



Trafikle İlgili Seslerin İşitsel Modeller ve Konvolüsyonel Sinir Ağları Kullanılarak Sınıflandırılması

Classification of Traffic-Related Sounds Using Auditory Models and Convolutional Neural Networks

¹Mariam Mine Cheikh Mohamed FADEL , ²Zeynep ÖZER 

¹Bandırma Onyediy Eylül Üniversitesi, Fen Bilimleri Enstitüsü, Akıllı Ulaşım Sistemleri ve Teknolojileri, Bandırma/Balıkesir, Türkiye

²Bandırma Onyediy Eylül Üniversitesi, Ömer Seyfettin Uygulamalı Bilimler Fakültesi, Yönetim Bilişim Sistemleri, Bandırma/Balıkesir, Türkiye

¹minebeyah@gmail.com, ²zozer@bandirma.edu.tr

Araştırma Makalesi/Research Article

ARTICLE INFO

Article history

Received : 20 April 2023

Accepted : 8 June 2023

Keywords:

Traffic Sounds, Auditory Model, Convolutional Neural Networks

ABSTRACT

This study proposes a new approach for the detection of acoustic sources on highways in order to reduce noise pollution in urban areas. The proposed method uses the features of rate map on a logarithmic scale based on the characteristics of the human ear and includes a threshold function to focus the regions where spectral energy is concentrated. Four different inner hair-cell methods were compared in the proposed model, and the Joergensen IHC method provided a significant improvement in classification performance compared to other models. The proposed model achieved approximately 10% improvement in the F1 value compared to previous studies. Overall, this study presents a promising approach to acoustic traffic monitoring using machine learning techniques and auditory models.

© 2023 Bandırma Onyediy Eylül University, Faculty of Engineering and Natural Science. Published by Dergi Park. All rights reserved.

MAKALE BİLGİSİ

Makale Tarihleri

Gönderim : 20 Nisan 2023

Kabul : 8 Haziran 2023

Anahtar Kelimeler:

Trafik Sesleri, İşitsel Model, Konvolüsyonel Sinir Ağları

ÖZET

Bu çalışma, kentsel bölgelerdeki gürültü kirliliğini azaltmak amacıyla otoyollardaki akustik kaynakların tespiti için yeni bir yaklaşım önermektedir. Önerilen metod, insan kulağının özelliklerine dayanan logaritmik ölçekte hız haritaları özelliklerini kullanır ve spektral enerjinin yoğunlaştığı bölgelere odaklanmak için bir eşik işlevini içermektedir. Önerilen modelde dört farklı iç tüy hücresi yöntemi karşılaştırıldı ve Joergensen IHC yöntemi diğer modellere göre sınıflandırma performansında önemli bir gelişme sağladı. Önerilen model, önceki çalışmalara kıyasla F1 değerinde yaklaşık %10'luk bir iyileşme sağlamıştır. Genel olarak, bu çalışma, makine öğrenimi teknikleri ve işitsel modeller kullanarak akustik trafik izlemeye umut verici bir yaklaşım sunmaktadır.

© 2023 Bandırma Onyediy Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi. Dergi Park tarafından yayınlanmaktadır. Tüm Hakları Saklıdır.

1. GİRİŞ

Dünya nüfusundaki artış, şehirleşme eğiliminin yükselmesi ve insanların kırsal alanlardan şehirlere taşınması şehir nüfuslarının büyümesine yol açmaktadır. Bu durum şehirlerdeki gürültü seviyelerinin artmasına ve insan sağlığına zarar veren düzeylere ulaşmasına neden olmaktadır. Bu bağlamda, gürültü kirliliği şehirlerde yaşayanlar için önemli bir sorun teşkil etmektedir. Gürültü sorunu şehir sakinleri arasında en yaygın şikayetlerden biridir [1]. Şehirlerdeki araç trafiği, gürültünün ana kaynaklarından biridir ve nüfus yoğunluğu arttıkça daha fazla araç kullanımı, uzun seyahat süreleri, gürültü ve hava kirliliği yaşanmaktadır.

Akıllı ulaşım sistemleri (AUS), trafik akışını hızlandırmak ve optimize etmek için yerel trafik koşullarına göre uyum sağlayarak etkili çözümler sunar. Son on yılda, trafik ışığı kontrolü, trafik kazalarının tespiti, lojistik takip ve diğer akıllı şehir uygulamaları için otomatik yöntemler geliştirilmiştir. Bu yöntemler, kamera sensörlerinden manyetik döngü sensörlerine, radar ve lidar sistemlerine kadar farklı sensör türlerini kullanarak trafik akışını ölçer. Bu sistemler, geniş kapsamlı sensör ağları olarak kurulabilir; ancak kurulum ve bakım maliyetleri yüksektir. Bu sebeple, daha ekonomik ve müdahaleci olmayan trafik ölçümleri için Akustik Trafik İzleme (ATİ), son yıllarda popülerlik kazanmıştır.

Gürültü kirliliği sorunu nedeniyle, son yıllarda ses ortamlarını izlemeye yönelik teknolojilerin geliştirilmesi hızlanmıştır [2][3]. Otomatik ses tanıma (OST) teknolojisi, sinyal işleme ve makine öğrenimi tekniklerini kullanarak ses olaylarını otomatik olarak tanımlamak için önemli bir araç haline gelmiştir. OST sistemleri, karayollarındaki akustik olayların kaynaklarını belirlemek amacıyla kullanılabilir. Ancak, bu görev birçok zorluk içermektedir. Bir yoldaki işitilebilir sesler, araçların motorları, egzozları, tekerlekleri ve hava türbülansı gibi çeşitli ses kaynaklarından kaynaklanır [1]. Aynı tür ses kaynakları (örneğin otomobiller) için farklı iç değişkenler, konumlar ve durumlar ile akustik ortamın etkileri (örneğin yankı, uzaklık) göz önünde bulundurulduğunda, her kategori için akustik nitelikler önemli ölçüde değişiklik gösterebilir. Bu sebeple, ses kaynaklarının çeşitlerini ayırt etmek oldukça güçtür.

Ses tanıma bir örüntü problemidir ve üç temel bileşenden oluşur. Bunlar sinyal ön işleme, nitelik çıkarma ve sınıflandırmadır. Bu yöntem, ses kaynaklarının tanımlanması, konuşma tanıma ve doğal dil işleme gibi çeşitli uygulama alanlarında kullanılır. Sinyal işleme evresinde, gelen sinyal parçalara ayrılır ve nitelik elde etme süreci için uygun hale getirilir. Özellik çıkarma aşamasında, verinin boyutu azaltılarak bir özellik vektörü olarak temsil edilir. Son olarak sınıflandırma aşamasında, elde edilen özellik vektörleri, belirli bir sınıfa atanmak üzere bir sınıflandırma algoritması kullanarak işlenir. Bu yaklaşım, ses tanıma teknolojisinin geniş bir uygulama yelpazesine yayılmasına olanak tanır.

Ses tanıma görevlerinde prozodik özellikler [1], ses kalitesi [2], Teager enerji operatörü (TEO) [3] 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 konvolüsyonel sinir ağları (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.

Son zamanlarda ses [4][5][6] ve konuşma [7][8][9] ile ilgili olaylarda zaman-frekans temsilleri başarıyla kullanılmaktadır. [9]'da yapılan çalışmada, SER görevi için CNN, SVM ve uzun kısa süreli bellek ağı (LSTM) ile birlikte 3B log Mel-spektrogram öznelikleri kullanılmıştır. Çalışmada [9], 64 Mel filtre bankası ve 25 ms uzunluğunda ve 15 ms örtüşme süresi olan Hamming penceresini kullanılmıştır. Önerilen modelin testleri, halka açık dört farklı veri seti üzerinde gerçekleştirilmiş ve EMO-DB veri seti üzerinde konuşmalardan duygu tespiti görevinde %86,44'lük bir sınıflandırma performansı elde edilmiştir. Buna benzer şekilde başka bir çalışmada [1], CNN ve LSTM ağları ile log Mel-spektrogram özelliklerini kullanmıştır. Ancak 3D CNN ağları yerine 1D ve 2D CNN ağları kullanılarak gerçekleştirilmiştir. Bir diğer grup çalışmalarda da spektrogram resim özellikleri kullanılarak gürültüye karşı dirençli ses olayı tespiti gerçekleştirilmiştir [10].

İnsan işitme sistemi çevresel sesleri ayırt etme konusunda oldukça başarılıdır. Dolayısıyla insan işitme sistemini modelleyen yaklaşımlar bir ATİ görevi için başarılı olabilir. Bu çalışmada [11] IHC modeli tabanlı oran haritası (rate map) özelliklerini kullanan bir ATİ sistemi önerilmiştir. İnsan kokleasındaki mekanik transdüksiyon sürecinin bir sonucu olarak, işitsel sinir ateşlemesi meydana gelir ve oran haritası özellikleri, bu ateşleme işlemi hızının spektro-zamansal temsiliyi kodlar. Oran haritası özellikleri dört temel aşamada oluşturulur [12]. İlk olarak, ses sinyaline ön vurgulama işlemi uygulanır. Daha sonra Koklear filtreleme gamaton 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 [13]. Alt frekans aralığında sağlanan daha kesin frekans çözünürlüğü, üst frekans aralığında önemli spektral bilgileri kaybetmeden daha fazla spektral bilgi elde etmede faydalıdır [13][14]. 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 [13][15]. Bu çalışmada ilk olarak WAV dosyaları normalleştirmeden önceki durumlarına ölçüklendi. Daha sonra yukarıda bahsedildiği gibi oran haritası özellikleri

elde edilmiştir. Bu doğrusal hız haritalarının dinamik aralığı log fonksiyonu kullanılarak sıkıştırılmış ve log-güç oran haritaları elde edilmiştir. Bu adımı takiben, oran haritalarının güçlü olduğu bölgelere odaklanmak için doymun ve 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 [16]. Spektral vadilerdeki bilgiler, daha önce gürültü ile dolduruldukları için sağlam olmayacaktır. Elde edilen oran haritası özellikleri OST görevi için bir CNN kullanılarak sınıflandırıldı. CNN, son yıllarda görüntü ile ilgili birçok uygulamada oldukça başarılı bir şekilde kullanılmıştır. Önerilen yaklaşım, kamuya açık olarak yayınlanmış bir veri seti üzerinde test edildi ve oldukça başarılı sonuçlar elde edildi.

2. MATERYAL VE METOT

Bu bölümde çalışmada kullanılan oran haritası temsillerinin elde edilişi, kullanılan CNN mimarisi, performans kıyaslama metrikleri ve veri seti hakkında bilgi verilmektedir.

2.1. Oran Haritası Temsilleri

Oran haritası gösterimi, insan işitsel sisteminden esinlenmiştir [12][17]. Temel olarak, işitsel sinir ateşleme hızının spektro-zamansal modelini kodlar [17]. Her frekans kanalının zarfı, bir işitsel sinir lifinin anlık ateşleme oranı olarak açıklanabilir [18][19]. Bu, kokleadaki mekanik transdüksiyon işlemine dayanmaktadır [17].

Oran haritası gösterimini elde etme süreci dört temel adımdan oluşur [12]. 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 [20] kullanılarak modellenmiştir. Bir sonraki aşamada, IHC modeli kullanılarak nöromekanik transdüksiyon gerçekleştirilir. Son olarak, sinyale Hamming penceresi uygulandıktan sonra uyartım oluşma olasılığı hesaplanır.

2.1.1. Ön vurgulama süreci

Dış ve orta kulak, kulak zarının ses basıncını farklı frekanslara göre artıran ve azaltan karmaşık bir akustik boşluk yaratır [12]. 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 sinyali. Bu durumda ön vurgu süreci şu şekilde ifade edilebilir:

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

2.2. Koklear filtreleme

Ön vurgulama işleminden sonra, insan kokleasının frekans seçim özelliği bir filtre bankası aracılığıyla modellenir. Bu işlem için bir GTF kullanılır [20][21], 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:

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

Faz, farklı frekans kanallarının her birinin başlangıç ve ofset sürelerinin karşılaştırılmasında kritik öneme sahiptir. Filtre bankası fazı telafi edilirse, frekans geçiş haritasında performans artar [12]. [22] 'de önerilen kompanzasyon için iki farklı yöntem kullanılmıştır. İlk yöntemde, zaman farkı, her bir dürtü yanıtının zarflarının tepe noktalarının hizalanması için filtrenin çıkışına uygulanır. Bu şu şekilde ifade edilebilir:

$$t_c = \frac{j-1}{2\pi B} \quad (3)$$

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

$$\phi_c = -2\pi f_c t_c \quad (4)$$

Denklem (3) ve (4), Denklem (2)'ye ikame edilirse, faz dengeli bir GTF elde edilir. Bu şu şekilde ifade edilebilir:

$$h(t) = A(t+t_c)^{j-1}e^{-2\pi B(t+t_c)} \cos(2\pi f_c t) \quad (t \geq -t_c) \quad (5)$$

GTF sıra yanıtı, düşük frekans değerlerinde ve aynı frekans çözünürlüğünde geleneksel hızlı Fourier dönüşümü (FFT) filtre bankasından daha hızlıdır. GTF bankası, [23]'de önerilen eşdeğer bir dikdörtgen bant genişliği (ERB) ölçeği ile kullanıldı. Bu aslında MFCC'nin (Mel-Frequency Cepstral Coefficients)[24] biyolojik olarak esinlenilmiş bir modifikasyonudur ve şu şekilde ifade edilebilir:

$$E(f) = 21.4 \log_{10}(4.37f + 1) \quad (6)$$

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

2.3. Nöromekanik Transdüksiyon

Koklea, ses basıncını baziler membranın titreşim hareketlerine dönüştürür ve daha sonra bu titreşimlerden nöral impulslar oluşturulur [25]. 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 [25]. Koklea şeklini yaklaşık 3.5 cm genişliğinde ve 1 cm uzunluğunda küçük bir tüp olarak tanımlamak mümkündür [25]. Baziler membran, kokleayı oluşturan ana unsurdur ve genişliği ve sertliği koklea boyunca değişir [25]. Aynı zamanda içi sıvı dolu iki tüpü birbirinden ayırır [25] ve oldukça sofistike bir yapıya sahip olan korti organını içerir. Bu organ, baziler membran titreşimlerine yanıt verir ve sinir uyarılarına iletilmesine izin verir [25]. Korti organı boyunca üç outer hair-cell (OHC) ve bir IHC vardır [25]. IHC'ler, saç hücrelerinin etrafındaki hareketi algılayan ve mekanik transdüksiyon yoluyla nöral uyarılar üreten ana sensörlerdir [25].

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 [26]. Bu çalışmada 4 farklı IHC modeli kullanılmıştır.

Dau modelinde, spesifik baziler membran segmentinin çıkışındaki sinyal, yarım dalga doğrultulmuş ve 1 kHz'de düşük geçişli filtrelendirilmiştir. Bu aşama, baziler membranın mekanik salınımlarının IHC'de 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. 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 [11].

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 [27].

Bir diğer yöntemde IHC'lerin etkin sinyal işleme, 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 [28][29][30][31].

Kullanılan son yöntem Hilbert dönüşümüdür (Hilbert Transform-HT). 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 $s(t)$ gamaton filtresinin çıkışındaki sinyal olmak üzere HT şu şekilde ifade edilir [32].

$$HT[s(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{(t - \tau)} d\tau \quad (7)$$

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

$$z(t) = s(t) + jp(t) = a(t)e^{j\theta(t)} \quad (8)$$

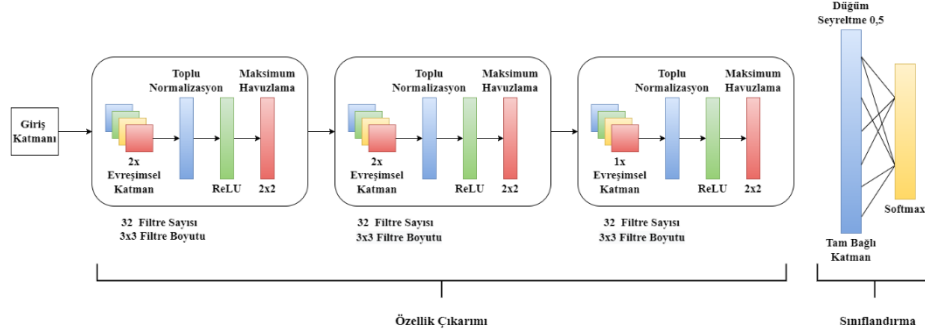
2.4. Oran Haritası

Oran 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.

3. 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 düşüm 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 1. CNN Mimarisi.

CNN son yıllarda pek çok farklı alanda kullanılan bir derin öğrenme modelidir [33][34][35]. Bu modelin zaman karmaşıklığı filtre sayısı, filtre boyutu ve giriş verisinin boyutu gibi çok sayıda parametreye bağlıdır. Burada l evrişim katmanının indeksi, d evrişim katmanlarının toplam sayısı, f_l l 'inci katmanın filtre sayısı, f_{l-1} giriş kanallarının sayısı, s_l filtre uzunluğu ve m_l özellik haritasının uzamsal boyutu olmak üzere zaman karmaşıklığı şu şekilde ifade edilebilir.

$$O\left(\sum_{l=1}^d f_{l-1} \cdot s_l^2 \cdot f_l \cdot m_l^2\right) \quad (9)$$

Yukarıda verilen eşitlik test aşamasındaki zaman karmaşıklığı için geçerlidir [36]. Eğitim aşamasındaki zaman karmaşıklığı bunun yaklaşık üç katıdır [36]. Bu durumda e eğitim tur sayısını belirtmek üzere, eğitim aşamasındaki zaman karmaşıklığı şu şekilde yazılabilir:

$$O\left(\left(\sum_{l=1}^d f_{l-1} \cdot s_l^2 \cdot f_l \cdot m_l^2\right) 3e\right) \quad (10)$$

Eğitim sürecinde modellerin kayıp fonksiyonu olarak kategorik çapraz entropi kullanıldı ve optimizasyon amacıyla Adam yöntemi seçildi. Tüm modellerde veri seti, eğitim, test ve doğrulama olmak üzere üç bölüme ayrıldı. Beş parçalı çapraz doğrulama testlerinde, eğitim setinin %10'u doğrulama verisi olarak kullanıldı. Geleneksel bölme yöntemi ile yapılan testlerde ise, eğitim, test ve doğrulama bölümleri daha güvenilir bir karşılaştırma sağlamak adına orijinal çalışmadaki şekilde uygulandı.

Tüm modeller, 250 döngü süresince eğitildi ve doğrulama setinde en iyi performansı sergileyen model ağırlıkları, test sürecinde kullanıldı. Bulgular bölümünde rapor edilen sonuçlar, test veri seti üzerinde elde edilen skorları yansıtmaktadır.

4. PERFORMANS DEĞERLENDİRME METRİKLERİ

Makine öğrenimi modellerinin performansını karşılaştırmak için kullanılan ölçütler, sonuçların doğruluğunu değerlendirmede oldukça kritiktir. Bu kapsamda, bu çalışmada sınıflandırma doğruluğu (ACC), duyarlılık (Sens), özgüllük (Spec), kesinlik (Prec) ve F1 gibi çeşitli metrikler kullanılmıştır. Bu ölçütler, makine öğrenimi algoritmalarının performansını nesnel bir şekilde değerlendirmek amacıyla yaygın olarak kullanılır.

Doğruluk değeri, doğru sınıflandırılmış örneklerin sayısının toplam örnek sayısına bölünmesi ile hesaplanır. Bu değer, sınıflandırma çalışmalarında sıkça başvurulan bir karşılaştırma ölçütüdür ve aşağıdaki şekilde hesaplanabilir:

$$\text{Doğruluk} = \frac{TN + TP}{TN + TP + FP + FN} \quad (11)$$

Performans değerlendirmesi için sadece doğruluk parametresine dayanmak yetersiz olabilir. Bu sebeple, özgüllük, duyarlılık, kesinlik ve F1 gibi ek istatistiksel performans ölçütleri kullanılması önerilir. Özgüllük, doğru şekilde tahmin edilmiş gerçek negatiflerin oranını belirtirken, duyarlılık doğru şekilde tahmin edilmiş gerçek pozitiflerin oranını ifade eder. Kesinlik, doğru tahmin edilen pozitif örneklerin oranını temsil eder. F1 ise, duyarlılık ve kesinlik değerlerinin harmonik ortalamasını gösterir. Bu değerlerin hesaplanması için kullanılan formüller şöyledir:

$$\text{Özgüllük} = \frac{TN}{TN + FP} \quad (12)$$

$$\text{Duyarluluk} = \frac{TP}{TP + FN} \quad (13)$$

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (14)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (15)$$

5. VERİ SETİ

Bu çalışmada, [29] numaralı kaynakta belirtilen IDMT-Trafik veri seti kullanılmıştır. Bu veri seti, Almanya'nın Ilmenau kentindeki üç şehir trafiği bölgesi ve bir köy yolu bölgesi olmak üzere toplam dört farklı kayıt konumunda, yüksek kaliteli sE8 mikrofonlar ve düşük maliyetli mikroelektromekanik sistem (MEMS) mikrofonları ile gerçekleştirilmiş eş zamanlı stereo ses kayıtlarını içerir. Kayıt senaryoları, çeşitli hız sınırları (30, 50 ve 70 km/s) ve hem kuru hem de ıslak yol koşullarını kapsamaktadır.

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. Araç türlerinin doğal olarak dengesiz dağıldığı yaygın trafik senaryolarını yansıtan bir dağılım elde edilmiştir. Veri setindeki örneklerin dağılımı Tablo 1'de yer almaktadır. Tablo 1'de verilen değerler orijinal çalışmayla birebir aynı olacak şekilde geleneksel bölme işleminin yapıldığı testlerde kullanılmıştır.

Tablo 1. Sınıf bazlı dağılımı.

Sınıf	Kamyon	Araba	Motosiklet	Taşıt Yok
Doğrulama Seti	32	275	15	266
Eğitim Seti	290	2471	132	2393
Test Seti	189	1157	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. Ham veri 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. Ham veri 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 araba sesinin sol taraftan saatte 30 km 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 2'de veri setinin nihai durumundaki sınıf ve olay sayıları verilmiştir.

Tablo 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 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 geleneksel test yapılmıştır.

6.1. Çapraz doğrulama testleri

Tablo 3'de 4 farklı IHC metodunun sınıf bazlı değerlendirme sonuçları verilmiştir. Çapraz doğrulama testleri veri setinde bulunan tüm kayıtlar kullanılarak gerçekleştirilmiştir. Tablo 3'de de görüldüğü gibi taşıt yok, motosiklet ve otobüs sınıfları için en yüksek F1 değerleri 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 değerleri 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 değeri %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 değeri farkı yaklaşık olarak %17'dir.

Şekil 2'de tüm sınıflar için 5-fold çapraz doğrulama ile elde edilmiş kıyaslama metriklerinin ortalaması görülmektedir. Hilbert ve Joergensen metodları geneli itibarıyla diğer metotlara göre daha üstün bir performans sergilemiştir. Esasında her iki yöntemde Hilbert dönüşümüne (HT) 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 [37]. 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 [38].

Tablo 3. IHC metodunun sınıf bazlı değerlendirme sonuçları.

		Dau	Joergensen	Hilbert	Breebart
Duyarlılık	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
Özgüllük	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
Kesinlik	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

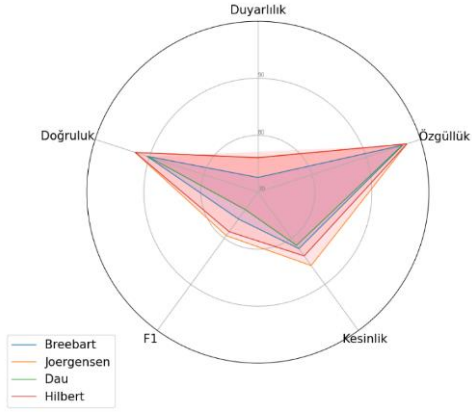
Çalışmada kullanılan veri setinde sınıflar arasındaki veri miktarı ciddi bir biçimde dengesizdir. Bu nedenle daha doğru bir kıyaslama sunabilmek adına duyarlılık, özgüllük, kesinlik, F1 ve doğruluk metrikleri de elde edilmiştir. Şekil 2 tüm bu metrikler için ortalama skorları göstermektedir. Tablo 3 ve Şekil 2 bir arada incelendiğinde duyarlılık metriğinin diğer metriklere kıyasla belirgin bir şekilde daha düşük olduğu görülmektedir. Duyarlılık metriği TP örnekler ile FN örnekler arasındaki oranı göstermektedir. Burada kamyon ve otobüs sınıfı için veri setinde bulunan örnek sayısı araba ve taşıt yok sınıflarına kıyasla bariz bir şekilde daha düşüktür. Makine öğrenmesi modelleri genel olarak çoğunluk sınıfına eğilimlidir [38]. Buna bağlı olarak kamyon ve bu sınıflarında FN değerlerinin daha yüksek olarak gerçekleştiği ve bununla ortalama duyarlılık skorunu düşürdüğü değerlendirilmektedir. Burada en düşük duyarlılık değeri %27.79 ile Dau IHC metodunda kamyon sınıfına aittir. En yüksek duyarlılık değeri ise %98.18 ile taşıt yok sınıfında Joergensen IHC metodu tarafından elde edilmiştir. Özgüllük ve kesinlik metriğinde de benzer şekilde en iyi skorlar Joergensen IHC metodu ile elde edilmiştir. Özgüllük için en iyi skorlar %99.91 değerleri ile motosiklet ve otobüs sınıfları için elde edilmiştir. Kesinlik metriğinde ise en yüksek skor %97.05 ile taşıt yok sınıfında gerçekleşmiştir.

7. DİĞER MODELLER İLE KIYASLAMA

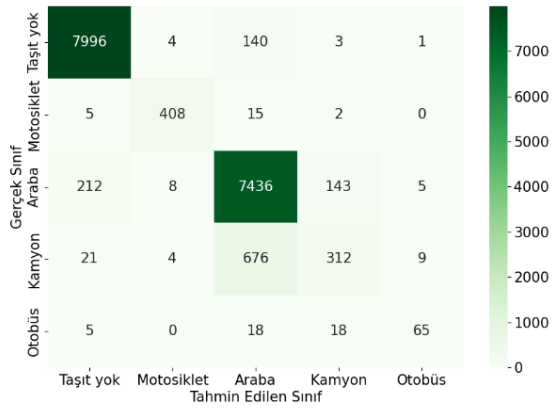
Bu bölümde önerilen modellerden elde edilen sonuçlar literatürde aynı veri seti ile gerçekleştirilen başka bir çalışmayla kıyaslanmaktadır. Sonuçları birebir kıyaslayabilmek için diğer çalışmada kullanılan verilerin birebir aynıları kullanılmıştır [39]. Diğer çalışmada olduğu gibi burada da geleneksel test işlemi gerçekleştirilmiştir.

Tablo 4'te bu çalışmada geleneksel test yöntemiyle elde edilen sonuçlar görülmektedir. Tablo 5'te ise birebir aynı veri seti ile yapılan çalışmaya ilişkin F1 değerleri görülmektedir [39].

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 F1 değerinde en iyi sonucu veren VGGNet modeline kıyasla bu çalışmada Joergensen IHC modeli ile F1 değerinde neredeyse %10'a varan belirgin bir iyileşme sağlanmıştır. Ayrıca Joergensen, Hilbert ve Breebart IHC modelleri motosiklet sınıfı içinde F1 değerinde yaklaşık %1.5'lik bir iyileşme sağlamıştır. Araba sınıfı için ise Joergensen IHC metodu F1 değeri %0.76'lık bir iyileşme sağlamıştır.



Şekil 2. Tüm sınıflar için kıyaslama metriklerinin ortalaması.



Şekil 3. Beş kat çapraz doğrulama karışıklık matrisi.

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 belirgin 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.

Tablo 4. Geleneksel test yöntemiyle elde edilen sonuçlar.

		Dau	Joergensen	Hilbert	Breebart
Duyarlılık	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
Özgüllük	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
Kesinlik	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 5. Literatürdeki F1 Değerleri [39].

	Model	Araba	Kamyon	Motosiklet	Taşıt Yok
NB=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
NB=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
NB=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
NB=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. VGGNet Karışıklık matrisi [39].

	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

8. SONUÇ

Son dönemlerde, gürültü kirliliğinin azaltılması amacıyla şehirlerdeki ses ortamlarını izlemek için kullanılan teknolojilerde önemli ilerlemeler kaydedilmiştir. Makine öğrenimi ve sinyal işleme yöntemlerinin uygulanmasıyla

otomatik ses tanıma sistemleri bu alanda dikkate değer gelişmeler göstermiştir. Bu tür sistemler, ses olaylarını makine öğrenimi ve sinyal işleme teknikleri yardımıyla otomatik olarak tanımaktadır. Bu yetenekler, karayollarındaki akustik olayların kaynaklarını saptamak için kullanılabilir. Bununla birlikte, bu görevin çeşitli zorlukları da vardır. Yol üzerinde duyulan sesler, geçen araçların motorları, egzozları, tekerlekleri ve hava türbülansı gibi farklı ses kaynaklarının bir araya gelmesiyle oluşur. Aynı tür ses kaynaklarının doğal değişkenliği ve akustik ortamın farklı konumlar ve koşullar için olan etkisi göz önünde bulundurulduğunda, her kategori için akustik özellikler oldukça çeşitli olabilir. Bu durum, ses kaynağı türünü sınıflandırmayı güçleştirir.

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. Kamuya açık bir veri seti üzerinde Joergensen IHC yönteminin kullanıldığı model hem çapraz doğrulama testlerinde hem geleneksel test yönteminde diğer modellere kıyasla belirgin bir iyileşme sağlamıştır.

Yazar Katkıları

Yazarlar çalışmaya eşit oranlı katkı sunmuşlardır.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

KAYNAKÇA

- [1] K.S. Rao, S.G. Koolagudi, and R.R. Vempada "Emotion recognition from speech using global and local prosodic features," *Int. J. Speech Technol.*, vol. 16, no. 2, pp. 143–160, 2013.
- [2] M. Valstar et al. "AVEC 2016 - Depression, mood, and emotion recognition workshop and challenge," in *AVEC 2016 - Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, co-located with ACM Multimedia, pp. 3–10, 2016.
- [3] S.R. Bandela and T.K. Kumar "Speech emotion recognition using semi-NMF feature optimization," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 27, no. 5, pp. 3741–3757, 2019.
- [4] O. Martin, I. Kotsia, B. Macq, and I. Pitas "The eNTERFACE'05 Audio-Visual emotion database," in *ICDEW 2006 - Proceedings of the 22nd International Conference on Data Engineering Workshops*, 2006.
- [5] Y. Wang and L. Guan "Recognizing human emotional state from audiovisual signals," *IEEE Trans. Multimed.*, vol. 10, no. 5, pp. 936–946, 2008.
- [6] J.B. Alonso, J. Cabrera, M. Medina, and C.M. Travieso "New approach in quantification of emotional intensity from the speech signal: Emotional temperature," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9554–9564, 2015.
- [7] C.K. Yogesh et al. "A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal," *Expert Syst. Appl.*, vol. 69, pp. 149–158, 2017.
- [8] M. Huzairah "Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Networks," 2017.
- [9] G. Lu, L. Yuan, W. Yang, J. Yan, and H. Li "Speech emotion recognition based on long short-term memory and convolutional neural networks," *Nanjing Youdian Daxue Xuebao (Ziran Kexue Ban)/Journal Nanjing Univ. Posts Telecommun. (Natural Sci.)*, vol. 38, no. 5, pp. 63–69, 2018.
- [10] I. Ozer, Z. Ozer, and O. Findik "Noise robust sound event classification with convolutional neural network," *Neurocomputing*, vol. 272, 2018.
- [11] T. Dau, D. Püschel, and A. Kohlrausch "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3615–3622, 1996.
- [12] G.J. Brown and M. Cooke "Computational auditory scene analysis," *Comput. Speech Lang.*, vol. 8, no. 4, pp. 297–336, 1994.
- [13] R.V. Sharan and T.J. Moir "Acoustic event recognition using cochleagram image and convolutional neural networks," *Appl. Acoust.*, vol. 148, pp. 62–66, 2019.
- [14] R.V. Sharan and T.J. Moir "Subband Time-Frequency Image Texture Features for Robust Audio Surveillance," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 12, pp. 2605–2615, 2015.
- [15] R.V. Sharan, S. Berkovsky, and S. Liu "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 and Biology Society (EMBC), pp. 998–1001, 2020.
- [16] L. Josifovski “Robust Automatic Speech Recognition with Missing and Unreliable Data,” 2002.
- [17] H. Meutzner, N. Ma, R. Nickel, C. Schymura, and D. Kolossa “Improving audio-visual speech recognition using deep neural networks with dynamic stream reliability estimates,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 5320–5324, 2017.
- [18] H.E. Romero, N. Ma, G.J. Brown, A.V Beeston, and M. Hasan “Deep Learning Features for Robust Detection of Acoustic Events in Sleep-disordered Breathing,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 810–814, 2019.
- [19] D. Wang and G.J. Brown “Fundamentals of computational auditory scene analysis,” in Computational Auditory Scene Analysis: Principles, Algorithms, and Applications, Wiley-IEEE Press, pp. 1–44, 2006.
- [20] R.D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand “Complex Sounds and Auditory Images,” in Auditory Physiology and Perception, CNBH, pp. 429–446, 1992.
- [21] E. de Boer “On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,” J. Acoust. Soc. Am., vol. 63, no. 1, p. 115, 1978.
- [22] R.D. Patterson “SVOS final report, part B: Implementing a gammatone filterbank,” Appl. Psychol. Unit Rep. 2341, 1988.
- [23] B.R. Glasberg and B.C. Moore “Derivation of auditory filter shapes from notched-noise data,” Hear. Res., vol. 47, no. 1–2, pp. 103–138, 1990.
- [24] S. Das, S. Pal, and M. Mitra “Supervised model for Cochleagram feature based fundamental heart sound identification,” Biomed. Signal Process. Control, vol. 52, pp. 32–40, 2019.
- [25] M. Russo, M. Stella, M. Sikora, and V. Pekić “Robust cochlear-model-based speech recognition,” Computers, vol. 8, no. 1, p. 5, 2019.
- [26] A.V Beeston “Perceptual compensation for reverberation in human listeners and machines,” University of Sheffield, 2015.
- [27] S. Jørgensen and T. Dau “Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing,” J. Acoust. Soc. Am., vol. 130, no. 3, pp. 1475–1487, 2011.
- [28] D.H. Johnson “The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,” J. Acoust. Soc. Am., vol. 68, no. 4, pp. 1115–1122, 1980.
- [29] T.F. Weiss and C. Rose “A comparison of synchronization filters in different auditory receptor organs,” Hear. Res., vol. 33, no. 2, pp. 175–179, 1988.
- [30] L.R. Bernstein and C. Trahiotis “The normalized correlation: Accounting for binaural detection across center frequency,” J. Acoust. Soc. Am., vol. 100, no. 6, pp. 3774–3784, 1996.
- [31] A. Breebaart, J. Van De Par, S. and Kohlrausch “Binaural processing model based on contralateral inhibition. I. Model structure,” J. Acoust. Soc. Am., vol. 110, no. 2, pp. 1074–1088, 2001.
- [32] D. Hilbert “Framework for a General Theory of Linear Integral Equations”, New York, 1953.
- [33] I. Ozer, S. B. Efe, and H. Ozbay, “A combined deep learning application for short term load forecasting,” Alexandria Eng. J., vol. 60, no. 4, pp. 3807–3818, 2021.
- [34] İ. Özer, S.B. Efe, and H. Özbay “CNN/Bi-LSTM-based deep learning algorithm for classification of power quality disturbances by using spectrogram images,” Int. Trans. Electr. Energy Syst., vol. 31, no. 12, p. e13204, 2021.
- [35] M. Bayram, Ö. İlyas “Deep learning methods for autism spectrum disorder diagnosis based on fMRI images,” Sakarya University Journal of Computer and Information Sciences, vol. 4, no. 1, 2021.
- [36] K. He and J. Sun “Convolutional Neural Networks at Constrained Time Cost.” Accessed: Jan. 27, 2021. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/He_Convolutional_Neural_Networks_2015_CVPR_paper.html.
- [37] P. Agarwal and S. Kumar “Imagined word pairs recognition from non-invasive brain signals using Hilbert transform,” Int. J. Syst. Assur. Eng. Manag., vol. 13, no. 1, pp. 385–394, 2022.
- [38] I. Ozer “Pseudo-colored rate map representation for speech emotion recognition,” Biomed. Signal Process. Control, vol. 66, 2021.
- [39] J. Abeber, S. Gourishetti, A. Katai, T. Claub, P. Sharma, and J. Liebetrau “IDMT-Traffic: An Open Benchmark Dataset for Acoustic Traffic Monitoring Research,” in 2021 29th European Signal Processing Conference (EUSIPCO), pp. 551–555, 2021.