: Bu tür görevlerde bilgiler, örneğin müşterileri satın alma davranışlarına göre gruplandırmak gibi gruplara bölünür.

İlişkilendirme: algoritmalar, verilerin geniş bitini açıklayabilecek standartları anlamaya çalışıyor, örneğin müşteri gömlek satın alıyor, genel olarak pantolon da satın alıyor.

Denetimsiz öğrenme hesaplamalarına bazı örnekler vermek gerekirse:

- K-ortalama kümeleme
- Apriori algoritma ilişkilendirmesi

4.6.3 Güçlendirme Makine Öğrenimi

Takviyeli öğrenme temelde bir çocuğun hayatının başlangıç döneminde öğrenmesi gibi çalışır. Bir çocuğun mükemmel çalıştığı noktada, birey uyarılır ve bir çocuk korkunç bir işi tamamlarken, sonuç şu ya da bu şekilde disiplinler ya da bunu tekrar etmemeyi tavsiye eder. Bu algoritmalar, burada çocukken doldurduğu bir aracı tarafından aynı görevi yapıyor. Acente çevre ile işbirliği yapar, taahhüdü doğru bir şekilde yerine getirdiği için ücret ve kötü yaptığı için ceza alır. Temsilci ödülleri artırmaya ve cezaları sınırlamaya çalıştığında.

Kendi kendini süren bir araç düşünün, eğer araç kaza yapmadan, caddeden çıkmadan veya korkunç durmadan hedefine ulaştıysa, ödül alacak, ancak atıfta bulunulan taahhütlerden herhangi birini yaptıysa ceza alacaktır. Bu sayede bir dahaki sefere araç ceza aldığı kazaları yapmayacaktır. Bu algoritmalara ayrıca dinamik programlama adı verilir ve kapsamlı çalışmalar ve araştırmalar bu algoritmaları geliştirmeye yönlendirilmektedir, yakın gelecekte çok sayıda görev için erişilebilir olacaklardır.

4.6.4 Destek Vektör Makinesi (SVM)

Destek Vektör Makinesi, sınıflandırma ve regresyon görevleri için kullanılabilecek denetimli bir makine öğrenimi algoritmasıdır.

Bunlar temelde hiper düzleme yakın olan veri noktalarıdır, örneğin nokta veri setinden kaldırılacaksa hiperdüzlem ayırıcının rolü de dahil olmak üzere birçok şeyi değiştirecektir. Bu, veri setlerindeki önemli faktörlerden biridir.

4.6.5 Karar Ağaçları

Bu tür algoritmalar, araştırmacılar tarafından en çok kullanılan denetimli algoritmalarlardır. Geliştiricinin, diğer ML modellerine kıyasla iyi bir basitliğe sahip yüksek doğrulukta modeller oluşturmasını sağlar. Karar ağacındaki temel sonuç “EVET” veya “HAYIR”dır.

Bu modeller, yukarıdaki şekilde gösterildiği gibi aynı şekilde bir ağaç olarak çalışmaktadır. Temel olarak tüm veri setini daha küçük parçalara böler ve süreç kökten alt bölümlerde devam eder. Yaprak düğümleri ve karar düğümleri ile bir ağaca ulaşmaya gider. İkincisi, kök düğüm adı verilen bir ağaç türüdür. Avantajı, kategorik ve sayısal veriler dahil olmak üzere her iki veri türünü de işleyebilmesidir.

4.7 REGRESYON ALGORİTMALARI

Çıktıyı temsil etmek için sürekli bir değer üretmek için girdi öznitelikleri tarafından alınan değerler üzerinde denklemlere veya matematiksel işlemlere dayalı bir model geliştiren algoritmalara regresyon algoritmaları denir. Bu algoritmaların girdisi, algoritmaya bağlı olarak hem sürekli hem de ayrık değerler alabilirken, çıktı sürekli bir değerdir.

Regresyon algoritmaları biyolojik çeşitlilik tahmini, orman yangını tespiti, akarsu akışı, finans, sağlık vb. birçok alanda kullanılmaktadır. Orman yangınları, insan hayatını tehlikeye atarken ekonomik ve ekolojik hasar yaratan önemli bir çevre sorunudur. Bu yangınları tespit etmek, böyle bir fenomeni kontrol etmede kilit bir unsurdur.



5. VERİ SETİ VE YÖNTEM

Bu bölümde bu tez çalışması kapsamında kullanılan akustik özellikler, derin öğrenme mimarisi ve veri seti tanıtılmaktadır.

5.1 AKUSTİK ÖZELLİKLER

Bu tez çalışması kapsamında farklı 2 boyutlu görsel temsil özelliklerinin sınıflandırma performansına olan etkileri değerlendirilmiştir. Bu doğrultuda periodogram, spektrogram ve oran haritası özellikleri kullanılarak özellik çıkarımı gerçekleştirilmiştir. Oran haritası özellikleri çıkartılırken 4 farklı IHC metodunun sınıflandırma performansına olan etkisi kıyaslanmıştır.

5.1.1 PERİODOGRAM VE SPEKTROGRAM ÖZELLİKLERİ

Esasında hem periodogram hem de spektrogram yöntemi bir sinyalin frekans spektrumunun görsel bir temsilini vermektedir. Periodogram ile dikdörtgen bir pencere kullanılarak giriş sinyali x 'in güç spektral yoğunluğu hesaplanır. Periodogram, otokorelasyon dizisinin yanlı (biased) tahmininin Fourier dönüşümüdür. x_n giriş sinyali, w_n bir çerçeve fonksiyonu ve Δt örnekleme aralığı olmak üzere periodogram şu şekilde tanımlanabilir;

$$\hat{P}(f) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} w_n x_n e^{-j2\pi f \Delta t n} \right|^2, \quad -1/2\Delta t < f \leq 1/2\Delta t \quad (1)$$

Spektrogram, hızlı Fourier dönüşümü (HFD) ile elde edilen büyüklük spektrumundan oluşan bir yığından meydana gelir. HFD, belirli bir ses vektörünün, uzunlukları w_s olan yüksek oranda örtüşen parçalara bölünmesiyle gerçekleştirilir ve spektral vektör f elde edilir. Bu durumda, δ ardışık iki çerçeve arasındaki artış olmak üzere mevcut çerçeve F 'nin spektral büyüklük vektörü f_F aşağıdaki gibi elde edilir:

$$s_F(N) = s(F\delta + n)w_n \quad \text{için } n = 0 \dots (w_s - 1) \quad (2)$$

$$f_F(k) = \sum_{n=0}^{w_s-1} s_F(n) e^{-j\pi n k w_s} \quad \text{için } k = 1 \dots (w_s 2 - 1) \quad (3)$$

5.1.2 ORAN HARİTASI TEMSİLLERİ

Oran haritası gösterimi, insan işitsel sisteminden esinlenmiştir (Brown vd., 1994; Meutznervd., 2017). Temel olarak, işitsel sinir ateşleme hızının spektro-zamansal modelini kodlar (Meutznervd., 2017). Her frekans kanalının zarfı, bir işitsel sinir lifinin anlık ateşleme hızı olarak açıklanabilir (Romero vd., 2019; Wang vd., 2006). Bu, kokleadaki mekanik transdüksiyon işlemine dayanmaktadır (Meutznervd., 2017).

Hız haritası gösterimini elde etme süreci dört temel adımdan oluşur (Brown vd., 1994). İlkinde, ses sinyaline ön vurgulama işlemi uygulanır. Daha sonra elde edilen sinyal üzerinde koklear filtreleme gerçekleştirilir. Bu işlem için, insan kokleasının frekans seçim özelliği GTF (Patterson vd, 1992) kullanılarak modellenmiştir. Bir sonraki aşamada, IHC modeli kullanılarak nöromekanik transdüksiyon gerçekleştirilir. Son olarak, sinyale Hamming penceresi uygulandıktan sonra uyartım oluşma olasılığı hesaplanır.

5.1.2.1 Ön Vurgulama Süreci

Dış ve orta kulak, kulak zarının ses basıncını farklı frekanslara göre artıran ve azaltan karmaşık bir akustik boşluk yaratır (Brown vd., 1994). Küçük ılıman ses yoğunlukları için dış ve orta kulakların yaklaşık olarak lineer olduğu söylenebilir. Bu durumda, rezonansları basit bir lineer filtre ile modellemek mümkündür; burada t zaman, $x(t)$ bir giriş sinyali ve $y(t)$ filtrelenmiş çıkış sinyalidir. Bu durumda ön vurgu süreci şu şekilde ifade edilebilir:

$$y(t) = x(t) - 0.97x(t - 1) \quad (4)$$

5.1.2.2 Koklear Filtreleme

Ön vurgulama işleminden sonra, insan kokleasının frekans seçim özelliği bir filtre bankası aracılığıyla modellenir. Bu işlem için bir GTF kullanılır (Patterson vd, 1992; De Boer vd, 1978), burada A genlik, j filtre sırası, f_c filtrenin merkez frekansı, ϕ faz, B bant genişliği ve t ise zamandır. Bu durumda filtre şu şekilde tanımlanabilir:

$$h(t) = At^{j-1}e^{-2\pi Bt} \cos(2\pi f_c t + \phi) \quad (5)$$

Faz, farklı frekans kanallarının her birinin başlangıç ve ofset sürelerinin karşılaştırılmasında kritik öneme sahiptir. Filtre bankası fazı telafi edilirse, frekans geçiş haritasında performans artar (Brown vd., 1994). (Patterson, 1988) 'de önerilen kompanzasyon için iki farklı yöntem kullanılmıştır. İlk yöntemde, zaman farkı, her bir dürtü yanıtının zarflarının tepe noktalarının hizalanması için filtrenin çıkışına sunulur. Bu şu şekilde ifade edilebilir:

$$t_c = \frac{j-1}{2\pi B} \quad (6)$$

İkinci yöntemde, faz düzeltme yoluyla, zamansal ince formdaki bir üst nokta, zarftaki bir üst nokta ile hizalanır. Bu şu şekilde ifade edilebilir:

$$\varphi_c = -2\pi f_c t_c \quad (7)$$

Denklem (6) ve (7), Denklem (5)'ye ikame edilirse, faz dengeli bir GTF elde edilir. Bu şu şekilde ifade edilebilir:

$$h(t) = A(t + t_c)^{j-1} e^{-2\pi B(t+t_c)} \cos(2\pi f_c t) \quad (t \geq -t_c) \quad (8)$$

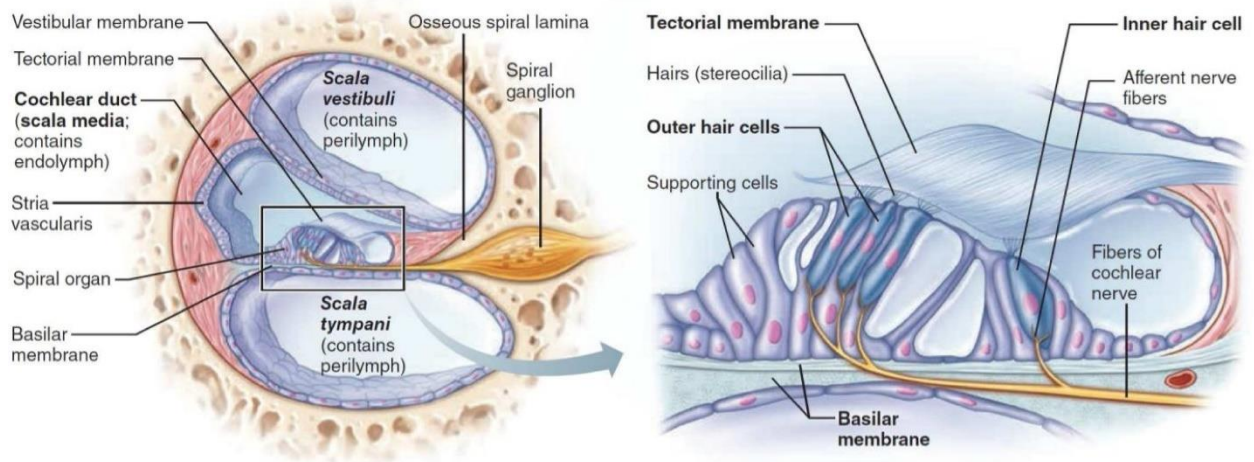
GTF sıra yanıtı, düşük frekans değerlerinde ve aynı frekans çözünürlüğünde geleneksel hızlı Fourier dönüşümü (FFT) filtre bankasından daha hızlıdır. GTF bankası, (Glasberg vd., 1990)'de önerilen eşdeğer bir dikdörtgen bant genişliği (ERB) ölçeği ile kullanıldı. Bu aslında MFCC'nin (Das vd., 2019) biyolojik olarak esinlenilmiş bir modifikasyonudur ve şu şekilde ifade edilebilir:

$$E(f) = 21.4 \log_{10}(4.37f + 1) \quad (9)$$

burada ERB sayısı $E(f)$ ve f kHz cinsinden frekanstır.

5.1.2.3 Nöromekanik Transdüksiyon

Koklea, ses basıncını baziler membranın titreşim hareketlerine dönüştürür ve daha sonra bu titreşimlerden nöral impulslar oluşturulur (Russo vd., 2019). Kokleayı karmaşık seslerin frekans bileşenlerini analiz eden bir sistem olarak düşünmek mümkündür çünkü baziler membran boyunca her pozisyonun belirli bir frekansa karşılık geldiği bir frekans analizörü işlevi gerçekleştirir (Russo vd., 2019). Koklea şeklini yaklaşık 3.5 cm genişliğinde ve 1 cm uzunluğunda küçük bir tüp olarak tanımlamak mümkündür (Russo vd., 2019). Baziler membran, kokleayı oluşturan ana unsurdur ve genişliği ve sertliği koklea boyunca değişir (Russo vd., 2019). Aynı zamanda içi sıvı dolu iki tüpü birbirinden ayırır (Russo vd., 2019) ve oldukça sofistike bir yapıya sahip olan korti organını içerir. Bu organ, baziler membran titreşimlerine yanıt verir ve sinir uyarılarına iletilmesine izin verir (Russo vd., 2019). Korti organı boyunca üç dış tüy hücresi (OHC) ve bir iç tüy hücresi (IHC) vardır (Russo vd., 2019). IHC'ler, saç hücrelerinin etrafındaki hareketi algılayan ve mekanik transdüksiyon yoluyla nöral uyarılar üreten ana sensörlerdir (Russo vd., 2019). Şekil 5-1, koklea ve korti yapılarını göstermektedir.



Şekil 5-1: Koklea (solda) ve (sağda) korti (büyütülmüş) görüntüleri [Marieb vd., 2016].

İşitsel modellerde, tüy hücresi iletimi, bir dizi sinir aktivitesinin bir ölçümünü elde etmek için bir girdi olarak baziler membranın hızı veya yer değiştirmesi kullanılarak temsil edilir (Beeston, 2015). Bu çalışmada 4 farklı IHC modeli kullanılmıştır.

Dau modelinde, spesifik baziler membran segmentinin çıkışındaki sinyal, yarım dalga doğrultulmuş ve 1 kHz'de düşük geçişli filtrelenmiştir. Bu aşama, baziler membranın mekanik salınımlarının iç tüy hücrelerinde reseptör potansiyellerine dönüşümünü kabaca simüle eder. Alçak geçiren filtreleme, esasen yüksek taşıyıcı frekansları için sinyalin zarfını korur.

Adaptasyonun etkileri geri besleme döngüleriyle simüle edildi. Bu aşama, sabit sinyalleri neredeyse logaritmik olarak sıkıştırırken, girişin hızlı dalgalanmaları daha doğrusal olarak dönüştürülür. Bu geri besleme döngüleri, geçici maskeleye etkilerini açıklamak için araştırılmıştır ve aşağıdaki alt bölümde daha kesin olarak açıklanacaktır. Geri besleme döngülerini takip eden aşamada, sinyal, 20 ms'lik bir zaman sabitine karşılık gelen 8 Hz'de düşük geçişli filtrelendirilmiştir. Bu değer, eşlik eden Dau ve diğerleri, 1996 belgesinde açıklanan simülasyonların sonucu ile önerilmiştir. (Dau et al., 1998)

Kullanılan bir diğer IHC metodu Joergensen yöntemidir. Bu yöntemin en temel özelliklerinden biri Hilbert dönüşümünün her gamaton filtresinin çıktısının zarfına uygulanmasıdır. Hilbert dönüşümü gerçekleştirildikten sonra elde edilen zarf, kesme frekansı 150 Hz olan tek bir modülasyonlu alçak geçiren filtreye girdi olarak uygulanır. Burada modülasyon-frekans alanında seçiciliğin olmadığı varsayılır. (Jørgensen & Dau, 2011)

Bir diğer yöntemde IHC'lerin etkin sinyal işlemesi, bir yarım dalga doğrultucu ve ardından 23 dB 770 Hz'lik bir kesme frekansına sahip beşinci dereceden bir alçak geçiren filtre tarafından modellenmiştir. Yaklaşık 770 Hz'nin altındaki frekanslar için alçak geçiren filtrenin çıkış üzerinde neredeyse hiçbir etkisi yoktur. Bu nedenle dalga formunun sadece negatif fazı kaybolur ve bu nedenle çıkışta dalga formunun ince yapısındaki zamanlama bilgisi korunur. 2000 Hz'nin üzerindeki frekanslar için, alçak geçiren filtreden sonra hemen hemen tüm faz bilgileri kaybolur ve bu aşamanın çıkışında sadece gelen sinyallerin zarfı bulunur. Aradaki frekanslar için kademeli bir faz bilgisi kaybı gözlemlenir. Bu şekilde model, işitsel sinir Kiang, 1975'te gözlemlenen artan frekansla faz kilitlenmesinin azalmasını etkin bir şekilde simüle eder (Johnson, 1980; Weis ve Rose, 1988; Bernstein ve Trahiotis, 1996; Breebaart ve diğerleri, 2001).

Kullanılan son yöntem Hilbert dönüşümüdür. Bu dönüşüm, bir spektral analizden minimum faz tepkisini elde etmek için kullanılan bir tekniktir. Ayrıca HT anlık frekansı bulmada kullanılabilen basit bir lineer operatördür. Burada $s(t)$ gammatone filtresinin çıkışındaki r sinyal olmak üzere HT şu şekilde ifade edilir (Hilbert, 1953).

$$HT[s(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{(t - \tau)} d\tau \quad (10)$$

$a(t) = \sqrt{s^2 + p^2}$ anlık genlik ve $\theta(t) = \arctan(p/s)$ faz fonksiyonu olmak üzere HT aracılığıyla analitik sinyali şu şekilde tanımlamak mümkündür:

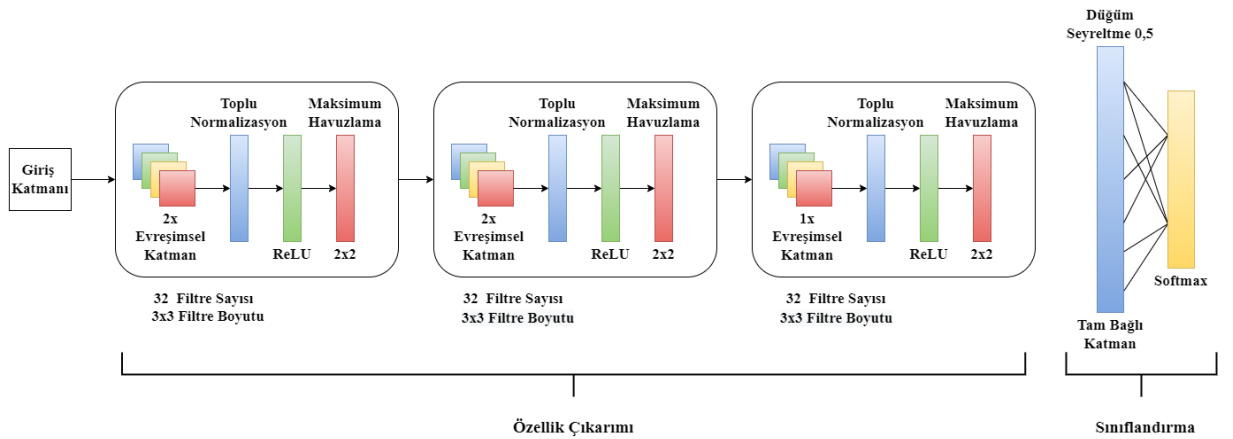
$$z(t) = s(t) + jp(t) = a(t)e^{j\theta(t)} \quad (11)$$

5.1.2.4 Oran Haritası

Hız haritası, her bir frekans kanalı için ayrı ayrı hesaplanır. Bu işlem için, IHC ölçüleri önce sızdıran bir entegratör kullanılarak düzeltilir. Burada tipik olarak zaman sabiti 8 ms'dir. Daha sonra her kanal örtüşen periyotlara bölünür ve Hamming penceresi uygulanır. Son olarak, her bir zaman çerçevesindeki örneklerin ortalaması alınır.

5.2 DERİN ÖĞRENME MODELİ

Çalışma kapsamında kullanılan CNN modeli 5 evrişimsel katman içermektedir. Tüm evrişimsel katmanlardaki filtre boyutu 3x3 olarak belirlenmiştir. İlk iki evrişimsel katmandaki filtre sayısı 32 olarak belirlenmiştir. Bu katmanları takiben sırasıyla bir toplu normalizasyon (batch normalization) katmanı, ReLU aktivasyon katmanı ve 2x2 boyutlu maksimum havuzlama katmanı yer almaktadır. Bu katmanlardan sonra tekrardan 2 ayrı evrişimsel katmanı yer almaktadır. Bu katmanlardaki filtre sayısı 64'tür. Bu evrişimsel katmanlardan sonra da ilk iki evrişimsel katmanı takip eden katmanlar yer almaktadır. Sonrasında 128 filtreden oluşan bir evrişimsel katman yer almaktadır. Bu katmandan sonra diğer evrişimsel katmanlardakilere ilave olarak düğüm seyreltme (dropout) işlemi uygulanmaktadır. Ayrıca 1024 düğümünden oluşan bir tam bağlı katman bulunur. Bu katmanın aktivasyon fonksiyonu olarak ReLU kullanılmıştır. Sonrasında ise sınıflandırma işleminin yapıldığı çıkış katmanı yer almaktadır. Bu katmanın düğüm sayısı sınıf sayısı ile aynıdır ve aktivasyon fonksiyonu olarak softmax kullanılmaktadır.



Şekil 5-2: CNN Mimarisi

Modellerin eğitimi aşamasında kayıp fonksiyonu olarak kategorik çapraz entropi kullanılmıştır. Optimizasyon için ise Adam metodu tercih edilmiştir. Tüm modellerde veri seti eğitim, test ve doğrulama olmak üzere 3 parçaya ayrılmıştır. 5 parça çapraz doğrulama testlerinde eğitim setinin %10'u doğrulama verisi olarak kullanılmıştır. Geleneksel bölme yöntemiyle yapılan testlerde ise eğitim, test ve doğrulama parçaları daha tutarlı bir kıyaslama yapabilmek için orijinal çalışmada belirtilen şekliyle kullanılmıştır. Tüm modeller 250 tekrar boyunca eğitilmiş ve doğrulama setinde en yüksek başarımlarını sağlayan model ağırlıkları test işlemi için kullanılmıştır. Bulgular kısmında rapor edilen sonuçların tamamı test veri setine ait skorları göstermektedir.

5.3 PERFORMANS DEĞERLENDİRME METRİKLERİ

Makine öğrenimi modellerini kıyaslamak için seçilen metrikler oldukça önemlidir. Bu çalışmada, farklı makine öğrenimi algoritmalarının performansını kıyaslamak amacıyla birkaç farklı metot kullanıldı. Sınıflandırma doğruluğu (ACC), hassasiyet (SENS), özgüllük (SPEC), kesinlik (PREC) ve F-skoru kullanılan kıyaslama metrikleridir.

ACC değeri doğru sınıflandırılmış örneklerin sayısının toplam örnek sayısına olan oranını verir. Sınıflandırma çalışmalarında yaygın olarak kullanılan bir kıyaslama metriğidir. ACC değeri şu şekilde hesaplanabilir:

$$ACC = \frac{TN + TP}{TN + TP + FP + FN} \quad (12)$$

ACC performans kıyaslaması için tek başına yeterli bir parametre olmayabilir. Bu nedenle, istatistiksel performans ölçüleri olan özgüllük, duyarlılık, kesinlik ve F-skor terimlerinin kullanılması birçok durumda doğru bir performans kıyaslaması için faydalı olur. "özgüllük" terimi, doğru tahmin edilen gerçek negatiflerin oranını ifade eder. Öte yandan "duyarlılık" terimi, doğru tahmin edilen gerçek pozitiflerin oranını gösterir. Doğru olan pozitif tahminlerin oranı ise "kesinlik" olarak ifade edilir. F skoru, SENS ve PREC değerlerinin harmonik ortalamasıdır. SPEC, SENS, PREC ve F-skorunu hesaplamak için kullanılan formüller şu şekildedir:

$$SPEC = \frac{TN}{TN + FP} \quad (13)$$

$$SENS = \frac{TP}{TP + FN} \quad (14)$$

$$PREC = \frac{TP}{TP + FP} \quad (15)$$

$$F - skor = \frac{2TP}{2TP + FP + FN} \quad (16)$$

Burada, gerçek pozitif (TP), doğru sınıflandırılmış pozitif örneklerin sayısını, gerçek negatif (TN) ise doğru şekilde sınıflandırılan negatif örneklerin sayısını ifade eder. Öte yandan yanlış pozitif (FP), pozitif olarak sınıflandırılan negatif örneklerin sayısı ve yanlış negatif (FN): negatif olarak sınıflandırılan pozitif örneklerin sayısıdır.

5.4 VERİ SETİ

Çalışmada IDMT-Trafik veri seti kullanılmıştır (Abeşer vd., 2021). Veri seti, Almanya Ilmenau içinde ve çevresinde üç şehir trafiği konumu ve bir köy yolu konumu dahil olmak üzere dört farklı kayıt konumundan hem yüksek kaliteli sE8 mikrofonları hem de düşük bütçeli mikroelektro-mekanik sistem (MEMS) mikrofonları ile yapılan zamanla senkronize edilmiş stereo ses kayıtlarını içerir. Kayıt senaryoları, farklı hız limitlerini (30, 50 ve 70 km/s) ve ayrıca ıslak ve kuru yol koşullarını içerir.

Veri setinin ilk rapor edilen hali dört sınıf içermektedir. Bunlar araba (3903 olay), kamyon(511 olay), otobüs (53 olay), motosiklet(251 olay) ve taşıt yok (4071) sınıflarıdır. Bu dağılım, yaygın trafik senaryolarında araç türlerinin doğal dengesizliğini yansıtır. Tablo 6-1’de veri setindeki örneklerin dağılımı görülmektedir. Tablo 5-1’de verilen değerler orijinal çalışmayla birebir aynı olacak şekilde geleneksel bölme işleminin yapıldığı testlerde kullanılmıştır.

Tablo 5-1: Veri setinin dağılımı

Veri Seti	Araba	Kamyon	Motosiklet	Taşıt Yok
Eğitim Seti	2471	290	132	2393
Doğrulama Seti	275	32	15	266
Test Seti	1157	189	99	1412

Veri seti örnek ve sınıf sayısı sonraki versiyonlarında artırılmıştır. Yukarıdaki sınıflara ilave olarak otobüs sınıfı eklenmiştir. Nihai durumda 30-50-70 km/s ve bilinmeyen hızlarda geçen 17506 aracın iki saniye uzunluğunda kaydedilmiş sesleri mevcuttur. Dosya metadatalarında kayıt tarihini yıl-ay-gün saat ve dakika olarak mevcuttur. Kayıtlar [Fraunhofer-IDMT, Schleusinger-Allee, Langewiesener-Strasse, Hohenwarte] adresinde oluşturulmuştur. Metadata olarak araç sınıfları otobüs, araba, motosiklet ve kamyon olarak ayrılmıştır. Kayıt sırasında hava şartları kuru ve ıslak olacak şekilde (D-W) sınıflandırılmıştır. Araçların hareket yönleri sağ ve sol olarak ayrılmıştır. Metadatada SE, sE8 tipi; ME ise MEMS tipi ICS-43434 tipi mikrofonlar kullanılıp 1+2 veya 3+4 olacak şekilde stereo kanallara kaydedilmiştir.

`2019-10-23-16-20\_Fraunhofer-IDMT\_30Kmh\_990828\_A\_D\_CL\_SE\_CH34.wav` bu örnekte, dosya ismine bakarak metadatayı kullanıp araba sesinin sol taraftan saatte 30km hızla sE8 3+4 stereo kanalı 2019.10.23 saat 16:20' de öğleden sonra hava kuru iken Fraunhofer IDMT lokasyonunda kaydedildiği çıkarımı yapılabilmektedir. Tablo 5-2'de veri setinin nihai durumundaki sınıf ve olay sayıları verilmiştir.

Tablo 5-2: Veri setinin nihai halindeki sınıflar ve olay sayıları

Sınıf	Olay Sayısı
Otobüs	106
Araba	7804
Motosiklet	430
Taşıt Yok	8144
Kamyon	1022

6. BULGULAR VE TARTIŞMA

Bu tez çalışması kapsamında öncelikli olarak spektrogram ve periodogram özellikleri test edilmiştir. Bu testler kapsamında lineer ve logaritmik ölçekte testler gerçekleştirilmiştir. Buradan elde edilen sonuçlar doğrultusunda oran haritası özellikleri logaritmik ölçek kullanılarak test edilmiştir. Oran haritası özellikleri elde edilirken 4 farklı IHC metodunun performansı kıyaslanmıştır.

6.1 SPEKTROGRAM VE PERİODOGRAM TESTLERİ

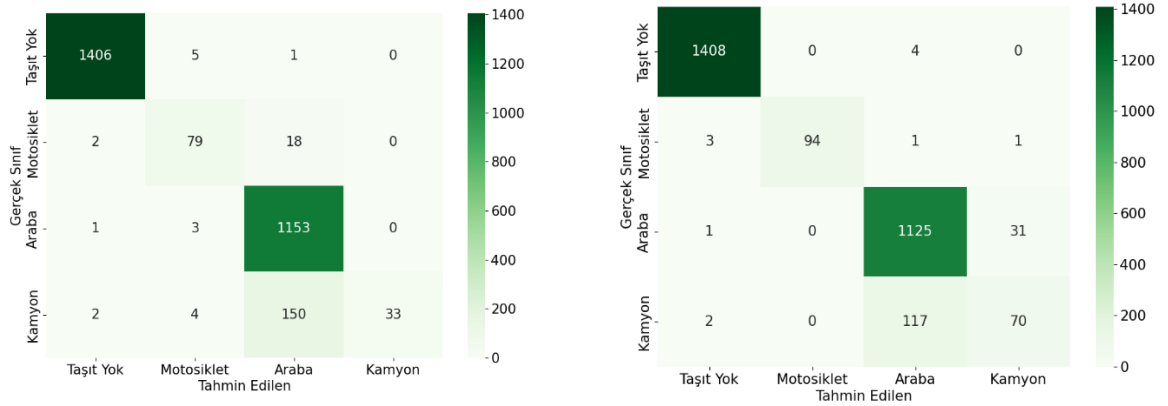
Çalışma kapsamında 4 farklı model test edilmiştir. Çalışmada kullanılan veri seti stereo olduğu için iki ayrı ses kanalına ait veri içermektedir. Bu çalışmadaki tüm modellerde ikinci ses kanalına ait veriler kullanılmıştır. Ses kanalının seçilmesinin ardından ilgili kanaldaki veriye denormalizasyon işlemi uygulanmıştır. Sonra da periodogram ve spektrogram özellikler çıkarılmıştır. Her iki özellik çıkarma işleminde de çerçeve fonksiyonu olarak Hamming kullanılmıştır. Spektrogram özelliklerinin elde edilmesi aşamasında çerçeve uzunluğu 1024 ve örtüşme oranı %50 olarak kullanılmıştır. Her iki özellikle de hem lineer hem de logaritmik ölçekte testler gerçekleştirilmiştir. Tüm modeller ile 50 devir boyunca eğitim gerçekleştirilmiştir. Bunların içerisinde doğrulama veri setinde en yüksek başarımlarını gösteren ağırlık değerleriyle test işlemi gerçekleştirilmiştir. Eğitim, doğrulama ve test setine ayırma aşamasında orijinal çalışmayla (Abeßer vd., 2021) birebir aynı örnekler kullanılmıştır. Bu bölümde rapor edilen sonuçlar test setine ait değerlerdir.

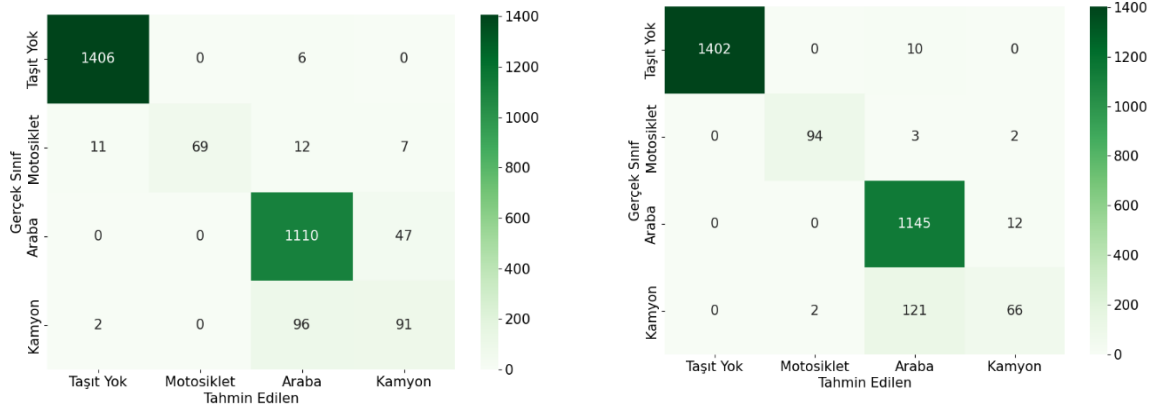
Tablo 6-1’de sınıf bazlı elde edilen sonuçlar görülmektedir. Tüm modeller için Sens sonuçları incelendiğinde örnek sayılarının az olduğu motosiklet ve kamyon sınıflarında elde edilen skorların taşıt yok ve araba sınıflarına kıyasla belirgin bir şekilde düşük olduğu görülmektedir. Dört model içinde en düşük Sens değerleri kamyon sınıfında elde edilmiştir. Burada en düşük değer lineer periodogram modelinde %17.46 olarak gerçekleşmiştir. Motosiklet sınıfında spektrogram modellerinin Sens değeri taşıt yok ve araba sınıflarına kıyasla yaklaşık olarak %3-5 daha düşük olarak gerçekleşmiştir. Bununla birlikte periodogram modellerinde motosiklet sınıfının Sens performansı taşıt yok ve araba sınıfına kıyasla belirgin bir şekilde daha düşüktür. Burada motosiklet ve kamyon sınıfları için Sens performansının düşük olmasının nedeni azınlık sınıf olmalarından kaynaklandığı değerlendirilmektedir. Makine öğrenmesi modelleri geneli itibarıyla çoğunluk sınıfına eğilimlidir [Ozer vd., 2021].

Tablo 6-1: Sınıf bazlı sonuçlar

		Spektrogram (Lineer)	Spektrogram (Logaritmik)	Periodogram (Lineer)	Periodogram (Logaritmik)
Sens	Taşıt Yok	99.72	99.29	99.58	99.58
	Motosiklet	94.95	94.95	79.80	69.70
	Araba	97.23	98.96	99.65	95.94
	Kamyon	37.04	34.92	17.46	48.15
Spec	Taşıt Yok	99.58	100.00	99.65	99.10
	Motosiklet	100.00	99.93	99.56	100.00
	Araba	92.82	92.12	90.06	93.29
	Kamyon	98.80	99.48	100.00	97.98
Prec	Taşıt Yok	99.58	100.00	99.65	99.08
	Motosiklet	100.00	97.92	86.81	100.00
	Araba	90.22	89.52	87.22	90.69
	Kamyon	68.63	82.50	100.00	62.76
F1	Taşıt Yok	99.65	99.64	99.61	99.33
	Motosiklet	97.41	96.41	83.16	82.14
	Araba	93.59	94.01	93.02	93.24
	Kamyon	48.11	49.07	29.73	54.49

Spec metriği için tüm modellerde elde edilen skorlar Sens metriğine kıyasla daha dengelidir. Burada tüm modeller için araba sınıfında elde edilen değerlerin diğer sınıflara kıyasla düşük olduğu görülmektedir. Şekil 6-1’de verilen karışıklık matrisleri incelendiğinde araba sınıfı için HP değerlerinin belirgin bir şekilde yüksek olduğu görülmektedir. Buna bağlı olarak bu sınıf için Spec değeri daha düşük olarak gerçekleşmiştir.



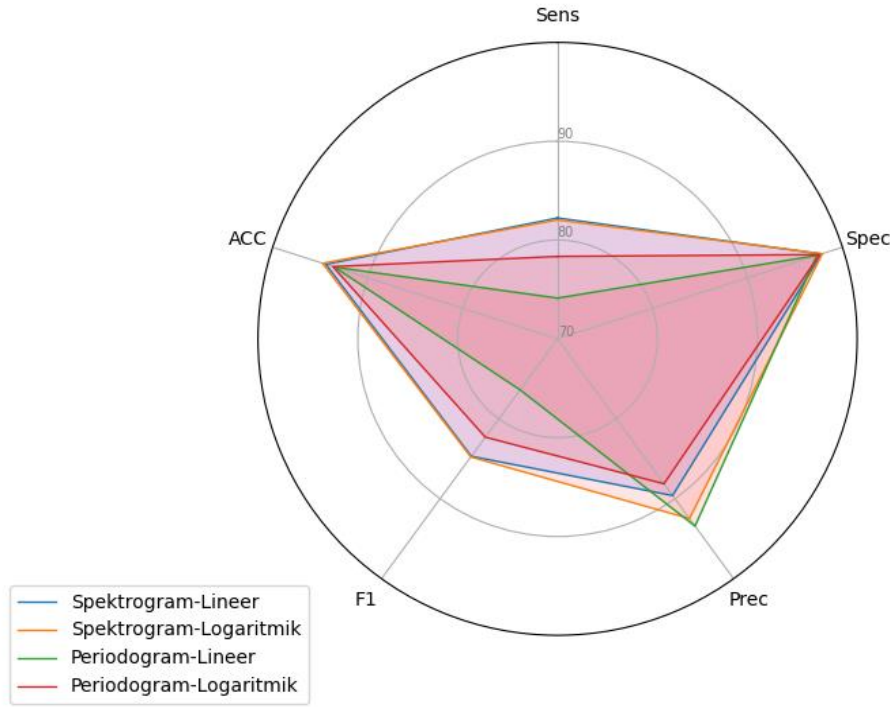


Şekil 6-1: Karışıklık matrisleri

Prec metriği için her bir modelin davranışı birbirinden ayrılmıştır. Öte yandan yine taşıt yok ve araba sınıfları için tüm modellerde diğer sınıflara kıyasla birbirine yakın skorlar sağlanmıştır. Bu metrik için elde edilen değerler %62.76 ile %100 arasında değişim göstermektedir.

F1 skoru dengeli dağılım göstermeyen veri setlerinin sınıflandırma performanslarının kıyaslanmasında yaygın olarak kullanılan bir metriktir. Bu arada kamyon sınıfının skorları diğer tüm sınıflara göre daha düşük olarak gerçekleşmiştir.

Şekil 6-1’de tüm metrikler için elde edilen makro ortalama değerleri görülmektedir. Lineer spektrogram, logaritmik spektrogram, lineer periodogram ve logaritmik periodogram modelleri için ACC skorları sırasıyla %94,40, %94,75 %93.49 ve %93.66 olarak gerçekleşmiştir. F1 değerleri de yine sırasıyla %84.69, %84.78, %76.38 ve %82.30 olarak gerçekleşmiştir. Elde edilen sonuçlar spektrogram modellerinin performansının periodogram modellerine kıyasla bir miktar daha iyi olduğunu göstermektedir. Spektrogram yönteminde sinyal küçük parçalara bölünerek HFT işlemi gerçekleştirilmektedir. Periodogram yönteminde ise HFT işlemi sinyalin tamamına tek seferde uygulanmaktadır. Buda ses olaylarında yerelliğin kaybolmasına neden olur. Buna bağlı olarak periodogram yönteminde performansın bir miktar daha düşük olarak gerçekleştiği değerlendirilmektedir. Ayrıca her iki yöntemde de logaritmik ölçek lineer ölçeğe kıyasla bir miktar daha iyi sonuç sağlamıştır. Burada da özellikler arasındaki varyansı azaltmanın sınıflandırma performansına olumlu katkı sağladığı değerlendirilmektedir (Ozer, 2021).



Şekil 6-2: Makro ortalama değerler

Makro ortalama ACC ve F1 değerlerinde spektrogram modelleri periodogram modellerine kıyasla daha iyi bir performans göstermiştir. Burada spektrogram özelliklerinin periodograma kıyasla yerellikleri koruyabilmesinin sınıflandırma performansını etkilediği değerlendirilmektedir. Ayrıca logaritmik ölçekte elde edilen skorların da lineer ölçeğe kıyasla bir miktar daha iyi olduğu görülmüştür. Burada da özellikler arasındaki varyansı azaltmanın sınıflandırma performansına olumlu katkı sunduğu değerlendirilmektedir. Sonuç olarak logaritmik spektrogram modeliyle makro ortalama ACC ve F1 değerleri için sırasıyla %94,75 ve %84.78 skoları elde edilmiştir.

6.2 ORAN HARİTASI TESTLERİ

Bu bölümde sonuçlar iki ayrı şekilde değerlendirilmektedir. İlk olarak 4 farklı IHC metodunun performansı 5 parça çapraz doğrulama ile verilmektedir. Sonrasında aynı veri seti üzerinde literatürde daha önce yapılmış çalışmayla sonuçlar kıyaslanmıştır. Literatürdeki çalışmayla kıyaslama yapılırken holdout test yapılmıştır.

6.2.1 ÇAPRAZ DOĞRULAMA TESTLERİ

Tablo 6-2’de 4 farklı IHC metodunun sınıf bazlı değerlendirme sonuçları verilmiştir. Çapraz doğrulama testleri veri setinde bulunan tüm kayıtlar kullanılarak gerçekleştirilmiştir. Tablo 1’de de görüldüğü gibi taşıt yok, motosiklet ve otobüs sınıfları için en yüksek F1 skorları Joergensen IHC metoduna aittir. Bu üç sınıf için sırasıyla %97.61 , %95.55 ve %69.89 skorları elde edilmiştir. Araba ve kamyon sınıfları için ise en yüksek F1 skorları sırasıyla %92.5 ve %91.96 olmak üzere Hilbert metoduna aittir. Bununla birlikte Joergensen IHC metodu ile de araba ve kamyon sınıfları için bu skorlara oldukça yakın değerler elde edilmiştir. Araba sınıfı için F1 skoru %92.44 kamyon için %41.6’dır. Dau ve Breebart IHC modellerinin performansı Joergensen ve Hilbert modellerinin bir miktar gerisinde kalmıştır. Breebart modelinde taşıt yok, motosiklet, araba, kamyon ve otobüs sınıfları için F1 değeri sırasıyla %95.71, %92.56, %90.10, %37.63 ve %62.98’dir. Dau IHC modelinde ise aynı sınıflar için sırasıyla %95.26, %91.9, %90.03, %38.77 ve %52.87 olarak gerçekleşmiştir. Özellikle otobüs sınıfı için Joergensen IHC metodu Dau IHC’ye göre belirgin bir şekilde daha iyi bir performans sergilemiştir. Otobüs sınıfı için iki metod arasındaki F1 skoru farkı yaklaşık olarak %17’dir.

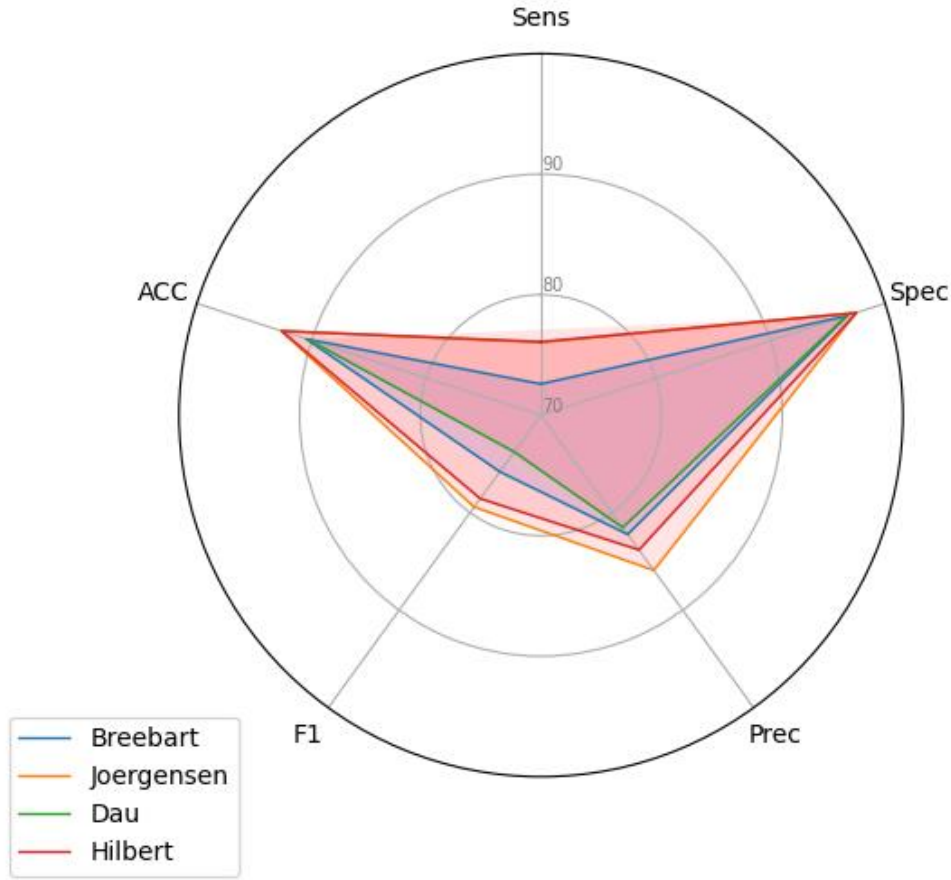
Şekil 6-3’de tüm sınıflar için 5-fold çapraz doğrulama ile elde edilmiş kıyaslama metriklerinin ortalaması görülmektedir. Hilbert ve Joergensen metotları geneli itibariyle diğer metotlara göre daha üstün bir performans sergilemiştir. Esasında her iki yöntemde Hilbert dönüşümüne dayanmaktadır ve sınıflandırma performansındaki iyileşmenin bu dönüşümden kaynaklandığı değerlendirilmektedir. Bunun başlıca sebebinin HT tabanlı özelliklerin durağan olmayan ve doğrusal olmayan sinyalleri analiz etme konusundaki başarımlarından kaynaklandığı değerlendirilmektedir (Agarwal vd., 2022). HT’ye ilave olarak Joergensen IHC metodu alçak geçiren filtre içermektedir. Bu durumda spectral enerjinin daha düşük frekanslarda toplandığı ses sınıflandırma gibi görevler için avantaj sağladığı değerlendirilmektedir (Ozer, 2021).

Tablo 6-2: IHC metodunun sınıf bazlı değerlendirme sonuçları

		Dau	Joergensen	Hilbert	Breebart
Sens	Taşıt Yok	97.62	98.18	98.17	97.94
	Motosiklet	89.77	94.88	94.19	92.56
	Araba	91.38	95.28	95.21	91.30
	Kamyon	27.79	30.53	30.63	27.40
	Otobüs	43.40	61.32	62.26	53.77
Spec	Taşıt Yok	93.62	97.40	97.20	94.16
	Motosiklet	99.86	99.91	99.89	99.81

	Araba	90.65	91.25	91.43	90.86
	Kamyon	99.04	98.99	99.05	98.87
	Otobüs	99.87	99.91	99.84	99.90
Prec	Taşıt Yok	93.02	97.05	96.83	93.58
	Motosiklet	94.15	96.23	95.52	92.56
	Araba	88.72	89.75	89.94	88.93
	Kamyon	64.11	65.27	66.60	60.09
	Otobüs	67.65	81.25	70.21	76.00
F1	Taşıt Yok	95.26	97.61	97.49	95.71
	Motosiklet	91.90	95.55	94.85	92.56
	Araba	90.03	92.44	92.50	90.10
	Kamyon	38.77	41.60	41.96	37.63
	Otobüs	52.87	69.89	66.00	62.98

Çalışmada kullanılan veri setinde sınıflar arasındaki veri miktarı ciddi bir biçimde dengesizdir. Bu nedenle daha doğru bir kıyaslama sunabilmek adına Sens, Spec, Prec, F1 ve ACC metrikleri de elde edilmiştir. Şekil 6-3 tüm bu metrikler için ortalama skorları göstermektedir. Tablo 6-2 ve Şekil 6-3 bir arada incelendiğinde Sens metriğinin diğer metriklerle kıyasla belirgin bir şekilde daha düşük olduğu görülmektedir. Sens metriği TP örnekler ile FN örnekler arasındaki oranı göstermektedir. Burada kamyon ve otobüs sınıfı için veri setinde bulunan örnek sayısı taşıt yok ve araba sınıflarına kıyasla belirgin şekilde daha düşüktür. Makine öğrenmesi modelleri genel olarak çoğunluk sınıfına eğilimlidir (Ozer, 2021). Buna bağlı olarak kamyon ve bu sınıflarında FN değerlerinin daha yüksek olarak gerçekleştiği ve bununda ortalama Sens skorunu düşürdüğü değerlendirilmektedir. Burada en düşük Sens değeri %27.79 ile Dau IHC metodunda truck sınıfına aittir. En yüksek Sens değeri ise %98.18 ile taşıt yok sınıfında Joergensen IHC metodu tarafından elde edilmiştir. Spec ve Prec metriğinde de benzer şekilde en iyi skorlar Joergensen IHC metodu ile elde edilmiştir. Spec için en iyi skorlar %99.91 değerleri ile motosiklet ve otobüs sınıfları için elde edilmiştir. Prec metriğinde ise en yüksek skor %97.05 ile taşıt yok sınıfında gerçekleşmiştir.



Şekil 6-3: Tüm sınıflar için kıyaslama metriklerinin ortalaması

6.2.2 DİĞER MODELLER İLE KIYASLAMA

Bu bölümde önerilen modellerden elde edilen sonuçlar literatürde aynı veri seti ile gerçekleştirilen başka bir çalışmayla kıyaslanmaktadır. Sonuçları birebir kıyaslayabilmek için diğer çalışmada kullanılan verilerin birebir aynıları kullanılmıştır [Abeßer vd., 2021]. Diğer çalışmada olduğu gibi burada da geleneksel test işlemi gerçekleştirilmiştir.

Tablo 6.3'te bu çalışmada geleneksel test yöntemiyle elde edilen sonuçlar görülmektedir. Tablo 6.4'te ise birebir aynı veri seti ile yapılan çalışmaya ilişkin F-skor değerleri görülmektedir (Abeßer vd., 2021).

Her iki çalışmada da örnek sayısının az olmasına bağlı olarak kamyon sınıfının performansı diğer modellere kıyasla belirgin şekilde daha düşüktür. Bununla birlikte diğer çalışmada ortalama F-skor değerinde en iyi sonucu veren VGGNet modeline kıyasla bu çalışmada Joergensen IHC modeli ile F-skor değerinde neredeyse %10'a varan belirgin bir iyileşme sağlanmıştır. Ayrıca Joergensen, Hilbert ve Breebart IHC modelleri motosiklet sınıfı içinde F-skor değerinde yaklaşık

%1.5'lik bir iyileşme sağlamıştır. Araba sınıfı için ise Joergensen IHC metodu F-skor değeri %0.76'lık bir iyileşme sağlamıştır.

Elde edilen sonuçlar diğer çalışmada kullanılan mel-spektrogram ve CNN yöntemi ile elde edilen değerler ile kıyaslandığında özellikle Joergensen IHC metodunun kullanıldığı oran haritası özelliklerinin belığın bir şekilde iyileşme sağladığı görülmektedir. Buna ilave olarak burada kullanılan CNN mimarisi diğer çalışmaya kıyasla daha az karmaşıklığa sahiptir. Bu da eğitim aşamasındaki hesaplama maliyetlerinin azalmasına katkı sunmaktadır. Benzer şekilde önerilen model Bölüm 6.1'de verilen spektrogram ve periodogram özelliklerine kıyasla da daha iyi sonuç sağlamıştır.

Tablo 6-3: Geleneksel test yöntemiyle elde edilen sonuçlar

		Dau	Joergensen	Hilbert	Breebart
Sens	Taşıt Yok	99.86	99.79	99.79	99.86
	Motosiklet	93.94	95.96	96.97	94.95
	Araba	98.53	97.67	98.18	98.70
	Kamyon	40.21	49.21	43.91	38.10
Spec	Taşıt Yok	99.45	99.93	99.86	99.86
	Motosiklet	99.93	99.96	99.93	100.00
	Araba	93.41	94.24	93.71	92.94
	Kamyon	99.40	98.88	99.18	99.36
Prec	Taşıt Yok	99.44	99.93	99.86	99.86
	Motosiklet	97.89	98.96	97.96	100.00
	Araba	91.05	92.02	91.39	90.49
	Kamyon	82.61	75.61	79.05	80.90
F1	Taşıt Yok	99.65	99.86	99.82	99.86
	Motosiklet	95.88	97.44	97.46	97.41
	Araba	94.65	94.76	94.67	94.42
	Kamyon	54.09	59.62	56.46	51.80

Tablo 6-4: Literatürdeki F-Skor Değerleri (Abeßer vd., 2021).

	Model	Araba	Kamyon	Motosiklet	Taşıt Yok
$N_B=16$	VGGNet	0.94	0.5	0.96	1.0
	ResNet	0.94	0.49	0.96	1.0
	SqueezeNet	0.92	0.48	0.9	1.0
$N_B=32$	VGGNet	0.94	0.46	0.96	1.0
	ResNet	0.94	0.44	0.97	1.0
	SqueezeNet	0.94	0.42	0.9	1.0
$N_B=64$	VGGNet	0.94	0.49	0.97	1.0
	ResNet	0.94	0.49	0.97	1.0
	SqueezeNet	0.94	0.5	0.95	1.0
$N_B=128$	VGGNet	0.94	0.44	0.96	1.0
	ResNet	0.94	0.45	0.95	1.0
	SqueezeNet	0.91	0.53	0.97	1.0

Tablo 6-5: VGGNet Karışıklık matrisi (Abeßer vd., 2021).

	Araba	Kamyon	Motosiklet	Taşıt Yok
Araba	97.29	2.62	0.02	0.09
Kamyon	60.21	38.84	0.63	0.32
Motosiklet	3.23	1.21	95.35	0.2
Taşıt Yok	0.23	0.01	0.11	99.65

7. SONUÇ

Dünya üzerinde yaşanan nüfus artışı ve kentleşme gibi unsurlara bağlı olarak şehirlerdeki gürültü seviyeleri de rahatsız edici ve hatta insan sağlığına zarar verecek şekilde artmaktadır. Karayollarındaki taşıtlardan kaynaklanan sesler özellikle kentsel alanlar için önemli bir gürültü kaynağıdır. Trafik sıkışıklıklarının artması da karayollarındaki gürültüyü artıran önemli bir etmendir. Bu doğrultuda AUS, yerel trafik durumlarına uyum sağlayarak trafik sıkışıklıklarını azaltmak ve optimize etmek için etkili karşı önlemler almaktadır. Geçtiğimiz on yılda, trafik ışığı döngülerini kontrol etme, trafik kazası algılama, lojistik izleme ve diğer akıllı şehir uygulamaları gibi uygulama senaryoları için trafik izleme için çeşitli otomatik yöntemler geliştirilmiştir.

Yukarıdaki bölümlerde de belirtildiği gibi trafik izleme sistemleri, kamera sensörlerinden, manyetik döngü sensörleri, radyo dalgalarına (Radar) ve ışık dalgalarına (Lidar) dayalı ölçüm sistemlerine kadar trafik akışını ölçmek için çeşitli sensör modalitelerini kullanmaktadır. Bu tür sistemler, geniş alanları kapsayacak şekilde dağıtılmış sensör ağları olarak kurulabilirken, kurulum ve bakım maliyetleri genellikle yüksektir. ATM modeli ise , müdahaleci olmayan trafik ölçümleri için daha ucuz bir alternatif sağlar ve bu tez çalışmasının odaklandığı nokta olmuştur.

Yukarıda sayılan sebepler doğrultusunda gürültü kirliliğinin azaltılması amacıyla kentlerdeki ses ortamlarını izlemeye yönelik teknolojilerin gelişimi son yıllarda hız kazanmıştır(Ozer vd., 2018; Lecun vd., 1998). Sinyal işleme ve makine öğrenimi tekniklerini kullanarak gerçekleştirilen otomatik ses tanıma bu alanda son yıllarda dikkat çekici bir gelişme kaydetmiştir. Bu tarz sistemlerde, ses olayları, sinyal işleme ve makine öğrenimi yöntemleri kullanılarak otomatik olarak tanımlanır. Bu sistemlerinin kabiliyetleri karayollarındaki akustik olayların kaynaklarını belirlemek için kullanılabilir. Öte yandan bu görev çeşitli zorlukları da bünyesinde barındırmaktadır. Bir yoldaki işitilebilir sesler, araçlar geçerken oluşan motorlar, egzozlar, tekerlekler ve hava türbülansı gibi çeşitli ses kaynaklarının birleşimidir (Hubel vd., 1968). Aynı tipteki ses kaynaklarının içsel değişkenliği ve farklı konumlar ve durumlar için akustik ortamın etkisi göz önüne alındığında, her sınıfın akustik özellikleri büyük çeşitlilik gösterebilir. Bu da ses kaynağının türünü sınıflandırmayı zorlaştırır.

SER görevi için kullanılacak el işi özelliklerin belirlenmesi, uzmanlık gerektiren zor ve zaman alıcı bir süreçtir. Son yıllarda bu sorunun üstesinden gelmek için CNN'ler kullanılarak 2B zaman-frekans gösterimlerinden elde edilen özniteliklerin kullanımı giderek artmıştır. Burada en uygun 2B gösterimi belirlemek çok önemlidir. Spektrogram ve Mel-spektrogram tabanlı özellikler geçmiş çalışmalarda yaygın olarak kullanılmıştır. Spektrogramlar eşit bant genişliği ve eşit aralıklı

frekans bileşenleri sunduğundan, enerjinin düşük frekanslı bileşenlerde yoğunlaştığı görevler için ideal değildir. Mel spektrogramının performansı genellikle benzer görevler için biyolojik olarak esinlenilmiş modellerin performansından daha düşüktür. Ayrıca Mel ölçekli özneliklerdeki düzgünlük kısıtı, öznelikler arasındaki varyansı artırarak sınıflandırma performansını olumsuz yönde etkiler.

Bu çalışma, yukarıda bahsedilen problemleri çözmek için logaritmik ölçekte oranı haritası özelliklerinin kullanılması önerilmiştir. Bir işitsel model olan oran haritası, frekans bileşenlerinin oluşturulması sırasında insan işitsel sisteminin frekans seçim özelliğini modellemek için bir gamaton filtresi kullanır. Bu filtre kullanılarak üst frekans bölgesinde geniş frekans bileşenleri, alt frekans bölgesinde dar frekans bileşenleri elde edilir. Böylece yüksek frekanslarda bilgi tamamen kaybolmadan düşük frekans aralığında daha iyi frekans çözünürlüğü sağlanarak daha fazla spektral bilgi elde edilir. Ek olarak, oran haritası yumuşatma özellikleri sağlar. Bunlar özellikler arasındaki farkı azaltarak sınıf ayrılabilirliğini artırır. Ayrıca önerilen model spektral enerjinin yoğun olduğu bölgelere odaklanmak amacıyla bir eşik fonksiyonunun kullanımını da içermektedir. Önerilen yöntemde öncelikle oran haritası gösterimleri oluşturulur. Bu gösterimlerin dinamik aralığı daha sonra logaritma fonksiyonu kullanılarak sıkıştırılır. Bu işlemden sonra eşik fonksiyonu uygulanır. Ayrıca önerilen modelde 4 farklı IHC yönteminin sınıflandırma performansına olan etkisi kıyaslanmıştır. Bunlara ilave olarak periodogram ve spektrogram özellikleri kullanılarak da testler gerçekleştirilmiştir. Kamuya açık bir veri seti üzerinde Joergen IHC yönteminin kullanıldığı model hem çapraz doğrulama testlerinde hem geleneksel test yönteminde diğer modellere kıyasla belirgin bir iyileşme sağlamıştır.

8. KAYNAKÇA

- [1] Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, 22(10), 1533-1545.
- [2] Abeßer, J., Gourishetti, S., Káta, A., Clauß, T., Sharma, P., & Liebetrau, J. (2021, August). IDMT-traffic: an open benchmark dataset for acoustic traffic monitoring research. In *2021 29th European Signal Processing Conference (EUSIPCO)* (pp. 551-555). IEEE.
- [3] Agarwal, P., & Kumar, S. (2022). Imagined word pairs recognition from non-invasive brain signals using Hilbert transform. *International Journal of System Assurance Engineering and Management*, 13(1), 385-394.
- [4] Aktürk, N., Akdemir, O. ve Üzkurt, İ. (2003). Trafik ışık sürelerinin neden olduğu çevresel taşıt gürültüsü. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 18(1), 71-87.
- [5] Alam, M. J., Attabi, Y., Dumouchel, P., Kenny, P., & O'Shaughnessy, D. D. (2013, August). Amplitude modulation features for emotion recognition from speech. In *INTERSPEECH* (pp. 2420-2424).
- [6] Bandela, S. R., & Kumar, T. K. (2019). Speech emotion recognition using semi-NMF feature optimization. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(5), 3741-3757.
- [7] Beeston, A. V. (2015). Perceptual compensation for reverberation in human listeners and machines (Doctoral dissertation, University of Sheffield).
- [8] Bertero, D., Kampman, O., & Fung, P. (2019). Towards Universal End-to-End Affect Recognition from Multilingual Speech by ConvNets. *arXiv preprint arXiv:1901.06486*.
- [9] Bregman, A. S., McAdams, S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound. *The Journal of the Acoustical Society of America* 95, 1177–1178
- [10] Briggs, F., Raich, R., & Fern, X. Z. (2009, December). Audio classification of bird species: A statistical manifold approach. In *2009 Ninth IEEE international conference on data mining* (pp. 51-60). IEEE.
- [11] Brigham, E. O. (1988). *The fast Fourier transform and its applications*. Prentice-Hall, Inc..
- [12] Brown, G. C. Martin.(1994). Computational auditory scene analysis. *Computer Speech and Language*, 8, 297-336.

- [13] Chu, S., Narayanan, S., & Kuo, C. C. J. (2009). Environmental sound recognition with time–frequency audio features. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6), 1142-1158.
- [14] Chu, S., Narayanan, S., & Kuo, C. C. J. (2008, March). Environmental sound recognition using MP-based features. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1-4). IEEE.
- [15] Chu, S., Narayanan, S. S., & Kuo, C. C. J. (2006, October). Content Analysis for Acoustic Environment Classification in Mobile Robots. In *AAAI Fall Symposium: Aurally Informed Performance* (pp. 16-21).
- [16] Colonna, J. G., Ribas, A. D., dos Santos, E. M., & Nakamura, E. F. (2012, June). Feature subset selection for automatically classifying anuran calls using sensor networks. In *The 2012 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
- [17] Cowling, M., & Sitte, R. (2003). Comparison of techniques for environmental sound recognition. *Pattern recognition letters*, 24(15), 2895-2907.
- [18] Das, S., Pal, S., & Mitra, M. (2019). Supervised model for Cochleagram feature based fundamental heart sound identification. *Biomedical Signal Processing and Control*, 52, 32-40.
- [19] De Boer, E., & De Jongh, H. R. (1978). On cochlear encoding: Potentialities and limitations of the reverse-correlation technique. *The Journal of the Acoustical Society of America*, 63(1), 115-135.
- [20] Deng, K., Shen, H. T., Xu, K., & Lin, X. (2006, April). Surface k-NN query processing. In *22nd International Conference on Data Engineering (ICDE'06)* (pp. 78-78). IEEE.
- [21] Dennis, J., Tran, H. D., & Li, H. (2010). Spectrogram image feature for sound event classification in mismatched conditions. *IEEE signal processing letters*, 18(2), 130-133.
- [22] Duan, S., Towsey, M., Zhang, J., Truskinger, A., Wimmer, J., & Roe, P. (2011, December). Acoustic component detection for automatic species recognition in environmental monitoring. In *2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (pp. 514-519). IEEE.
- [23] Duan, Z., Wang, Y., & Xing, Y. (2015). Sound quality prediction of vehicle interior noise under multiple working conditions using back-propagation neural network model. *Journal of transportation Technologies*, 5(02), 134.
- [24] Fagerlund, S. (2007). Bird species recognition using support vector machines. *EURASIP Journal on Advances in Signal Processing*, 2007, 1-8.

- [25] Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1-2), 103-138.
- [26] Griffin, D., & Lim, J. (1984). Signal estimation from modified short-time Fourier transform. *IEEE Transactions on acoustics, speech, and signal processing*, 32(2), 236-243.
- [27] Halim, H., & Abdullah, R. (2014). Equivalent noise level response to number of vehicles: a comparison between a high traffic flow and low traffic flow highway in Klang Valley, Malaysia. *Frontiers in Environmental Science*, 2, 13.
- [28] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- [29] Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., & Zhu, X. (2016). Semi-supervised active learning for sound classification in hybrid learning environments. *PloS one*, 11(9), e0162075.
- [30] Hilbert, D. (1953). *Framework for a General Theory of Linear Integral Equations*. New York, NY
- [31] Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215-243.
- [32] Jiang, P., Fu, H., Tao, H., Lei, P., & Zhao, L. (2019). Parallelized convolutional recurrent neural network with spectral features for speech emotion recognition. *IEEE Access*, 7, 90368-90377.
- [33] Josifovski, L. (2002). *Robust automatic speech recognition with missing and unreliable data* (Doctoral dissertation, University of Sheffield).
- [34] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [35] Lamel, L., Rabiner, L., Rosenberg, A., & Wilpon, J. (1981). An improved endpoint detector for isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(4), 777-785.
- [36] Langley, P., & Sage, S. (1994). Induction of selective Bayesian classifiers. In *Uncertainty Proceedings 1994* (pp. 399-406). Morgan Kaufmann.
- [37] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [38] Luengo, I., Navas, E., & Hernáez, I. (2010). Feature analysis and evaluation for automatic emotion identification in speech. *IEEE Transactions on Multimedia*, 12(6), 490-501.

- [39] Marieb, E. N., & Hoehn, K. *Anatomy & Physiology: Fourth Edition*. (2016). Benjamin Cummings, San Francisco, CA, USA
- [40] May, T. (2012). Binaural scene analysis: localization, detection and recognition of speakers in complex acoustic scenes.
- [41] Mesaros, A., Diment, A., Elizalde, B., Heittola, T., Vincent, E., Raj, B., & Virtanen, T. (2019). Sound event detection in the DCASE 2017 challenge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(6), 992-1006.
- [42] Mesaros, A., Heittola, T., Benetos, E., Foster, P., Lagrange, M., Virtanen, T., & Plumbley, M. D. (2017). Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 challenge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(2), 379-393.
- [43] Meutzner, H., Ma, N., Nickel, R., Schymura, C., & Kolossa, D. (2017, March). Improving audio-visual speech recognition using deep neural networks with dynamic stream reliability estimates. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5320-5324). IEEE.
- [44] Mitrović, D., Zeppelzauer, M., & Eidenberger, H. (2009, September). On feature selection in environmental sound recognition. In *2009 International Symposium ELMAR* (pp. 201-204). IEEE.
- [45] Nor, M. J. M., Fouladi, M. H., Nahvi, H., & Ariffin, A. K. (2008). Index for vehicle acoustical comfort inside a passenger car. *Applied Acoustics*, 69(4), 343-353.
- [46] Ozer, I. (2021). Pseudo-colored rate map representation for speech emotion recognition. *Biomedical Signal Processing and Control*, 66, 102502.
- [47] Ozer, I., Cetin, O., Gorur, K., & Temurtas, F. (2021). Improved machine learning performances with transfer learning to predicting need for hospitalization in arboviral infections against the small dataset. *Neural Computing and Applications*, 33(21), 14975-14989.
- [48] Ozer, I., Ozer, Z., & Findik, O. (2018). Noise robust sound event classification with convolutional neural network. *Neurocomputing*, 272, 505-512.
- [49] Papakostas, M., Spyrou, E., Giannakopoulos, T., Siantikos, G., Sgouropoulos, D., Mylonas, P., & Makedon, F. (2017). Deep visual attributes vs. hand-crafted audio features on multidomain speech emotion recognition. *Computation*, 5(2), 26.
- [50] Patterson, R. D. (1988). SVOS final report, part B: Implementing a gammatone filterbank. *Applied Psychology Unit report 2341*.

- [51] Patterson, R. D., Robinson, K. E. N., Holdsworth, J., McKeown, D., Zhang, C., & Allerhand, M. (1992). Complex sounds and auditory images. In *Auditory physiology and perception* (pp. 429-446). Pergamon.
- [52] Paulraj, M. P., Adom, A. H., Sundararaj, S., & Rahim, N. B. A. (2013). Moving vehicle recognition and classification based on time domain approach. *Procedia Engineering*, 53, 405-410.
- [53] Peng, Z., Li, X., Zhu, Z., Unoki, M., Dang, J., & Akagi, M. (2020). Speech emotion recognition using 3D convolutions and attention-based sliding recurrent networks with auditory front-ends. *IEEE Access*, 8, 16560-16572.
- [54] Pfeiffer, S., & Vincent, T. (2001). Formalisation of MPEG-1 compressed domain audio features. Technical Report 01/196.
- [55] Piedad E. J., Le, T. T., Aying, K., Pama, F. K., & Tabale, I. (2019, October). Vehicle count system based on time interval image capture method and deep learning mask R-CNN. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 2675-2679). IEEE.
- [56] Pohjalainen, J. (2007). Methods of automatic audio content classification.
- [57] Qiu, J., Liu, X., & Liu, R. (2018). Short-term traffic state prediction based on support vector machine. In *Proc., 2018 World Transport Convention*.
- [58] Rao, K. S., Koolagudi, S. G., & Vempada, R. R. (2013). Emotion recognition from speech using global and local prosodic features. *International journal of speech technology*, 16(2), 143-160.
- [59] Rocchesso, D. (2003). *Introduction to sound processing*. Mondo estremo.
- [60] Rokach, L., & Maimon, O. (2005). Top-down induction of decision trees classifiers-a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35(4), 476-487.
- [61] Romero, H. E., Ma, N., Brown, G. J., Beeston, A. V., & Hasan, M. (2019, May). Deep learning features for robust detection of acoustic events in sleep-disordered breathing. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 810-814). IEEE.
- [62] Russo, M., Stella, M., Sikora, M., & Pekić, V. (2019). Robust cochlear-model-based speech recognition. *Computers*, 8(1), 5.
- [63] Sajjad, M., & Kwon, S. (2020). Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM. *IEEE Access*, 8, 79861-79875.

- [64] Sharan, R. V., & Moir, T. J. (2015). Subband time-frequency image texture features for robust audio surveillance. *IEEE Transactions on Information Forensics and Security*, 10(12), 2605-2615.
- [65] Sharan, R. V., & Moir, T. J. (2019). Acoustic event recognition using cochleagram image and convolutional neural networks. *Applied Acoustics*, 148, 62-66.
- [66] Sharan, R. V., Berkovsky, S., & Liu, S. (2020, July). Voice command recognition using biologically inspired time-frequency representation and convolutional neural networks. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 998-1001). IEEE.
- [67] Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PloS one*, 3(12).
- [68] Towsey, M. (2013). Noise removal from waveforms and spectrograms derived from natural recordings of.
- [69] Trowitzsch, I. (2020). Robust sound event detection in binaural computational auditory scene analysis.
- [70] Valstar, M., Gratch, J., Schuller, B., Ringeval, F., Lalanne, D., Torres Torres, M., ... & Pantic, M. (2016, October). Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 6th international workshop on audio/visual emotion challenge* (pp. 3-10).
- [71] Vergel, K. N., Cacho, F. T., & Capiz, C. L. E. (2004). A study on roadside noise generated by tricycles. *Philippine Engineering Journal*, 25(2).
- [72] Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.
- [73] Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.
- [74] Wang, D., & Brown, G. J. (2006). *Computational auditory scene analysis: Principles, algorithms, and applications*. Wiley-IEEE press.
- [75] Warren, R. M., Bashford, J. A., Healy, E. W., & Brubaker, B. S. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Perception & psychophysics*, 55(3), 313-322.
- [76] Zhu, Z., Xu, G., Yang, B., Shi, D., & Lin, X. (2000). VISATRAM: A real-time vision system for automatic traffic monitoring. *Image and Vision Computing*, 18(10), 781-794.

ÖZGEÇMİŞ

Kişisel Bilgiler		
Adı Soyadı		
Doğum Yeri		
Doğum Tarihi		
Uyruğu		
Telefon		
E-Posta Adresi		
Web Adresi		

	Eğitim Bilgileri
	Lisans
Üniversite	
Fakülte	
Bölümü	
Mezuniyet Yılı	

Yüksek Lisans	
Üniversite	
Enstitü Adı	
Anabilim Dalı	
Programı	

Makale ve Bildiriler