



Doğal dil işleme ve derin öğrenme algoritmaları ile makine dili modellemesi

Modelling the machines' language with natural language processing and deep learning algorithms

Ali Dayan¹, Atıncı Yılmaz^{2*}

¹ Beykent Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği A.B.D., alidayan93@gmail.com
ORCID: <https://orcid.org/0000-0001-5924-2299>

² Beykent Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, atinciyilmaz@beykent.edu.tr
ORCID: <https://orcid.org/0000-0003-0038-7519>

MAKALE BİLGİLERİ

Makale Geçmişi:

Geliş 16 Haziran 2022
Revizyon 6 Temmuz 2022
Kabul 20 Temmuz 2022
Online 30 Eylül 2022

Anahtar Kelimeler:

Makine Dili,
Derin Öğrenme,
Doğal Dil İşleme,
Evrişimsel Sinir Ağları (CNN)

ÖZ

İnsanlar ve tüm canlılar için dil, iletişim konusunda en önemli unsurlardan biridir. Literatürde dillerin oluşumu ve doğal dil işleme süreçleri ile ilgili oldukça fazla sayıda çalışmalar yapılmış; bu çalışmalarda analiz, yaklaşım ve yöntem açısından farklılıklar içerdiği görülmüştür. Çalışmada literatüre katkı olarak makinelerin kendi dilini oluşturabilmesi için bir model önerilmiştir. Bu bakış açısı ile makinelerin kendi aralarında insanlar gibi kendi dilleri ile iletişim kurarak tüm süreçlerin verimliliğinde artış olacağı düşüncesine katkıda bulunulmuştur. Makinelerin kendi dillerini üretebilmesi adına bir yaklaşım geliştirilmesi amacı ile çalışmada evrişimli derin sinir ağları (CNN) yöntemi ile canlıların seslerini sınıflandırarak ayırt etmesi ve yeni sesler üretebilmesi hedeflenmiştir. Çalışmada, kullanılacak olan alfabenin ne olacağı sistem tarafından karar verilmesi sağlanmış; daha sonrasında Yinelemeli Yapay Sinir Ağları, Mel Frekans Cepstral Katsayısı ve Dinamik Zaman Çözgü metodolojileri ile birlikte kullanılarak benzer sesler oluşturulmuş ve canlılar ilgili sesler ile isimlendirilmiştir. Benzer çalışmalardan farklı olarak Kaggle açık veri deposundan "Audio Cats and Dogs" veri setindeki ses dosyaları üzerinden elde edilen MFCC görsel bir veri olarak kullanılmıştır. Bunun yanında CNN ağları ile model desteklenerek daha iyi performans elde edilmesi sağlanmıştır. Çalışmada, önerilen yöntem kullanılarak %91 oranında doğruluk oranına ulaşılmıştır.

ARTICLE INFO

Article history:

Received 16 June 2022
Received in revised form 6 July 2022
Accepted 20 July 2022
Available online 30 September 2022

Keywords:

Machine Language,
Deep Learning,
Natural Language Processing,
Convolutional Neural Networks (CNN)

Doi: 10.24012/dumf.1131565

* Sorumlu Yazar

ABSTRACT

Language is one of the most important elements of communication for people and all living things. A large number of studies have been conducted in the literature on the formation of languages and natural language processing processes. It has been observed that these studies contain differences in terms of analysis, approach and method. In the study, a model has been proposed for machines to create their own language as a contribution to the literature. With this point of view, it has been contributed to the idea that the efficiency of all processes will increase by communicating with each other in their own language like humans. With the aim of developing an approach for machines to produce their own languages, it is aimed to distinguish the sounds of living things by classifying them and to generate new sounds by using convolutional deep neural network (CNN) method. In the study, the applied alphabet was decided by the system. In next step, Regenerative Artificial Neural Networks were used together with Mel Frequency Cepstral Coefficient and Dynamic Time Warp methodologies to create similar sounds and the living things were named with their respective sounds. Unlike similar studies, a visual data was applied which was converted from the audio files that was obtained from Kaggle open data repository "Audio Cats and Dogs" dataset. In addition, the model was supported with CNN networks to achieve effective performance. In the study, an accuracy rate of 91% was achieved by using the proposed method.

Giriş

Canlılar tüm organizasyonların sağlanması adına haberleşmeye ihtiyaç duymaktadır. Her canlı bu durumu gerçekleştirmek adına farklı şekilde yöntem uygular. İnsanlar için iletişimi ve haberleşmeyi sağlayan en etkili yol, dili kullanarak konuşmaktır. Bu beceri yazının icadı, gelişmesi ile kazanılmıştır. Doğal dil işleme (NLP), konuşmaların ve/veya metinlerin işlenerek makineler tarafından anlamlandırılabilme sürecini ifade etmektedir. Tüm doğal diller farklı özelliklere sahip olduğundan ötürü her dilin farklı analize ve yaklaşıma ihtiyacı bulunmaktadır. Makineler ile doğal dil işleme yapılabilmesi için danışmanlı, yarı-danışmanlı, danışmansız ve takviyeli öğrenme yöntemlerinden birinin kullanılması gerekmektedir. Alan Turing'in çalışmasında ortaya koyduğu "Makineler düşünebilir mi?" sorusu doğal dil işleme çalışmalarının başlangıcı olarak sayılmaktadır. Turing "Makine ve Zeka" isimli makalesinde makinelerin ne kadar iyi rasyonel davranış sergileyebildiğini sıyanan Turing Testi'ni anlatmıştır [1]. 60'lı yıllarda ise Terry Winograd SHRDLU isimli doğal dilleri anlayan bir program tasarlamıştır [2]. Searle, Çin Odası isimli deney ile Çince bilgisi olmayan bir makinenin Çince çıktı üretmesini sağlamıştır [3]. Doğal dil işleme ile ilgili çalışmalar, yapay zeka ve makine öğrenmesi yöntemlerinin gelişmesi ve popülerleşmesi ile birlikte daha yaygın hale gelmiştir. Cihazların birbiri ile iletişime geçerek etkileşim halinde olabilmeleri birçok karmaşık işlemi çok daha kolay hale getirebilecektir. Buzdolabının içerisindeki ürünlerin bitme noktasına geldiğini saptadıktan sonra ilgili cihazlar ile iletişime geçerek bitmek üzere olan ürünlerin siparişinin verilmesini sağlaması, bu tip bir sürece örnek olabilecektir. Bu çalışmada, önerilen yapay zeka destekli doğal dil işleme metodolojisi ile ses dosyalarının işlenerek örnek bir vaka çalışması uygulanmıştır. Ayrıca çalışmada, "Doğal Dil İşleme ile Makinelerin Kendi Dilini Modellemesi" isimli çalışmada kullanılan algoritmalar ve yaklaşım bu makalenin ortaya konmasını sağlamıştır [4]. Çalışmada, kedi ve köpek seslerinin farklı varyasyonlarını içeren "Kaggle Audio Cats and Dogs" veri seti kullanılmıştır [5]. Ayrıca, özellik çıkarımının yapılması ve sınıflandırma için Dinamik Zaman Bükme Algoritması (DTW), Yinelemeli Yapay Sinir Ağları (RNN) ve Mel-Frekansı Kepstral Katsayısı (MFCC) yöntemleri çalışmada uygulanmıştır.

Literatürdeki benzer çalışmaların temelinde ses dalgalarının çözümlenerek konuşmadan yazıya dönüştürülmesi çalışmalarının yoğun olarak bulunduğu gözlemlenmiştir. Ayrıca yapılan çalışmalarda odağın daha çok yazım yanlışlarının ortaya konması, kontrol edilmesi ve düzeltilmesi, sesli asistan, dil çevirisi gibi anlamsal ve biçimsel çözümleme alanlarında olduğu görülmektedir.

Rani ve ark., Doğal Dil İşleme (NLP) ve Nesnelerin İnterneti (IoT) Kullanan Ses Kontrollü Ev Otomasyon Sistemi isimli çalışmalarında, artırılmış gerçeklik ve doğal

dil işleme kullanarak ev gereçlerinin birlikte uyum içerisinde çalışmasını sağlayacak, maliyet odaklı tam ses kontrolü sağlayan proje geliştirmişlerdir [6]. Var olan akıllı ev sistemlerinin çoğunun ev gereçlerine sadece açma, kapama komutu verdiğini belirtmişlerdir. Alexakis ve ark., çok katmanlı bir mimari ile doğal dil işleme, ses tanıma ve IoT teknolojilerini kullanarak akıllı ev kontrolü sağlamışlardır [7]. IoT Aracısı, doğal dil işlemeyi (NLP) kullanarak metin veya sesli komutları anlayabilen bir sohbet botu entegre etmektedir. Mass ve ark., cihazdan cihaza iletişimi kolaylaştırmak, mobil ağı gereksiz veri aktarımından kurtarmak, uçtan uca uygulamaların yanıt verme sürecini hızlandırmak ve geleneksel hizmetlerin sağlanmasını merkezi olmayan hale getirmek için potansiyel bir strateji üzerine çalışmışlardır [8]. Almahdi, yaptığı çalışmada; denetimsiz öğrenme mantığını kullanarak dil benzerliğinin otomatik olarak saptanmasını ortaya koymuştur [9]. Toprak, çalışmasında doğal dil işleme yöntemlerini kullanarak Helmholtz Prensibi ve Terim Frekansı - Ters Belge Frekansı (TF-IDF - Term Frequency Inverse Document Frequency) ile İngilizce bir doküman üzerinden İngilizce sözlüğün otomatik şekilde ortaya çıkartılmasını sağlamıştır [10]. Büyükoğuz, çalışmasında haber metinlerini ve kullanıcı yorumlarının sınıflandırılması için doğal dil işleme ile birlikte derin öğrenme yöntemlerini birlikte kullanmıştır [11]. Özcan, seyahat ve turizm alanında doğal dil işleme süreci yöntemlerini kullanarak seyahat asistanı oluşturulmasını çalışmasında hedeflemiştir [12]. Gupta ve ark., görme engelli kişilere yardımcı olunması amacı ile derin öğrenme ve doğal dil işleme tekniklerini kullanarak görüntü girdisine karşılık kısa açıklama çıktısı verilmesini sağlamışlardır [13]. Anand ve ark., sosyal ağlarda rahatsız edici dil kullanımının tespit edilmesi için özellik seçimi ve grup sınıflandırma için derin öğrenme ve doğal dil işleme yöntemlerini birlikte kullanmışlardır [14]. Çıkarılan özellikler, naive bayes ve destek vektör makinelerinin birlikte kullanımı ile ortaya konan hibrit bir model ile sınıflandırılmıştır. Kasthuri ve ark., uzun-kısa süreli bellek algoritması ve doğal dil işleme algoritmaları ile sohbet robotu ortaya koymuşlardır [15]. Chotirat ve ark., konuşma etiket kategorisine dayalı özellik seçimi problemi için lojistik regresyon, naive bayes ve destek vektör makinesi yöntemleri uygulayarak bu yöntemlerin etkinliklerini karşılaştırmıştır [16]. Shahi ve ark., çocukların fiziksel istismarının tespit edebilmek için doğal dil işleme ve derin öğrenme yöntemlerini birlikte kullanmışlardır [17]. Alhogail ve ark., sahte e-postaların tespiti için makine öğrenmesi ve doğal dil işleme yöntemlerini birlikte kullanarak grafiksel evrişim ağı önermişlerdir [18]. Mumcuoğlu ve ark., yüksek mahkeme sonuçlarının tahmini için hukuk alanında doğal dil işleme ile birlikte makine öğrenmesi ve derin öğrenme mimarilerini kullanarak, modellerin performans sonuçları karşılaştırılmıştır [19]. Mendez ve ark., doğal dil işleme ve makine öğrenmesi algoritmalarını birlikte kullanarak finans haberleri üzerine zamansallık tespitini gerçekleştirmişlerdir [20]. Çalışma neticesinde finansal alanda karar verme konusunda önerdikleri yöntemin başarılı bir piyasa taraması yapabileceğini ifade

etmişlerdir. Hodorog ve ark., yerleşim yerlerinde olay tespiti için makine öğrenme algoritmaları ile birlikte doğal işleme yöntemini birlikte kullanmıştır [21]. Çalışmada regresyon analizi ile olaylar ile memnuniyet arasındaki ilişkiler doğrulanırken; en yüksek doğruluk oranı %88.5 ile elde edilmiştir. Takahashi ve ark. derin evrişimli sinir ağları kullanarak akustik olay tanıma için uzun süreli frekans yapısına uygun büyük giriş yapısına sahip bir mimari önermişlerdir [22]. Çalışmada, akustik olay sınıflandırması için 28 olaydan oluşan veri seti kullanılmıştır. Elde edilen verilerin bir kısmı olan kedi ve köpek sesleri ile Kaggle veri deposundaki veri seti meydana getirilmiştir [5]. Önerilen mimari ile veri büyüme olmaksızın 80.3%, veri büyüme ile 92.8% doğruluk oranı performansına ulaşıldığı, çalışmada ifade edilmiştir.

Araştırılan benzer çalışmalar neticesinde, çalışmada ses işleme yöntemleri ile doğal dil işleme metodolojilerinin harmanlanma düşüncesi ile RNN, MFCC ve DTW yöntemlerinin beraber uygulanması hedeflenmiştir. Bu uygulama sonucunda, canlı seslerine karşılık yeni seslerin oluşturulması ve türetilmesi düşünülmüştür. Literatürdeki çalışmalardan farklı olarak ses dosyaları görsel veriye dönüştürülerek evrişimsel sinir ağları yöntemleri ile model desteklenecektir. Bunun yanında seslerin görselleştirilmesinden önce Fourier Dönüşümünden faydalanılarak seslerin ayrıştırılması sağlanacaktır.

İlk olarak benzer çalışmalar incelenerek literatürde yer alan ses işleme ve doğal dil işleme ile ilgili çalışmalar incelenmiştir. Çalışmanın diğer kısımlarında ise yapay zeka ile doğal işleme süreçleri ile birlikte ses işleme adımları açıklanmış; son olarak makinenin kendi alfabetini tercih etmesi ile birlikte özellik seçimi uygulanarak köpek ve kedi sesleri için ses türetilmesini sağlayan modelleme ortaya konmuştur.

Materyal ve Metot

Veri seti

Çalışmada “Kaggle Audio Cats and Dogs” veri seti kullanılmıştır [5, 22]. Veri seti WAV formatındaki ses dosyalarından oluşmaktadır. Veri setinde 164 adet kedi sesi ve 113 adet köpek sesi bulunmaktadır. Kedi seslerinin toplam uzunluğu 1323 saniye ve köpek sesleri ise 598 saniyedir. Tüm ses dosyaları değişken uzunlukta ve 16 KHz ses içermektedir. Çalışmada eğitim ve test seti için K çapraz sorgulama yöntemi uygulanarak K=5 için %70 eğitim, %30 test seti belirlenerek en iyi sonuçlar elde edilmiştir.

Yapay zekâ ve doğal dil İşleme metodları

Yapay zeka, doğadaki herhangi bir rasyonel davranışı benzetim yapmayı amaçlayan; ayrıca içerisinde bir çok farklı yaklaşım ve algoritmayı barındıran yöntemler topluluğudur [23]. Makine Öğrenmesi ve Derin Öğrenme yöntemleri yapay zekanın alt dallarından iki farklı yaklaşımdır. Makine Öğrenmesi ve Derin Öğrenme arasındaki en temel fark derin öğrenmenin öznelilik çıkarımını kendisinin yapabilme becerisidir. Doğal dil işleme (NLP), yapay zeka metodolojilerinin yardımı ile ortaya konan yazılım vasıtası ile ses veya metin gibi dilsel

fonksiyonların analizi ile bir sonuç alınmasını sağlayan bilim dalıdır [24]. Yapay zeka ve doğal dil işleme yöntemlerinin bütünleşik olarak kullanılması Anlamsal Ağ ve Ayrıştırma Ağacı yaklaşımları ile kullanılmaktadır. Anlamsal ağ, cümlelerin anlamsal karşılığına odaklanmaktadır [24]. Ayrıştırma ağacı yaklaşımında ise biçimsel ve dilbilgisel olarak doğru sıralamaya ulaşmak amaçlanmaktadır [25]. Doğal dil işleme sürecindeki temel hedef dilin makine tarafından anlaşılabilir şekilde bu duruma tepki ile cevap verebilmesinin sağlanmasıdır.

Derin Öğrenme

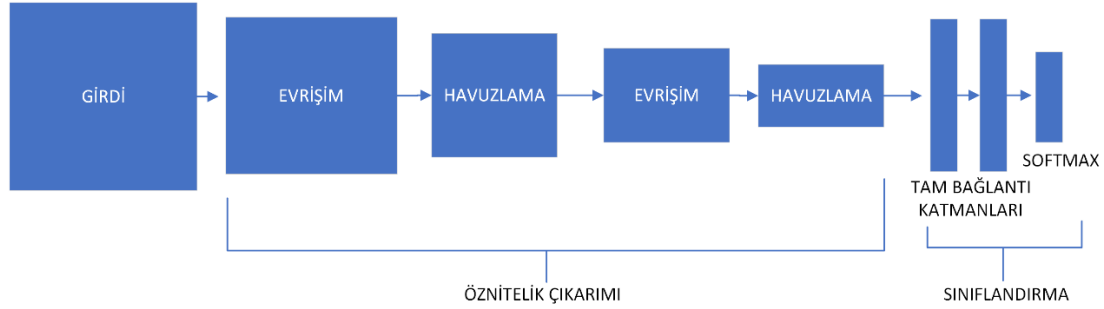
Derin öğrenme, yapay zeka disiplinin altında yer alan yaklaşımlardan biridir. Makine öğrenme algoritmalarında öğrenme süreci içerisinde öznelilik bilgisinin sisteme aktarılması gerekirken, derin öğrenmede farklı olarak öznelilik çıkarımını yöntem kendisi yapmaktadır. Derin öğrenme mimarisi içerisinde en önemsiz özellikler ile başlayarak, katman katman ilerleyerek en önemli özelliklerin çıkarımını ile model öznelilik çıkarımını tamamlamaktadır. Yapısal olarak her katmanda çok katmanlı yapay sinir ağını barındırmaktadır. Bu nedenle öğrenme yeteneği oldukça güçlü; büyük ve karmaşık verilere sahip kompleks problemlerin çözümünde oldukça başarılıdır.

Evrişimsel Sinir Ağları - CNN

Teknolojinin gelişmesi ile birlikte hem verinin yapısı değişmekte hem de bilgisayarların fonksiyonları ve donanımsal gücü artmaktadır. Yapısal olmayan verilerin analizinin çok daha zor olması ile birlikte; video, görüntü gibi yapısal olmayan verilerin başarılı şekilde algılanması, sınıflandırılması Evrişimsel Sinir Ağları (Convolutional Neural Network - CNN) ile mümkündür. Evrişimsel sinir ağı, bir girdi üzerinden görüntü veya nesne tespitini yapabilen derin öğrenme algoritmalarından biridir [26]. Yöntem kullanılarak görüntü içerisinde ortak özellikler yakalanarak kümeleme veya etiket bilgisi üzerinden sınıflandırma yapılabilmektedir. Mimari olarak evrişimsel sinir ağları görüntü ya da videoları girdi verisi halinde modele almaktadır. Bu girdinin modele sunulması için belirli format halinde ve matris formunda olması gerekmektedir. Görüntüye uygulanan filtre ile renkli veya siyah-beyaz olarak görüntü üzerinden veriler elde edilmektedir. Filtrelerin uygulanması neticesinde çıktı olarak farklı yeni matrisler ortaya çıkmaktadır. Elde edilen bu veriler girdi olarak alınarak katmanlarda elde edilen özellik çıkarımları ile modelin tespit neticesi hesaplanmaktadır.

Evrişimli sinir ağları evrişim, havuzlama, tam bağlantı katmanlarından oluşmaktadır (Şekil 1). Evrişim katmanında filtre, girdi üzerinde tarama yaparak evrişimin uygulanmasını sağlar. Evrişim işlemi neticesinde öznelilik haritası elde edilmektedir. Havuzlama katmanı ise evrişimden sonra ki örnekleme aşamasıdır. Bu katman içerisinde elde edilen değerler üzerinden aritmetik ortalama alma veya en büyük değeri alma gibi seçimler ile özellik seçimi için örnekleme yapılmaktadır. Tam bağlantı katmanları ise hedef sınıfın en iyi şekilde hesaplanması ve optimizasyonun yapılması için kullanılmaktadır. Bunun

yanında, doğrusal birim aktivasyon fonksiyonu (ReLU) ve Softmax aktivasyon fonksiyonu evrişimsel sinir ağlarında en sık kullanılan aktivasyon fonksiyonlarıdır.



Şekil 1. Evrişimsel sinir ağı-CNN mimarisi

Bir evrişimsel sinir ağı oluştururken ardışık şekilde evrişim katmanlarını ve aktivasyon fonksiyonlarını mimariye yerleştirmek gerekmektedir. Daha sonrasında ise havuzlama katmanları ve tam bağlantı katmanı mimariye eklenir (Şekil 1).

Uzun Kısa Süreli Bellek - LSTM

LSTM (Uzun Kısa Süreli Bellek – Long Short Term Memory), geri bildirim bağlantılarına sahip olan yinelemeli sinir ağı mimarili bir derin öğrenme algoritmasıdır [27]. Bu yaklaşımda, öğrenme süreci içerisinde saklanan parametrik değerler değiştirilmez.

Bu gibi yetenekleri sayesinde özellikle zaman serisi problemleri için iyi sonuçlar üreten bir derin öğrenme yöntemidir.

Mimari olarak LSTM, önemli ve önemsiz bilginin belirlendiği kapılara sahiptir. Bu kapılar sayesinde önemli olanların hatırlanması, önemsiz olan bilgilerin ise unutulması sağlanmış olur. Unutma kapısı, unutulacak olan bilginin kararını vermektedir. Eğer ağ, önemsiz bir bilgi olarak öngörüyor ise ilgili girdinin ağırlığı 0 olarak atanır. Bu durumun neticesinde önceki katmanda bulunan aktivasyon fonksiyonundan elde edilen sonuçlar 1'e doğru yakınsanıyorsa tutulacak olan; aksi durumda 0'a doğru ise unutulacak olan bilgi olarak tanımlanmaktadır. LSTM mimarisinde, unutma kapısının ardından girdi kapısı gelmektedir. Tıpkı unutma kapısında yapılan işlem gibi ilk olarak bilginin tutulma kararı elde edilir. Sonrasında ise tutulan bilgi ile ağırlık düzenlenmesi sağlanır. LSTM mimarilerinde bilginin taşınması görevini hücre durumu birimi yapmaktadır. Hücre durumu birimi, tutulan bilginin diğer hücrelere taşınmasını sağlayarak veri akışını meydana getirir. Son olarak ise çıktı kapısı sayesinde tahmin yapılarak sıradaki katmana iletilecek olan değer karar verilir.

DeneySEL Çalışmalar

Dile ait biçimsel kurallar, dilbilgisi ve alfabe dil işleme süreçlerinde makinenin kendi dilini oluşturabilmesi için gereklidir. Çalışmada uygulama için python programlama dili kullanılarak, jupyter notebook ortamında modeller geliştirilmiştir. Çalışmada kullanılan alfabe tercihi için uygulanan algoritma Şekil 2'de gösterilmiştir.

Şekil 2'de gösterilen ve uygulanan algoritma ile Latin, Kiril, Yunan, Hangil, Hanzi, Kana, Arap, İbrani, Kuzey Hint, Güney Hint, Etiyopya, Tana, Kanada Hece gibi alfabeler arasından tercih yapılması sağlanmıştır. Bu süreçte veriler binlik olarak bölütlenmiş ve ağırlık değerleri verilmiş; bölüm sonlarında ise ağırlıklar güncellenmiştir. Bin iterasyon sonucunda 0.17 ağırlık değeri ile makinenin Latin alfabetesini tercih ettiği belirlenmiştir.

Çalışmada ses dosyalarında eşik noktaları 0.2 ile -0.2 olarak kararlaştırılmıştır. Bu değerlerden büyük veya küçük olan sesler başlangıç noktaları olmuştur. Ses işleme için t zamanındaki genlik denklem (1) ile bulunmaktadır. Denklemde faz ofseti, (ϕ) sembolü ile, frekans, (f) notasyonu ile ve t anındaki genlik (y(t)) olarak ifade edilmiştir.

$$y(t) = A \sin(2\pi f t + \phi) \quad (1)$$

Ses dosyası üzerinden özellik çıkarımlarının yapılabilmesi için Python programlama dili ile kullanılabilen Librosa kütüphanesi tercih edilmiş; veri setinde bulunan 22050 ses 0.4 alfa değeri ile dalga formları hesaplanmıştır. Bu işlem adımından sonra ise Fourier Dönüşümü kullanılarak seslerin şiddet spektrumunun çıkartılması yapılmaktadır. Denklem 2'de faz ofseti (ϕ), frekans (f) ve t anındaki genlik (y(t)) olmak üzere Fourier dönüşüm denklemi verilmiştir. Fourier Dönüşümü kullanılarak ses verisi frekans tabanlı hale getirilmiştir.

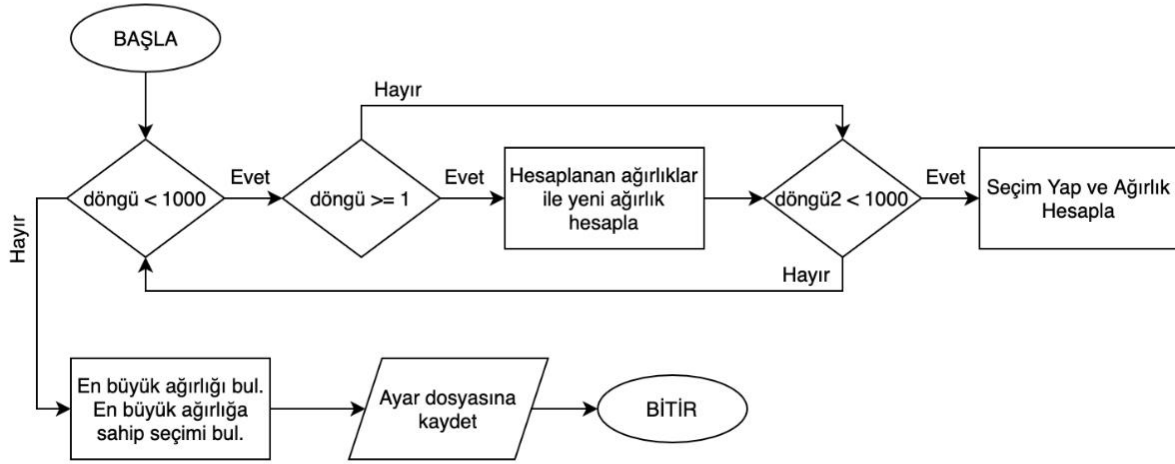
$$s = A_1 \sin(2\pi f_1 t + \phi_1) + A_2 \sin(2\pi f_2 t + \phi_2) \quad (2)$$

Kısa Zaman Fourier Dönüşümü adımının ardından seslere MFCC katsayısı yöntemi uygulanmıştır. MFCC katsayı yöntemi konuşma ve dil tanıma problemleri için en nitelikli özellik olduğu ifade edilmektedir [11]. Son olarak ise genlik zarfı, spektral ağırlık merkezi, 0 geçiş oranı ve spektral akış hesaplamalarının ardından ses verilerinin işlenmesi için makine öğrenmesi yöntemleri uygulanmaktadır. Şekil 3'de MFCC algoritması, Şekil 4'te ise makine öğrenme algoritmaları uygulanma algoritması gösterilmiştir.

Karmaşık periyodik sesin farklı frekanslarda salınan sinüs dalgalarının toplamına ayrıştırılmasını sağlayan Fourier dönüşümü, makine öğrenmesi algoritması ile desibel/frekans bilgisi elde edilmiştir. Fourier dönüşümü sırasında zaman bilgisi kaybolduğu için farklı aralıklarla birden çok Fourier dönüşümü hesaplanarak, Kısa Zamanlı Fourier Dönüşümü

(STFT) kullanılmış ve desibel/frekans bilgisi ile zaman bilgisi de elde edilmiştir. Çalışmada geleneksel yaklaşımların aksine görüntü işleme algoritmaları, ses işleme algoritmaları ile kullanılmıştır. STFT uygulan ses dosyalarının genlik

spektrumlarının logaritması alınmış, Mel ölçeklemesi ve kosinüs dönüşümü uygulanarak MFCC hesaplanmıştır.



Şekil 2. Alfabe seçim algoritması

MFCC bilgisi ile ses dosyasının bir görüntüsü elde edilmiş olmaktadır. Bu görüntüye daha iyi yakınsama ve degradenin kaybolma olasılıklarının azaltılması için doğrultulmuş lineer birim (Rectified Linear Unit-ReLU) uygulanmıştır.

Bunun yanı sıra çok katmanlı yapay sinir ağlarından daha yüksek hesaplama gücü olan; ayrıca daha az parametre sayısına sahip görüntü ve ses işleme konularında oldukça güçlü olan evrişimli sinir ağları (CNN) önerilen mimariye eklenmiştir.

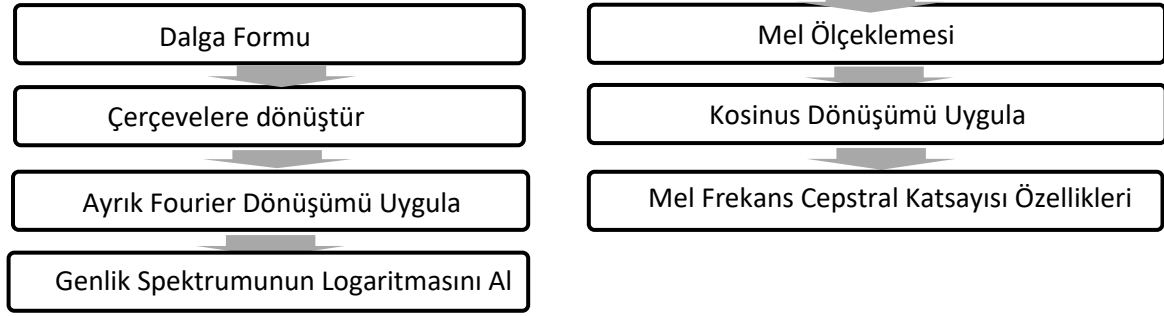
CNN algoritmasında görüntü olarak Spektrogram/MFCC, x ve y eksenleri olarak zaman ve frekans, piksel değeri yerine ise genlik kullanılmıştır. CNN modelinin kullanımının ardından mimarinin üzerine sonraki adımları tahmin edebilen, ses ve müzik uygulamaları için ideal bir çözüm olan takviyeli yapay sinir ağı (recurrent neural network-RNN) eklenmiştir. RNN ağlarını eğitme güçlüğü, ReLU katmanının ve degradelerin hata alma ihtimalini azaltmak için aktivasyon fonksiyonu olarak tanh (hiperbolik tanjant) kullanılmıştır. RNN ağları uzun dönem hafızaya sahip değildir ve uzak geçmiş bilgisini tutmamaktadır. Uzun dönem desenleri öğrenebilmesi için RNN ağlarının özel bir türü olan uzun-kısa dönem hafıza (long-short time memory-LSTM) modele eklenmiştir. Modelin eğitilmesi sona erdikten sonra canlıları ayırt etme ve yeni gelen sesleri anlamlandırabilmesi için seslerin karşılaştırılması dinamik zaman çözgü yaklaşımı (dynamic time wrapping-DTW) algoritması ile gerçekleştirilmiştir. Tablo 1'de evrişimli

yapay sinir ağları destekli ağ topolojisi katman özellik ve değerleri gösterilmiştir.

Bulgular

Özellik çıkarımı yapılan veri setinden seçilen eğitim verisine ilk uygulanan katman ReLU katmanıdır. Veri setinin eğitim öncesi doğruluk oranı 0.2380 ve kayıp oranı 52.8339 iken eğitim sonrası doğruluk oranı 0.9819 ve kayıp oranı 0.0572 olarak elde edilmiştir. Sinir ağındaki aşırı uyum sorunu olduğu görülmektedir. Sinir ağındaki aşırı uyum sorununun çözümü için her bir ReLU katmanından sonra nöron bırakma uygulanmıştır. Her eğitim esnasında bilgisayardan rastgele eğitim seti seçmesi istenmiştir. Nöron bırakma uygulaması öncesi doğruluk oranı 0.2206 ve kayıp oranı 60.2963 iken eğitim sonrası doğruluk oranı 0.8507 ve kayıp oranı 1.0290 olarak değişmiştir. Şekil 5'de ReLU katmanlarının test ve hata değerlendirme grafikleri gösterilmiştir.

Özellik çıkarımı algoritması tamamlanan verilere CNN mimarisinin uygulanması ile kayıp oranlarında düşüş yaşandığı gözlemlenmiştir. CNN ağları öncesi doğruluk oranı 0.1165 ve kayıp oranı 2.5465 iken CNN mimarisi uygulandıktan sonra doğruluk oranı 0.8939 ve kayıp oranı 0.2994 olmuştur. Şekil 6'da CNN modelinin, Şekil 7'de ise LSTM modeline ait test ve hata değerlendirme grafikleri gösterilmiştir.



Şekil 3. Mel frekans cepstral katsayısı oluşturma algoritması

Tablo 1. Evrişimli yapay sinir ağları destekli ağ topolojisi katman özellik ve değerleri.

Katman	Çıktı, Şekil	Değer
Dönüşüm 2D (Conv2D)	(Yok, 11, 11, 32)	320
Maks_Havuzlama2D (MaxPooling)	(Yok, 6, 6, 32)	0
İş_Parçacığı_Normalleştirme (Batch)	(Yok, 6, 6, 32)	128
Dönüşüm_2D_1 (Conv2D)	(Yok, 4, 4, 32)	9048
Maks_Havuzlama2D_1 (MaxPooling)	(Yok, 2, 2, 32)	0
İş_Parçacığı_Normalleştirme1 (Batch)	(Yok, 1, 1, 32)	128
Dönüşüm_2D_2 (Conv2D)	(Yok, 1, 1, 32)	4128
Maks_Havuzlama2D_2 (MaxPooling)	(Yok, 1, 1, 32)	0
İş_Parçacığı_Normalleştirme2 (Batch)	(Yok, 1, 1, 32)	128
Düzleştirilmiş (Flatten)	(Yok, 32)	0
Yoğunluk (Dense)	(Yok, 64)	16448
Bırakma (Dropout)	(Yok, 64)	0
Yoğunluk 1 (Dense)	(Yok, 10)	650

Toplam değer: 16842

Eğitilebilir değer: 16650

Eğitilemez değer: 192

Elde edilen bulgular incelendiğinde ReLU katmanlarının aşırı uyumluluk sorunu giderildiğinde, doğruluk oranlarının düştüğü ve hata oranlarının çok fazla yükseldiği görülmüştür. ReLU katmanları CNN mimarisi ile desteklendiğinde doğruluk oranında kısmi bir artış hata oranında ise büyük ölçüde düşüş olduğu gözlemlenmiştir.

CNN destekli önerilen model RNN'nin özel bir türü olan LSTM ile desteklendiğinde doğruluk oranlarının yükseldiği kayıp oranında ise kısmi bir düşüş olduğu gözlemlenmiştir. 30 iterasyon sonucunda CNN mimarisi ile desteklenen model için doğruluk oranı 0.8939 ve kayıp oranı 0.2994 olarak elde edilmiştir. LSTM ile desteklenen modelde ise aynı iterasyon sonunda doğruluk oranı 0.9151 ve kayıp oranı 0.2358 olmuştur (Tablo 2).

Tablo 2. Yöntemlerin performans karşılaştırması.

1. Sütun	Doğruluk	Kayıp Oranı
LSTM	0.91	0.23
CNN	0.89	0.29

Bulgular sonucunda, makine dili ve doğal dil işleme modellemesinde ses dosyalarında daha kararlı sınıflandırma yapılabilmesi için ReLU ile özellik çıkarımı yapılan veri setlerinin, görüntü işlemede kullanılan CNN ve ses işlemede kullanılan LSTM algoritmalarının birlikte kullanılarak bir modelin önerilmesine ve ardından önerilen

modelin eğitilmesine karar verilmiştir. Veri seti modele sunulup modelin eğitimi bitirildikten sonra, bilgisayarın kendi seçtiği alfabe ile ürettiği sesler ve eğitilen veri seti, DTW algoritması ile karşılaştırılmıştır. Bilgisayarın ürettiği sesler İngiliz alfabesinde bulunan A-Z arası harfler a, aa, aaa, aaaa şeklinde P(32,1), P(32,2), P(32,3) ve P(32,4) permütasyonları ile sınırlandırılmıştır. Sesler karşılaştırılırken, Öklid eşleşmesinde zaman/frekans değerleri karşılıklı olarak eşleştirilirken, DTW algoritmalarında birçok nokta tek bir nokta ile kıyaslanmaktadır. Böylece özellikle konuşma sırasındaki duraklamalar ses şiddetinin yükselip, alçalması, yavaş veya hızlı konuşmalardan daha az etkilenmektedir.

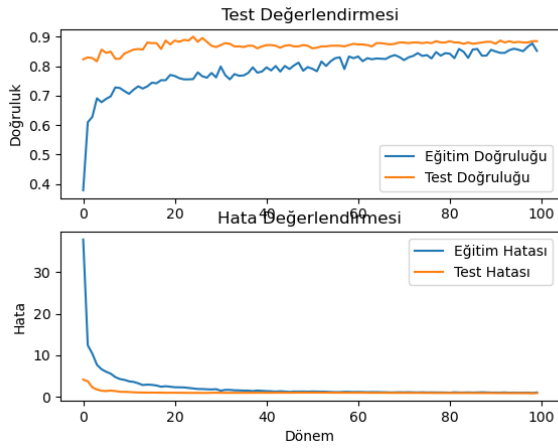
Son olarak çalışmada kullanılan "Kaggle Audio Cats and Dogs" veri setini literatürde modellemelerde kullanan çalışmalar incelenmiş; çalışmalarda elde edilen doğruluk oranları karşılaştırılması Tablo 3'de gösterilmiştir.

Tablo 3. Literatürdeki çalışmalar ile karşılaştırılması.

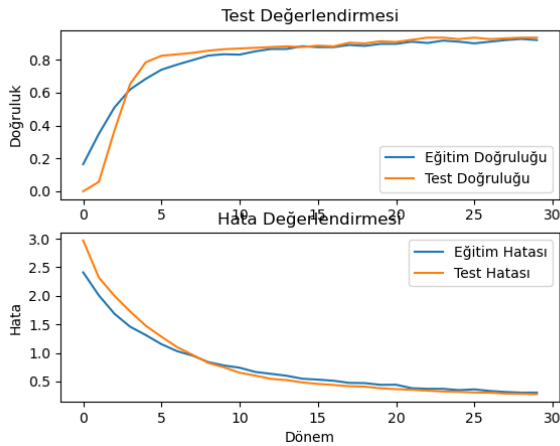
Çalışma	Yöntem	Doğruluk Oranı
Waasiu ve ark. [28]	YSA	0.88
Cengil ve ark. [29]	CNN	0.83
Pandeya ve ark. [30]	CNN	0.91
Önerilen yöntem	LSTM-NLP	0.91



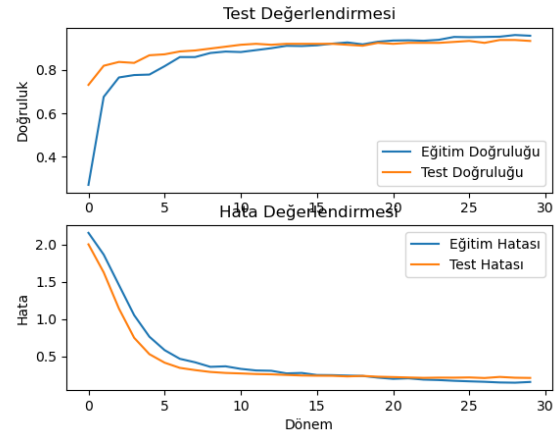
Şekil 4. Ses verileri için geleneksel makine öğrenmesi akışı



Şekil 5. ReLU katmanları değerlendirme grafikleri



Şekil 6. CNN değerlendirme grafikleri



Şekil 7. LSTM değerlendirme grafikleri

Tartışma ve Sonuç

Çalışmada ilk olarak literatürdeki benzer çalışmalarda olduğu gibi ses dosyaları üzerinden özellik çıkarımı yapılmış; DTW ile seslerin grafikleri oluşturulmuştur. Dil işleme ve ses işleme eğitim süreçleri neticesinde makine Latin alfabesini tercih etmiş ve bu dil üzerinden makinenin kelimeler üretmesi sağlanmıştır. Bunun yanında uygulanan modelde seslerin ayrıştırılmasının sağlanması için Fourier Dönüşümü uygulanmıştır. Bu formül neticesinde kayıp olan zaman verisi Kısa Zaman Fourier Dönüşümü ile tekrar elde edilmiştir. Çalışmada benzer çalışmalardan farklı olarak ses dosyaları üzerinden bulunan MFCC, görsel bir veri olarak kullanılmıştır. Bunun yanında CNN ağları ile model desteklenerek daha iyi performans elde edilmesi sağlanmıştır. Ayrıca MFCC katmanında yer alan ardışık şekilde olan verilerin daha efektif olarak işlenmesi için uzun dönem hafıza yeteneği olan RNN-LSTM ağları, mimariye katılmıştır.

Sonuç olarak kedi sesine en yakın ses makine tarafından "QTKP" karakterleri ile üretilmiştir. Netice olarak türetilen ses ile benzerlik farkı 56823.36 olarak elde edilmiştir. "XOXO" karakterlerinden üretilen ses ise köpek sesine 26862.43 fark değeri ile en fazla benzerliği olan ses olmuştur. Çalışmada DTW, CNN, RNN, MFCC yöntemleri kullanılarak ses işleme, dil işleme ve makine öğrenmesi algoritmaları birlikte uygulanmış ve makinenin kendi dili için kelime üretmesi sağlanmıştır. Bu minvalde, makine

veri setinde öğretilen kedi ve köpek seslerini tanıyarak sınıflandırmıştır. Bunun yanında kedi ve köpek seslerine yaklaşık sesler türetmiştir. Çalışmada uygulanan modeller arasında en yüksek doğruluk oranı ve en az kayıp değeri RNN-LSTM ile uygulanan önerilen mimari ile ulaşılmıştır.

Elde edilen bulgular neticesinde makine dilinin oluşturulabilmesi için derin öğrenme destekli doğal dil işleme yönteminin uygulanabilirliği ortaya konmuştur. Uygulanan her iki derin öğrenme mimarisi de doğal dil işleme ile elde edilen sonuçların iyileştirilmesini desteklemiştir. Ayrıca ses dosyalarının görsel veriye dönüştürülerek CNN ve LSTM ağları üzerinde modellemede kullanılması da literatüre katkılardan biri olmuştur.

Çalışmanın kısıtlarından biri kullanılan kütüphanenin sadece İngilizce ile ilgili kelimeler türetebilmesidir. Bunun yanında çalışmada çoğul ekler belirlenmediğinden ötürü tekil kelimeler üretilmiştir. Ayrıca ses dosyaları üzerinden canlı sayısı tespiti, gramer özelliklerinden sıfat, zarf, bağlaç tespitleri de bulunmamaktadır. Daha büyük veri setleri, daha farklı canlılara ait sesler kullanılması halinde çalışma daha da geliştirilerek doğal dil ve yapay zeka alanlarına yeni katkılar eklenebileceği düşünülmektedir.

Etik kurul onayı ve çıkar çatışması beyanı

Hazırlanan makalede etik kurul izni alınmasına gerek bulunmamaktadır. İlgili makale, yüksek lisans tezinden üretilmiştir. Ayrıca Kaggle açık veri deposunda bulunan hazır veri seti üzerinden veriler ile modelleme yapılmıştır.

Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.

Teşekkür

Yazarlar, çalışmaya değerli zamanlarını ayırarak katkılarını sunan dergi editörleri ve hakemlere teşekkür etmektedir. Çalışma, Dr.Öğr.Üyesi Atınç Yılmaz'ın danışmanlığında yürütülen, Ali Dayan'ın Beykent Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği A.B.D. "Doğal Dil İşleme ile Makinelerin Kendi Dilini Modellemesi" isimli yüksek lisans tezinden türetilmiştir.

Yazar Katkıları

Tüm yazarlar makaleye yeterli derecede katkıda bulunmuştur. İlgili makale yüksek lisans tezinden oluşturulmuştur. Yazarlar yüksek lisans tez öğrencisi ve tez danışmanıdır.

- Çalışma konsepti ve tasarım (D, Y)
- Veri toplama (D)
- Verilerin analizi ve yorumlanması (D, Y)
- Taslağın oluşturulması (D)
- Revizyon (Y)

Kaynakça

- [1] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433-460, 1950.
- [2] T. Winograd, "Five lectures on artificial intelligence," 1st ed., Stanford University, 1974.
- [3] J. Searle, "Minds, brains and programs," *The Behavioral and Brain Sciences*. 1980.
- [4] A. Dayan, "Doğal dil işleme ile makinelerin kendi dilini modellemesi," M.S. thesis, Dept. Computer. Eng., Beykent Univ., İstanbul, Türkiye, 2022.
- [5] Kaggle audio cats and dogs dataset [Online]. Available: <https://www.kaggle.com/datasets/mmmoreaux/audio-cats-and-dogs>, Accessed on: Aug. 10, 2021
- [6] P. J. Rani, J. Bakthakumar, B. P. Kumar, U. P. Kumaar and S. Kumar, "Voice controlled home automation system using natural language processing(NLP) and internet of things(IoT)," in *Proc. ICONSTEM*, 2017, pp. 368-373.
- [7] G. Alexakis, S. Panagiotakis, A. Fragakakis, E. Markakis, E. and K. Vassilakis, "Control of smart home operations using natural language processing, voice recognition and IoT technologies in a multi-tier architecture," *Designs*, vol. 3, no. 3, pp. 32, 2019.
- [8] J. Mass, S. N. Srirama, H. Flores, and C. Chang, "Proximal and social-aware device-to-device communication via audio detection on cloud," In *Proc. 13th International Conference on Mobile and Ubiquitous Multimedia*, 2014, pp. 143-150.
- [9] M. E. Almahdi, "Doğal dillerin zaman ve konuma bağlı değişimlerinin nicel olarak ölçülmesi," M.S. thesis, Fen Bilimleri Enstitüsü, Gebze Teknik Üniversitesi, Kocaeli, Türkiye, 2020.
- [10] A. Toprak, "Doğal dil işleme ile ingilizce otomatik sözlük oluşturma," M.S. thesis, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye, 2019.
- [11] B. Büyüköz, "Doğal dil işlemede bağlam odaklı derin öğrenme yöntemlerinin transfer kapasitesinin metin sınıflandırması problemi üzerinden incelenmesi," M.S. thesis, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye, 2016.
- [12] D. G. Özcan, "Doğal dil işleme tekniklerinin incelenmesi ve seyahat-turizm sesli asistanı için Türkçe varlık ismi tanıma aracı geliştirilmesi," M.S. thesis, Akdeniz Üniversitesi Fen Bilimleri Enstitüsü, Antalya, Türkiye, 2020.
- [13] S. C. Gupta, N .R. Singh, T. Sharma, A. Tyagi and R. Majumdar, "Generating image captions using deep learning and natural language processing," in *Proc. 9th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*, 2021, pp. 1-4.

- [14] M. Anand, K. B. Sahay, M. A. Ahmed, D. Sultan, R. R. Chandan and B. Singh, "Deep learning and natural language processing in computation for offensive language detection in online social networks by feature selection and ensemble classification techniques," *Theoretical Computer Sciences*, in Press, 2022.
- [15] E. Kasthuri and S. Balaji, "Natural language processing and deep learning chatbot using long short term memory algorithm," *Materials Today: Proceedings*, in Press, 2022.
- [16] S. Chotirat, and P. Meesad, "Part-of-speech tagging enhancement to natural language processing for Thai wh-question classification with deep learning," *Heliyon*, vol. 7, no. 10, e08216, 2021.
- [17] N. Shahi, A. K. Shahi, R. Phillips, G. Shirek, D. M. Lindberg, and S. L. Mpulton, "Using deep learning and natural language processing models to detect child physical abuse," *Journal of Pediatric Surgery*, vol. 56, no. 12, pp. 2326-2332, 2021.
- [18] A. Alhogail and A. Alsabih, "Applying machine learning and natural language processing to detect phishing email," *Computers & Security*, vol. 110, 102414, 2021.
- [19] E. Mumcuoğlu, C. E. Öztürk, H. M. Özaktaş and A. Koç, "Natural language processing in law: Prediction of outcomes in the higher courts of Turkey," *Information Processing & Management*, vol. 58, no. 5, 102684, 2021.
- [20] S. G. Mendez, F. A. Perez, A. B. Villa and F. J. G. Castano, "Detection of temporality at discourse level on financial news by combining Natural Language Processing and Machine Learning," *Expert Systems with Applications*, vol. 197, 2022.
- [21] A. Hodorog, I. Petri, and Y. Rezgui, "Machine learning and natural language processing of social media data for event detection in smart cities," *Sustainable Cities and Society*, In Press, 2022.
- [22] N. Takahashi, M. Gygli, B. Pfister, and L.V. Gool, "Deep convolutional neural networks and data augmentation for acoustic event recognition," In *Proc. 17th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 2982-2986, 2016.
- [23] A. Yılmaz, "Yapay Zekâ,". Kodlab Yayınevi, İstanbul, Türkiye, 2020.
- [24] E. Brill, and R. J. Mooney, "An Overview of Empirical Natural Language Processing," *AI Magazine*, vol. 18, no. 4, 1997.
- [25] K. R. Chowdhary, "Fundamentals of Artificial Intelligence," Springer, 2020.
- [26] M. Momeny, A. A. Neshat, M. A. Hussain, S. Kia, M. Marhamati, A. Jahanbakhshi and G. Hamarneh, "Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep CNN for the detection of Covid-19 in X-ray images," *Computers in Biology and Medicine*, vol. 136, 104704, 2021.
- [27] H. Ma, and S. Liang, "Development of the GLASS 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model," *Remote Sensing of Environment*, vol. 273, 112985, 2022.
- [28] A. Waasiu, B. A. Ilham and A. Lawi, Klasifikasi Audio "Cats and Dogs Menggunakan Model Artificial Neural Network Multi-perceptron," *Proceeding KONIK*, vol. 5, no. 1, pp. 56-61, 2021.
- [29] E. Cengil, A. Çınar and Z. Güler, "A GPU-based convolutional neural network approach for image classification," In *Proc. 2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, pp. 1-6, 2017.
- [30] Y. R. Pandeya, D. Kim, and J. Lee, "Domestic cat sound classification using learned features from deep neural nets," *Applied Sciences* vol. 8, 1949, 2018.