T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ACİL DURUM TABANLI TELSİZ KONUŞMALARINDAKİ GÜRÜLTÜLÜ SES VERİLERİNİN DERİN ÖĞRENME YÖNTEMLERİ KULLANILARAK TANIMLANMASI

Celalettin ARSLAN

Danışman: Doç. Dr. Volkan KAYA

YAPAY ZEKA VE ROBOTİK
ANABİLİM DALI

ERZİNCAN
2024

ÖZET

Yüksek Lisans Tezi

ACİL DURUM TABANLI TELSİZ KONUŞMALARINDAKİ GÜRÜLTÜLÜ SES VERİLERİNİN DERİN ÖĞRENME YÖNTEMLERİ KULLANILARAK TANIMLANMASI

Celalettin ARSLAN

Erzincan Binali Yıldırım Üniversitesi
Fen Bilimleri Enstitüsü
Yapay Zeka ve Robotik Anabilim Dalı

Danışman: Doç. Dr. Volkan KAYA

Ses, kişiler arası iletişimde kullanılan temel unsurdur ve özellikle acil durum iletişiminde hayati bir rol oynamaktadır. Acil durum iletişiminde, konuşma seslerinin anlaşılabilirliği hızlı ve doğru bilgi akışı için kritik öneme sahiptir. Geleneksel iletişim yöntemlerinin başarısız olduğu durumlarda, telsiz iletişimi bağımsız yapısı sayesinde güvenilir bir araç olarak öne çıkmaktadır. Telsiz iletişiminde karşılaşılan en büyük sorun gürültülü seslerin anlaşılamamasıdır. Son yıllarda öne çıkan derin öğrenme yöntemlerinin kullanımı, gürültülü seslerin doğru tanımlanması için büyük faydalar sağlamaktadır. Bu tez kapsamında, acil durum kontrollerinde kullanılmak üzere ses verilerini tanıyan derin öğrenme tabanlı yeni bir model geliştirilmiştir. Geliştirilen modelin eğitimi ve test edilmesi için çeşitli çevresel gürültü koşulları altında elde edilen telsiz kayıtları kullanılarak altı farklı Türkçe ses verisi içeren yeni bir veri seti hazırlanmıştır. Bu veri seti kullanılarak, geliştirilen modelle %96,40 başarı oranıyla altı farklı acil durum ses verisi tanımlanmıştır.

2024, 64 Sayfa

Anahtar Kelimeler: Derin öğrenme, Evrişimsel sinir ağı, Gürültülü ses tanıma, Ses analizi

ABSTRACT

MSc Thesis

IDENTIFICATION OF NOISE VOICE DATA IN EMERGENCY-BASED RADIO CONVERSATIONS USING DEEP LEARNING METHODS

Celalettin ARSLAN

Erzincan Binali Yıldırım University
Graduate School of Natural and Applied Sciences
Department of Artificial Intelligence and Robotics

Supervisor: Assoc. Prof. Dr. Volkan KAYA

Voice is the basic element used in interpersonal communication and plays a vital role especially in emergency communication. In emergency communication, the intelligibility of speech sounds is of critical importance for fast and accurate information flow. In cases where traditional communication methods fail, radio communication stands out as a reliable tool thanks to its independent structure. The biggest problem encountered in radio communication is the inability to understand noisy sounds. The use of deep learning methods, which have become prominent in recent years, provides great benefits for the correct identification of noisy sounds. Within the scope of this thesis, a new deep learning-based model that recognizes voice data has been developed to be used in emergency controls. For training and testing of the developed model, a new dataset containing six different Turkish voice data was prepared using radio recordings obtained under various environmental noise conditions. Using this dataset, six different emergency voice data were identified with 96.40% success accuracy with the developed model.

2024, 64 Pages

Keywords: Audio analysis, Convolutional neural network, Deep learning, Noisy voice recognition

TEŞEKKÜR

Yüksek lisans eğitimim sürecinde bilgi ve tecrübeleriyle her zaman yanımda olan, tez çalışmam boyunca hoşgörü ve rehberliği ile değerli katkılar sağlayan saygıdeğer danışmanım Doç. Dr. Volkan KAYA' ya içtenlikle teşekkür ederim. Saygılarımı sunarım.

Veri toplama aşamasında desteklerini esirgemeyen Türkiye Radyo Amatörleri Cemiyeti (TRAC) Erzurum Şubesi'ne, özellikle Ömer Faruk ÖZER (TA9A), Recep TURGUT (TA9L) ve Gökhan BULUT (TA9M)' a teşekkür ederim.

Yüksek lisans eğitimim ve tez yazım sürecinde desteğini hiçbir zaman esirgemeyen kıymetli eşime teşekkür ederim.

Celalettin ARSLAN
Temmuz, 2024

İÇİNDEKİLER

Sayfa

ŞEKİLLER LİSTESİ

Sayfa

TABLOLAR LİSTESİ

Sayfa

SİMGELER ve KISALTMALAR

Simgeler

α	Alfa (Öğrenme Oranı)
ε	Epsilon
e	Euler Sabiti
$\sum$	Toplam Sembolü
log	Logaritma
∇	Nebla (Gradyan)

Kısaltmalar

ADAM	Uyarlanabilir Moment Tahmini (Adaptive Moment Estimation)
BCE	İkili Çapraz Entropi (Binary Cross Entropy)
CCE	Kategorik Çapraz Entropi (Categorical Cross Entropy)
ÇKYSA	Çok Katmanlı Yapay Sinir Ağları
DSA	Derin Sinir Ağı
ESA	Evrişimli Sinir Ağları
GD	Gradyan İniş(Gradient Descent)
GKM	Gauss Karışım Modeli
GMM	Gizli Markov Modeli
MFCK	Mel Frekans Cepstral Katsayıları
MSE	Ortalama Kare Hatası(Mean Squared Error)
OST	Otomatik Ses Tanıma
RELU	Düzeltmeli Doğrusal Ünite (Rectified Lineer Unit)
REVERB	Yankılanan Ses Geliştirme ve Tanıma Ölçütü (REverberant Voice Enhancement and Recognition Benchmark)
RF	Radyo Frekansı

SGD	Stokastik Gradyan Azaltma (Stochastic Gradient Descent)
TIMIT	Texas Instruments/Massachusetts Teknoloji Enstitüsü (Texas Instruments/Massachusetts Institute of Technology)
TSA	Tekrarlayan Sinir Ağları
UKVB	Uzun Kısa Vadeli Bellek
YGE	Yapay Gürültü Eklemesi

1. GİRİŞ

Ses insanlar arası iletişimde kullanılan başlıca unsurdur. İletişim, insanlar arasında bilgi alışverişi ve etkileşimi sağlayan çeşitli yöntemlerden oluşurken, sesli iletişim bu yöntemler arasında önemli bir yere sahiptir. Sesli iletişim, konuşma ve ses yoluyla iletişimi içerir ve genellikle yüz yüze konuşma, telefon görüşmesi veya sesli mesaj gibi biçimlerde gerçekleşmektedir.

Sesli iletişimin en çok gerekli olduğu alanlardan birisi ise acil durum iletişimidir. Acil durum iletişimi, zamanında ve doğru bilgi akışının hayati önem taşıdığı kritik bir alandır. Acil durum yönetiminde, etkin iletişim yönetimi, kötü durumların en aza indirilebilmesi için en güçlü unsur olarak önemli bir rol oynar (Yaman, 2020). Günümüzde acil durum iletişimi, çeşitli yöntemler aracılığıyla gerçekleştirilmektedir. Bunlar arasında, telefon, Kısa Mesaj Hizmeti (Short Message Service - SMS), internet, sosyal medya platformları gibi teknolojik araçlar önemli bir rol oynamaktadır. Ancak, özellikle afet durumlarında veya altyapısal sıkıntıların yaşandığı kriz zamanlarında, geleneksel iletişim sistemleri sıklıkla çökmekte veya etkisiz hale gelebilmektedir. Afet durumlarında, geleneksel iletişim sistemlerinin çökmesi durumunda, telsiz haberleşmesi güvenilir bir iletişim çözümü olarak öne çıkmaktadır (Kütükcü ve Eren, 2017). Telsiz iletişimi, altyapıya bağımlı olmayan bir iletişim çözümü sunar ve genellikle afet bölgelerindeki acil durum ekipleri, amatör telsizciler veya kurtarma ekipleri tarafından kullanılmaktadır. Telsiz cihazları, hızlı ve etkili bir şekilde haberleşme imkanı sunarak, koordinasyonu sağlamak ve acil durum yönetimini kolaylaştırmak için kritik bir rol oynamaktadır. Bu nedenle, telsiz iletişiminin acil durum iletişimindeki önemi, özellikle kriz anlarında iletişim altyapısının zayıf veya çökmüş olduğu durumlarda belirgin bir şekilde ortaya çıkmaktadır.

Telsiz kullanımı sırasında ses verilerinin anlaşılabilirliği büyük önem taşımaktadır. Bu nedenle telsiz iletişimi genellikle zorlu ve gürültülü ortamlarda gerçekleşmektedir. Özellikle acil durum iletişiminde, çevresel gürültüler veya cihazın kendi içsel gürültüsü ses verilerinin net anlaşılmasını zorlaştırabilmektedir. Gürültülü ses verileri sesin tanımlanma başarısını önemli ölçüde azaltmaktadır (Ahmed, 2020). Bu durum, iletişim verimliliğini ve etkinliğini ciddi şekilde azaltarak, acil durum ekiplerinin

koordinasyonunu engellemektedir. Dolayısıyla, ses verilerinin net ve anlaşılır olması, telsiz iletişiminin güvenilirliği ve etkinliği için kritik bir faktördür. Ses verilerinin gürültülü olabileceği durumlar, çevresel faktörlerden kaynaklanabileceği gibi, cihazın kendisinden veya iletim sırasında yaşanan teknik sorunlardan da kaynaklanabilmektedir. Bu durum iletişim verimliliğini önemli ölçüde olumsuz etkilemektedir. Bu bağlamda, gürültülü ses verilerinin etkin bir şekilde tanımlanması, acil durum iletişim sisteminin güvenilirliğini artırmak için kritik bir adım olmaktadır.

Bu tez çalışmasında, acil durum tabanlı telsiz konuşmalarındaki gürültülü ses verilerinin derin öğrenme yöntemleri kullanılarak tanımlanması üzerine odaklanmaktadır. Acil durum yönetiminde iletişimin doğru ve hızlı olması, hayati önem taşıyan bir gerekliliktir. Ancak, özellikle acil durum koşullarında telsiz iletişimi sıklıkla zorlu ve gürültülü ortamlarda gerçekleşir. Bu durumda, ses verilerinin net ve anlaşılır olması, iletişimin etkinliğini ve güvenilirliğini belirleyen kritik bir faktördür. Bununla birlikte çevresel gürültüler, cihazların kendi içsel gürültüsü veya iletim sırasında yaşanan teknik sorunlar gibi faktörler, ses verilerinin doğru bir şekilde tanımlanmasını zorlaştırabilmektedir. Bu tez çalışması, bu soruna çözüm getirerek acil durum telsiz iletişiminin güvenilirliğini artırmayı ve etkinliğini maksimize etmeyi amaçlamaktadır.

Geleneksel yöntemler gürültülü ses verilerini etkin bir şekilde işlemekte yetersiz kalmaktadır. Geleneksel yöntemler genellikle klasik makine öğrenmesi ve sinyal işleme tekniklerini içermektedir. Bu yöntemler, genellikle önceden tanımlanmış özellikler veya filtreler kullanarak ses verilerini analiz eder ve sınıflandırır. Ancak, gürültülü ortamlarda bu tür yöntemlerin performansı sınırlıdır. Bu nedenle karmaşık ve değişken akustik koşullarda etkili olmaları zordur. Derin öğrenme yöntemleri ise karmaşık veri yapılarını analiz etme ve öğrenme kapasitesine sahip olduğu için bu tür problemleri çözmekte daha etkilidir. Dolayısıyla, acil durum tabanlı telsiz konuşmalarındaki gürültülü ses verilerinin doğru bir şekilde tanımlanması için derin öğrenme yöntemlerine başvurulması gerekliliği ortaya çıkmaktadır. Bu yöntemler, ses verilerini daha doğru bir şekilde işleyebilir ve iletişim verimliliğini artırabilir. Bu bağlamda, geliştirilecek olan derin öğrenme yöntemleri, telsiz konuşmalarında geçen ses verilerini işleyerek gürültülü ortamlarda net ve anlaşılır ses iletişimi sağlayacaktır.

Ayrıca çalışma, acil durum iletişim sistemlerinin güvenilirliğini artırmak ve kriz anlarında daha etkin bir iletişim sağlamak adına önemli bir katkı sunacaktır. Bununla birlikte, ses tanımlama alanında derin öğrenme yöntemlerinin potansiyelini ortaya koyarak, gelecekteki araştırmalara da yol gösterecektir. Ek olarak, bu çalışma, acil durum iletişiminde yaşanan zorluklarını azaltarak güvenilir ve etkin bir iletişim ortamı sağlamayı amaçlamakta ve bu alanda önemli bir ilerleme sağlamayı hedeflemektedir.

Çalışmanın ikinci bölümünde çalışmanın odaklandığı konuyla ilgili olarak ses tanımlama alanında önceden yapılmış derin öğrenme araştırmalarının detaylı literatür taraması yer almaktadır. Üçüncü bölümde çalışmanın teorik temellerini oluşturan ses tanımlama, derin öğrenme ve acil durum iletişimi gibi konuların literatürdeki bağlamı ve kavramsal çerçevesine yer verilmiştir. Dördüncü bölümde tez kapsamında yapılan materyal, veri toplama yöntemleri, veri toplama araçları ve geliştirilen derin öğrenme algoritmaları ve bunların nasıl uygulandığıyla ilgili detaylar anlatılmaktadır. Beşinci bölümde geliştirilen derin öğrenme modeli ile elde edilen ses tanımlama deneysel sonuçlarına yer verilmiştir. Altıncı bölümde ise tez kapsamında yapılan bulgular kısaca değerlendirilerek elde edilen sonuçların önemi ve bu araştırma neticesinde sunulan öneriler yer almaktadır.

2. KAYNAK ÖZETLERİ

Son yıllarda, ses sınıflandırma alanında yapılan çalışmaların çoğunda derin öğrenme tekniklerinin kullanımı yaygınlaşmıştır. Bu teknikler, özellikle evrişimli sinir ağları (ESA) ve tekrarlayan sinir ağları (TSA) gibi derin öğrenme modellerini içermektedir. Bu modeller, çeşitli ses sınıflandırma alanlarında başarılı sonuçlar elde etmişlerdir.

Elyousseph ve Altamimi (2021) yaptıkları çalışmada, derin öğrenme tekniklerinin Radyo Frekansı (RF) sinyallerinin sınıflandırılması üzerindeki etkisi incelenmektedir. Araştırmanın hipotezi, hem zaman hem de frekans domain temsillerini birleştiren hibrit görüntülerin, yalnızca tek bir domain temsiline dayanan klasik ön işleme yöntemlerine kıyasla daha yüksek sınıflandırma doğruluğu sağlamasıdır. Çalışmada, farklı ön işleme adımları uygulanarak, RF sinyalleri ile eğitilen sabit bir ESA mimarisi kullanılmış ve %100 doğruluk oranına ulaşılmıştır. Sonuçlar, hibrit görüntülerin, zaman domain görüntüleri (%92,5 doğruluk), güç spektral yoğunluk görüntüleri (%80 doğruluk) ve spektrum görüntülerine (%75 doğruluk) göre üstün olduğunu göstermektedir. Bu bulgular göre, ESA'nın her kanalın güçlü yönlerini kullanarak kayıp fonksiyonunu en aza indirdiğini ve RF sinyallerinin sınıflandırılmasında hibrit ön işleme adımlarının daha fazla sağlamlık sağladığını ortaya koymaktadır.

Graves vd. (2013) yapmış oldukları çalışmalarında ses tanımlama alanında derin öğrenme tekniklerinin kullanımını ele almıştır. Özellikle, Derin Uzun-Kısa Vadeli Bellek (UKVB) TSA'nın bu alandaki performansını incelemişlerdir. Geleneksel yöntemlerin aksine, end-to-end eğitim metotları, giriş-çıkış hizalamasının bilinmediği sıralı veri problemlerinde TSA'ları eğitmek için kullanılmaktadır. Derin UKVB TSA'ların, Texas Instruments/Massachusetts Teknoloji Enstitüsü (Texas Instruments/Massachusetts Institute of Technology - TIMIT) fonem tanıma benchmarkında elde ettiği %17.7'lik test seti hatası, ses tanımlama alanında derin öğrenme yöntemlerinin ne kadar etkili olabileceğini göstermiştir. Bu sonuçlar, geleneksel yöntemlerle karşılaştırıldığında önemli bir gelişme sağladığını ve ses tanımlama alanındaki potansiyelini vurgulamaktadır. Bu çalışma, ses tanımlama alanındaki derin öğrenme yöntemlerinin etkisini daha iyi anlamak için önemli bir adım olarak değerlendirilebilir.

Schwarz vd. (2015) çalışmalarında gürültülü ve yankılı ortamlarda tanınabilirlik doğruluğunu artırmak için derin sinir ağı (DSA) tabanlı otomatik konuşma tanıma için bir uzaysal yayılım özelliği önermektedir. Bu özellik, geliştirilmiş konuşma özelliklerine dayanarak gerçek zamanlı olarak çoklu mikrofon sinyallerinden hesaplanarak, gürültülü ve yankılı ortamlarda tanınabilirlik doğruluğunu artırmaktadır. Ayrıca bu özellik, varış yön bilgisine veya tahminine gerek kalmadan her bir zaman ve frekans aralığında yayılan gürültü miktarını temsil etmektedir. Önerilen yayılım özelliğinin, Yankılanan Ses Geliştirme ve Tanıma Ölçütü (REverberant Voice Enhancement and Recognition Benchmark - REVERB) meydan okuma veri setinde kelime hata oranını azalttığı gösterilmiştir. Bu, hem gürültülü sinyallerden çıkarılan logmel özelliklerine hem de spektral çıkarma ile artırılmış özelliklere kıyasla daha düşük bir hata oranına yol açmaktadır. Bu çalışma, insanların gürültülü ve yankılı ortamlarda konuşma tanıma için benzer uzaysal bilgileri kullandığını ve DSA tabanlı bir akustik modelde başarıyla kullanılabileceğini göstermektedir.

Abdel-Hamid vd. (2014) yaptıkları çalışmada, geleneksel Gaussian Karışım Modeli (GKM)- Gizli Markov Modeli (GMM)'ne kıyasla hibrit DSA-GMM kullanarak konuşma tanıma performansını önemli ölçüde artırdığını göstermişlerdir. DSA'nın karmaşık korelasyonları modelleme yeteneği, performans iyileştirmesinin bir kısmına katkıda bulunmaktadır. Ancak, hata oranı daha fazla azaltmak için ESA kullanılmaktadır. Bu çalışmada, temel bir ESA'nın kısa bir tanımı sunulmakta ve konuşma tanıma için nasıl kullanılabileceği açıklanmaktadır. Ayrıca, konuşma özelliklerini daha iyi modelleyebilen sınırlı ağırlık paylaşımı şeması önerilmektedir. Yapılan deneysel sonuçlara göre, ESA'ların TIMIT telefon tanıma ve ses arama büyük kelime dağarcığı konuşma tanıma görevlerinde DSA'lara kıyasla hata oranını %6-10 arasında azalttığını göstermektedir. Özetle, bu çalışma ESA'ları konuşma tanıma görevlerinde etkili bir şekilde kullanmanın yolunu göstermektedir.

Shin vd. (2020) yaptıkları çalışmada, apartman dairelerinde komşular arası gürültü sorunlarını çözmek amacıyla ESA modeli kullanılmış ve katlar arası gürültü kaynaklarının sınıflandırılması hedeflenmiştir. 24 saatlik kayıtlar sonucunda elde edilen 1.515 adet ses verisi, altı farklı gürültü kaynağına (adımlar, mobilya sürükleme, çekiçleme, anlık darbe, elektrik süpürgesi, duyuru sistemi) ayrılmış ve ESA modelleriyle

analiz edilmiştir. Çalışmada kullanılan ResNet modeli ile en yüksek doğruluk oranı (%95,27) ve en iyi performans metriklerine ulaşarak diğer modellerden üstün olduğu sonucuna varılmıştır. Bu çalışma, komşular arası gürültü problemlerinin çözümünde otomatik sınıflandırma sistemlerinin uygulanabilirliğini göstermekte ve gelecekte iç mekan ses ortamlarının izlenmesine yönelik çalışmalar için temel oluşturmaktadır.

Park ve Lee (2020) tarafından yapılan çalışmada, işitme cihazları için çevresel gürültü sınıflandırma algoritması geliştirmek amacıyla ses sinyallerinden dönüştürülen görüntü sinyallerini ve ESA'ları kullanmaktadır. 10 farklı yaşam ortamı gürültüsünden elde edilen verilerle geliştirilen algoritmada, ses verilerinden dönüştürülen spektrogram görüntüleri, keskinleştirme maskesi ve medyan filtre ile işlenerek ESA'ya girdi olarak kullanılmıştır. Önerilen algoritma, diğer gürültü sınıflandırma yöntemleri ile karşılaştırılmış ve 1 saniyelik spektrogram süresi için %99,25 başarı oranı ile en yüksek sınıflandırma doğruluğuna ulaşılmıştır. Spektrogram süresi arttıkça doğruluk oranı azalmış, ancak 8 saniyelik süre için %98,73 doğruluk oranı elde edilmiştir. Bu oran, geleneksel yöntemin 1 saniyelik süre için elde ettiği %98,79 doğruluk oranı ile karşılaştırılabilir düzeydedir. Önerilen algoritma, performanstan ödün vermeden düşük hesaplama karmaşıklığı sunarak işitme cihazlarında gürültü sınıflandırma için etkili bir çözüm sunmaktadır.

Salamon ve Bello (2017) bu çalışmada, çevresel ses sınıflandırması için derin ESA mimarisi önerilmekte ve veri yetersizliği problemini aşmak için ses veri artırımı kullanımı araştırılmaktadır. Önerilen ESA mimarisi, spektral-zamansal kalıpları etkili bir şekilde öğrenebilme kapasitesi sayesinde çevresel ses sınıflandırmasında üstün performans göstermektedir. Veri artırımı yöntemleri ile birlikte kullanılan bu model, veri artırımı uygulanmayan haliyle ve artırma kullanılan yüzeysel bir modelle kıyaslandığında daha yüksek doğruluk oranları elde etmektedir. Araştırmada, farklı veri artırma tekniklerinin her bir sınıf için modelin doğruluğu üzerindeki etkileri incelenmiş ve her bir sınıfın farklı artırma yöntemlerinden farklı şekillerde etkilendiği gözlemlenmiştir. Bu bulgular, sınıfa özel veri artırma uygulamalarının model performansını daha da artırabileceğini önermektedir.

Cheng vd. (2023) yaptıkları çalışmada, Hong Kong gibi yoğun nüfuslu şehirlerdeki gereksiz gürültü kirliliği sorunlarına odaklanılmıştır. Araştırmacılar, özellikle gece geç saatlerde, modifiye edilmiş yüksek sesli egzozlara sahip araçların neden olduğu aşırı

gürültünün, konut alanlarına yakın bölgelerdeki yoğun trafik yollarında insanları rahatsız ettiğini belirtmektedir. Mevcut teknolojilerin, gürültülü araçları tespit etmede etkisiz olduğunu göstermiştir. Bu nedenle, araştırmacılar düşük maliyetli bir tasarıma sahip spektrogram tabanlı bir ses tanıma prototipi geliştirmişlerdir. Bu prototip, makine öğrenimi için ses spektrogramlarını kullanarak, modifiye edilmiş yüksek sesli egzozlara sahip araçları tanımlamak için bir ESA modeli kullanmışlardır. Prototip ayrıca nesne tanıma için YOLO v4 Tiny algoritmasını kullanarak, geçen araçları yolcu aracı, ticari araç ve motosiklet olarak ayırt etmektedir. Sonuçlar, prototipin, modifiye edilmiş egzozlara sahip yolcu araçlarını %96 başarı doğruluğuyla tespit ettiğini göstermektedir.

Al-Allaf (2015) bu çalışmada, FitNet, NARX, TSA ve Cascaded-ForwardNet olmak üzere dört farklı yapay sinir ağı (YSA) modeli, her biri gürültüyü herhangi bir konuşma sinyalinden çıkarmak için ayrı ayrı inşa edilmiş ve eğitilmiştir. Her model, giriş, gizli ve çıkış katmanlarından oluşmaktadır. Konuşma sinyalini ve ilişkili gürültüyü temsil eden iki nöron içeren giriş katmanı bulunmaktadır. Çıkış katmanı, gürültü çıkarıldıktan sonra geliştirilmiş sinyali temsil eden bir nöron içermektedir. Dört model, temiz (gürültüsüz) ve kirli (gürültülü) ses sinyalleri üzerinde ayrı ayrı eğitilmiştir. Her model için farklı mimari, optimizasyon eğitim algoritmaları ve öğrenme parametreleri ile deneyler gerçekleştirilmiştir. Yapılan çalışmada incelenen 4 modelin eğitilmesi için Levenberg-Marquardt algoritması (TrainLM) kullanılmıştır. Deneysel sonuçlara göre en iyi sonuçlar sırasıyla FitNet ve NARAX modellerinden elde edilmiştir. Elde edilen sonuçlara göre, önerilen mimarinin, dört modelin de eğitimli ve eğitimsiz konuşma sinyallerinden gürültüyü başarıyla çıkarma yeteneğine sahip olduğunu göstermektedir.

Choi vd. (2019) yapmış oldukları çalışmada, apartman veya çok katlı yapıların duvarları veya tavanları aracılığıyla bir kattan diğerine iletilen ara kat gürültüsünün doğru bir şekilde türünü ve konumunu belirlemek için derin ESA'ları kullanan bir yaklaşım önermektedirler. Bu çalışmada bir mobil cihaz kullanılarak, bir kampüs binasındaki üç kat üzerinde 9 farklı konumda üretilen 5 farklı ara kat gürültüsü içeren yaklaşık 2000 olaylık bir veri seti toplanmıştır. Önceden eğitilmiş ESA modelleri, tür ve konum sınıflandırması için ayrı ayrı tasarlanıp değerlendirilerek, doğrulama veri setlerinde sırasıyla %99,5 ve %95,3 tür ve konum sınıflandırma doğruluğu elde etmişlerdir. Ayrıca, modelin gürültü türü sınıflandırmasının sağlamlığı yeni bir test veri seti ile kontrol

edilmiştir. Elde edilen sonuçlara göre, ara kat gürültüsünün türünü ve konumunu belirlemede derin öğrenme tekniklerinin etkili bir şekilde kullanılabileceğini göstermektedir.

Özer vd. (2018) yapmış oldukları çalışmada, otomatik ses tanımanın (OST) karmaşık bir ses ortamında gerçekleştirilmesi gereken zorluğunu ele almaktadır. Geleneksel yöntemlerin gürültülü koşullarda yetersiz performans gösterdiği gerçeğiyle karşılaşıldığında, son yıllarda spektrogram görüntü özelliklerinin gürültü altında daha iyi performans göstermiştir. Bu çalışmada, spektrogramların doğrusal kuantize edilmiş görüntülere dönüştürülmesi ve çeşitli boyut azaltma yöntemlerinin uygulanmasıyla otomatik konuşma tanıma performansının artırılması hedeflenmiştir. Deneysel sonuçlara göre, önerilen yöntemin mevcut tekniklere göre %4,5'lik bir performans artışı sağladığı ve %63,4'lük bir göreceli hata azalması elde ettiğini göstermektedir. Bu sonuçlar göre, gürültülü ortamlarda OST'nin etkinliğini artırmak için spektrogram görüntü özelliklerinin ve CNN'nin kullanımının önemli olduğunu göstermektedir.

Zaman ve Direkoğlu (2020) tarafından yapılan çalışmada, işitme kaybı yaşayan engelli bireyler için işitme cihazlarının önemi vurgulanmaktadır. Genellikle, esnek ve programlanabilir oldukları için dijital işitme cihazları tercih edilmektedir. İşitme cihazı uygulamaları genellikle zararlı gürültüleri filtrelemek ve gelen konuşma sinyallerini yükseltmek için düşük geçiş filtreleri, adaptif filtreler ve spektral analiz kullanmaktadır. Bu çalışmada, gelen konuşma sinyallerini farklı sınıflara ayırmaya odaklanılmaktadır. Bu sınıflardan biri temiz konuşma sinyali sınıfı iken, diğerleri beş farklı gürültü tipine göre oluşturulan zararlı gürültülü konuşma sinyali sınıflarıdır. Gelen konuşma sinyallerinin spektrogram görüntüleri ve bir ESA kullanılarak doğru sınıflandırılması önerilmektedir. Farklı gürültü tipleriyle bozulmuş konuşma sinyallerini içeren bir veri seti oluşturulmuş ve farklı ESA mimarilerinin performansları karşılaştırılmıştır. Sonuçlar, önerilen yöntemin farklı gürültü tipleriyle bozulmuş konuşma sinyallerini çok doğru bir şekilde sınıflandırabildiğini göstermektedir.

Zulfiqar vd. (2021) yapmış oldukları çalışmada, solunum sesi özellikleri ve analizleri, pnömonik patolojinin temel bir parçasını oluşturmakta ve hastanın akciğerleri hakkında belirleyici bilgiler sağlamaktadır. Bu çalışmada, anormal solunum seslerinin görsel incelemesi için Fourier analizi uygulanmıştır. Spektrum analizi, yapay gürültü eklemesi

(YGE) ve farklı derin ESA'ları ile birlikte gerçekleştirilerek, yedi anormal solunum sesini hem sürekli hem de sürekli olmayan şeklinde sınıflandırmak için kullanılmıştır. Önerilen çerçeve, sağlıksız solunum seslerine benzer türde gürültü ekleyen uyarlanabilir bir mekanizma içermektedir. YGE, solunum seslerini, YGE olmadan olanlardan daha doğru bir şekilde tanımlanabilir hale getirir. Önerilen çerçeveden elde edilen sonuçlar, önceki tekniklerden üstün olduğu ve yedi farklı anormal solunum sesini aynı anda ele aldığı görülmektedir. Bu çalışma, solunum sesi anormalliklerinin sınıflandırılmasında derin evrişimli sinir ağları ve Fourier analizi gibi yöntemlerin etkili olduğunu göstermektedir.

Jin vd. (2021) yapmış oldukları çalışmada, otomotiv şanzıman gürültüsünün ses kalitesini tahmin etmek için, yarı yansıyan odada iki otomotiv şanzımanıyla birlikte bir düzleştirilmiş yavaşlama gürültü testi gerçekleştirmişlerdir. Kaydedilen tüm şanzıman gürültü sinyalleri 5 saniyelik segmentlere ayrılmış ve derecelendirme ölçekleri testiyle subjektif olarak değerlendirilmiştir. Ayrıca, Mel-Frekans Cepstral Katsayıları tabanlı evrişimli sinir ağları (MFCK-ESA) adlı yeni bir tahmin yöntemi önerilmiştir. Bu yöntem, genel ESA yapısının çıkışında "Softmax" sınıflandırma katmanını doğrusal dönüşüm tahmin katmanıyla değiştirmekte ve giriş olarak MFCK özellik haritasını almaktadır. MFCK'nın ses kalitesini ayırt etme performansı doğrulanarak, MFCK-ESA modelinin parametre seçimi, bir ızgara araması kullanılarak karşılaştırılmış ve incelenmiştir. Ayrıca, yeni geliştirilen MFCK-ESA ile performansın karşılaştırılabilmesi için üç geleneksel makine öğrenimi tabanlı yöntem tanıtılmıştır. Elde edilen sonuçlara göre; farklı şanzıman dişlilerinde, MFCK özellikleri farklı ses kalitesi gürültülerini ayırt edebilir ve önerilen MFCK-ESA ses kalitesi tahmin yaklaşımının doğruluğu, diğer 3 referans yöntemininkinden daha iyidir. Bununla birlikte MFCK-ESA'dan gelen tahmin değerinin korelasyon katsayısı 0,95'ten fazladır ve MFCK-ESA'dan gelen tahmin değerinin ortalama mutlak hatası 0,55'ten azdır, bu da mühendislik ihtiyacını tam olarak karşılamaktadır. Son olarak, yeni önerilen MFCK-ESA yaklaşımının gelecekte diğer araç gürültülerine de uygulanabileceği düşünülmektedir.

Stabellini (2023) tarafından yapılan çalışmada, çevresel seslerin sınıflandırılmasında derin öğrenme algoritmalarının gürültüye karşı dayanıklılığı incelenmektedir. Çalışma, cam kırılması, araba korna sesi, çocuk ağlaması gibi çevresel seslerin sınıflandırılmasının toplumsal güvenlik ve sağlık açısından önemli olduğunu vurgulamaktadır. Çeşitli derin

öğrenme modelleri, farklı doğadaki arka plan gürültüsüne maruz bırakılarak değerlendirilmiştir. Deneylerde, önceden eğitilmiş modellerin gürültü varlığında daha iyi genelleme yeteneğine sahip olduğu ve daha hassas sınıflandırma yaptığı gözlemlenmiştir. HTS_AT ve mn40_as modelleri, özellikle kendi kendine dikkat mekanizmaları sayesinde gürültüye karşı daha dayanıklı sonuçlar vermektedir. Sonuçlar, önceden eğitim (pre-training) ve dikkat mekanizmalarının kullanımının, modelin gürültüye karşı sınıflandırma kapasitesini artırdığını ve gelecekteki ağların geliştirilmesine katkıda bulunabileceğini göstermektedir. Bu bulgular, çevresel seslerin sınıflandırılması için daha dayanıklı ve etkili derin öğrenme tekniklerinin geliştirilmesine yönelik önemli bilgiler sunmaktadır.

3. KURAMSAL TEMELLER

3.1. Derin Öğrenme

3.1.1. Derin öğrenmenin tanımı ve tarihçesi

Derin öğrenme, yapay sinir ağlarının çok katmanlı ve karmaşık yapılarını kullanarak veri içindeki bilgileri otomatik olarak keşfetmek ve temsil etmek için kullanılan bir makine öğrenme yaklaşımıdır (LeCun vd., 2015). Bu tanım, yapay sinir ağlarının, desen tanımanın ve otomatik öğrenmenin kesişiminde bulunan ve derin öğrenmenin temel prensiplerini ortaya koyan bir ifadedir.

Derin öğrenme, özellikle son yıllarda büyük veri ve yüksek hesaplama gücü ile birlikte popülerlik kazanmıştır. Derin öğrenme modelleri, veri içindeki karmaşık ilişkileri ve yapıları otomatik olarak keşfetmek için kullanılır (Schmidhuber, 2015). Bu modeller, genellikle çok katmanlı yapılara sahip olan ve her katmanın birbirini takip eden katmanlardan öğrenilen özelliklerle temsil edildiği yapay sinir ağlarıdır.

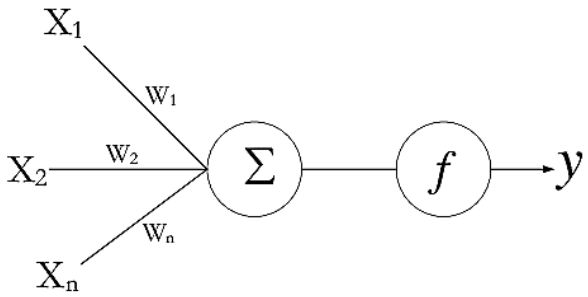
Derin öğrenmenin temelinde, veriler arasındaki ilişkileri temsil etmek için ağın ağırlıklarının iteratif olarak ayarlandığı bir öğrenme süreci bulunur. Bu süreç, gerçek ve tahmin edilen çıktılar arasındaki farkı minimize etmek için gerçekleştirilir ve genellikle "geri yayılım" (backpropagation) adı verilen bir optimizasyon algoritması kullanılır (Rumelhart vd., 1986).

Derin öğrenme, pek çok alanda başarıyla kullanılmaktadır. Özellikle görüntü ve ses tanıma, doğal dil işleme, oyun oynama ve tıbbi görüntüleme gibi alanlarda büyük başarılar elde etmiştir (Esteva vd., 2017; Hinton vd., 2012). Derin öğrenme, karmaşık problemleri ele alabilen ve büyük miktarda veriye dayalı öğrenme yeteneği sunan güçlü bir araçtır. Bu nedenle, endüstride ve akademik araştırmalarda yaygın bir şekilde kullanılmaktadır.

3.1.1.1. McCulloch-Pitts hücresi

Derin öğrenmenin temelleri, 1943 yılında Warren McCulloch ve Walter Pitts tarafından geliştirilen bir model olan McCulloch-Pitts hücresi ile başlamıştır (McCulloch ve Pitts,

1943). Bu model, yapay sinir ağlarının temelini oluşturmakta ve bir tür basit sinir hücresini simüle etmektedir. Bu model, biyolojik sinir hücrelerinin işlevselliğini basitleştirerek matematiksel bir şekilde tanımlanır. McCulloch-Pitts hücresi, girdileri alır, bu girdilere ağırlıklar atar, toplar ve bir eşik değeriyle karşılaştırır. Eğer toplam girdi eşik değerinden büyükse, hücre aktive olur ve bir çıkış üretir (Şekil 3.1).



Şekil 3.1. McCulloch-Pitts hücresi modeli (Sooksil ve Benjaoran, 2021).

Matematiksel olarak, McCulloch-Pitts hücresi, girdileri $x_{1,}\, x_{2,\ldots}, x_n$ ve bu girdilere karşılık gelen ağırlıkları $w_{1,}\, w_{2,\ldots}, w_n$ almaktadır. Bu girdi-ağırlık çiftlerinin çarpımı toplanır. Elde edilen sonuç belirlenen aktivasyon fonksiyonu ile işleme konulduktan sonra bir eşik değeri ile karşılaştırılır. Eğer toplam, belirlenen eşik değerinden büyükse, hücre aktif olur ve çıktı olarak 1 üretir; aksi halde, çıktı 0 olur. Bu yapıya ait matematiksel ifade denklem 3.1'de görülmektedir.

$$y = f\left(\sum_{j=1}^{n} w_j * x_j\right) \quad (3.1)$$

Bu model basit olmasına rağmen, sinir ağlarının temelini oluşturan ilk matematiksel modeldir. McCulloch-Pitts hücresi genel anlamda basit bir etkinleştirme mekanizması sağlar ancak öğrenme yeteneğine sahip değildir. Bu model, yapay sinir ağlarının gelişiminde ve modern derin öğrenme modellerinin temellerinde önemli bir rol oynamaktadır.

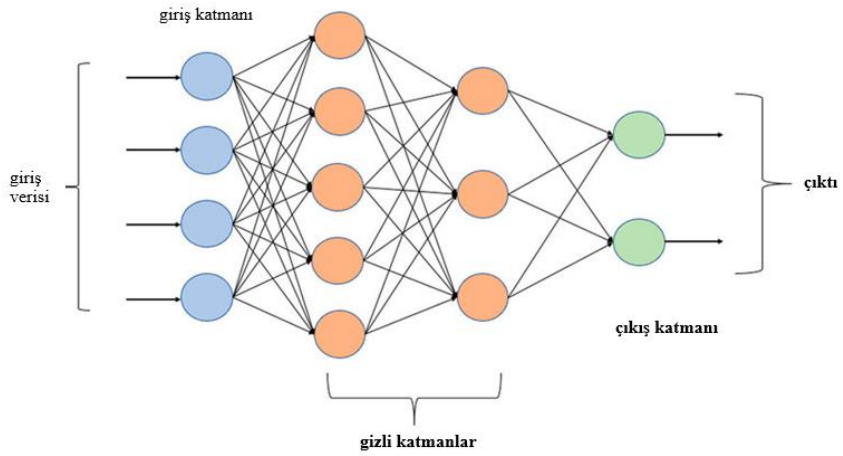
3.1.1.2. Perceptron modeli

1958 yılında Frank Rosenblatt, McCulloch-Pitts Hücresi modelini temel alarak yapay sinir ağlarının ilk pratik uygulamalarından biri olan Perceptron'u geliştirmiştir (Rosenblatt, 1958).

Perceptron modeli, McCulloch-Pitts hücresi modelini kullanarak girişleri ve bu girişlere ait ağırlıkları kullanarak bir çıkış üretmektedir. Perceptron'un bu modele katkısı ise ağırlıkların eğitim süreci boyunca otomatik olarak güncellenmesidir. Perceptron'un eğitim süreci, genellikle "geri yayılım (back-propagation)" olarak adlandırılan bir süreç kullanılarak gerçekleştirilmektedir. Bu süreçte, Perceptron'un tahminleri gerçek etiketlerle karşılaştırılır ve tahmin hatalarının bir ölçüsü olan bir kayıp fonksiyonu kullanılarak hata hesaplanır. Ardından, bu hata geriye doğru ağa yayılır ve ağırlıkların nasıl güncelleneceğini belirlemek için kullanılır. Bu şekilde, Perceptron eğitilir ve giriş verilerine daha iyi uyum sağlayacak şekilde ağırlıkları ayarlar. Bu nedenle, Perceptron, McCulloch-Pitts hücresinden farklı olarak öğrenme yeteneğine sahip bir modeldir.

3.1.1.3. Çok katmanlı yapay sinir ağları

Derin öğrenmenin ilk adımları, 1960'larda çok katmanlı yapay sinir ağlarının (ÇKYSA) geliştirilmesiyle atılmıştır. ÇKYSA, yapay sinir ağlarının bir türüdür ve en temel derin öğrenme mimarilerinden biridir. ÇKYSA'ları, en az bir gizli katmana sahip olan ve birçok giriş ve çıkış düğümü arasında bağlantıları bulunan bir yapay sinir ağı türüdür. ÇKYSA mimarisini ait örnek bir yapı Şekil 3.2'de görülmektedir.



Şekil 3.2. ÇKYSA modeli (Afan vd., 2021).

ÇKYSA'nın etkinliğinin artmasındaki temel unsur geri yayılım işlemidir. 1974 yılında Paul Werbos tarafından yapılan çalışma geri yayılım algoritmasının temeli olarak kabul edilebilir. Werbos, geri yayılım algoritmasını, yapay sinir ağlarının eğitiminde ve genel

olarak tahmin ve analiz problemlerinde kullanılabilir bir araç olarak tanımlamıştır (Werbos, 1974). Bu algoritma sayesinde öğrenme süreci daha etkin bir hale gelmiştir.

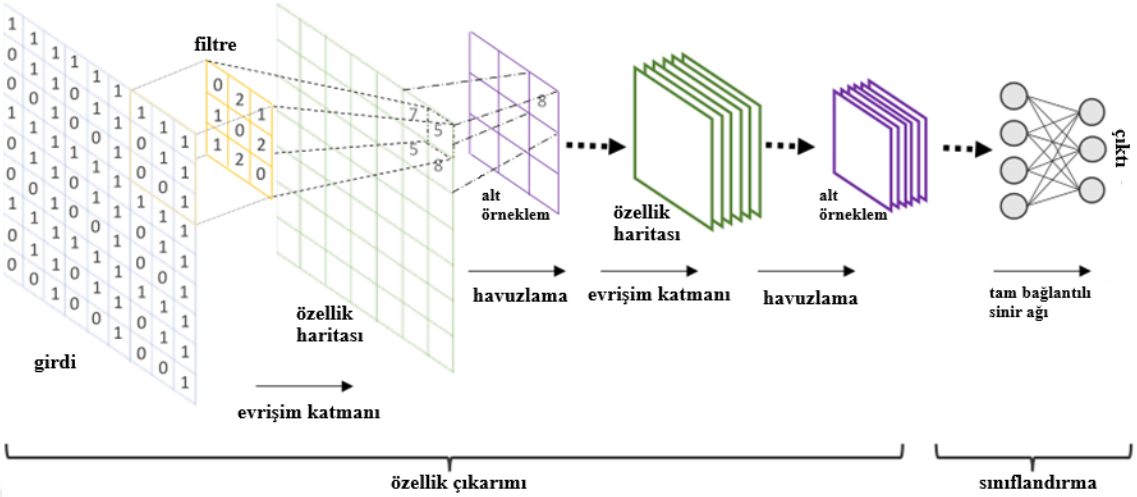
Rumelhart, Hinton ve Williams'ın 1986'da yayınladıkları makale geri yayılım algoritmasının modern anlamda tanıtımını ve kullanımını içermektedir (Rumelhart vd., 1986). Bu algoritma, bir sinir ağının eğitiminde kullanılarak ağırlıkların ve biasların güncellenmesini sağlar. Ağın çıkışı ile gerçek değer arasındaki farka göre bir kayıp fonksiyonu tanımlanır ve bu kayıp fonksiyonunun ağın parametrelerine göre türevi hesaplanarak geriye doğru yayılarak ağırlıkların ve biasların güncellenmesi sağlanmaktadır.

Bu makale, sinir ağlarının eğitiminde geri yayılım algoritmasının etkinliğini göstermekte ve birçok problemde başarılı sonuçlar elde edilmesini sağlamaktadır. Aynı zamanda, derin öğrenme alanının gelişimine büyük bir katkı sağlayarak, yapay zekâ ve makine öğrenimi alanlarında önemli bir dönüm noktası olmuştur.

3.1.2. Evrişimli sinir ağları

Evrişimli sinir ağları (ESA), özellikle görüntü işleme alanında kullanılan derin öğrenme modelleridir. Bu modeller, biyolojik görme sisteminden ilham alınarak tasarlanmıştır ve görüntü verilerinin etkili bir şekilde işlenmesini sağlar. ESA'ların temel bileşenleri arasında evrişim katmanları, aktivasyon fonksiyonları, havuzlama katmanları ve tam bağlantılı katmanlar yer almaktadır. Evrişim katmanları, giriş verisinden öznitelik haritaları çıkarırken, havuzlama katmanları öznitelik haritalarının boyutunu azaltmaktadır. Bu mimari özellikler, özellikle görüntü sınıflandırma, nesne tespiti ve doğal dil işleme gibi alanlarda başarıyla kullanılmaktadır (He vd., 2016; Krizhevsky vd., 2012). Ayrıca, ESA'lar ses sınıflandırması gibi ses işleme görevlerinde de başarılı sonuçlar elde etmektedir. Ses verisine gerekli ön işlemler uygulanarak her ses çerçevesinden çıkarılan özelliklerin birleştirilmesiyle oluşturulan giriş görüntüsü, ESA tarafından ses sınıflandırması için etkili bir şekilde kullanılabilmektedir (Lim vd., 2018).

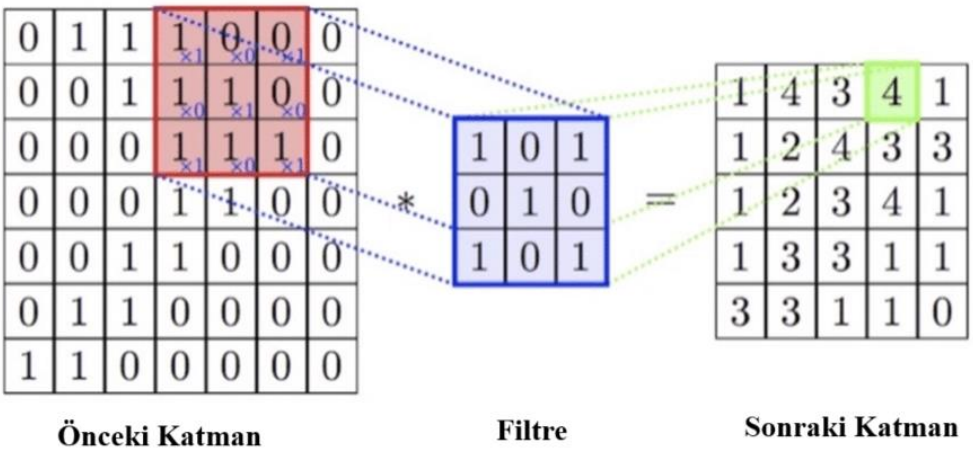
ESA’nın model yapısı Şekil 3.3’ te görülmektedir.



Şekil 3.3. ESA model yapısı (Cunha vd., 2023).

3.1.2.1. Evrişim katmanları

ESA'ların temel yapı taşlarından biri evrişim katmanlarıdır. Evrişim katmanı, girdi görüntüsünden özellik haritalarını çıkarmak için bir dizi filtre kullanarak karmaşık görsel tanıma görevlerinde etkili bir şekilde kullanılmaktadır (Albawi vd., 2017). Bu filtreler, görüntüdeki özellikleri vurgulamak veya çıkarmak için tasarlanmıştır. Örneğin, bir filtre kenarları belirlemek veya belirli bir deseni tanımak için kullanılmaktadır. Evrişim katmanları, girdi verisinin özelliklerini korurken boyutunu azaltır ve böylece ağın ölçeklenebilirliğini artırır. Evrişim işlemine ait örnek model Şekil 3.4’te görülmektedir.

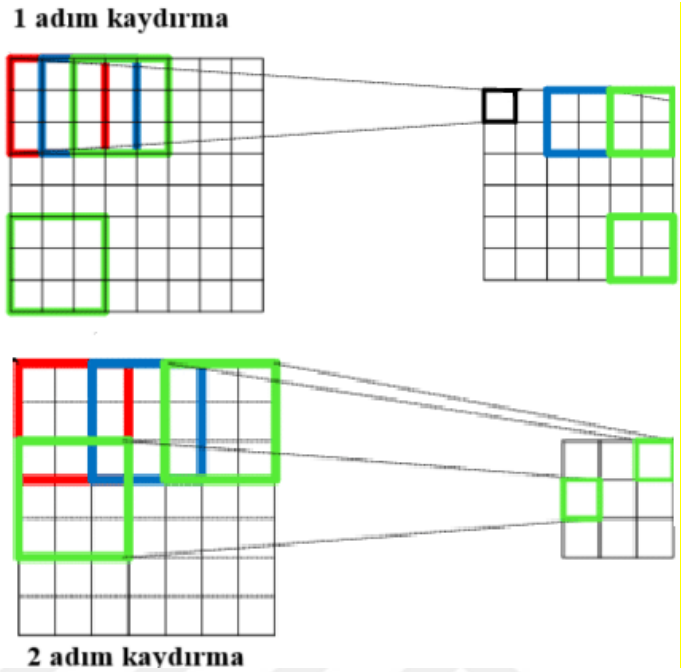


Şekil 3.4. Evrişim işlemi model örneği (Sun vd., 2020).

Evrişim işlemi için kullanılan filtrenin boyutu ve özellikleri, ağın öğrenme kapasitesini belirleyen temel unsurlardandır (Krizhevsky vd., 2012). Filtreler, bir öğrenme modelinde belirli özellikleri belirlemek için kullanılan matris veya ağırlık setleridir. Kullanılan filtreler, girdi verisi üzerinde gezinerek belirli özellikleri vurgular veya çıkarır. Filtreler, eğitim süreci boyunca iteratif olarak ayarlanır, böylece ağın hedeflenen çıktıya yaklaşmasına yardımcı olur. Bu ayarlamalar, öğrenme modelinin belirli bir görevi başarıyla gerçekleştirmesi için gerekli özellikleri belirleyen filtrelerin önemini vurgular. Bu filtrelerin kullanımı, ağın veri üzerinde daha iyi performans göstermesine ve daha iyi sonuçlar elde etmesine olanak tanımaktadır.

Bir evrişimli sinir ağının filtrelerinin girdi verisi üzerinde kaydırılma miktarı adım büyüklüğü parametresi ile belirlenir. Bu parametre, filtrelerin girdi üzerinde kaç birimlik adımlarla hareket edeceğini kontrol eder. Örneğin, bir adım değeri olarak 1 belirlendiğinde, filtre girdi verisinin her pikselini tek tek ziyaret eder. Ancak, bir adım değeri olarak 2 belirlendiğinde, filtre girdinin her iki pikselini de atlar ve bu şekilde daha büyük bir adımla hareket eder. Bu, filtrelerin girdi verisi üzerinde daha hızlı hareket etmesini ve daha küçük bir çıktı boyutuna sahip olmasını sağlar. Adım değeri, modelin karmaşıklığını ve hesaplama maliyetini etkileyen önemli bir parametredir. Daha büyük bir adım, modelin daha az sayıda özellik haritası oluşturmasına ve dolayısıyla daha az hesaplama gerektirmesine neden olurken, aynı zamanda veri kaybına da yol açmaktadır.

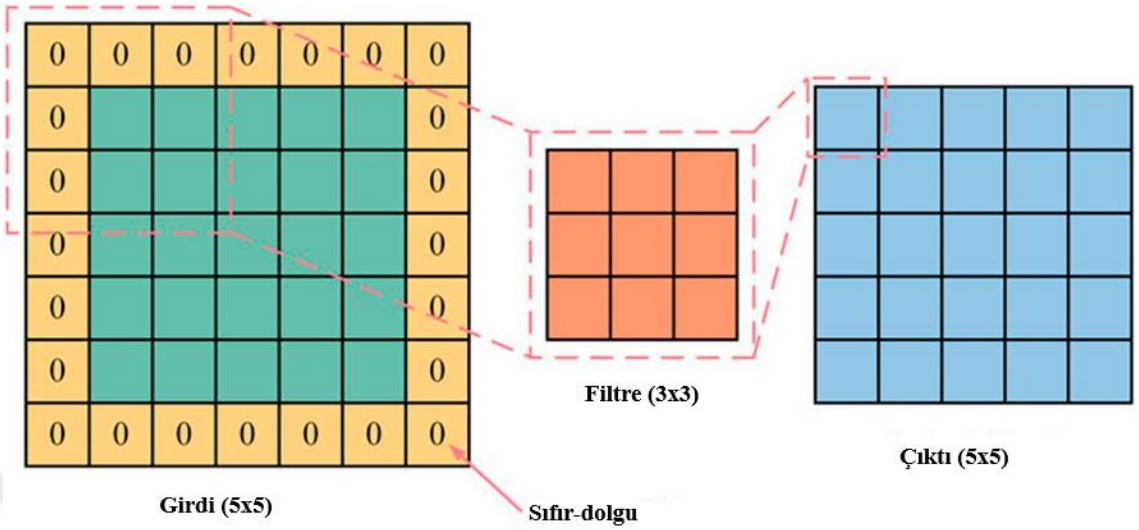
Şekil 3.5'te 1 ve 2 adım kaydırma evrişim işlemi görülmektedir.



Şekil 3.5. 1 ve 2 adım kaydırma işlemi (Syafeeza vd., 2015).

Evrişim veya diğer filtreleme teknikleri kullanılarak giriş verisi üzerinde işlem yapılırken, özellikle kenar piksellerinde, filtrelerin tam olarak uygulanamaması veya yeterli çevresel bilginin bulunmaması nedeniyle, giriş verisinin kenarlarında bilgi kaybı yaşanabilmektedir. Bu bilgi kaybını önlemek için kenarlık dolgusu (padding) tekniği kullanılmaktadır. Kenarlık dolgusu, giriş verisinin kenarlarına eklenen piksellerden oluşur ve genellikle sıfır değeriyle veya giriş verisinin kenarındaki değerlerin bir kopyasıyla doldurulmaktadır. Bu, filtrelerin giriş verisinin kenarlarına kadar uygulanabilmesini sağlar ve kenarlardaki bilgilerin korunmasını sağlar. Ayrıca, çıktı boyutunun kontrol edilmesine yardımcı olmaktadır. Özellikle, istenen bir ilişki varsa, kenarlık dolgusu bu ilişkinin korunmasını sağlar.

Kenarlık dolgusu işlemine ait örnek bir görüntü Şekil 3.6'da görülmektedir.



Şekil 3.6. Kenarlık dolgusu işlemi (Kong vd., 2020).

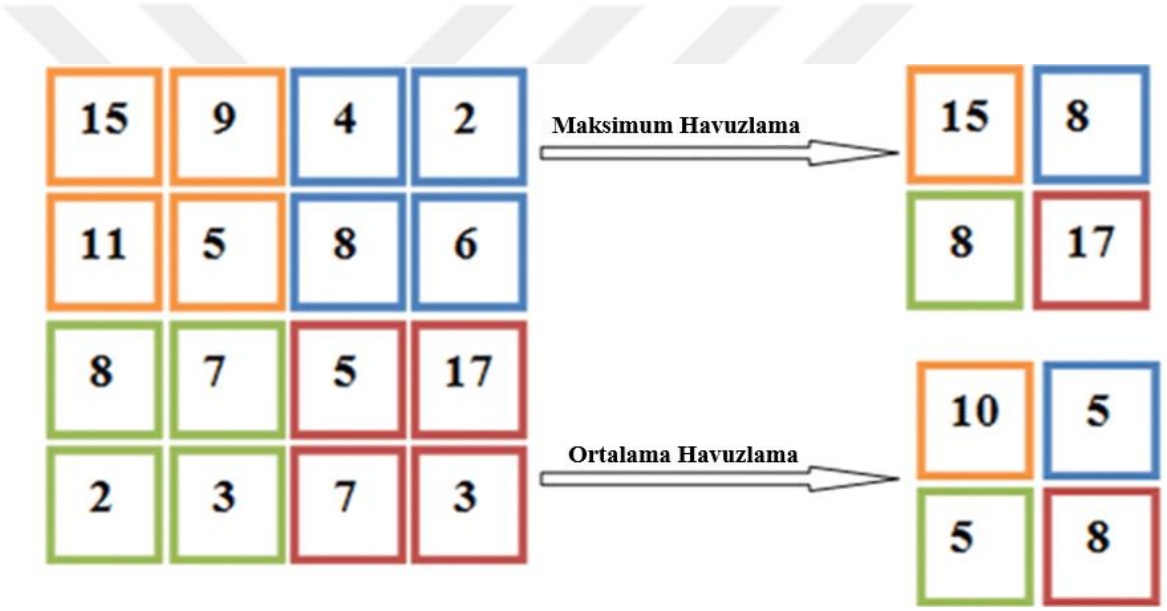
3.1.2.2. Havuzlama katmanları

Havuzlama katmanı, bir derin öğrenme modelinde bulunan bir yapılandırma katmanıdır. Bu katman, her bir küçük bölgenin içindeki bilgiyi birleştirmek ve ardından sonuçları örneklemlemek için havuzlama operatörünün uygulanmasıyla elde edilmektedir (Sun vd., 2017). Genellikle bir evrişim katmanından sonra gelir ve özellik haritalarının boyutunu küçültmek için kullanılır. Havuzlama katmanları, özellik haritasındaki küçük alanlardan özet bilgi toplar ve bu bölgeleri örnekler (Krizhevsky vd., 2012). Bu, ağın özelliklerini konumdan bağımsız hale getirir ve genelleme yeteneğini artırır. Havuzlama katmanları, ağın karmaşıklığını azaltırken önemli bilgileri korur ve nesne tanıma ve sınıflandırma gibi görevler için etkili modeller oluşturmaya yardımcı olmaktadır. Havuzlama katmaları olarak genellikle maksimum havuzlama ve ortalama havuzlama teknikleri kullanılmaktadır.

Maksimum havuzlama, ESA içinde sıkça kullanılan bir katman olarak, özellik haritalarının çözünürlüğünü azaltırken modelin pozisyon değişikliklerine, ölçek değişikliklerine ve küçük varyasyonlara karşı daha durağan ve genelleştirilebilir olmasını sağlar (Yu vd., 2014). Bu işlemde, önceki katmandan gelen özellik haritası, belirli boyutlarda ve adımda kaydırılan bir filtreyle taranır. Her pencerenin içindeki en büyük özellik değeri seçilir ve bu değerler, yeni bir özellik haritasına aktarılır. Bu şekilde,

önemli özellikler korunurken modelin hesaplama maliyeti azalır ve daha hızlı eğitim ve tahmin süreçleri sağlanmaktadır.

Ortalama havuzlama, belirli bölgelerdeki özellikleri gruplayarak her grubun ortalamasını alır (Bieder vd., 2021). Bu işlem, görüntünün boyutunu küçültürken önemli özellikleri korur. Ortalama havuzlama katmanı, özellik haritasındaki her bir bölgeyi gruplayarak bu grupların ortalamasını hesaplar ve bu değerleri yeni bir özellik haritasında oluşturur (Wang vd., 2017). Bu işlem, yerel değişikliklere karşı kararlılık sağlar ve aşırı öğrenmeyi azaltarak genelleme yeteneğini artırır. Bu nedenle, özellikle evrişimli sinir ağlarında görüntü sınıflandırma ve tanıma gibi görevler için yaygın olarak tercih edilmektedir. Şekil 3.7'de maksimum ve ortalama havuzlama örneği görülmektedir.

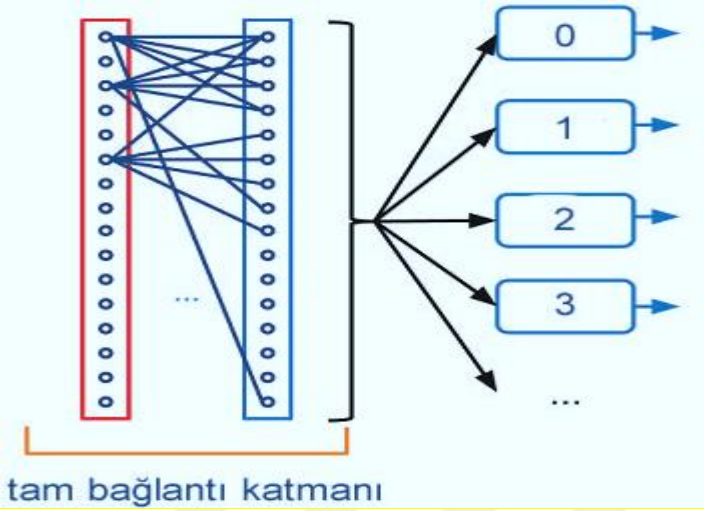


Şekil 3.7. Maksimum ve ortalama havuzlama (Rakshit vd., 2021).

3.1.2.3. Tam bağlantılı katmanlar

Tam bağlantılı katmanlar, evrişimli sinir ağlarının sonunda yer alır ve sınıflandırma yapmak için kullanılmaktadır. Bu katmanlar, özellik çıkarma aşamasından gelen özellik vektörlerini alarak, bunları sınıflandırma için uygun bir formata dönüştürmektedir. Böylece, önceki katmanlarda öğrenilen özellikleri kullanarak girdi görüntülerini doğru şekilde sınıflandırmaktadır (Krizhevsky vd., 2012). Tam bağlantılı katmanlar, yoğun katmanlar olarak da bilinir ve karmaşık özellikleri kullanarak veriler arasında ayırt edici özellikler bulur ve sınıflandırma yapar.

Şekil 3.8'de tam bağlantılı katman modeli örneği görülmektedir.



Şekil 3.8. Tam bağlantılı katman modeli (Kemaloğlu ve Sevli, 2019).

3.1.2.4. Aktivasyon fonksiyonları

Aktivasyon fonksiyonları, yapay sinir ağlarında nöronların çıktısını belirleyen matematiksel işlemlerdir. Farklı problem durumları ve matematiksel yaklaşımlara uygun aktivasyon fonksiyonları geliştirilmiştir. Aktivasyon fonksiyonları, sinir ağlarının performansı ve öğrenme kapasitesi üzerinde büyük bir etkiye sahiptir (Myers ve Hutchinson, 1989). Bu yüzden problemin doğasına uygun bir aktivasyon fonksiyonu seçilmelidir. Sinir ağlarında genellikle Sigmoid, ReLU, Leaky ReLU, Tanh ve Softmax aktivasyon fonksiyonları kullanılmaktadır.

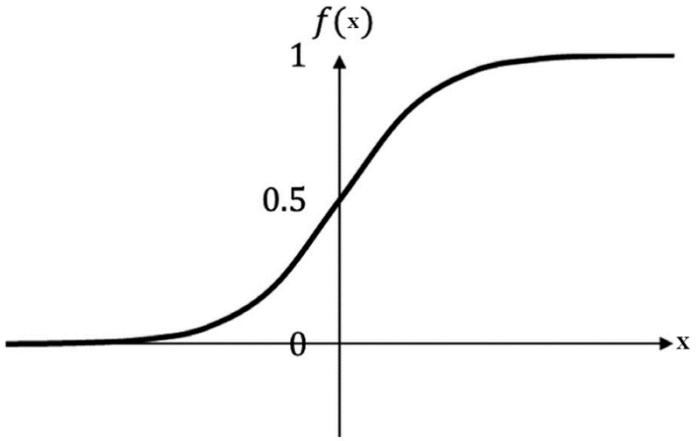
Sigmoid fonksiyonu

Sigmoid fonksiyonu, genellikle sınıflandırma problemlerinde kullanılan ve girişleri [0, 1] veya [-1, 1] aralığına sıkıştıran bir tür aktivasyon fonksiyonudur. En yaygın olarak kullanılan sigmoid fonksiyonu, lojistik sigmoid fonksiyonudur, matematiksel olarak denklem 3.2'deki gibi ifade edilmektedir.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.2}$$

Bu fonksiyon, negatif sonsuza veya pozitif sonsuza giden bir girdiyi [0, 1] aralığındaki bir çıktıya dönüştürmektedir. Sigmoid fonksiyonu, yapay sinir ağlarında özellikle çıkış

katmanında kullanılarak, çıktının olasılık veya aktivasyon seviyesi olarak yorumlanmasını sağlar. Sigmoid fonksiyonunun grafiği Şekil 3.9'da görülmektedir.



Şekil 3.9. Sigmoid fonksiyonu grafiği

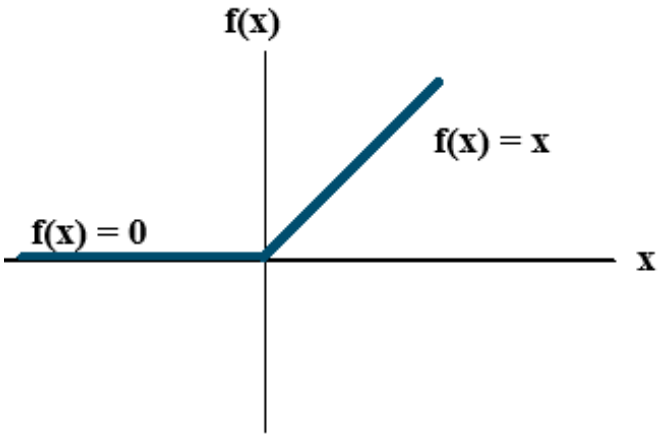
ReLU fonksiyonu

ReLU (Rectified Linear Unit), derin öğrenme ve yapay sinir ağlarında yaygın olarak kullanılan bir aktivasyon fonksiyonudur. ReLU fonksiyonu, giriş değeri negatif ise sıfırı, pozitif ise kendisini çıkış olarak verir. ReLU fonksiyonu ait matematiksel olarak denklem 3.3'te görülmektedir.

$$f(x) = \max(0, x) \tag{3.3}$$

Bu fonksiyon incelenen yapıdaki negatif değerlerin elenmesini sağlar (Fırıldak ve Talu, 2019). ReLU fonksiyonunun avantajı, diğer aktivasyon fonksiyonlarına kıyasla hesaplama açısından daha hızlı olması ve sıfır olmayan türevlere sahip olmasıdır. Bu özellikler sayesinde, ReLU fonksiyonu derin sinir ağlarının eğitimini hızlandırabilir ve genellikle gizli katmanlarda tercih edilir.

ReLU fonksiyonunun grafiği Şekil 3.10'da görülmektedir.



Şekil 3.10. ReLU fonksiyonu grafiği (Li vd., 2022).

Leaky ReLU fonksiyonu

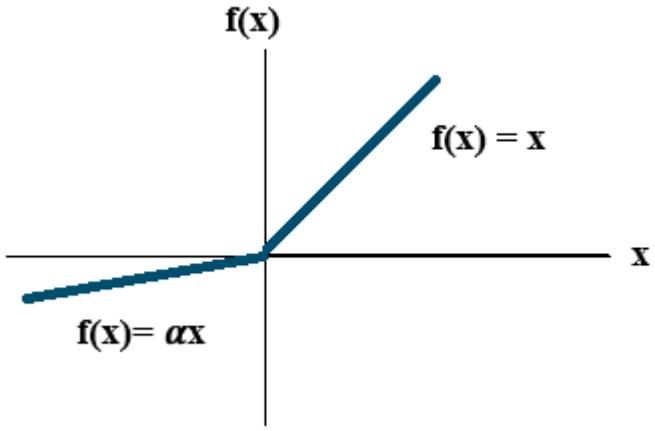
Bu fonksiyon ReLU fonksiyonunun genişletilmiş bir halidir. ReLU fonksiyonu negatif girişler için sıfır değerini döndürürken, pozitif girişler için aynı değeri döndürür. Ancak, ReLU fonksiyonunun dezavantajı, negatif girişler için sıfır döndürdüğü için "ölü nöron" problemine yol açabilir. Bu problemde, ağdaki bazı nöronlar eğitim sürecinde hiç aktive olmayabilir ve bu nöronlar artık güncellenmez, yani öğrenme durur. Leaky ReLU, negatif girişlerde sıfır yerine küçük bir eğime sahiptir. Bu durum ölü nöronları azaltabilir ve öğrenmeyi negatif aralıkta da devam ettirebilir (Xu vd., 2020). Leaky ReLU fonksiyonu matematiksel olarak denklem 3.4'deki gibi tanımlanmaktadır.

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \tag{3.4}$$

Burada α parametresi, negatif girişler için belirlenen küçük bir eğimi ifade etmektedir. Tipik olarak, α değeri küçük bir pozitif sabittir. Bu değer, negatif girişler için küçük bir eğim sağlayarak, Leaky ReLU'nun ölü nöron sorununu azaltmasına yardımcı olmaktadır.

Derin sinir ağlarında tercih edilen Leaky ReLU, sıfır olmayan türevlere sahip olduğu için geri yayılım sırasında daha istikrarlı ve etkili bir öğrenme sağlayabilir. Ayrıca, negatif girişler için belirlenen eğri, ağın daha geniş bir giriş aralığında stabil kalmasına ve daha çeşitli verilere uyum sağlamasına olanak tanır.

ReLU fonksiyonunun grafiği Şekil 3.11'de görülmektedir.



Şekil 3.11. Leaky ReLU fonksiyonu grafiği (Li vd., 2022).

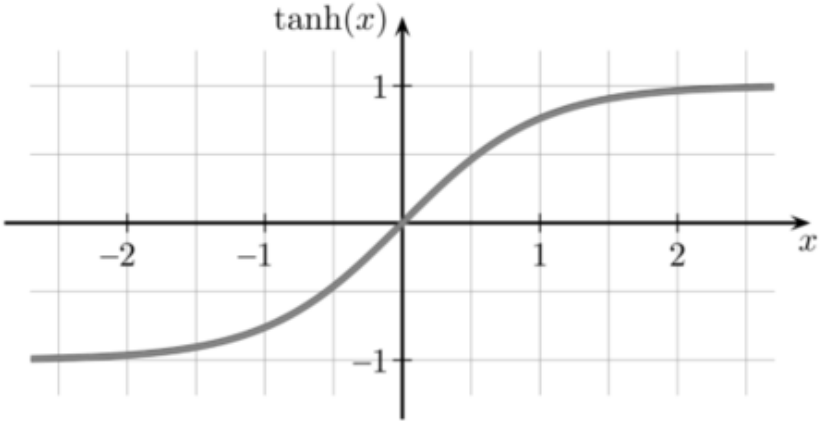
Tanh fonksiyonu

Tanh fonksiyonu, sigmoid fonksiyonuna benzer ancak orijin etrafında simetrik olarak dağılır, bu da önceki katmanlardan gelen çıkışların farklı işaretlere sahip olmasını sağlar. Bu fonksiyon sürekli, türevlenebilir ve değerleri -1 ile 1 arasındadır, sigmoid fonksiyonuna kıyasla gradyanı daha dik bir yapıya sahiptir ve sıfır merkezlidir. Matematiksel olarak, tanh fonksiyonu denklem 3.5'deki gibi tanımlanmaktadır.

$$\tan \mathrm{h}(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}} \tag{3.5}$$

Tanh fonksiyonu, özellikle simetrik bir yapıya sahip olduğu ve ortalama olarak sıfır merkezli olduğu için, veri normalleştirmede yaygın olarak kullanılmaktadır. Ayrıca, sınırlı bir çıkış aralığına sahip olması nedeniyle, aşırı uç değerlerle başa çıkmak için de tercih edilmektedir.

Tanh fonksiyonunun grafiği Şekil 3.12'de görülmektedir.



Şekil 3.12. Tanh fonksiyonu grafiği (Rojas vd., 2023).

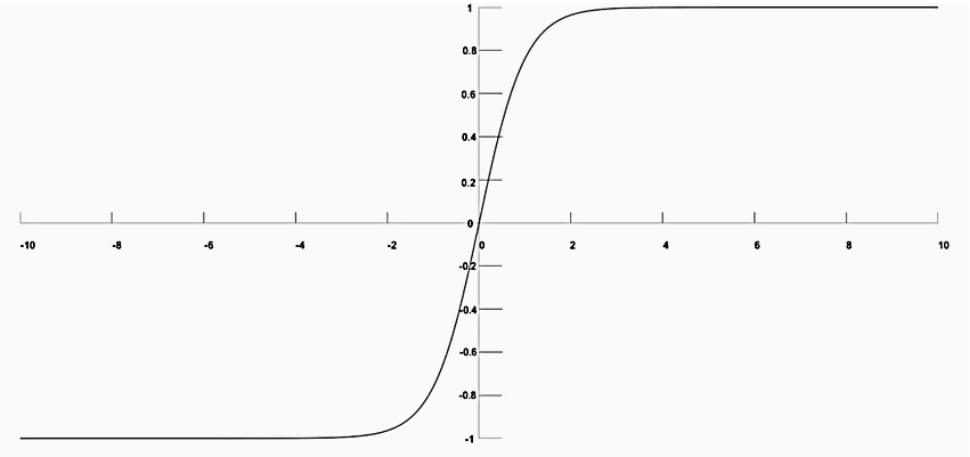
Softmax fonksiyonu

Softmax fonksiyonu, özellikle çok sınıflı sınıflandırma problemlerinde kullanılan bir aktivasyon fonksiyonudur. Herhangi bir sayı kümesini olasılık dağılımına dönüştürmek için kullanılmaktadır. Softmax fonksiyonu, bir olayın n farklı olay üzerindeki olasılık dağılımını hesaplayarak, her hedef sınıfın olasılığını belirler ve bu olasılıklar, girdiler için hedef sınıfın seçilmesine yardımcı olur (Altıparmak vd., 2019). Çıkış olasılıkları 0 ile 1 arasında değer alır ve toplamları 1'e eşittir (Goodfellow vd., 2016). Softmax fonksiyonunun matematiksel tanımı denklem 3.6'da görülmektedir.

$$f_{(x)_i} = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \tag{3.6}$$

Bu formülde x, giriş vektörünü; x_i bu vektörün i-inci elemanını; K sınıf sayısını temsil eder ve $f(x)_i$ ise i-inci sınıfa ait olasılığı ifade eder.

Softmax fonksiyonunun grafiği Şekil 3.13'te görülmektedir.



Şekil 3.13. Softmax fonksiyonu grafiği (Shen vd., 2019).

3.1.2.5. Kayıp fonksiyonları

Kayıp fonksiyonları (loss functions), makine öğrenimi ve yapay zeka modellerinin performansını ölçmek için kullanılan özel fonksiyonlardır. Bu fonksiyonlar, modelin tahminlerinin gerçek değerlerden ne kadar farklı olduğunu hesaplar ve bu farkı bir skor veya hata değeri olarak ifade eder. Kayıp fonksiyonları, modelin parametrelerini (ağırlıklarını) güncellemek için geriye doğru yayılma (backpropagation) sürecinde kullanılır ve modelin optimize edilmesinde önemli bir rol oynar (Goodfellow vd., 2016; LeCun vd., 2015; Rumelhart vd., 1986).

Kayıp fonksiyonları, verilerin işlenip katmanlardan geçirilmesiyle tahminsel bir çıktı elde edilen makine öğrenimi sürecinde önemli bir rol oynamaktadır. Bu tahminsel çıktı ile gerçek çıktı arasındaki fark, kayıp fonksiyonu tarafından analiz edilir ve modelin performansı incelenir. Kayıp fonksiyonları, optimizasyon fonksiyonları yardımıyla sürekli olarak güncellenir ve tahminsel hatayı azaltmayı öğrenir. Bu sürekli güncelleme ve iyileştirme, yüksek performanslı tahmin modellerinin geliştirilmesine olanak tanır. Problemlerin tekrar sorgulanması ve iteratif olarak çözülmesiyle bu süreç devam eder. Birçok kayıp fonksiyonu bulunmaktadır ve kullanılacak fonksiyon, problem yapısına uygun olarak seçilmelidir.

Ortalama kare hata fonksiyonu (Mean Squared Error - MSE)

Bu fonksiyon, modelin tahminlerinin gerçek değerlerden ne kadar saptığına dair bilgi sağlar. Sapma miktarı, tahmin edilen değerler ile gerçek değerler arasındaki farkların karelerinin toplamının ortalaması alınarak hesaplanır. Bu fonksiyona ait matematiksel ifade denklem 3.7'de görülmektedir.

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{3.7}$$

Bu formülde, n toplam veri sayısını, y_i gerçek değerleri, $\hat{y}_i$ ise modelin tahmin ettiği değerleri belirtmektedir.

İkili çapraz entropi fonksiyonu (Binary Cross Entropy-BCE)

Bu kayıp fonksiyonu, genellikle modelin iki sınıf arasındaki tahminlerin doğruluğunu değerlendirir, bu sınıf etiketleri genellikle 0 veya 1 olarak temsil edilir. Bu fonksiyona ait matematiksel ifade denklem 3.8'de görülmektedir.

$$BCE = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{3.8}$$

Bu formülde, n toplam veri sayısını, y_i gerçek değerleri, $\hat{y}_i$ ise modelin tahmin ettiği değerleri belirtmektedir.

Kategorik çapraz entropi fonksiyonu (Categorical Cross Entropy-CCE)

Bu kayıp fonksiyonu, yaklaşım olarak ikili çapraz entropiye benzemektedir. Farklı yönü ise sınıf sayısının ikiden fazla olmasıdır. Bu fonksiyon, modelin çoklu sınıflar arasında tahmin ettiği olasılıkları değerlendirir. Her bir sınıf için ayrı ayrı gerçek etiketler ve tahmin edilen olasılıklar arasındaki farkı hesaplar ve bu farkların ortalamasını alarak toplam kayıp miktarını belirler. Bu fonksiyona ait matematiksel ifade denklem 3.8'de görülmektedir.

$$CCE = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij}) \tag{3.9}$$

Bu formülde, n toplam veri sayısını, C sınıf sayısını, y_i gerçek değerleri, $\hat{y}_i$ ise modelin tahmin ettiği değerleri belirtmektedir.

3.1.2.6. Optimizasyon algoritmaları

Gradyan iniş algoritması (Gradient Descent-GD)

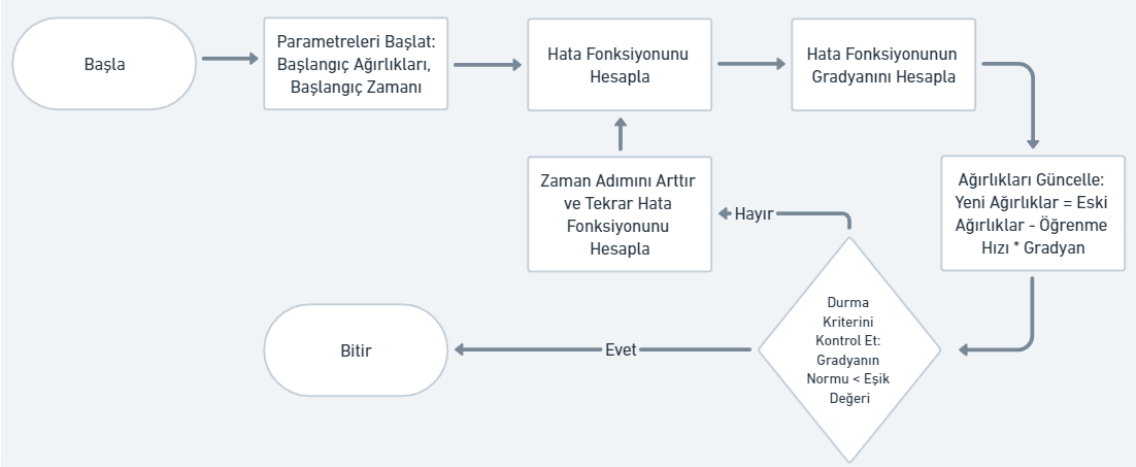
Gradient iniş algoritması, tüm parametreleri tek bir ağırlık vektörü şeklinde gruplayarak işlem yapar. Hata fonksiyonu, bu ağırlık vektörüne bağlı olarak ifade edilir. Algoritma, başlangıçta belirlenen bir ağırlık tahminiyle başlar ve hata fonksiyonunun en hızlı azaldığı yöne küçük adımlarla ilerleyerek ağırlık vektörünü günceller. Bu işlem, ağırlık vektörünün uzayında tekrarlanarak hatanın en aza indirildiği bir dizi nokta elde edilir. Küçük bir pozitif öğrenme hızı parametresi, güncellemenin ne kadar hızlı yapılacağını belirler. Son olarak, hata fonksiyonunun ağırlıklara göre türevleri hesaplanır ve modelin eğitimi için bu türevler kullanılır. Bu şekilde, model her bir eğitim örneği için ayrı ayrı güncellenir ve gerçek zamanlı uygulanabilirlik sağlanır (Bishop, 1995). Bu algoritmaya ait matematiksel ifade denklem 3.10'da görülmektedir.

$$w_{t+1} = w_t + \alpha \nabla E(w_t) \tag{3.10}$$

Bu formülde, w_t ağırlık vektörü, $E(w_t)$ hata fonksiyonunun w_t noktasındaki değeri, $\nabla E(w_t)$ ifadesi w_t noktasındaki hata fonksiyonunun gradyanını (türevini) belirtmektedir. α ise öğrenme hızıdır ve her adımda ne kadar ilerleyeceğimizi belirlemektedir.

Algoritmanın işleyişi, belirli adımlardan oluşan bir süreç izler. Başlangıçta, algoritma başlatılır ve ilk adımda başlangıç ağırlıkları ve başlangıç zamanı parametreleri belirlenir. Ardından, tüm veri seti kullanılarak hata fonksiyonu hesaplanır. Hata fonksiyonunun hesaplanmasından sonra, bu fonksiyonun gradyanı hesaplanır. Hesaplanan gradyan kullanılarak ağırlıklar güncellenir; bu işlem, yeni ağırlıkların eski ağırlıklardan öğrenme hızı ile gradyanın çarpımının çıkarılmasıyla gerçekleştirilir. Güncellenen ağırlıklarla birlikte, durma kriteri kontrol edilir. Bu kriter, gradyanın normunun belirli bir eşik değerinden küçük olup olmadığını kontrol eder. Eğer gradyanın normu eşik değerinden küçükse, algoritma sona erer. Aksi takdirde, zaman adımı artırılır ve hata fonksiyonunu yeniden hesaplama sürecine geri dönülür. Bu döngü, durma kriteri sağlanana kadar devam eder.

Şekil 3.14'de gradyan iniş algoritmasına ait akış diyagramı görülmektedir.



Şekil 3.14. Gradyan iniş algoritması akış diyagramı

Stokastik gradyan iniş algoritması (Stochastic Gradient Descent-SGD)

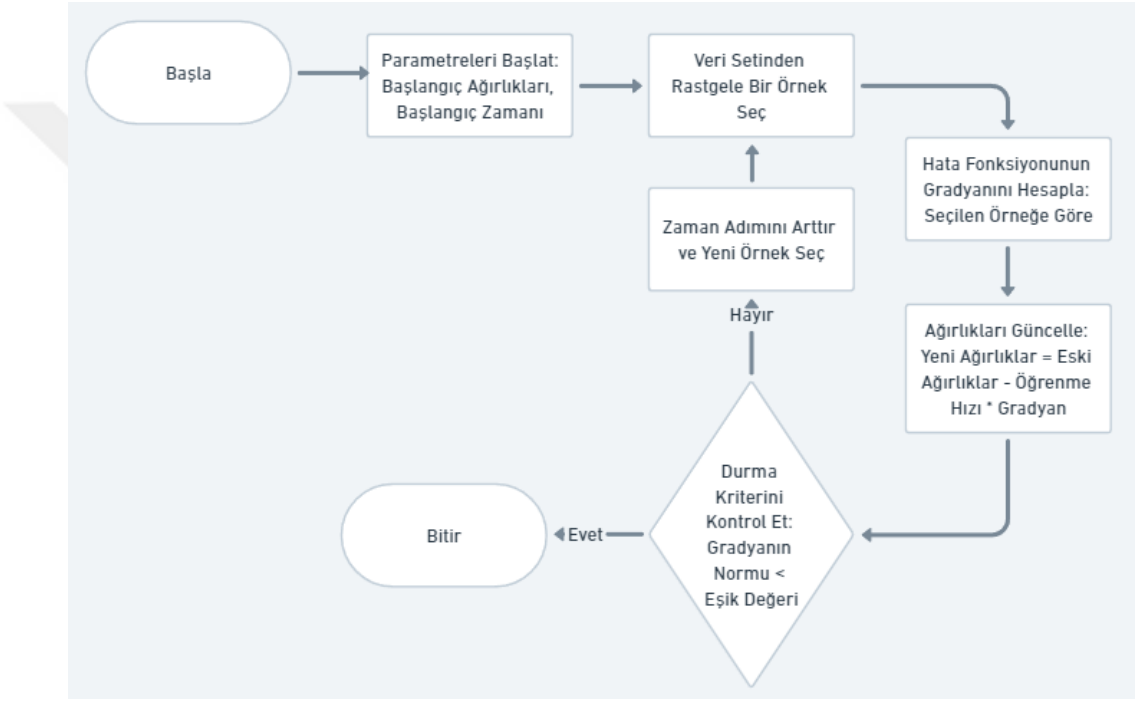
SGD, parametreli ağlarda kullanılan ve gradyan inişin stokastik bir versiyonu olan bir öğrenme yöntemidir. Bu yöntemin temel amacı, parametrelerin güncellenmesi sırasında her bir eğitim örneği için hata fonksiyonunun gradyanını hesaplayarak optimizasyon yapmaktır (Amari, 1993). Bu, büyük veri setlerinde daha hızlı ve verimli çalışmayı sağlar, çünkü her adımda tüm veri seti yerine rastgele seçilen tek bir örnek veya küçük bir örnekler kümesi kullanılır. Bu algoritmanın matematiksel ifadesi ait denklem 3.11'de görülmektedir.

$$\theta_{t+1} = \theta_t - \alpha_t \nabla l(x_t, y_t; \theta_t) \tag{3.11}$$

Bu formülde, θ_t ifadesi t adımdaki ağırlık vektörünü, α_t ifadesi t adımdaki öğrenme hızını, $\nabla l(x_t, y_t; \theta_t)$ ifadesi ise t adımındaki hata fonksiyonunun gradyanı temsil etmektedir.

Algoritmanın işleyişi, belirli adımlardan oluşan bir süreç izler. Başlangıçta, algoritma, başlangıç ağırlıkları ve başlangıç zamanını belirleyerek başlamaktadır. İlk adımda, tüm veri seti yerine, eğitim süreci için rastgele bir örnek seçilmektedir. Bu seçim, büyük veri setlerinde işlem hızını artırmak için yapılır. Seçilen örnek kullanılarak hata fonksiyonunun gradyanı hesaplanır. Gradyan, hata fonksiyonunun eğimidir ve ağırlıkları hangi yönde güncellememiz gerektiğini göstermektedir. Hesaplanan gradyan kullanılarak

ağırlıklar güncellenir. Bu güncelleme, öğrenme hızını da dikkate alarak hata fonksiyonunu minimize etmeye çalışır. Güncellenmiş ağırlıklarla birlikte durma kriteri kontrol edilmektedir. Bu kriter, gradyanın normunun belirli bir eşik değerden küçük olup olmadığını kontrol eder. Böylece, algoritmanın yeterince yakınsayıp yakınsamadığını belirlenir. Eğer gradyanın normu eşik değerden küçükse, algoritma durur ve süreç sona erer. Ancak, gradyanın normu eşik değerden büyükse algoritma devam eder. Bu durumda, zaman adımı artırılır ve yeni bir rastgele örnek seçilerek süreç tekrar edilir. Şekil 3.15'te stokastik gradyan iniş algoritmasına ait akış diyagramı görülmektedir.



Şekil 3.15. Stokastik gradyan iniş algoritması akış diyagramı

Uyarlanabilir moment tahmini algoritması (Adaptive Moment Estimation-ADAM)

ADAM, gradyan tabanlı stokastik optimizasyon problemlerinin çözümünde kullanılan bir algoritmadır. Özellikle yüksek boyutlu parametre uzaylarında ve değişken veri dağılımlarında etkili bir şekilde çalışmak üzere tasarlanmıştır. Adam, gradyanların ilk ve ikinci momentlerini (ortalama ve varyans) izleyerek adaptif öğrenme hızları hesaplar ve bu sayede daha stabil ve hızlı bir yakınsama sağlamaktadır (Kingma ve Ba, 2014). Algoritma, hem AdaGrad (Duchi vd., 2011) hem de RMSProp (Tieleman, 2012)

yöntemlerinin avantajlarını birleştirir. Bu algoritmaya ait matematiksel ifadeler denklem (3.12-3.15)'te görülmektedir.

$$m_t = \beta_1 m_{t-1} - (1-\beta_1) g_t \tag{3.12}$$

$$v_t = \beta_2 v_{t-1} - (1-\beta_2) {g_t}^2 \tag{3.13}$$

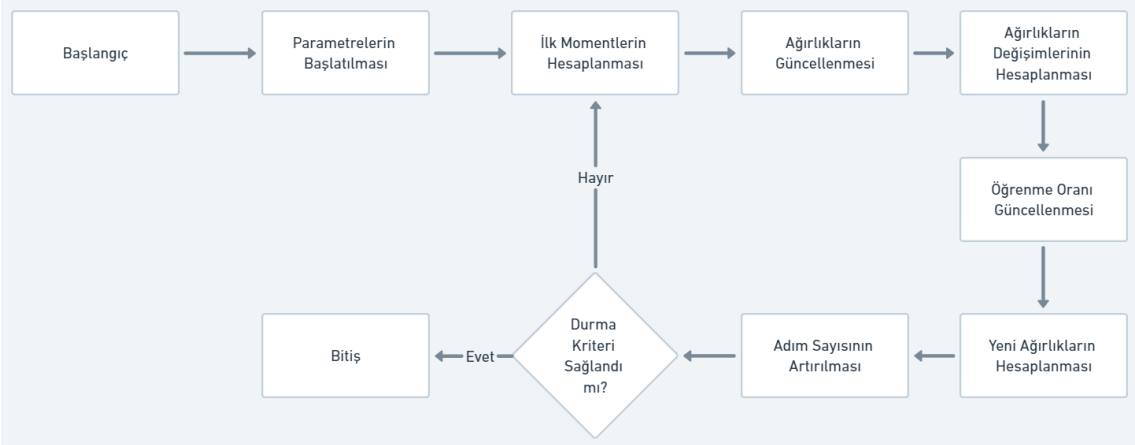
$$\widehat{m}_t = \left(\frac{m_t}{1-{\beta_1}^t}\right) \hat{v}_t = \left(\frac{v_t}{1-{\beta_2}^t}\right) \tag{3.14}$$

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v}_t + \varepsilon}} \widehat{m}_t \tag{3.15}$$

Bu formüllerde, m_t ifadesi t zamanındaki birinci moment tahmini, m_{t-1} ifadesi t zamanındaki birinci moment tahmini, β_1 ifadesi ikinci momentin üstel hareketli ortalama katsayısı, β_2 ifadesi ikinci momentin üstel hareketli ortalama katsayısı, v_t ifadesi t zamanındaki ikinci moment tahmini, v_{t-1} ifadesi t zamanındaki ikinci moment tahmini, g_t ifadesi t zamanındaki gradyan (türev), ${g_t}^2$ gradyanın her bileşeninin karesi, $\widehat{m}_t$ ifadesi önyargıdan arındırılmış birinci moment tahmini, $\hat{v}_t$ ifadesi önyargıdan arındırılmış ikinci moment tahmini, θ_t ifadesi t zamanındaki model parametreleri, θ_{t+1} ifadesi güncellenmiş model parametreleri, α ifadesi öğrenme oranı ve ε ifadesi ise küçük bir sabiti temsil etmektedir.

Algoritmanın işleyişi, belirli adımlardan oluşan bir süreç izler. Başlangıçta algoritma başlatılır ve ilk adımda parametreler başlatılır. Ardından, ilk momentler hesaplanır. Bu hesaplamaların ardından ağırlıklar güncellenir. Ağırlıkların güncellenmesi tamamlandıktan sonra, ağırlıkların değişimlerinin hesaplanmasına geçilir. Bu aşamada, öğrenme oranı da güncellenir. Sonraki adımda, yeni ağırlıklar hesaplanır. Bu hesaplamaların ardından, adım sayısı artırılır ve durma kriterinin sağlanıp sağlanmadığı kontrol edilir. Eğer durma kriteri sağlanmışsa, algoritma sona erer. Aksi takdirde, algoritma başa dönerek ağırlıkların güncellenmesi adımından itibaren süreci tekrar eder. Bu döngü, durma kriteri sağlanana kadar devam eder. Bu adımlar dizisi, her bir ağırlık güncelleme ve hesaplama işleminin ardışık olarak gerçekleştirilmesiyle, öğrenme sürecinin optimize edilmesini sağlar. Durma kriteri sağlanana kadar bu işlemler tekrar edilerek, ağırlıkların ve modelin sürekli iyileştirilmesi hedeflenir.

Şekil 3.16'da ADAM algoritmasına ait akış diyagramı görülmektedir.



Şekil 3.16. ADAM algoritması akış diyagramı

Kök ortalama kare yayılımı (Root Mean Square Propagation-RMSprop)

RMSProp, derin öğrenme modellerinin eğitimi sırasında yaygın olarak kullanılan bir optimizasyon algoritmasıdır. Bu algoritma, her parametre için uyarlanabilir bir öğrenme oranı kullanarak, gradyanların büyük değişimlerine karşı daha istikrarlı ve hızlı bir öğrenme süreci sağlar. RMSProp, gradyanların karelerinin hareketli ortalamasını hesaplayarak öğrenme oranlarını dinamik olarak ayarlar. Bu yöntem, gradyanların büyük olduğu durumlarda öğrenme oranını düşürürken, küçük olduğu durumlarda artırır. Ayrıca düşük bellek kullanımı ile avantaj sağlar ve özellikle derin sinir ağlarının eğitimi sırasında SGD gibi basit yöntemlerden daha iyi performans gösterir. RMSProp'un hiperparametrelerini doğru ayarlamak, belirli bir problem için en iyi performansı elde etmek açısından önemlidir. Bu özellikleri sayesinde, RMSProp modern makine öğrenmesi ve derin öğrenme uygulamalarında önemli bir rol oynamaktadır (Graves, 2013; Tieleman, 2012).

RMSProp, gradyanların hareketli ortalamasını kullanarak, öğrenme oranlarını uyarlamak için gradyanların karelerinin ortalamasını hesaplar. Bu yöntem, gradyanların büyük dalgalanmalar gösterdiği durumlarda faydalıdır ve öğrenme oranlarını her parametre için ayrı ayrı uyarlayarak daha dengeli bir öğrenme süreci sağlar. RMSProp'un çalışma prensibi, her parametre için öğrenme oranını dinamik olarak ayarlamaktır. Bu, gradyanların büyük olduğu durumlarda öğrenme oranını düşürerek, parametre

güncellemelerinin çok büyük olmasını engeller. Aynı zamanda, gradyanların küçük olduğu durumlarda öğrenme oranını artırarak, parametrelerin daha hızlı güncellenmesini sağlar. Bu algoritmaya ait matematiksel ifadeler denklem (3.16-3.17)'de görülmektedir.

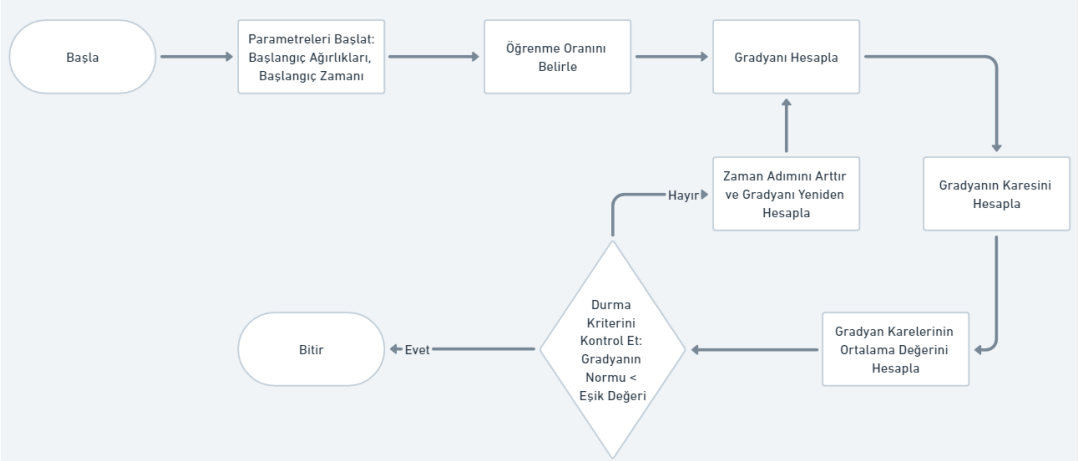
$$E[g^2]_t = \beta E[g^2]_{t-1} + (1-\beta){g_t}^2 \tag{3.16}$$

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{E[g^2]_t + \varepsilon}} g_t \tag{3.17}$$

Bu formüllerde, β hareketli algoritmanın ağırlığını belirleyen bir hiperparametre, g_t ifadesi t zamanındaki gradyan, ${g_t}^2$ gradyanın her bileşeninin karesi, θ_t ifadesi t zamanındaki güncellenen parametre, θ_{t+1} ifadesi güncellenmiş model parametreleri, $\sqrt{E[g^2]_t}$ ifadesi gradyan karelerinin ortalamasının karekökü, α ifadesi öğrenme oranı ve ε ifadesi ise küçük bir sabiti temsil etmektedir.

Algoritmanın işleyişi, belirli adımlardan oluşan bir süreç izler. Başlangıçta, ağın başlangıç ağırlıkları ve başlangıç zamanı belirlenir. Ardından, öğrenme oranı belirlenir. Her eğitim iterasyonunda, ağın mevcut ağırlıkları kullanılarak gradyanlar hesaplanır. Daha sonra, gradyanın karesi hesaplanır, ardından bu karelerin üstel hareketli ortalaması alınır. Ağırlıkların güncellenmesi için, yeni ağırlıklar, eski ağırlıklardan belirli bir öğrenme oranı ile gradyanlarla çarpılmış ve gradyan karelerinin ortalamasının tersi ile bölünmüş değerlerin çıkarılmasıyla hesaplanır. Bu aşamadan sonra, gradyanın normu belirli bir eşik değerinden küçük mü diye kontrol edilir. Eğer küçükse, eğitim sona erdirilir, aksi takdirde zaman adımı arttırılır ve gradyan yeniden hesaplanır. Bu süreç, belirli bir durma kriterine veya eşik değerine ulaşılana kadar tekrarlanır. Bu şekilde, RMSprop algoritması gradyanların değişkenliğini dikkate alarak modelin eğitimini gerçekleştirir.

Şekil 3.17'de RMSprop algoritmasına ait akış diyagram görülmektedir.



Şekil 3.17. RMSprop algoritması akış diyagramı

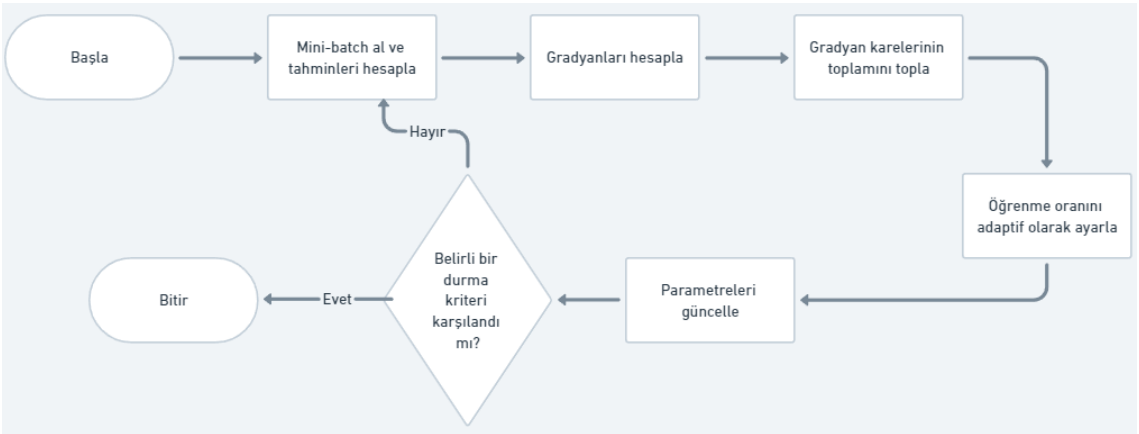
Uyarlanabilir gradyan algoritması (Adaptive Gradient Algorithm-AdaGrad)

Adagrad, gradient tabanlı bir optimizasyon algoritması olup, özellikle büyük boyutlu parametre uzaylarında ve seyrek veri setlerinde etkilidir (Duchi vd., 2011). Algoritmanın temel amacı, her parametrenin öğrenme oranını adaptif bir şekilde ayarlamak ve böylece güncellemelerin genel olarak daha dengeli ve verimli olmasını sağlamaktır (Zeiler, 2012). Adagrad, önceki gradyanların karelerinin toplamının karekköküne bölünerek öğrenme oranını günceller. Bu sayede, seyrek özellikler daha büyük güncellemeler alırken, sık görülen özellikler daha küçük güncellemeler alır. Bu adaptif öğrenme oranı ayarı, özellikler arasındaki önemli farklılıkları dengeleyerek eğitim sürecini optimize eder. Adagrad'ın bir diğer önemli özelliği, her parametre için ayrı ayrı öğrenme oranlarının hesaplanmasıdır. Bu, özellikle büyük boyutlu parametre uzayları veya seyrek veri setleri gibi durumlarda avantaj sağlamaktadır. Ancak, Adagrad'ın eğitim ilerledikçe öğrenme oranının azalması gibi dezavantajı vardır. Bu, eğitim sürecinin ilerleyen aşamalarında daha küçük güncellemelerin yapılmasına ve optimizasyonun yavaşlamasına yol açabilir. Adagrad optimizasyon algoritması, adaptif öğrenme oranı ayarıyla gradyanların önceki karelerinin toplamının karekköküne bölünerek, eğitim sürecinin dengeli ve etkin olmasını sağlar. Bu nedenle, büyük boyutlu parametre uzayları veya seyrek veri setleri gibi senaryolarda tercih edilir (Duchi vd., 2011; Zeiler, 2012).

Bu algoritmaya ait matematiksel ifade denklem 3.18'de görülmektedir.

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\alpha}{\sqrt{G_{t,ii}} + \varepsilon} g_{t,i} \tag{3.18}$$

Algoritmanın işleyişi, belirli adımlardan oluşan bir süreç izler. İlk adımda, algoritma başlar ve parametreler başlangıç değerleriyle belirlenir. Daha sonra, eğitim veri setinden bir mini-batch alınır ve bu mini-batch üzerinde ileri yayılım yapılarak tahminler üretilir. Ardından, geriye doğru yayılım kullanılarak gradyanlar hesaplanır. Elde edilen gradyanlar, her bir parametre için karelerinin toplamını biriktirmek üzere toplanır. Bu, gradyanların önceki karelerinin toplamını oluşturur. Sonraki adımda, bu birikimlenmiş gradyan karelerinin toplamı alınır ve bir küçük değer (epsilon) eklenerek sıfıra bölme hatası önlenir. Daha sonra, öğrenme oranı adaptif olarak ayarlanır. Bu adımda, her bir parametre için öğrenme oranı, gradyan karelerinin toplamının karekökü ile bölünerek hesaplanır. Son olarak, parametreler güncellenir ve belirli bir durma kriteri karşılanıp karşılanmadığı kontrol edilir. Eğer belirli bir durma kriteri karşılanmışsa, algoritma sona erer. Aksi takdirde, eğitim adımları tekrarlanır: yeni bir mini-batch alınır, tahminler hesaplanır ve eğitim adımları devam eder. Bu şekilde, Adagrad optimizasyon algoritması gradyanların karelerinin toplamını birikimleyerek, adaptif öğrenme oranı ile parametreleri güncelleyerek modelin eğitimini gerçekleştirir. Şekil 3.18'de AdaGrad algoritmasına ait akış diyagramı görülmektedir.



Şekil 3.18. AdaGrad algoritması akış şeması

3.2. Ses Analizi

3.2.1. Sesin oluşumu ve yapısı

Ses, bir ortamda (genellikle havada) yayılan basınç dalgaları olarak bilinen maddenin küçük değişimlerinin sonucudur (Rayleigh, 1896). Oluşan bu basınç dalgalanmaları, insan kulakları veya diğer ses algılayıcılar tarafından algılanabilir. Sesin algılanması, çevredeki seslerin işitsel sisteme iletilmesi ve sinirsel olarak işlenmesiyle başlar. Bu süreç, dış kulak yolundan gelen ses dalgalarının timpanik membrana çarpmasıyla başlar (Hall, 2015). İç kulağın içinde bulunan koklea, ses dalgalarını mekanik enerjiye dönüştürerek içindeki saç hücrelerine iletilmesini sağlar (Yost vd., 2007). Daha sonra, saç hücreleri tarafından işitme siniri boyunca elektrik sinyallerine dönüştürülür ve beyne gönderilir. Beyindeki işitme korteksi, bu sinyalleri işleyerek sesin kaynağını, yönünü, frekansını ve anlamını belirlememizi sağlar (Hickok ve Poeppel, 2007).

3.2.2. Ses sinyalinin özellikleri

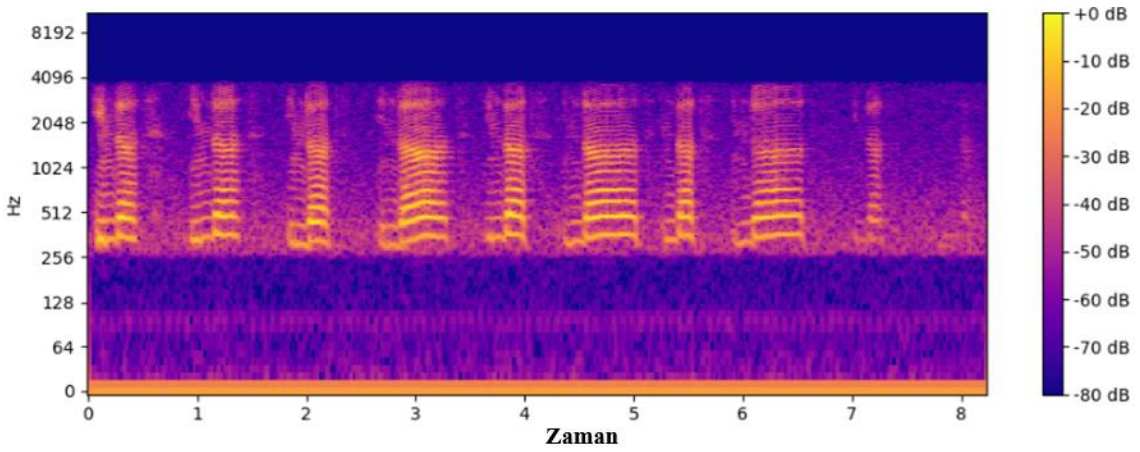
Ses sinyalinin özellikleri arasında sesin yoğunluğu (amplitüd), frekans bileşenleri, periyot ve faz gibi önemli özellikler bulunmaktadır. Ses sinyalinin amplitüdü, ses dalgasının yoğunluğunu ve şiddetini belirtir ve insanlar tarafından sesin yüksekliği veya genişliği gibi algılanan ses gücünü temsil eder. Frekans bileşenleri, sesin hangi frekanslarda bulunduğunu belirten önemli bir özelliktir ve insanlar tarafından sesin tonu olarak algılanır (Stevens ve Davis, 1938). Periyot, ses dalgasının tekrar eden deseninin uzunluğunu ifade eder ve sesin bir dönemi boyunca kaç tam dalga döngüsü geçtiğini belirler. Faz, ses dalgasının bir noktasının zaman içindeki konumunu ifade eder ve ses dalgasının diğer dalgalarla olan ilişkisini tanımlar (Kinsler vd., 2000). Bu özelliklerin insan algısı ile ilişkisi, psikofizik deneyler ve algısal psikofizik çalışmaları yoluyla incelenmiştir (Fletcher ve Munson, 1933). İnsanların algıladığı belirli amplitüd ve frekans aralıkları, genellikle bireyden bireye değişiklik gösterebilir ve çeşitli faktörlere bağlıdır. Ancak, genel olarak, insanların duyabildikleri ses frekans aralığı 20 - 20.000 Hz'dir (Barsties ve De Bodt, 2015).

3.3. Ses temsili ve özellik çıkarma

Ses sinyallerinin analizinde spektrogramlar, Mel spektrogramlar ve Mel-Frekans Kepstral Katsayıları (MFCK) gibi temsil yöntemleri ve özellik çıkarma teknikleri kullanılmaktadır. Spektrogramlar, sesin zaman ve frekans bileşenlerini görselleştirirken, Mel spektrogramlar insan kulağının frekans algısına uygun bir ölçekle bu temsili iyileştirir. MFCK ise ses sinyalinin kısa süreli güç spektrumunun logaritmik sıkıştırılması ve kepstral analiziyle elde edilen özelliklerdir ve konuşma tanıma ile müzik sınıflandırma gibi uygulamalarda yüksek doğruluk sağlamaktadır. Bu yöntemler, ses tabanlı uygulamaların daha verimli ve doğru çalışmasına imkân tanır.

3.3.1. Spektrogramlar

Spektrogram, bir ses sinyalinin zaman içinde frekans içeriğini gösteren bir grafik temsildir. Zaman-frekans analizi yaparak sesin çeşitli bileşenlerini görselleştirme imkânı sunar. Spektrogramlar, genellikle Fourier dönüşümü kullanılarak oluşturulur. Spektrogramda yatay eksen zamanı, dikey eksen ise frekansı gösterir. Belirli bir frekanstaki enerji yoğunluğu, renk veya gri tonlama ile temsil edilir; koyu renkler yüksek enerji seviyelerini, açık renkler düşük enerji seviyelerini göstermektedir. Ses sinyalinin spektrogram ile gösterimine ait bir örnek Şekil 3.19'da görülmektedir.



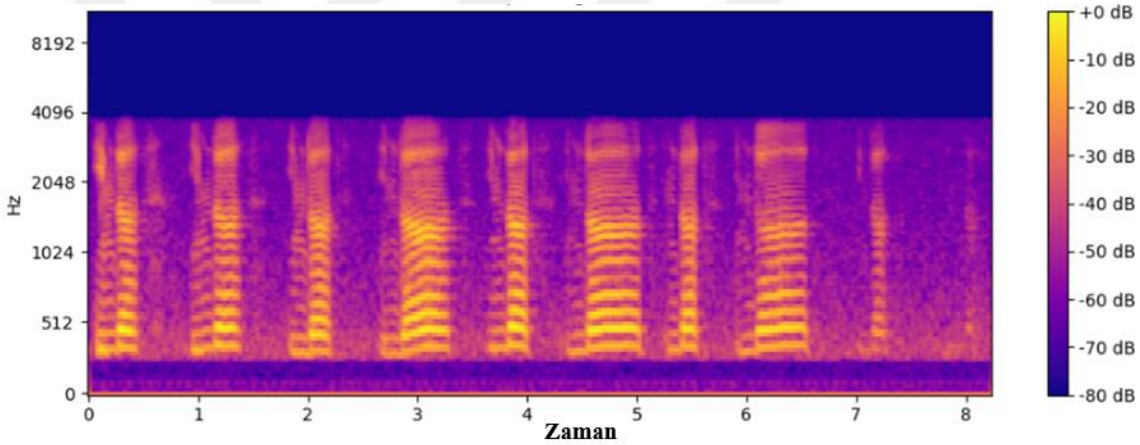
Şekil 3.19. Spektrogram örneği

Spektrogramlar, ses tanıma (Rabiner ve Juang, 1999), müzik analizi (Müller, 2015) ve biyomedikal sinyal analizi (Rajendra Acharya vd., 2006) gibi çeşitli alanlarda kullanılır. Bu yöntem konuşma ve ses tanıma sistemlerinde, ses sinyalinin zaman-frekans

özelliklerini incelemek için kullanılırken, müzik yapılarının analizi ve sınıflandırılması için de oldukça etkilidir.

3.3.2. Mel-Spektrogram

Mel-spektrogram, insan kulağının duyma algısına daha uygun bir şekilde ses frekanslarını gösteren spektrogram türüdür. Mel ölçeği, frekansların logaritmik olarak ölçeklendiği bir frekans ölçeğidir ve bu sayede insan kulağının duyarlılığına daha uygun bir temsiliyet sağlar. Mel-spektrogramlar, düşük frekanslarda daha yüksek, yüksek frekanslarda ise daha düşük çözünürlük sağlar. Spektrogram gibi, enerji yoğunluğu renk veya gri tonlama ile temsil edilir. Ancak Mel ölçeğinde frekans ekseni daha duyarlı bir temsile sahiptir. Ses sinyalinin spektrogram ile gösterimine ait bir örnek Şekil 3.20'de görülmektedir.



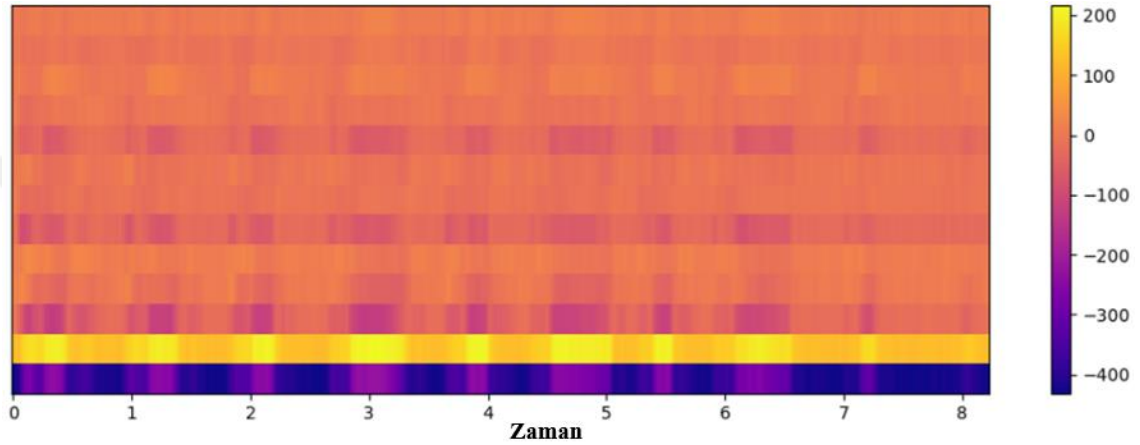
Şekil 3.20. Mel-spektrogram örneği

Mel-spektrogramlar, ses tanıma sistemlerinde yaygın olarak kullanılır (Humphrey vd., 2012). Ayrıca, çeşitli akustik olayların analizi ve sınıflandırılmasında da etkilidir (Bai vd., 2019). Bu özellikleri sayesinde, konuşma tanıma ve müzik bilgi erişimi gibi alanlarda sıkça tercih edilirler.

3.3.3. Mel-Frekans cepstral katsayıları (MFCK)

MFCK'ler, kısa süreli güç spektrumunu (veya logaritmik spektrumunu) Mel ölçeğinde temsil eden katsayılar setidir. MFCK'ler, ses sinyallerinin insan kulağı tarafından nasıl algılandığını modellemek amacıyla geliştirilmiştir. Bu katsayılar, ses sinyalinin önemli özelliklerini özetler ve cepstral analiz ile elde edilir. Cepstral analiz, sesin zaman-frekans

bileşenlerinin cepstral alanında analiz edilmesiyle gerçekleştirilir. Bu, frekans spektrumunun logaritması alındıktan sonra ters Fourier dönüşümü ile yapılır. Mel ölçeği kullanılarak frekans bileşenleri düzenlenir. Bu da insan işitme sistemine daha yakın bir temsiliyet sağlar. MFCK'ler, ses sinyalinin önemli özelliklerini özetleyen düşük boyutlu bir vektör sağlar, bu da hesaplama açısından verimlidir ve genellikle ilk birkaç MFCK, sesin en belirgin özelliklerini taşır. Bir ses verisine ait MFCK örneği Şekil 3.21'de görülmektedir.



Şekil 3.21. MFCK örneği

MFCK'ler, konuşma tanıma sistemlerinin temel bileşenlerinden biridir ve insan konuşmasının karakteristik özelliklerini etkili bir şekilde yakalar (Davis ve Mermelstein, 1980). Hoparlör tanıma ve doğrulama sistemlerinde de kullanılmaktadır (Reynolds, 1995). Ayrıca, müzik sinyallerinin tür sınıflandırmasında da yaygın olarak kullanılır (Tzanetakis ve Cook, 2002). Bu özellikleri sayesinde, MFCK'ler ses tanıma ve sınıflandırma sistemlerinde geniş bir kullanım alanına sahiptir.

3.4. Ses Sınıflandırma

Ses sınıflandırması, ses sinyallerini belirli kategorilere ayırma işlemidir ve ses işleme alanında temel bir rol oynamaktadır. Bu süreç, ses verilerinin analiz edilmesi, özelliklerinin çıkarılması ve belirli sınıflara atanması adımlarını içerir. Ses sınıflandırması, konuşma tanıma, müzik türü tespiti, ortam sesi analizi gibi birçok alanda yaygın olarak kullanılmaktadır. Ses sınıflandırma işlemi sırasıyla veri toplama ve ön işleme, özellik çıkarma, model eğitimi ve sınıflandırma aşamalarını içermektedir.

Veri toplama ve ön işleme adımları, ses sinyallerinin temizlenmesi ve hazırlanmasını içerir. Bu adımlar, gürültü temizleme, normalizasyon ve sinyal parçalarının belirlenmesi gibi işlemleri kapsar. Bu ön işleme adımları, verilerin kalitesini artırır ve sonraki işlemler için daha uygun hale getirmektedir (Zölzer, 2022).

Özellik çıkarma aşamasında, ses sinyalinin temsilini sağlayan özellikler çıkarılır. Ses özellikleri, genellikle spektrogramlar, Mel-spektrogramlar veya Mel-Frekans Cepstral Katsayıları (MFCK) gibi yöntemlerle elde edilir (Logan, 2000). Bu özellikler, ses sinyalinin frekans ve zaman içindeki dağılımını yakalar ve sınıflandırma işlemi için gerekli bilgiyi sağlamaktadır.

Model eğitimi adımında, ses özellikleri bir sınıflandırma modeli üzerinde eğitilir. Bu model, genellikle derin öğrenme yöntemleri kullanılarak eğitilir ve özellik vektörlerini belirli sınıflara atamak için kullanılır (LeCun vd., 2015). Model eğitimi sürecinde, büyük miktarda etiketli veri kullanılarak modelin doğruluğu artırılmaya çalışılmaktadır.

Sınıflandırma adımında, eğitilen model, yeni gelen ses sinyallerini belirli sınıflara sınıflandırır. Bu adımda, modelin öğrendiği bilgilere dayanarak sinyallerin hangi kategoriye ait olduğu tahmin edilir. Sınıflandırma doğruluğunu artırmak için çeşitli metrikler kullanılabilir ve modelin performansı değerlendirilir.

3.5. Derin Öğrenme ile Ses Tanıma

Geleneksel yöntemlerde ses tanıma için kullanılan özellik çıkarımı, genellikle el ile tasarlanmış özellik çıkarıcılar kullanılarak yapılırken, derin öğrenme yaklaşımı, ses verilerinden otomatik olarak özelliklerin çıkarılmasına olanak tanır. Derin öğrenme ile ses tanıma sürecinde, genellikle evrişimli sinir ağları ve tekrarlayan sinir ağları gibi derin mimariler tercih edilir. Evrişimli sinir ağları, özellikle ses spektrogramları gibi 2 boyutlu verilerle iyi çalışır. Bu mimariler, spektrogramlardan özellikler çıkarırken zaman-frekans ilişkilerini yakalayabilmektedir (LeCun vd., 2015).

Ses sınıflandırması üzerine yapılan çalışmalarda, derin öğrenme yöntemlerinin sıklıkla kullanıldığı görülmektedir. Özellikle, ESA ve TSA gibi derin öğrenme mimarileri, ses sınıflandırmasında yüksek doğruluk oranları sağlamaktadır (Deng vd., 2013). Ayrıca, ses özellik çıkarma tekniklerinin sınıflandırma performansını artırmak için kullanıldığı

görülmektedir. Ses tanıma alanındaki derin öğrenme çalışmaları, özellikle derin ağların karmaşık yapıları ve büyük veri setleri üzerindeki etkili performansları üzerine odaklanmıştır (Feurer ve Hutter, 2019). Bununla birlikte, derin öğrenme modellerinin eğitim süreçlerinin optimizasyonu ve hiperparametre seçimi konuları da derin öğrenme ile ses sınıflandırma başarısını önemli ölçüde etkilemektedir.

4. MATERYAL ve YÖNTEM

4.1. Veri Seti ve Ön İşleme

4.1.1. Veri setinin oluşturulması

Çalışma kapsamında, 50 gönüllüden altı farklı sınıfa ait ses kayıtları elde edilmiştir. Bu ses kayıtları ("acil", "imdat", "kimse yok mu", "kurtarın", "yaralıyım" ve "yardım"), telsiz cihazları ve diğer gerekli ekipmanlar kullanılarak toplanmıştır. Ses sinyalleri, telsiz cihazından bilgisayara bağlı bir antene iletilmiş ve bilgisayara ulaşan sinyaller, özel olarak tasarlanmış bir program aracılığıyla ses verisi olarak kaydedilmiştir. Verilerin elde edilme sürecinde, her bir gönüllüye, farklı çevresel gürültü koşulları altında acil durum ifadelerini birçok kez tekrarlatarak ses kayıtları alınmıştır.

Bu yöntem sayesinde, gerçek hayatta karşılaşılabilecek çeşitli senaryolar simüle edilerek veri setinin çeşitliliği artırılmıştır. Elde edilen ses veri sayıları Tablo 4.1'de görülmektedir.

Tablo 4.1. Veri seti eleman sayıları

Sınıf Adı	Eğitim Verisi	Test Verisi	Toplam
acil	621	174	795
imdat	674	173	847
kimse yok mu	607	147	754
kurtarın	737	172	909
yaralıyım	662	156	818
yardım	697	178	875
Toplam	**3.998**	**1.000**	**4.998**

Toplamda 4998 ses kaydından oluşan bu veri seti, modelin eğitim ve test aşamalarında %80 eğitim ve %20 test verisi olarak kullanılmıştır. Eğitim verileri modelin öğrenme sürecinde, test verileri ise modelin performansını değerlendirmek amacıyla kullanılmıştır.

4.1.2. Veri ön işleme

Çalışmada ilk olarak, ses dosyaları mel-spektrogramlarına dönüştürülmüştür. Mel-frekans spektrogramları, sesin zaman-frekans temsiline dayalı bir görüntüsünü oluşturur ve bu dönüşüm, belirli bir kütüphane içerisindeki bir fonksiyon aracılığıyla gerçekleştirilir. Bu spektrogramlar, sesin zaman ve frekans boyunca dağılımını temsil eden önemli özelliklerin çıkarılmasını sağlamaktadır.

Sonraki adımda, mel-spektrogramları belirli bir maksimum uzunluğa doldurulmuş veya bu uzunluktan kesilmiştir. Bu adımın amacı, sinir ağı modeline sabit bir boyut sağlamaktır. Eğer spektrogramun sütun sayısı (zaman boyutu), belirlenen maksimum uzunluktan daha küçükse, spektrogram sıfırlarla doldurulur veya kenarlara sıfır eklenir. Bu işlem, bir dizi işlevin kullanılmasıyla gerçekleştirilir. Eğer spektrogramun sütun sayısı, belirlenen maksimum uzunluktan büyük veya eşitse, sadece belirlenen maksimum uzunluktaki sütunlar alınır. Bu, spektrogramun dilimlenmesi yoluyla sağlanır. Çalışmamızda maksimum uzunluk (sütun sayısı) olarak 200 değeri belirlenmiştir.

4.2. Evrişimli Sinir Ağı Mimarisi

Bu tez çalışmasında ESA tabanlı bir derin öğrenme modeli geliştirilmiş ve eğitim süreci gerçekleştirilmiştir. Bu model, bir ESA olan "sequential" modelini temsil etmektedir. Bu model, 1 giriş katmanı, 8 evrişim katmanı, 4 havuzlama katmanı, 6 unutma katmanı, 1 küresel ortalama havuzlama katmanı, 2 tam bağlı ve 1 çıkış katmanı olmak üzere toplam 23 katmandan oluşmaktadır. Bu model toplamda 1.436.134 parametre içermektedir.

Tablo 4.2. Geliştirilen modelin katmanlarının yapısı

Katman Adı	İşlem	Giriş Boyutu	Çıkış Boyutu	Parametre Sayısı
Giriş Katmanı		(128, 200, 1)	(128, 200, 1)	
1. Evrişim Katmanı	(3, 3), 32	(128, 200, 1)	(128, 200, 32)	320
2. Evrişim Katmanı	(3, 3), 32	(128, 200, 32)	(128, 200, 32)	9.248
1. Maksimum Havuzlama Katmanı	2×2	(128, 200, 32)	(64, 100, 32)	0
1. Unutma Katmanı	(0.25)	(64, 100, 32)	(64, 100, 32)	0
3. Evrişim Katmanı	(3, 3), 64	(64, 100, 32)	(64, 100, 64)	18.496
4. Evrişim Katmanı	(3, 3), 64	(64, 100, 64)	(64, 100, 64)	36.928
2. Maksimum Havuzlama Katmanı	2×2	(64, 100, 64)	(32, 50, 64)	0
2. Unutma Katmanı	(0.25)	(32, 50, 64)	(32, 50, 64)	0
5. Evrişim Katmanı	(3, 3), 128	(32, 50, 64)	(32, 50, 128)	73.856
6. Evrişim Katmanı	(3, 3), 128	(32, 50, 128)	(32, 50, 128)	147.584
3. Maksimum Havuzlama Katmanı	2×2	(32, 50, 128)	(16, 25, 128)	0
3. Unutma Katmanı	(0.5)	(16, 25, 128)	(16, 25, 128)	0
7. Evrişim Katmanı	(3, 3), 256	(16, 25, 128)	(16, 25, 256)	295.168
8. Evrişim Katmanı	(3, 3), 256	(16, 25, 256)	(16, 25, 256)	590.080
4. Maksimum Havuzlama Katmanı	2×2	(16, 25, 256)	(8, 12, 256)	0
4. Unutma Katmanı	(0.25)	(8, 12, 256)	(8, 12, 256)	0
Küresel Ortalama Havuzlama Katmanı		(8, 12, 256)	256	0
1. Tam Bağlantılı Katman	(0.5)	256	512	131.584
5. Unutma Katmanı	(0.25)	512	512	0
2. Tam Bağlantılı Katman	(0.5)	512	256	131.328
6. Unutma Katmanı	(0.5)	256	256	0
Çıkış Katmanı		256	6	1.542

Giriş katmanı, 128×200 piksel boyutunda ses verisi almaktadır. Ardından, 32 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi gerçekleştirilir ve çıkış boyutu 128×200 piksel boyutunda olmaktadır. Bu işlem toplamda 320 parametreye sahiptir. Bir sonraki adımda tekrar 32 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi gerçekleştirilmektedir ve çıkış boyutu yine 128×200 piksel boyutunda olmaktadır. Bu katman toplamda 9.248 parametre içermektedir.

Önceki evrişim katmanının çıkışı olan 128×200 piksel boyutundaki ses verisi, 2×2 boyutunda bir havuzlama işlemine tabi tutulmaktadır ve çıkış boyutu 64×100 boyutunda olmaktadır. Ağın ezberlemesini önlemek için %25 oranında bir unutma işlemi uygulanmaktadır.

Daha sonra, 64 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi gerçekleştirilmektedir ve çıkış boyutu 64×100 piksel boyutunda olmaktadır. Bu katman toplamda 18.496 parametre içermektedir. Ardından, tekrar 64 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi gerçekleştirilmektedir ve çıkış boyutu yine 64×100 piksel boyutunda olmaktadır. Bu katman toplamda 36.928 parametre içermektedir.

Önceki evrişim katmanının çıkışı olan 64×100 piksel boyutundaki ses verisi, 2×2 boyutunda bir havuzlama işlemine tabi tutulmaktadır ve çıkış boyutu 32×50 boyutunda olmaktadır. Ağın ezberlemesini önlemek için %25 oranında bir unutma işlemi uygulanmaktadır.

Daha sonra, 128 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi gerçekleştirilmektedir ve çıkış boyutu 32×50 piksel boyutunda olmaktadır. Bu katman toplamda 73.856 parametre içermektedir. Ardından, tekrar 128 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi gerçekleştirilmektedir ve çıkış boyutu yine 32×50 piksel boyutunda olmaktadır. Bu katman toplamda 147.584 parametre içermektedir.

Önceki evrişim katmanının çıkışı olan 32×50 piksel boyutundaki ses verisi, 2×2 boyutunda bir havuzlama işlemine tabi tutulmaktadır ve çıkış boyutu 16×25 piksel boyutunda olmaktadır. Ağın ezberlemesini önlemek için %25 oranında bir unutma işlemi uygulanmaktadır.

Daha sonra, 256 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi uygulanmaktadır ve çıkış boyutu 16×25 piksel boyutunda olmaktadır. Bu katman toplamda 295.168 parametre içermektedir. Ardından, tekrar 256 filtre kullanılarak 3×3 boyutunda bir evrişim işlemi gerçekleştirilmektedir ve çıkış boyutu yine 16×25 piksel boyutunda olmaktadır. Bu katman toplamda 590.080 parametre içermektedir.

Önceki evrişim katmanının çıkışı olan 16×25 piksel boyutundaki ses verisi, 2×2 boyutunda bir havuzlama işlemine tabi tutulmaktadır ve çıkış boyutu 8×12 piksel boyutunda olmaktadır. Ağın ezberlemesini önlemek için %25 oranında bir unutma işlemi uygulanmaktadır.

Daha sonra, 256 kanal üzerinden küresel ortalama havuzlama işlemi uygulanmaktadır ve daha sonra 256 boyutlu bir vektör elde edilmektedir. Bu vektör, 512 nörona sahip bir tam bağlı (dense) katmanın girişi olarak kullanılmaktadır. Bu katman, toplamda 131.584 parametre içermektedir.

Ardından, aşırı öğrenmeyi önlemek için %50 unutma oranı ile bu katmandan çıkan çıktılar alınmaktadır ve giriş boyutuyla aynı olan 512 nörona sahip bir başka tam bağlı katmana iletilmektedir. Unutma işlemi sırasında hiçbir parametre güncellenmediği için bu katmanın parametre sayısı 0 olmaktadır.

Sonraki adımda, yine %50 unutma oranı ile bu katmandan çıkan çıktılar alınmaktadır ve giriş boyutu 256 olan bir başka tam bağlı katmana iletilmektedir. Bu katman da 131.328 parametre içermektedir. Son olarak, bu katmandan çıkan çıktılar, 6 nörona sahip bir çıkış katmanına iletilmektedir. Bu çıkış katmanında sınıflandırma işlemini gerçekleştirilerek, toplamda 1.542 parametre elde edilmektedir.

4.3. Model Eğitimi ve Hiperparametre Ayarları

Geliştirilen derin öğrenme modelinin evrişim katmanlarında aktivasyon fonksiyonu olarak “ReLU”, tam bağlantılı katman olan “Dense” katmanında ise aktivasyon fonksiyonu olarak “Softmax” kullanılmıştır. Modelin derlenmesi sürecinde kayıp fonksiyonu olarak Kategorik Çapraz Entropi, optimizasyon algoritması olarak ADAM fonksiyonu tercih edilmiştir.

Bununla birlikte eğitim tur sayısı (epoch), model eğitiminde veri kümesinin parçalarıyla yapılan her bir iterasyonu temsil eder, her iterasyonda modelin performansı değerlendirilir ve ağırlıklar güncellenerek en uygun ağırlık değerleri hesaplanır (Sevinç ve Kaya, 2021). Oluşturulan modelin eğitilme sürecince "epoch" sayısı 50 olarak belirlenmiştir.

Batch size, bir seferde ağa verilen örnek sayısını belirten bir hiperparametredir (LeCun vd., 1998). Eğitim verileri kümesinden alınan belirli bir örnek grubunun ağı aynı anda eğitmesini ifade etmektedir (Krizhevsky vd., 2012). Batch size'ın seçimi, eğitim sürecinin hızı ve modelin doğruluğu üzerinde önemli bir etkiye sahiptir (Sutskever vd., 2013). Küçük batch size'lar daha fazla bellek kullanır ve güncelleme adımları arasında daha fazla varyans oluşturabilirken, büyük batch size'lar daha az bellek kullanır ve daha kararlı gradyan tahminleri sağlar (Goodfellow vd., 2016). Ancak, büyük batch size'lar genellikle daha fazla bellek kullanır ve daha uzun süre alabilir (Hinton vd., 2012). Geliştirilen ESA modelinin eğitilme sürecince "batch size" sayısı 32 olarak kullanılmaktadır.

Öğrenme oranı, modelin eğitim sürecinde ağırlıkların güncelleme hızını belirleyen bir hiperparametredir. Yüksek öğrenme oranı daha hızlı öğrenmeyi teşvik ederken, düşük öğrenme oranı daha istikrarlı bir öğrenme süreci sağlamaktadır (Ruder, 2016). Tasarladığımız modelde, öğrenme oranı belirtilmemiş ve varsayılan değer kullanılmıştır. Modelin derlenme sürecinde "Adam" optimizasyon algoritması tercih edilmiş ve bu algoritmanın varsayılan öğrenme oranı 0,001 kullanılmıştır.

4.4. Performans Metrikleri

Model performansı değerlendirme sürecinde birçok farklı metrik yer almaktadır. Bu metrikler hesaplanırken temelde dört farklı değişken kullanılır. Bunlar:

DP (Doğru Pozitif): Gerçekte pozitif olan ve model tarafından doğru şekilde pozitif olarak tahmin edilen örneklerdir.

DN (Doğru Negatif): Gerçekte negatif olan ve model tarafından doğru şekilde negatif olarak tahmin edilen örneklerdir.

YP (Yanlış Pozitif): Gerçekte negatif olan ancak model tarafından yanlış şekilde pozitif olarak tahmin edilen örneklerdir.

YN (Yanlış Negatif): Gerçekte pozitif olan ancak model tarafından yanlış şekilde negatif olarak tahmin edilen örneklerdir.

Hassasiyet (Precision)

Hassasiyet, sınıflandırıcının doğru pozitif tahminlerin toplam tahminler içindeki oranını ifade etmektedir. Her sınıf için ayrı bir hassasiyet değeri hesaplanmaktadır (Powers, 2020). Hassasiyet hesaplaması doğru pozitif tahminlerin toplam pozitif tahminlere oranlanmasıyla hesaplanır. Hesaplamaya ait matematiksel ifade denklem 4.1'de görülmektedir.

$$\text{Hassasiyet} = \frac{DP}{DP + YP} \tag{4.1}$$

Duyarlılık (Recall)

Duyarlılık, gerçek pozitiflerin tüm pozitif örnekler içindeki oranını ifade etmektedir. Her sınıf için ayrı bir duyarlılık değeri hesaplanır (Powers, 2020). Bu metrik, gerçek pozitiflerin (DP) model tarafından doğru bir şekilde sınıflandırılan pozitif örneklerin toplam pozitif örnekler içindeki oranını ifade eder. Hesaplamaya ait matematiksel ifade denklem 4.2'de görülmektedir.

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \tag{4.2}$$

F1 Skoru (F1-Score)

F1 skoru, hassasiyet (precision) ve duyarlılık (recall) arasında bir denge sağlayan bir ölçüdür. Hassasiyet ve duyarlılık arasında bir denge kurar ve bu nedenle dengesiz sınıf dağılımlarında daha güvenilir bir performans ölçüsüdür. F1-score, harmonik ortalama kullanılarak hesaplanır (Powers, 2020). Bu metrik, hassasiyet ve duyarlılık arasında bir denge sağlayan bir ölçüdür ve bu iki metriğin harmonik ortalaması kullanılarak hesaplanır. Harmonik ortalama, en küçük değerlerin etkisini daha fazla öne çıkarır, bu da F1-score'un hem hassasiyetin hem de duyarlılığın düşük olduğu durumları

cezalandırmasını ve her iki metriğin de yüksek olduğu durumları ödüllendirmesini sağlar. Hesaplamaya ait matematiksel ifade denklem 4.3'te görülmektedir.

$$\text{F1 Skoru} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.3}$$

Destek (Support)

Destek, her bir sınıf için gerçek değerlerin sayısını ifade eder. Bu ifade, her sınıfın veri setinde ne kadar temsil edildiğini gösterir.

Doğruluk (Accuracy)

Doğruluk, modelin doğru olarak sınıflandırdığı örneklerin toplam örnekler içindeki oranını ifade eder. Doğruluk metriği genellikle basit bir performans ölçüsü olarak kullanılır, ancak dengesiz sınıf dağılımlarında yanıltıcı olabilir (Japkowicz ve Shah, 2011). Doğru pozitifler ve doğru negatiflerin toplamını tüm pozitif ve negatif örneklerin toplamına bölerek ifade eden bir metriktir. Hesaplamaya ait matematiksel ifade denklem 4.4'te görülmektedir.

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \tag{4.4}$$

Özgüllük (Specificity)

Özgüllük, bir testin negatif sonuçları doğru bir şekilde tanımlama yeteneğini ölçer. Özgüllük, negatif sınıfa ait örneklerin doğru bir şekilde tanımlandığı orandır. Özgüllük, 0 ile 1 arasında bir değer alır. Yüksek bir özgüllük değeri, testin yanlış pozitif sonuç üretme olasılığının düşük olduğunu gösterir.

$$\text{Özgüllük} = \frac{DN}{DP + YP} \tag{4.5}$$

Makro ortalama (Macro Avg)

Makro ortalama, her bir sınıf için hesaplanan metrik değerlerinin, sınıfa ait destek (support) sayısına göre eşit ağırlıklı olarak ortalaması alınarak hesaplanır. Bu, her bir

sınıfın performansının eşit olarak değerlendirildiği bir ortalama sağlar (Takahashi vd., 2022). Hesaplamaya ait matematiksel ifade denklem 4.5'te görülmektedir.

$$Macro\ Avg = \frac{1}{n}\sum_{i=1}^{n} m_i \tag{4.6}$$

Bu formülde , n sınıf sayısını ve m_i ise her sınıf için hesaplanan metrik değerini temsil etmektedir.

Ağırlıklı ortalama (Weighted Avg)

Ağırlıklı ortalama, her bir sınıfın desteğine (support) göre ağırlıklandırılmış metrik değerlerinin ortalaması alınarak hesaplanır. Bu, her sınıfın performansının, sınıfın veri setinde ne kadar temsil edildiğine bağlı olarak ağırlıklı olarak hesaplandığı bir ortalama sağlar (Sokolova ve Lapalme, 2009). Hesaplamaya ait matematiksel ifade denklem 4.6'da görülmektedir.

$$\text{Weighted } Avg = \frac{1}{n}\sum_{i=1}^{n} m_i \frac{\sum_{i=1}^{n}(s_i \times m_i)}{\sum_{i=1}^{n} s_i} \tag{4.7}$$

Bu formülde , n sınıf sayısını, m_i her sınıf için hesaplanan metrik değerini ve s_i ise her sınıfın desteğini temsil etmektedir.

4.5. Kullanılan Araçlar ve Teknolojiler

Tez kapsamında geliştirilen derin öğrenme modelinin uygulanması için Python programlama dilinin 3.11.9 versiyonu kullanılmıştır. Kullanılan bilgisayarda ise i7-13620H 2.40 GHz işlemci, 32GB DDR5 RAM, 64 Bit Windows 11 işletim sistemi, NVIDIA GeForce RTX 4050 2,40 GHz DDR6 6GB ekran kartı özellikleri mevcuttur.

Ses verilerin elde edilme sürecinde telsiz cihazı olarak TYT-UV99 model telsiz cihazı kullanılmıştır. Bu cihaz frekans aralığı VHF için 136-174 MHz, UHF için 400-470 MHz; güç seviyeleri VHF/UHF için sırasıyla 2W/5W; kanal kapasitesi 200 kanal; alıcı duyarlılığı VHF için <-120 dBm, UHF için <-122 dBm; ara frekans 38.0 MHz; ara frekans bant genişliği VHF için 12.5 kHz, UHF için 25 kHz; modülasyon türleri FM, AM, NFM, WFM; verici gücü VHF/UHF için sırasıyla 2W/5W; maksimum sapma VHF/UHF için

sırasıyla 5 kHz/10 kHz; yan bant bastırma VHF/UHF için sırasıyla >60 dB/>60 dB özelliklerine sahiptir.

Kullanılan telsiz cihazı ile üretilen ses verileri “EchoLink” adlı uygulama ile kaydedilmiştir. Ses verilerinin bilgisayara iletilmesi için EchoLink arabirim devresi kullanılmıştır. EchoLink arabirim devresi, genellikle telsiz cihazı ile bilgisayar arasında ses sinyallerini iletmek için kullanılan bir donanım bileşenidir. Bu devreler, genellikle analog ses sinyallerini dijital formata dönüştüren analog-dijital dönüştürücüler içerir. Ayrıca, bilgisayarın ses kartına bağlanabilmesi için uygun bağlantı noktaları da bulunmaktadır. Bu arabirim devreleri genellikle USB veya seri bağlantı gibi standart arabirimlerle bilgisayara bağlanır ve EchoLink uygulamasının kullanımını mümkün kılmaktadır.

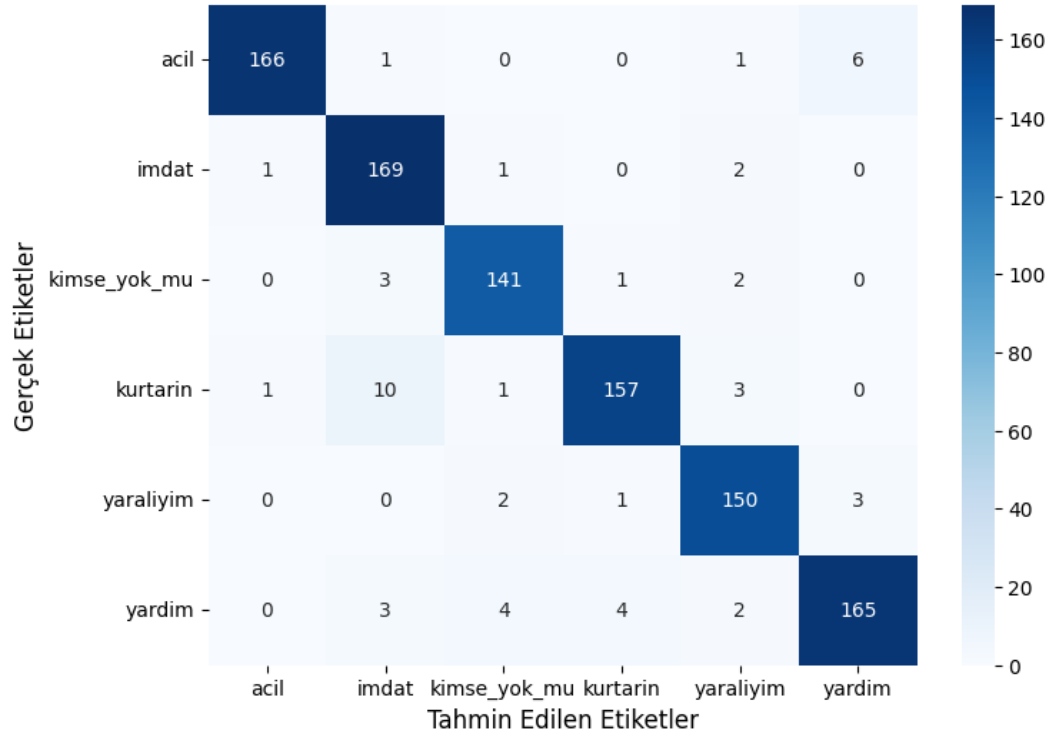
5. ARAŞTIRMA BULGULARI

Bu tez çalışmasında 6 farklı ses türünü sınıflandırabilen bir evrişimli sinir ağı modeli geliştirilmiştir. Bu modelin sınıflandırma performansı, telsiz cihazı kullanılarak 50 gönüllüden elde edilen ses kayıtları ile test edilmiştir. Testte, 6 farklı sınıfa ait toplam 4.998 ses verisi kullanılmıştır. Bu ses verilerinin %80'i eğitim ve %20'si test verisi olarak kullanılmıştır. Eğitim verisi olarak ayrılan kısım, modelin öğrenme sürecini gerçekleştirmesi için kullanılmış ve modelin parametreleri bu verilerle optimize edilmiştir. Test verisi olarak ayrılan kısım ise, modelin genel performansını değerlendirmek için kullanılmıştır. Modelin sınıflandırma başarısı, doğru sınıflandırılan ses kayıtlarının toplam test verisi içindeki oranı ile ölçülmüştür.

Modelin başarısının ölçülmesi için yapılan işlemlerle elde edilen deneysel sonuçlara bakıldığında modelin doğruluk oranının % 96,40 olduğu görülmüştür. Geliştirilen modele ait performans değerlendirme sonuçları Tablo 5.1'te sunulmuş, performans değerlendirme sürecinde elde edilen karmaşıklık matrisi ise Şekil 5.1'de gösterilmiştir.

Tablo 5.1. Geliştirilen modele ait performans değerlendirme sonuçları

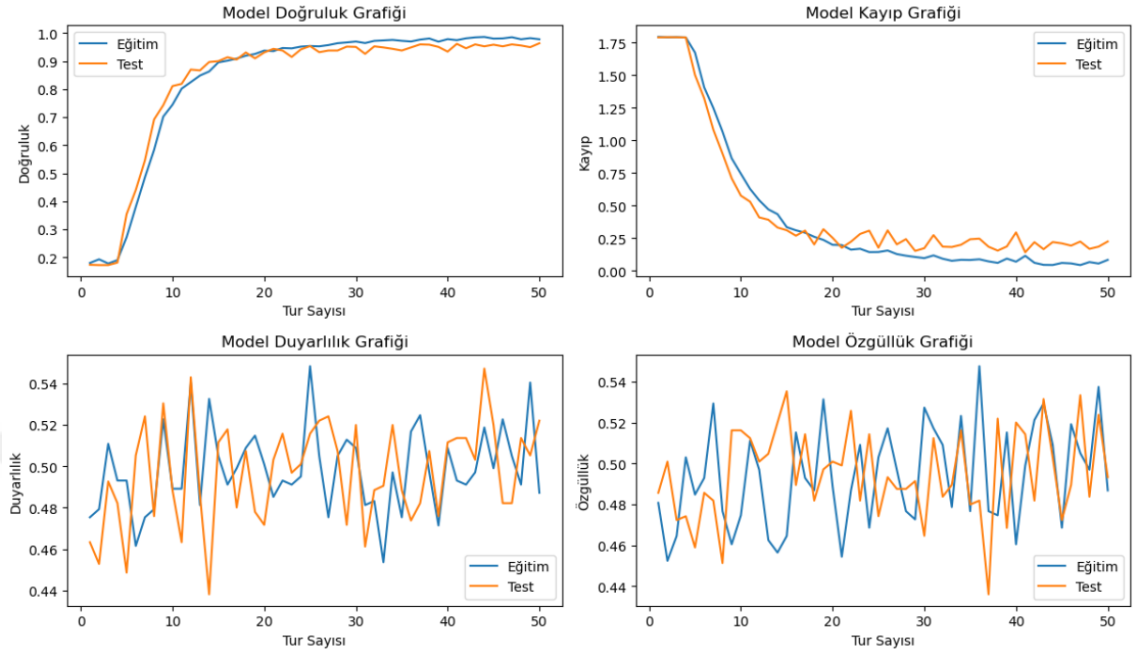
Veri Sınıfı	Precision	Recall	F1-Score	Support
acil	0,99	0,96	0,98	174
imdat	0,96	0,97	0,97	173
kimse yok mu	0,99	0,97	0,98	147
kurtarın	0,96	0,97	0,97	172
yaralıyım	0,94	0,96	0,95	156
yardım	0,94	0,95	0,95	178
Accuracy			0,96	1.000
Macro Avg	0,96	0,96	0,96	1.000
Weighted Avg	0,96	0,96	0,96	1.000



Şekil 5.1. Geliştirilen modele ait karmaşıklık matrisi sonuçları

Geliştirilen modelde eğitim işleminin başarısını değerlendirmek için doğruluk, kayıp, duyarlılık ve özgüllük gibi önemli metrikler kullanılmıştır. Bu metriklerin değişimini görselleştiren grafikler, modelin eğitim sürecindeki performansını detaylı bir şekilde analiz etmemize olanak sağlamaktadır. Doğruluk grafiği, modelin doğru sınıflandırma oranını eğitim ve test veri setlerine göre gösterirken, kayıp grafiği ise modelin ne kadar hızlı öğrendiğini ve ne zaman durağanlaştığını göstermektedir. Duyarlılık ve özgüllük grafikleri ise modelin pozitif ve negatif sınıfları ne kadar doğru tanımladığını göstermektedir. Bu grafiklerin birlikte analizi, modelin aşırı uyum veya yetersiz uyum gibi sorunları olup olmadığını değerlendirmemize, iyileştirme gerektiren alanları belirlememize ve genel olarak modelin güvenilirliğini sağlamamıza yardımcı olmaktadır.

Şekil 5.2'de sunulan grafikler, modelin performansını kapsamlı bir şekilde değerlendirmemize olanak sağlamaktadır.



Şekil 5.2. Doğruluk, Kayıp, Duyarlılık ve Özgüllük değişim grafikleri

Modelin doğruluk ve kayıp grafikleri incelendiğinde, eğitim doğruluğunun 40. epoch sayısında %98 seviyesine ulaştığı ve test doğruluğunun %95-97 aralığında sabit kaldığı görülmektedir. Bu, modelin eğitim verileri üzerinde oldukça başarılı olduğunu ve genelleme yeteneğinin iyi olduğunu göstermektedir. Eğitim kaybı ise başlangıçta 1,75 seviyelerinden hızlıca düşerek 50. epoch sayısında 0,1'in altına inmiştir. Test kaybı da benzer bir düşüş göstererek, eğitim kaybına yakın seviyelere inmektedir, bu da aşırı uyum probleminin olmadığını göstermektedir.

Duyarlılık ve özgüllük grafikleri incelendiğinde, eğitim duyarlılığının 0,44 ile 0,56 arasında dalgalandığı, test duyarlılığının ise eğitim duyarlılığına yakın seyrettiği gözlemlenmektedir. Benzer şekilde, eğitim özgüllüğü 0,46 ile 0,52 arasında dalgalanmakta, test özgüllüğü ise eğitim özgüllüğüne yakın değerlere sahiptir. Bu dalgalanmalar genellikle kabul edilebilir seviyelerdedir ve modelin pozitif ve negatif sınıfları ayırt etme kapasitesinin yüksek olduğunu göstermektedir.

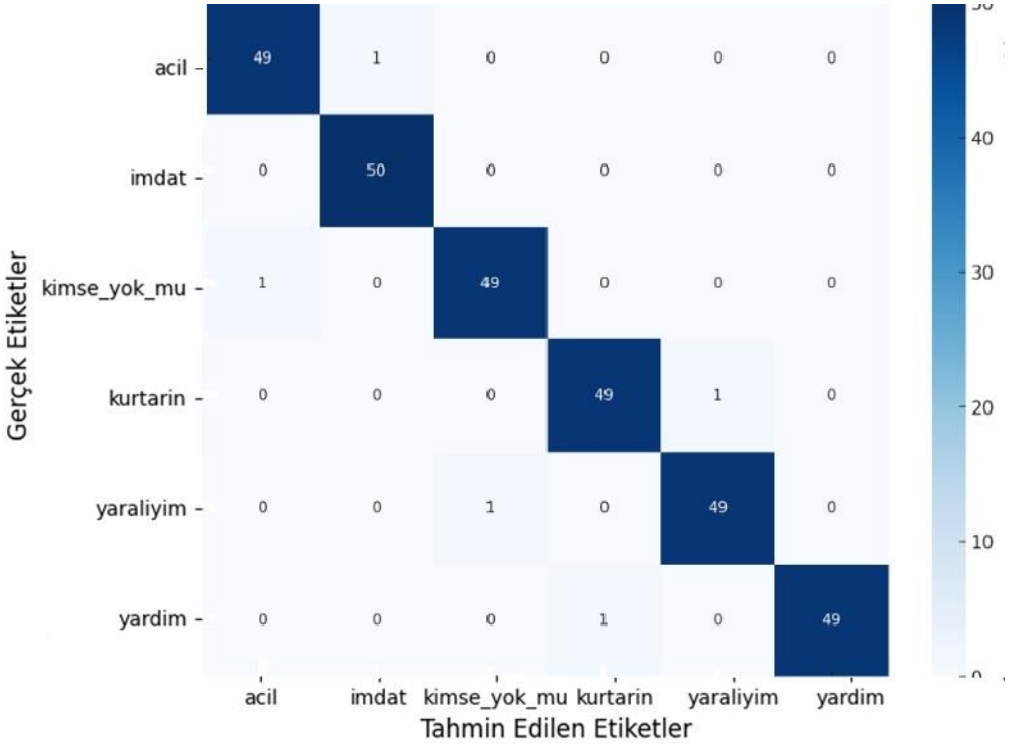
Genel olarak, modelin hem doğruluk hem de kayıp değerlerinde iyi bir performans sergilediği ve aşırı uyum probleminin olmadığı görülmektedir. Modelin genelleme yeteneğinin iyi olduğunu ve hem pozitif hem de negatif sınıfları ayırt edebilme kapasitesinin yüksek olduğunu görülmektedir.

5.1. Modelin Gerçek Hayattaki Başarımının İncelenmesi

Geliştirilen modelin gerçek hayatta kullanılabilmesi için eğitim ve test verileri dışında, sonradan elde edilen her sınıfa ait 50'şer adet yeni ses verisi kullanılarak performans testi yapılmıştır. Bu test sürecinde, modelin performansını ve sınıflandırma yeteneğini değerlendirmek amacıyla her bir sınıf için toplanan 50'şer adet yeni ses verisi model tarafından sınıflandırılmıştır. Modelin tahminleri ve gerçek sınıf etiketleri karşılaştırılarak %97,67 oranında doğruluk elde edilmiştir. Yeni veriler kullanılarak yapılan sınıflandırma sonuçlarına ait performans değerlendirme sonuçları Tablo 5.2'te verilmiş, performans değerlendirme sürecinde elde edilen karmaşıklık matrisi ise Şekil 5.3'de gösterilmiştir.

Tablo 5.2. Performans değerlendirme sonuçları

Veri Sınıfı	Precision	Recall	F1-Score	Support
acil	0,98	0,98	0,98	50
imdat	1,00	1,00	1,00	50
kimse yok mu	0,96	0,98	0,97	50
kurtarın	1,00	0,96	0,98	50
yaralıyım	0,94	0,98	0,96	50
yardım	0,98	0,96	0,97	50
Accuracy			0,98	300
Macro Avg	0,98	0,98	0,98	300
Weighted Avg	0,98	0,98	0,98	300



Şekil 5.3. Performans test sonuçlarına ait karmaşıklık matrisi sonuçları

6. SONUÇLAR

Acil durum iletişimi, hayati bilgi akışının hızlı ve doğru bir şekilde sağlanmasını gerektiren kritik bir alandır. Geleneksel iletişim yöntemlerinin başarısız olduğu afet durumlarında, telsiz iletişimi bağımsız yapısı sayesinde güvenilir bir araç olarak öne çıkmaktadır. Ancak telsiz iletişiminde karşılaşılan en büyük sorun, çevresel gürültüler ve cihaz içsel gürültüleri nedeniyle seslerin anlaşılabilirliğinin azalmasıdır. Bu durum, acil durum ekiplerinin koordinasyonunu ve etkinliğini ciddi şekilde olumsuz etkileyebilmektedir.

Bu tez çalışmasında, acil durum telsiz iletişimindeki gürültülü ses verilerinin tanımlanması için derin öğrenme tabanlı yeni bir model geliştirilmiştir. Modelin eğitimi ve test edilmesi için altı farklı acil durum ses verisi ("acil", "imdat", "kimse yok mu", "kurtarın", "yaralıyım" ve "yardım") içeren bir veri seti hazırlanmıştır. Bu veri seti, çeşitli çevresel gürültü koşulları altında elde edilen telsiz kayıtları kullanılarak oluşturulmuştur.

Geliştirilen model, ses sınıflandırma görevinde yüksek bir başarı elde etmiş ve %96,40 doğruluk oranıyla altı farklı acil durum ses verisini tanımlamada oldukça etkili olmuştur. Modelin gerçek hayatta elde edilen veriler üzerindeki performansı da benzer şekilde etkili olup, %97,67 başarı oranı ile yüksek bir tutarlılık sergilemiştir. Bu yüksek başarı oranları, modelin ses türlerini doğru bir şekilde tanımlama yeteneğinin etkili olduğunu ortaya koymaktadır.

Bu tez çalışmasının sonuçları, acil durum telsiz iletişiminin etkinliğini ve güvenilirliğini artırma potansiyeline sahiptir. Gürültülü ortamlar için derin öğrenme tabanlı bu yaklaşım, ses verilerinin doğru ve hızlı bir şekilde tanımlanmasını sağlayarak acil durum ekiplerinin koordinasyonunu ve bilgi akışını iyileştirmektedir. Ayrıca, bu çalışma ile derin öğrenme yöntemlerinin gürültülü ses tanımlama alanındaki potansiyelini de ortaya konmuş ve gelecekteki araştırmalara yol gösterici bir nitelik kazanmıştır.

Geliştirilen modelin yüksek performans sergilemesi, temel mimarisinin ve özelliklerinin etkinliğini de göstermektedir. Veri artırma veya gürültü azaltma teknikleri kullanılmadan elde edilen bu yüksek başarı oranları, modelin doğru özellikleri öğrenme yeteneği ve karmaşıklığı ele alma kabiliyetinin güçlü olduğunu göstermektedir. Bu durum, geliştirilen

modelin gelecekte daha karmaşık ve gerçekçi senaryolarda da etkin bir şekilde kullanılabileceğini göstermektedir.

Sonuç olarak, bu tez çalışması ile geliştirilen derin öğrenme tabanlı model, acil durum telsiz iletişiminde önemli bir ilerleme sağlamış ve bu alanda yaşanan zorlukların üstesinden gelme konusunda değerli bir katkı sunmuştur. Gelecekte, modelin farklı acil durum senaryolarında test edilmesi ve performansının daha da iyileştirilmesi için ek araştırmalar yapılabilir. Ayrıca, modelin farklı dil ve aksanlarda da etkili bir şekilde çalışabilmesi için uyarlamalar yapılabilir. Bu tür geliştirmeler, acil durum iletişiminde daha geniş bir uygulama yelpazesi sunarak, kriz anlarında daha güvenilir ve etkin bir iletişim sağlamaya yardımcı olacaktır.

KAYNAKLAR

Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., Deng, L., Penn, G. and Yu, D. (2014). "Convolutional neural networks for speech recognition". ***IEEE/ACM Transactions on audio, speech, and language processing***, *22*(10), 1533-1545.

Afan, H. A., Ibrahem Ahmed Osman, A., Essam, Y., Ahmed, A. N., Huang, Y. F., Kisi, O., Sherif, M., Sefelnasr, A., Chau, K.-w. and El-Shafie, A. (2021). "Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques". ***Engineering Applications of Computational Fluid Mechanics***, *15*(1), 1420-1439.

Ahmed, M. J. (2020). "*Çağrı merkezleri için derinöğrenme tabanlı interaktif konuşma tanıma*". ***Selçuk Üniversitesi Fen Bilimleri Enstitüsü***.

Al-Allaf, O. (2015). "Removing noise from speech signals using different approaches of artificial neural networks". ***Int. J. Inf. Technol. Comput. Sci***, *7*, 8-18.

Albawi, S., Mohammed, T. A. and Al-Zawi, S. (2017). "Understanding of a convolutional neural network". ***2017 international conference on engineering and technology (ICET)***,

Altıparmak, F., Akyön, F. Ç., Özmen, E., Çoğun, F. and Bayri, A. (2019). "Towards cognitive sensing: Radar function classification using multitask learning". ***2019 27th Signal Processing and Communications Applications Conference (SIU)***,

Amari, S.-i. (1993). "Backpropagation and stochastic gradient descent method". ***Neurocomputing***, *5*(4-5), 185-196.

Bai, X., Du, J., Wang, Z.-R. and Lee, C.-H. (2019). "A Hybrid Approach to Acoustic Scene Classification Based on Universal Acoustic Models". ***Interspeech***,

Barsties, B. and De Bodt, M. (2015). "Assessment of voice quality: Current state-of-the-art". ***Auris Nasus Larynx***, *42*(3), 183-188.

Bieder, F., Sandkühler, R. and Cattin, P. C. (2021). "Comparison of methods generalizing max-and average-pooling". ***arXiv preprint arXiv:2103.01746***.

Bishop, C. M. (1995). "Neural networks for pattern recognition". ***Oxford university press***.

Cheng, K. W., Chow, H. M., Li, S. Y., Tsang, T. W., Ng, H. L. B., Hui, C. H., Lee, Y. H., Cheng, K. W., Cheung, S. C. and Lee, C. K. (2023). "Spectrogram-based classification on vehicles with modified loud exhausts via convolutional neural networks". ***Applied Acoustics***, *205*, 109254.

Choi, H., Yang, H., Lee, S. and Seong, W. (2019). "Classification of inter-floor noise type/position via convolutional neural network-based supervised learning". ***Applied Sciences***, *9*(18), 3735.

Cunha, B. Z., Droz, C., Zine, A.-M., Foulard, S. and Ichchou, M. (2023). "A review of machine learning methods applied to structural dynamics and vibroacoustic". ***Mechanical Systems and Signal Processing***, *200*, 110535.

Davis, N. and Suresh, K. (2018). "Environmental sound classification using deep convolutional neural networks and data augmentation". ***2018 IEEE recent advances in intelligent computational systems (RAICS)***,

Davis, S. and Mermelstein, P. (1980). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences". ***IEEE transactions on acoustics, speech, and signal processing***, *28*(4), 357-366.

Deng, L., Hinton, G. and Kingsbury, B. (2013). "New types of deep neural network learning for speech recognition and related applications: An overview". ***2013 IEEE international conference on acoustics, speech and signal processing***,

Duchi, J., Hazan, E. and Singer, Y. (2011). "Adaptive subgradient methods for online learning and stochastic optimization". ***Journal of machine learning research***, *12*(7).

Elyousseph, H. and Altamimi, M. L. (2021). "Deep learning radio frequency signal classification with hybrid images". ***2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)***,

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. and Thrun, S. (2017). "Dermatologist-level classification of skin cancer with deep neural networks". ***Nature***, *542*(7639), 115-118.

Feurer, M. and Hutter, F. (2019). "Hyperparameter optimization". ***Automated machine learning: Methods, systems, challenges***, 3-33.

Fırıldak, K. ve Talu, M. F. (2019). "Evrişimsel sinir ağlarında kullanılan transfer öğrenme yaklaşımlarının incelenmesi". ***Computer Science***, *4*(2), 88-95.

Fletcher, H. and Munson, W. A. (1933). "Loudness, its definition, measurement and calculation". ***Bell System Technical Journal***, *12*(4), 377-430.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). "Deep learning". ***MIT press***.

Graves, A. (2013). "Generating sequences with recurrent neural networks". ***arXiv preprint arXiv:1308.0850***.

Graves, A., Mohamed, A.-r. and Hinton, G. (2013). "Speech recognition with deep recurrent neural networks". ***2013 IEEE international conference on acoustics, speech and signal processing***,

He, K., Zhang, X., Ren, S. and Sun, J. (2016). "Deep residual learning for image recognition". ***Proceedings of the IEEE conference on computer vision and pattern recognition***,

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P. and Sainath, T. N. (2012). "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". ***IEEE Signal processing magazine***, *29*(6), 82-97.

Humphrey, E. J., Bello, J. P. and LeCun, Y. (2012). "Moving beyond feature design: Deep architectures and automatic feature learning in music informatics". ***Ismir***,

Japkowicz, N. and Shah, M. (2011). "Evaluating learning algorithms: a classification perspective". ***Cambridge University Press***.

Jin, S., Wang, X., Du, L. and He, D. (2021). "Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN". ***Applied Acoustics***, *172*, 107562.

Kemaloğlu, N. ve Sevli, O. (2019). "Evrişimsel Sinir Ağları ile İşaret Dili Tanıma". ***Proceedings on 2nd International Conference on Technology and Science***,

Kingma, D. P. and Ba, J. (2014). "Adam: A method for stochastic optimization". ***arXiv preprint arXiv:1412.6980***.

Kinsler, L. E., Frey, A. R., Coppens, A. B. and Sanders, J. V. (2000). "Fundamentals of acoustics". ***John wiley & sons***.

Kong, C., Chang, J.-T., Li, Y.-F. and Chen, R.-Y. (2020). "Deep learning methods for super-resolution reconstruction of temperature fields in a supersonic combustor". ***AIP Advances***, *10*(11).

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks". ***Advances in neural information processing systems***, *25*.

Kütükcü, A. Z. ve Eren, T. (2017). "Acil Durum Haberleşmesinde Kullanılan El Telsizinin Çok Ölçütü Karar Verme Yöntemleri İle Seçilmesi". ***Selçuk Üniversitesi Mühendislik, Bilim ve Teknoloji Dergisi***, *5*(2), 183-203.

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

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). "Gradient-based learning applied to document recognition". ***Proceedings of the IEEE***, *86*(11), 2278-2324.

Li, Z., Nash, W., O'Brien, S., Qiu, Y., Gupta, R. and Birbilis, N. (2022). "cardiGAN: A generative adversarial network model for design and discovery of multi principal element alloys". ***Journal of Materials Science & Technology***, *125*, 81-96.

Lim, M., Lee, D., Park, H., Kang, Y., Oh, J., Park, J.-S., Jang, G.-J. and Kim, J.-H. (2018). "Convolutional Neural Network based Audio Event Classification". ***KSII Transactions on Internet & Information Systems***, *12*(6).

Logan, B. (2000). "Mel frequency cepstral coefficients for music modeling". ***Ismir***,

Lopac, N., Hržić, F., Vuksanović, I. P. and Lerga, J. (2021). "Detection of non-stationary GW signals in high noise from Cohen's class of time–frequency representations using deep learning". ***IEEE access***, *10*, 2408-2428.

McCulloch, W. S. and Pitts, W. (1943). "A logical calculus of the ideas immanent in nervous activity". ***The bulletin of mathematical biophysics***, *5*, 115-133.

Müller, M. (2015). "Fundamentals of music processing: Audio, analysis, algorithms, applications (Vol. 5)". ***Springer***.

Myers, D. and Hutchinson, R. (1989). "Efficient implementation of piecewise linear activation function for digital VLSI neural networks". ***Electronics Letters***, *25*, 1662.

Özer, I., Özer, Z. and Fındık, O. (2018). "Noise robust sound event classification with convolutional neural network". ***Neurocomputing***, *272*, 505-512.

Park, G. and Lee, S. (2020). "Environmental noise classification using convolutional neural networks with input transform for hearing aids". ***International journal of environmental research and public health***, *17*(7), 2270.

Piczak, K. J. (2015). "Environmental sound classification with convolutional neural networks". ***2015 IEEE 25th international workshop on machine learning for signal processing (MLSP)***,

Powers, D. M. (2020). "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation". ***arXiv preprint arXiv:2010.16061***.

Rabiner, L. R. and Juang, B.-H. (1999). "Fundamentals of speech recognition". ***Tsinghua University Press***.

Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M. and Suri, J. S. (2006). "Heart rate variability: a review". ***Medical and biological engineering and computing***, *44*, 1031-1051.

Rakshit, R. D., Kisku, D. R., Gupta, P. and Sing, J. K. (2021). "Cross-resolution face identification using deep-convolutional neural network". ***Multimedia Tools and Applications***, *80*, 20733-20758.

Reynolds, D. A. (1995). "Speaker identification and verification using Gaussian mixture speaker models". ***Speech communication***, *17*(1-2), 91-108.

Rojas, O. C., Montoya, M. H. and Contero, M. (2023). "Applıcatıon Of A Predıctıve Model Created Wıth Deep Learnıng To Increase Student Retentıon: A Comparatıve Study". ***INTED2023 Proceedings***,

Rosenblatt, F. (1958). "The perceptron: a probabilistic model for information storage and organization in the brain". ***Psychological review***, *65*(6), 386.

Ruder, S. (2016). "An overview of gradient descent optimization algorithms". ***arXiv preprint arXiv:1609.04747***.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). "Learning representations by back-propagating errors". ***Nature***, *323*(6088), 533-536.

Salamon, J. and Bello, J. P. (2017). "Deep convolutional neural networks and data augmentation for environmental sound classification". ***IEEE Signal processing letters***, *24*(3), 279-283.

Schmidhuber, J. (2015). "Deep learning in neural networks: An overview". ***Neural networks***, *61*, 85-117.

Schwarz, A., Huemmer, C., Maas, R. and Kellermann, W. (2015). "Spatial diffuseness features for DNN-based speech recognition in noisy and reverberant environments". ***2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)***,

Sevinç, A. ve Kaya, B. (2021). "Derin öğrenme ve istatistiksel modelleme yöntemiyle sıcaklık tahmini ve karşılaştırılması". ***Avrupa Bilim ve Teknoloji Dergisi***(28), 1222-1228.

Shen, L., Zhang, Q., Cao, G. and Xu, H. (2019). "Fall detection system based on deep learning and image processing in cloud environment. ***Complex, Intelligent, and Software Intensive Systems: Proceedings of the 12th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2018)***,

Shin, H.-k., Park, S. H. and Kim, K.-w. (2020). "Inter-floor noise classification using convolutional neural network". ***Plos one***, *15*(12), e0243758.

Sokolova, M. and Lapalme, G. (2009). "A systematic analysis of performance measures for classification tasks". ***Information processing & management***, *45*(4), 427-437.

Sooksil, N. and Benjaoran, V. (2021). "Non-linear modelling of construction workers' behaviors for accident prediction". ***Songklanakarin Journal of Science & Technology***, *43*(2).

Stabellini, V. (2023). "Impact of Noise on Different Neural Network Architectures for Environmental Sound Classification". ***Politecnico di Torino***

Stevens, S. S. and Davis, H. (1938). "Hearing: its psychology and physiology". ***John Wiley & Sons, Inc.*** New York, London

Sun, M., Song, Z., Jiang, X., Pan, J. and Pang, Y. (2017). "Learning pooling for convolutional neural network". ***Neurocomputing***, *224*, 96-104.

Sun, X., Peng, J., Shen, Y. and Kang, H. (2020). "Tobacco plant detection in RGB aerial images". ***Agriculture***, *10*(3), 57.

Sutskever, I., Martens, J., Dahl, G. and Hinton, G. (2013). "On the importance of initialization and momentum in deep learning". ***International conference on machine learning***,

Syafeeza, A., Khalil-Hani, M., Liew, S. S. and Bakhteri, R. (2015). "Convolutional neural networks with fused layers applied to face recognition". ***International Journal of Computational Intelligence and Applications***, *14*(03), 1550014.

Takahashi, K., Yamamoto, K., Kuchiba, A., and Koyama, T. (2022). "Confidence interval for micro-averaged F_1 and macro-averaged F_1 scores". ***Applied Intelligence***, *52*(5), 4961-4972.

Tieleman, T. (2012). "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude". ***COURSERA: Neural networks for machine learning***, *4*(2), 26.

Tzanetakis, G. and Cook, P. (2002). "Musical genre classification of audio signals". ***IEEE Transactions on speech and audio processing***, *10*(5), 293-302.

Wang, S., Jiang, Y., Hou, X., Cheng, H. and Du, S. (2017). "Cerebral micro-bleed detection based on the convolution neural network with rank based average pooling". ***IEEE access***, *5*, 16576-16583.

Werbos, P. (1974). "Beyond regression: New tools for prediction and analysis in the behavioral sciences". ***PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA***.

Xu, J., Li, Z., Du, B., Zhang, M. and Liu, J. (2020). "Reluplex made more practical: Leaky ReLU". ***2020 IEEE Symposium on Computers and communications (ISCC)***,

Yaman, F. (2020). "Stratejik insan kaynakları yönetimi penceresinden afet ve acil durum yönetiminde iletişim". ***Academic Perspective Procedia***, *3*(2), 920-927.

Yu, D., Wang, H., Chen, P. and Wei, Z. (2014). "Mixed pooling for convolutional neural networks". ***Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings 9***,

Zaman, K. and Direkoğlu, C. (2020). "Classification of Harmful Noise Signals for Hearing Aid Applications using Spectrogram Images and Convolutional Neural Networks." ***2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)***,

Zeiler, M. D. (2012). "Adadelta: an adaptive learning rate method". ***arXiv preprint arXiv:1212.5701***.

Zölzer, U. (2022). "Digital audio signal processing. ***John Wiley & Sons***.

Zulfiqar, R., Majeed, F., Irfan, R., Rauf, H. T., Benkhelifa, E. and Belkacem, A. N. (2021). "Abnormal respiratory sounds classification using deep CNN through artificial noise addition". ***Frontiers in Medicine***, *8*, 714811.