Çağrı Merkezlerinde Olumsuzluk İçeren Çağrıların Evrişimsel Sinir Ağları ile Tespiti

Araştırma Makalesi/Research Article

Ali Fatih KARATAŞ[1], Öykü Berfin MERCAN[1], Umut ÖZDİL[2], Şükrü OZAN[1]

[1]AdresGezgini A.Ş. Ar-Ge Merkezi, İzmir, Türkiye
[2]Elektrik-Elektronik Mühendisliği, İzmir Demokrasi Üniversitesi, İzmir, Türkiye
alifatih449@gmail.com, oykumercan@adresgezgini.com, 2117108001@std.idu.edu.tr, sukruozan@adresgezgini.com
(Geliş/Received:04.08.2022; Kabul/Accepted:09.11.2022)
DOI: 10.17671/gazibtd.1156330

Özet— Bu çalışmada çağrı merkezi çalışanları ile müşteriler arasındaki telefon konuşmalarının otomatik olarak olumlu veya olumsuz şeklinde değerlendirilmesi üzerine odaklanılmıştır. Çalışmada kullanılan veri seti firma bünyesinde gerçekleştirilen telefon görüşmelerinden oluşmaktadır. Veri seti üçer saniyelik 10411 adet ses kaydını içermekte olup bu kayıtların 5408 tanesi olumlu kayıtlardan 5003 tanesi münakaşa, öfke ve hakaret içeren olumsuz kayıtlardan oluşmaktadır. Çağrı merkezi kayıtlarından duygu tanıma için anlamlı öznitelikler elde etmek amacıyla her bir ses kaydından MFCC öznitelikleri çıkarılmıştır. Çağrı merkezi kayıtlarını olumlu olumsuz olarak sınıflandırmak için önerilen CNN mimarisi MFCC öznitelikleriyle eğitilmiştir. Önerilen CNN modeli %86,1 eğitim başarısı, %77,3 doğrulama başarısı göstermiş olup test verileri üzerinde %69,4 sınıflandırma başarısı elde edilmiştir. Bu çalışma ile çağrı merkezlerinde gerçekleşen konuşmaların otomatik analizi yapılıp olumsuz durumların kalite yöneticilerine bildirilmesiyle gerekli önlemlerin alınarak müşteri memnuniyetinin artırılması amaçlanmaktadır.

Anahtar Kelimeler— ses sinyalinden duygu tanıma, çağrı merkezi, MFCC, CNN.

Detection of Negative Calls in Call Centers with Convolutional Neural Networks

Abstract— In this study, it is focused on the automatic evaluation of telephone conversations between call center employees and customers as positive or negative. The dataset used in the study include telephone conversations between call center employees and customers in the company. The data set contains 10411 three-second call center records; 5408 of them are positive records and 5003 of them are negative records that include arguments, anger and insults. In order to obtain meaningful features for emotion recognition from voice records, MFCC features were extracted from each call center records. The proposed CNN architecture is trained with MFCC features to classify call center records as positive or negative. The proposed CNN model showed 86.1% training accuracy, 77.3% validation accuracy and it achieved 69.4% classification accuracy on the test data. This study aimed to increase customer satisfaction by automatic analysis of conversations in call centers and notifying quality managers of negative records.

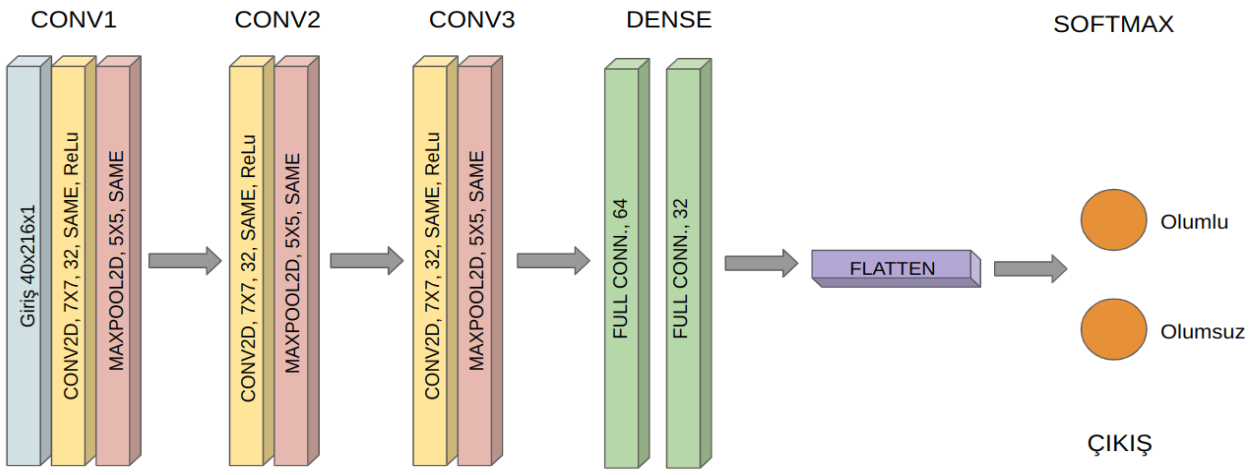
Keywords— emotion recognition from audio signal, call center, MFCC, CNN

1. GİRİŞ (INTRODUCTION)

Müşteri ilişkilerinin oldukça önem kazandığı günümüzde firmalar müşterileri ile iletişim kurmak amacıyla çevrimiçi yöntemler ve telefon görüşmelerini yaygın olarak kullanmaktadırlar. Teknolojik gelişmelerin sağladığı imkanlarla çevrimiçi görüşmelerdeki kalite ve standartların iyileştirilmesi hedeflenerek müşteri memnuniyetinin üst düzeye çıkarılması hizmet veren kurumların öncelikli hedefi haline gelmiştir. Bu kapsamda çevrimiçi görüşmelerdeki sahtekarlık, anomali ve duygu tespitinin gerçekleştirilmesi amaçlanmaktadır. Dijitalleşme ile birlikte çevrimiçi görüşmelerdeki artış veri yığını oluşturmakta olup görüşmelerde anomali ve duygu tespitinin hızlı ve etkili bir şekilde gerçekleştirilmesinde zorluklarla karşılaşılmaktadır. Son yıllarda genel olarak yapay zeka alanında ve özel olarak da dijital ses işleme alanında yaşanan gelişmelerle birlikte otomatik konuşma tanıma ve ses verisinden duygu tanıma konularında elde edilen başarılı sonuçlar ses verilerinin geleneksel yöntemlerle analizinde karşılaşılan zorluklara pratik ve hızlı çözümler sunmaktadır [1-3]. Özlan vd. derin evrişimli sinir ağları ile sahte çağrı merkezi görüşmelerini otomatik olarak tespit eden yöntemi önermişlerdir. Önerilen yöntem konuşma tanıma motoru kullanılarak metne çevrilmiş çağrı merkezi görüşmelerini metin sınıflandırma algoritmasıyla sınıflandırarak sahte görüşmeleri otomatik algılamaktadır [1]. Iheme vd. doğrusal olmayan güç dönüşümü, sinirsel özellik öğrenme ve kümeleme içeren yarı denetimli yöntem ile çağrı merkezi temsilcisi hatasını tespit eden bir çalışma sunmuşlardır [2]. Çağrı süresindeki sessizlik miktarı önemli bir performans göstergesi olarak belirlenmiş olup, önerilen sistem ile kalite kontrol yöneticilerinin ve çağrı merkezi çalışanlarının performansında artış gözlemlendiği belirtilmiştir. Pappas vd. konuşma sinyalinden doğrudan çıkarılan Mel Frekans Cepstral Katsayısı (Mel Frequency Cepstral Coefficient- MFCC) öznitelikleri ile eğitilen Lojistik Regresyon sınıflandırıcısıyla çağrı merkezi diyaloğunda %70 başarı ile öfke tespiti gerçekleştirmiştir [3].

Ses işleme araştırmalarında MFCC önemli bir yere sahip olup yüksek başarı oranıyla ses verisinden öznitelik elde edilmesinde yaygın olarak kullanılan bir tekniktir [4]. Ses verilerinden çıkarılan MFCC'nin makine öğrenmesi ve derin öğrenme algoritmalarında öznitelik olarak kullanılmasıyla dijital ses işleme alanında birçok çalışma gerçekleştirilmiştir [5-11]. Turnbull vd. ses kayıtlarının müzik türlerine göre otomatik sınıflandırılmasını amaçlayan çalışmada MFCC ve ses sinyalinin kısa süreli Fourier dönüşümüne (STFT) dayalı öznitelikleri ile Radyal Temelli Fonksiyon (RBF) eğitmiş ve önerilen yöntemin insan performansına yakın başarı gösterdiği gözlemlenmiştir [5]. Iheme vd. ses verisinin sessizlik, konuşma ve müzik olmak üzere üç sınıftan biri olarak sınıflandırmak amacıyla ses verilerinin MFCC katsayılarını, türevlerini ve ikinci türevlerini SVM (Support Vector Machine) ve Naive Bayes algoritmalarının eğitiminde öznitelik olarak kullanmış ve bu özniteliklerin başarıya etkilerini araştırmışlardır [6]. MFCC öznitelikleri konuşma duygusu tanımada yaygın olarak kullanılmaktadır. Likitha vd. konuşmacıların ses sinyallerinden duygunun tespit edilmesi için önerdikleri çalışma, MFCC kullanarak öznitelik çıkarımına ve standart sapma kullanarak karar vermeye dayanmaktadır. Çalışmada kullanılan veri seti farklı duygulara sahip 60 farklı kişinin ses kaydını içermektedir. Ses kayıtlarından MFCC öznitelikleri elde edilmiş ardından MFCC'nin ortalama değeri ve ortalama değerinin standart sapması bulunmuştur. Elde edilen standart sapma, farklı duygular için optimize edilmiş standart sapma değerleriyle "if-else" yapısında karşılaştırılarak sese ait duygu tespit edilmiştir [7]. Bir diğer bir duygu tespiti çalışmasında Milton vd. Berlin EmoDB veri setinde bulunan 7 farklı duyguyu sınıflandırmak için 535 tane ses verisinden elde edilen MFCC öznitelikleri SVM sınıflandırıcısının eğitiminde kullanılmıştır. Performans analizi sonucu da %68'dir [8]. Waghmare vd. yapay duygusal Marathi konuşma veri tabanından konuşma duygusunu analiz etmek ve tanımak için MFCC öznitelikleri çıkarmış, çıkarılan öznitelikler LDA (Lineer Diskriminant Analiz) sınıflandırıcısının eğitiminde kullanılmıştır [9]. Demircan vd. Berlin EmoDB veri setiyle gerçekleştirdikleri çalışmada ses kliplerinden MFCC'leri çıkardıktan sonra konuşma duygusunu sınıflandırmak için bir KNN (K En Yakın Komşu) algoritması önermişlerdir. 7 farklı duygunun sınıflandırıldığı bu çalışmada %50 sınıflandırma başarısı elde edilmiştir [10]. Nalini vd. YSA (Yapay Sinir Ağları), SVM, RBFNN (Radyal Temel Fonksiyonu Sinir Ağları) kullanarak müzikte duyguyu tanımak için MFCC kullanmışlardır [11].

Makine öğrenmesi ve ses verisinden öznitelik çıkarma yöntemleri ile gerçekleştirilen çalışmalar ile başarılı sonuçlara ulaşılmış ve ses işleme alanına olan ilgi artmıştır. Yeterli sayıdaki ses verisi ile yapay sinir ağları eğitilerek Konuşma Duygusu Tanıma (KDT) alanında birçok çalışma gerçekleştirilmiştir [12-14]. CNN'ler (Evrişimsel Sinir Ağları (Convolutional Neural Network)) ve RNN'lere (Tekrarlayan Sinir Ağları (Recurrent Neural Network)) dayalı KDT algoritmasına değinilmiş ve birleştirilmiş CNN'ler ve RNN'lere dayalı bir KDT yöntemi önerilmiştir [12]. Bir diğer çalışmada konuşma duygusunu tanımak için 1B ve 2B CNN LSTM (Uzun Kısa Süreli Bellek (Long Short Term Memory)) ağları incelenmiştir [13]. Bu iki ağın kıyaslanması için iki farklı veri seti üzerinde çalışma gerçekleştirilmiştir. Tasarlanan iki CNN LSTM ağının, duygusal bilginin ayırt edici özelliklerini öğrenebileceği ve üst düzey soyutlamaları modelleyebileceği gösterilmiştir. Deneysel sonuçlar karşılaştırıldığında 2B CNN LSTM ağının performansının 1B CNN LSTM ağının performansına göre yüksek olduğu gözlemlenmiştir. Duygu tanıma amacıyla gerçekleştirilen başka bir çalışmada ise LSTM tabanlı sınıflandırma ağı ile öznitelik çıkarmak için paralel evrişimli katmanların ortaklaşa eğitimi önerilmiştir. Önerilen çalışma paralel çok katmanlı CNN ağının bir LSTM üzerinde istiflendiğini ve ham konuşma kullanan mevcut yöntemlerle karşılaştırıldığında daha iyi doğruluk sağladığını göstermiştir [14].



Şekil 1. Önerilen CNN model
(Proposed CNN model)

Ham ses sinyallerinden MFCC gibi yüksek seviyeli bilgilerin alınması ve ardından bu bilgilerin bir sinir ağından geçirilmesiyle konuşma duygusu tanıma alanında çalışmalar kaydedilmiştir. Wang vd. MFCC özniteliklerine ve ses sinyallerinden üretilen mel-spektrogramlarına dayalı olarak duyguları tahmin eden çift aşamalı bir yöntem önermişlerdir [15]. Önerilen yöntemde standart bir LSTM, MFCC özelliklerini işlerken, Dual-Sequence LSTM (DSLSTM) olarak belirtilen yeni bir LSTM mimarisi, iki mel-spektrogramı aynı anda işler. EmoDB veri seti ile gerçekleştirilen başka bir çalışmada ise spektrogram, Mel-spektrogram ve MFCC olmak üzere önerilen CNN+BLSTM mimarisi üç farklı öznitelik ile eğitilmiş ve konuşma duygusu tanımada MFCC özniteliklerinin başarılı sonuçlar verdiği gözlemlenmiştir [14, 16].

Literatürde yer alan çalışmalar incelendiğinde EmoDB veri setinin duygu tanımada yaygın olarak kullanıldığı görülmektedir [8, 10, 12, 14, 16]. Berlin EmoDB [17], 7 duygu içeren ve sınıflandırma doğruluğunu doğru bir şekilde değerlendirmek için her duyguyu neredeyse aynı sayıda içeren dengeli bir veri setidir. 10 profesyonel oyuncunun duygu ifadelerini öfkeli, can sıkıntısı, iğrenme, korku, mutlu, nötr ve hüzünlü bir şekilde söylemesiyle veri seti oluşturulmuştur. Günlük iletişimden gelen ve tüm duygularda yorumlanabilen 535 cümle vardır. Berlin EmoDB ve bu veri seti gibi konuşma duygusu tanıma çalışmalarında kullanılmak üzere oluşturulan veri setleri haricinde çağrı merkezi konuşmaları, röportajlar ve televizyon programları gibi kayıtlar gerçek duyguları içeren, duyguların doğal olarak elde edilebileceği veri kaynaklarını oluşturmaktadır. Spontane duyguları içeren, üç boyutlu bir duygu alanı çerçevesinde oluşturulan VAM veri tabanı [18, 19] bir Alman TV talk-show'u "Vera am Mittag" kaydedilmiştir. Veri tabanı talk-show'da tartışılan konular nedeniyle çoğunlukla nötr ve olumsuz duygular içerir. [3]'de kullanılan veri seti ise bir telefon sağlayıcı şirketin çağrı merkezinden toplam 9 saat 30 dakika uzunluğunda 137 kayıtlı müşteri temsilcisi görüşmesini içeren Yunanca görüşmelerden oluşmaktadır.

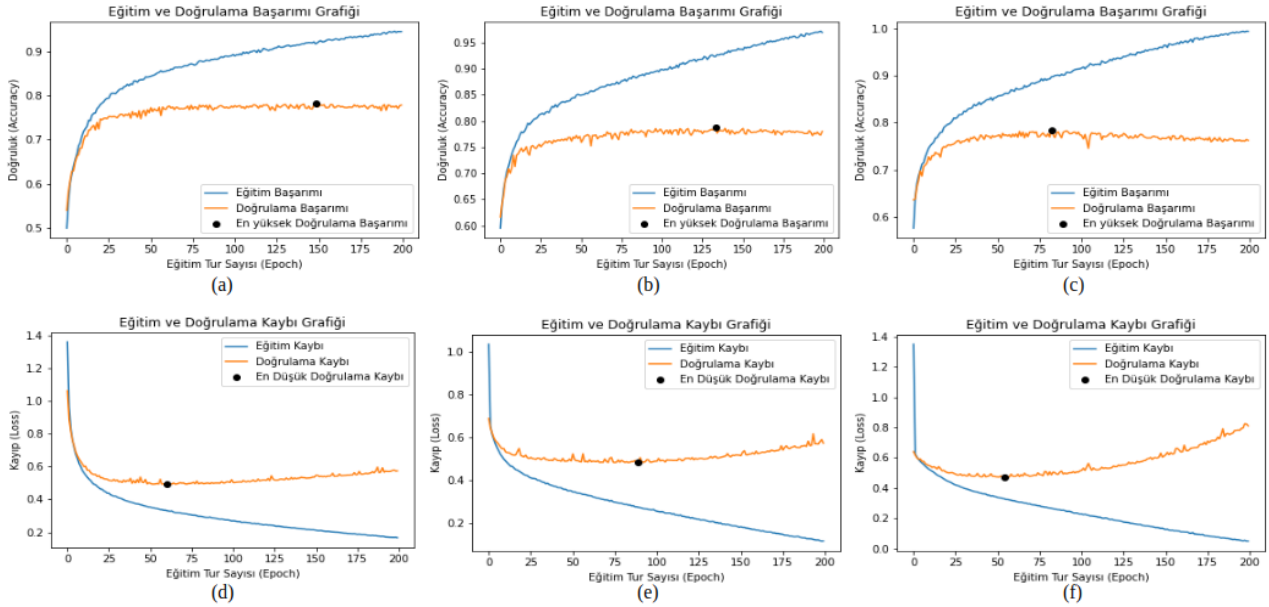
Bu çalışma ile çağrı merkezleri telefon görüşmelerinin otomatik olarak analizinin yapılmasıyla müşteri ve müşteri temsilcisi arasında gerçekleşen görüşmelerde müşteri memnuniyetini olumsuz etkileyen münakaşa durumlarının otomatik olarak tespiti sağlanması önerilmiştir. AdresGezgini A.Ş. bünyesindeki çağrı merkezi görüşmelerinden oluşturulan veri setinin MFCC öznitelikleri çıkarılmış ve CNN modeli bu özniteliklerle eğitilerek modelin başarısı değerlendirilmiştir. Önerilen yöntem ile çağrı merkezi görüşmelerini olumlu ve olumsuz olarak değerlendiren bir sistem oluşturularak firmaların müşteri memnuniyetini artırmak için kullanılması amaçlanmıştır.

Çalışmanın geri kalanı şu şekilde düzenlenmiştir; 2. Bölümde çalışmada kullanılan veri seti, öznitelik çıkarımı ve derin öğrenme mimarisi açıklanmıştır. 3. Bölümde elde edilen sonuçlar açıklanmıştır. 4. Bölümde gerçekleştirilen çalışma özetlenmiş ve gelecek çalışmalara yer verilmiştir.

2. MATERYAL VE METOD (METARIAL AND METHOD)

2.1. Veri Seti (Dataset)

Veri seti, şirket bünyesinde gerçekleştirilen çağrı merkezi görüşme kayıtlarından oluşmaktadır. Müşteri ve müşteri temsilcileri arasında gerçekleşen, uzunluğu 2 ile 5 dakika arasında değişen, içeriğinde küfür, hakaret, öfke gibi olumsuz duyguların bulunduğu farklı uzunluklardaki 100 görüşme kaydı kullanılarak oluşturulmuştur. Her bir görüşme birer saniye kaydırılarak oluşturulmuş üçer saniyelik ses kayıtlarına ayrılmıştır. Ardından üçer saniyelik bu kayıtlar münakaşa, küfür, hakaret içeriyorsa olumsuz, nötr veya olumlu seyir gösteren kayıtlar olumlu şeklinde etiketlenmiştir. Oluşturulan 3 saniyelik parçaların sadece çağrı merkezi çalışanı ve müşterinin aktif olarak konuştuğu kısımlardan veri seti oluşturulmuştur. 5408 adet olumlu 5003 adet olumsuz olarak ayrılan toplamda 10411 adet ses kaydı içeren bir veri seti elde edilmiştir.



Şekil 2. Modeller için eğitim ve doğrulama başarımı grafikleri Model 1 (a), Model 2 (b), Model 3 (c), modeller için eğitim ve doğrulama kaybı grafikleri Model 1 (d), Model 2 (e), Model 3 (f) (Training and validation performance graphs for models Model 1 (a), Model 2 (b), Model 3 (c), training and validation loss graphs for models Model 1(d), Model 2(e), Model 3(f))

2.2. Yöntem (Method)

2.2.1. Öznitelik Çıkarımı (Feature Extraction)

MFCC, sinyal işleme için popüler ve yüksek performanslı bir tekniktir. İnsan kulağının kritik frekans bant genişliğini temel almasıyla yaygın olarak kullanılan MFCC katsayıları, dijital ses işleme alanında önemli bir öznitelik çıkarma yöntemidir [4]. Sınırlı sayıda veriyle, frekans alanı özellikleri, ses sinyalinde potansiyel olarak sinyalin altında yatan duyguyu tanımlamamıza yardımcı olabilecek daha derin kalıpları ortaya çıkarır. MFCC öznitelik vektörleri oluşturulurken ilk olarak Hamming penceresi ile ses sinyalini küçük pencereler şeklinde tekrar şekillendirir ardından çerçevelere böler. Spektrum, Hızlı Fourier Dönüşümü ile her çerçeve için oluşturulur ve her biri filtre bankası kullanılarak ağırlıklandırılır. Ardından MFCC öznitelik vektörü Logaritma ve Ayrık Kosinüs Dönüşümü kullanılarak hesaplanır [20]. MFCC öznitelik vektörleri iki boyutlu matris veya ortalaması alınarak tek boyutlu matris şeklinde elde edilmektedir. Bu durum genellikle iki veya tek boyutlu matrislerin sınıflandırılması için kullanılan CNN modellerinde MFCC'lerin girdi olarak kullanılarak sınıflandırma çalışmaları yapılmasına imkan tanımaktadır. Bu kapsamda Python'ın Librosa kütüphanesi kullanılarak ses kayıtlarından elde edilen 2 boyutlu MFCC öznitelik matrisleri CNN modellerinde girdi verisi olarak kullanılmıştır. Veri seti içerisindeki ses kayıtlarının öznitelik sayısı literatürdeki çalışmaların başarıları dikkate alındığında 40 olarak belirlenmiştir [20, 21].

2.2.2. Derin Öğrenme Mimarisi (Deep Learning Architecture)

Evrişimli Sinir Ağı (Convolutional Neural Network (CNN)) ileri beslemeli yapay sinir ağıdır [22]. Çok sayıda evrişimli katman, havuzlama ve tamamen bağlı katmanlar başta olmak üzere farklı katman türleri ile derin evrişimsel sinir ağı oluşturulur. Verinin öznitelikleri evrişim katmanında elde edilir. Evrişim katmanında, girdi verisi üzerinde öznitelikleri çıkaran birden çok filtre (kernel) kayar. Filtrelerin ve girdinin her elemanın eleman elemana çarpımının toplamı bu katmanın çıktısı olan özellik haritasını verir. Evrişim katmanı, filtre boyutu, adım değeri (stride) ve dolgu (padding) ile özelleştirilebilir. Adım değeri filtrenin girdi verisi üzerinde kaç adım kaydırılacağını belirler. Dolgu ise evrişim katmanının çıktısı olan özellik haritasının boyutu ve orijinal girdi matrisinin boyutunu eşlemek için sıfırlardan oluşan satır ve sütunlar ekler. Sinir ağlarının yalnızca doğrusal bir fonksiyonu öğrenmesi ve hesaplaması değil aynı zamandan görüntü, video, ses, metin gibi kompleks veri türlerini modelleme görevini gerçekleştirmesi amaçlanmaktadır. Bir yapay sinir ağında, girdiler ve çıktılar arasındaki doğrusal olmayan eşlemelerin öğrenilmesi ve anlamlandırılması için aktivasyon fonksiyonları önem taşımaktadır. Aktivasyon fonksiyonu, bir giriş sinyalini, sırayla yığındaki bir sonraki katmana girdi olarak beslenen bir çıkış sinyaline dönüştürmek için özel olarak kullanılır [23]. Ağırlıklara göre kayıpları hesaplayıp kaybın azaltılması için bir optimizasyon tekniği kullanılarak ağırlıkların optimize edilmesi gerekmekte olup optimizasyon tekniğinin uygulanması için aktivasyon fonksiyonun türevlenebilir olması gerekmektedir. Literatürdeki çalışmalarda Sigmoid, Tanh, ReLu (Rectified Linear Unit), SoftMax yaygın olarak kullanılan aktivasyon fonksiyonlarıdır [24]. Havuzlama katmanı önemli bilgileri koruyup gereksiz detayları azaltmasıyla özellik haritasının boyutunu küçülterek ağın hesaplama karmaşıklığını azaltır [25]. Tam bağlantılı katmanda sinir ağlarıyla öğrenme işlemi gerçekleştirilir. CNN modellerinin, bilgisayarlı görü alanında uygulamaları yaygın olarak kullanılmasının yanı

Tablo 1. Modellerin eğitim, doğrulama ve test başarıları
(Training, validation and test accuracy of models)

Model	Evrişim Katmanı Sayısı	En Yüksek Doğrulama Başarısı Tur Sayısı	Eğitim Başarısı	Doğrulama Başarısı	Test Başarısı	En Düşük Doğrulama Kaybı Tur Sayısı	Eğitim Başarısı	Doğrulama Başarısı	Test Başarısı
Model 1	1	149 (Şekil 2 (a))	0.918	0.782	0.364	61 (Şekil 2 (d))	0.858	0.775	0.542
Model 2	2	134 (Şekil 2 (b))	0.923	0.789	0.453	90 (Şekil 2 (e))	0.887	0.778	0.622
Model 3	3	83 (Şekil 2 (c))	0.898	0.785	0.524	**55 (Şekil 2 (f))**	**0.861**	**0.773**	**0.694**

sıra ses ve sinyal verileri üzerinde de etkili uygulamalar geliştirilmiştir [12-15].

Çalışmada veri seti olarak kullandığımız AdresGezgini A.Ş. bünyesinde gerçekleştirilen çağrı merkezi görüşme kayıtlarından duyguyu tanımlamamıza yardımcı olabilecek daha derin kalıpları ortaya çıkarmak amacıyla MFCC öznitelikleri çıkarılmış olup 40x216 boyutundaki bu öznitelikler Şekil 1'de genel yapısı verilen CNN modelinin eğitim verisini oluşturmuştur.

Önerilen CNN mimarisinde, benzer şekilde yapılandırılmış üç evrişim katmanını iki tamamen bağlı katman izlemektedir. Tamamen bağlı katmandan gelen çıktı düzleştirme katmanı (flatten layer) ile tek boyutlu diziye çevrilir. Softmax aktivasyon fonksiyonlu sınıflandırma katmanı önceki katmalardan gelen girdilere bağlı olarak olumlu olumsuz olmak üzere iki sınıflı çıktı üretir. Her evrişim katmanının dolgu parametresi 'same', adım değeri 1, filtre boyutu (7,7) ve filtre sayısı 32 olarak belirlenmiştir. Evrişim katmanları boyunca ReLu aktivasyon fonksiyonu kullanılmıştır. Evrişim katmanlarını maksimum havuzlama katmanı takip etmektedir. Maksimum havuzlama katmanlarının çekirdek boyutu (kernel size) (5,5), dolgu 'same' şeklindedir. Düzleştirme katmanının bağlandığı Softmax katmanının olumlu ve olumsuz ses kayıtlarını sınıflandırmak için iki düğümü vardır. Önerilen modelin eğitim, doğrulama ve test verileri üzerindeki başarısı Bölüm 3'te açıklanmıştır.

3. SONUÇLAR VE TARTIŞMA (RESULT AND DISCUSSION)

Ses verisinden duyguyu tanımlamamıza yardımcı olabilecek daha derin kalıpları ortaya çıkarmak amacıyla MFCC öznitelikleri çıkarılarak 40x216 boyutundaki bu öznitelikler derin öğrenme mimarisinin eğitim verisi olarak kullanılmıştır. Az sayıdaki veri setiyle en yüksek doğruluktaki modeli elde etmek amacıyla model mimarisinde evrişim katmanlarının sayısı değiştirilerek oluşturulmuş farklı modeller eğitilmiştir. Deneysel çalışmaların gerçekleştirildiği bu modellerde sadece evrişim katmanlarının sayısı değişmekte olup model mimarisindeki tüm katmanlar ve parametreler sabit tutulmuştur. Model 1 tek evrişim katmanı, Model 2 iki evrişim katmanı, Model 3 üç evrişim katmanından oluşmakta olup bu katmanları maksimum havuzlama katmanları takip etmektedir. Her üç model MFCC öznitelikleriyle eğitilmiş olup veri setinin %80'i eğitim %20'si doğrulama verisi olarak ayrılmıştır. Model eğitim parametreleri, eğitim tur sayısı (epoch) 200, öğrenme oranı (learning rate) 0,001, batch boyutu 32 ve Adam optimizasyon algoritması olarak belirlenmiştir. Ayrıca Tablo 1'de verilen üç model dışında dört ve beş evrişim katmanından oluşan modellerde eğitilmiş ve performansları değerlendirilmiştir. Modellerin eğitim süreçlerinin başlangıcından itibaren aşırı öğrenme (overfitting) gösterdiği gözlemlenmiştir. Veri setindeki örnek sayısının az olması nedeniyle dört ve beş evrişim katmanından oluşan modellerin veri setine göre karmaşık yapıda olduğu sonucuna varılmıştır.

Çalışmada CNN modeli ile çağrı merkezi kayıtlarının olumsuz ve olumlu olarak sınıflandırması önerilmekte olup modellerin bu sınıflandırma problemindeki performansları sınıflandırma doğruluğu (Denklem 1), hassasiyet (Denklem 2), duyarlılık (Denklem 3) ve F1-puanı (Denklem 4) ile değerlendirilmiştir.

$$Doğruluk = \frac{DP + DN}{DP + DN + YP + YN} \quad (1)$$

$$Hassasiyet = \frac{DP}{DP + YP} \quad (2)$$

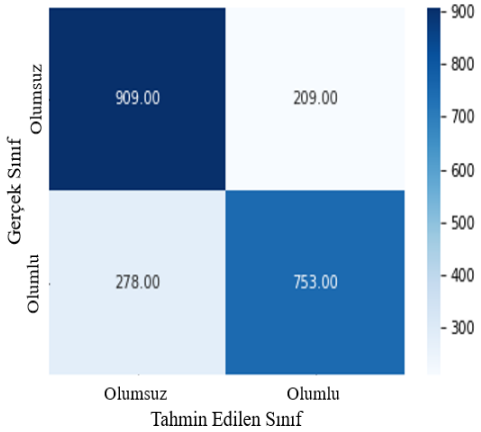
$$Duyarlılık = \frac{DP}{DP + YN} \quad (3)$$

$$F1\ Puanı = 2x\frac{Hassasiyet\ x\ Duyarlılık}{Hassasiyet + Duyarlılık} \quad (4)$$

DP (Doğru-Pozitif) ve DN (Doğru-Negatif) sırasıyla doğru tahmin edilen doğru pozitif ve doğru negatif çıktıların miktarını tanımlarken, YP (Yanlış-Pozitif) ve YN (Yanlış-Negatif) sırasıyla yanlış tahmin edilen yanlış pozitif ve yanlış negatif çıktıların sayısıdır. Hassasiyet, duyarlılık ve F1-puanı, tahmin edilen pozitif ve negatif çıktıların oranını hesaplayarak bir sınıflandırıcının performansını istatistiksel olarak değerlendirir. Hassasiyet, doğru tahmin edilen pozitiflerin toplam pozitif tahminlere oranıdır, duyarlılık ise doğru tahmin edilen pozitiflerin toplam gerçek pozitiflere ve yanlış negatiflere oranıdır. Son olarak, hassasiyet ve geri çağırmanın harmonik ortalaması, 1 en iyi ve 0 en kötü

olmak üzere [0,1] arasında bir değere sahip F1-puanı verir.

Her üç modelin eğitim ve doğrulama başarısı 200 eğitim tur sayısı boyunca izlenmiş olup modellerin en yüksek doğrulama başarısı ve en düşük doğrulama kaybı gösterdiği tur sayısı, Şekil 2'de verilen eğitim, doğrulama başarısı ve kayıp grafiklerinde siyah nokta ile işaretlenmiştir. Eğitilen modellerin 200 eğitim tur sayısı boyunca en yüksek doğrulama başarısı ve en düşük doğrulama kaybının elde edildiği tur sayısındaki doğrulama başarıları Tablo 1'de verilmiş ve aynı katman sayısındaki modellerin farklı tur sayılarında kaydedilen başarıları karşılaştırılmıştır. Tablo 1'de görüldüğü üzere en yüksek doğrulama başarısının elde edildiği turdaki başarı değerleri ve en düşük doğrulama kaybının elde edildiği tur sayısındaki doğrulama başarılarının benzer olduğu gözlemlenmektedir. Fakat modellerin gerçek hayat problemlerindeki başarılarının karşılaştırılması için eğitim ve doğrulama veri setinde bulunmayan çağrı merkezi kayıtları ile test edilmiştir. Test veri seti 7 farklı çağrı merkezi kaydının üçer saniyelik kayıtlara bölünmesiyle elde edilmiş 3600 ses kaydı içermektedir. Test verisiyle gerçekleştirilen başarı değerlendirmesi sonucunda en düşük doğrulama kaybının elde edildiği tur sayısında kaydedilen modellerin test verisi üzerindeki sınıflandırma başarısının daha yüksek olduğu görülmüştür. Modellerin evrişim katmanı sayısına bağlı performansı incelendiğinde evrişim katmanı sayısındaki artış ile modelin test başarısının arttığı gözlemlenmiştir. Test verileri üzerinde gerçekleştirilen deneyler sonucunda üç evrişim katmanı içeren Model 3, 55 eğitim tur sayısında %69,4 ile en yüksek test başarısını vermiştir.



Şekil 3. Model 3 karmaşıklık matrisi
(Confusion matrix of Model 3)

Tablo 2'de Model 3' ün hassasiyet, duyarlılık, F1-puanı performans metrikleri ile değerlendirilmesi verilmiştir. Hassasiyet, duyarlılık ve F1-puan için ortalama 0.77 olmasına rağmen olumsuz ve olumlu sonuçları için değerler farklılık göstermektedir. Modelin sınıf bazında değerlendirilmesi Şekil 3'te verilen karmaşıklık matrisi ile açıklanmıştır. Karışıklık matrisi, doğru etiket ve tahmin etiketleri arasındaki ilişkiyi göstermektedir. Modelin doğrulama verisi üzerinden oluşturulmuş Şekil 3'teki karmaşıklık matrisinde Model 3' ün 1118 olumsuz olarak sınıflandırılan çağrı merkezi kayıtlarından 909 tanesini olumsuz olarak doğru bir şekilde sınıflandırdığı olumlu sınıfındaki 1031 veriden ise 753'ünü olumlu olarak doğru bir şekilde sınıflandırdığı gözlenmektedir.

Tablo 2. Model 3'ün sınıflandırma başarısının hassasiyet, duyarlılık ve F1-puanı ile değerlendirilmesi (Evaluation of the Model 3 classification performance in terms of precision, recall and F1-score)

Sınıflar	Hassasiyet	Duyarlılık	F1-puanı
Olumsuz	0.76	0.81	0.79
Olumlu	0.78	0.73	0.76
Ağırlık Ortalama	0.78	0.78	0.78

Önerilen sistem Python programlama diliyle Tensorflow kütüphanesi için Keras arayüzü kullanılarak oluşturulmuştur. Verilerin düzenlenmesi için Numpy ve Pandas kütüphaneleri verilerin görselleştirmeleri için ise Matplotlib kütüphaneleri kullanılmıştır. Model eğitimleri, 16 çekirdekli 3.70GHz Intel(R) Xeon(R) W-2145 CPU, 64 GB 2666 MHz DDR4 RAM ve GeForce RTX 2080 8GB GPU' ya sahip bir Dell Precision 5820 iş istasyonunda gerçekleştirildi.

4. SONUÇ (CONCLUSION)

Gerçekleştirilen çalışma ile yoğun telefon görüşmelerinin gerçekleştirildiği çağrı merkezlerinde kayıt altına alınan telefon görüşmelerinin otomatik olarak analizi yapılarak, müşteri ve müşteri temsilcisi, arasında gerçekleşen görüşmelerde, müşteri memnuniyeti için çok büyük bir sorun teşkil eden olası münakaşa veya münakaşa benzeri duygu değişimlerinin otomatik olarak tespiti sağlanmaktadır. Münakaşa içeren görüşmelerin otomatik olarak tespiti yapılıp kalite yöneticilerine bilgi verilerek anında duruma müdahale şansı oluşturulacak sistemin yapay zeka modelinin oluşturulması çalışmanın önemli bir özelliğidir. Çalışmanın başarısı literatürdeki [3, 5, 8] çalışmalarla kıyaslandığında çok yakın [5] veya daha yüksek [3, 8] başarı gösterdiği görülmektedir. Çağrı merkezi kayıtlarından oluşturulan veri seti ile eğitilmiş CNN modelinin doğrulama başarısının %77,3 olması ve eğitim sonuçlarına ek olarak çağrı merkezi görüşmeleri üzerinde yapılan testlerde sistemin başarısının kullanılan az sayıda veriye rağmen %69,4 olması sistemin kullanılabilirliği açısından önem arz etmektedir. Ayrıca çağrı merkezi kayıtlarından oluşturulan özel Türkçe veri seti de çalışmanın özgünlüğünü göstermektedir. Veri seti büyüklüğü ile başarı oranı genellikle artma eğiliminde olduğu göz önünde bulundurularak ilerleyen çalışmada kullanılan veri setinin çağrı merkezi kayıtları ile genişletilmesiyle başarının artırılması ve konuşmacı tanıma çalışması gerçekleştirilerek konuşmadaki olumsuzluğun müşteri veya çağrı merkezi çalışanından mı kaynaklı olduğunun belirlenmesi hedeflenmektedir.

5. TEŞEKKÜR (ACKNOWLEDGEMENT)

Çalışma TÜBİTAK TEYDEB 1501 programıyla desteklenmekte olan 3200788 numaralı “Olumsuz Çağrı Merkezi Görüşmelerinin Derin Yapay Sinir Ağları Tabanlı Sınıflandırma Algoritmaları ile Otomatik Tespitini Sağlayan Sistemin Geliştirilmesi” adlı proje kapsamında geliştirilmiştir.

KAYNAKLAR (REFERENCES)

[1] B. Özlan, A. Haznedaroğlu, L. M. Arslan. "Automatic fraud detection in call center conversations", **Signal Processing and Communications Applications Conference (SIU)**, Sivas, Türkiye, 27, 2019.

[2] L. O. Iheme, Ş. Ozan, "A novel semi-supervised framework for call center agent malpractice detection via neural feature learning", *Expert Systems with Applications*, 118173, 2022.

[3] D. Pappas, I. Androutsopoulos, H. Papageorgiou, "Anger Detection in Call Center Dialogues", **6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)**, Macaristan, 139-144, 2015.

[4] Ş. Ozan, "Classification of Audio Segments in Call Center Recordings using Convolutional Recurrent Neural Networks", *arXiv preprint arXiv:2106.02422*, 2021.

[5] D. Turnbull, C. Elkan, "Fast Recognition of Musical Genres Using RBF Networks", *IEEE Transactions on Knowledge and Data Engineering*, 17(4), 580-584, 2005.

[6] L. O. Iheme, Ş. Ozan, "Multiclass Digital Audio Segmentation with MFCC Features using Naive Bayes and SVM Classifiers", **Innovations in Intelligent Systems and Applications Conference (ASYU)**, İzmir, Türkiye, 1-5, 2019.

[7] M. S. Likitha, S. S. R. Gupta, K. Hasitha, A. U. Raju, "Speech Based Human Emotion Recognition using MFCC", **International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)**, Hindistan, 2257-2260, 2017.

[8] A. Milton, S. S. Roy, S. T. Selvi, "SVM Scheme for Speech Emotion Recognition using MFCC Feature", *International Journal of Computer Applications*, 69(9),34-39, 2013.

[9] V. B Waghmare, R. R. Deshmukh, P. P. Shrishrimal, G. B. Janvale, "Emotion Recognition System from Artificial Marathi Speech using MFCC and LDA Techniques**", Fifth International Conference on Advances in Communication, Network, and Computing–CNC**, Hindistan, 2014.

[10] S. Demircan, H. Kahramanlı, "Feature Extraction from Speech Data for Emotion Recognition", *Journal of Advances in Computer Networks*, 28-30, 2014

[11] N. J. Nalini, S. Palanivel, "Music emotion recognition: The combined evidence of MFCC and residual phase", *Egyptian Informatics Journal*, *17*(1), 1-10, 2016.

[12] W. Lim, D. Jang, T. Lee. "Speech Emotion Recognition using Convolutional and Recurrent Neural Networks", **Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)**, Kore 1-4, 2016.

[13] J. Zhao, X. Mao, L Chen, "Speech Emotion Recognition using Deep 1D & 2D CNN LSTM Networks", *Biomedical Signal Processing and Control*, 47, 312-323, 2019.

[14] S. Latif, R. Rana, S. Khalifa, R. Jurdak, J. Epps, “Direct Modelling of Speech Emotion from Raw Speech”, *arXiv preprint ArXiv:1904.03833*, 2019.

[15] J. Wang, M. Xue, R Culhane, E. Diao, J. Ding, V. Tarokh, "Speech Emotion Recognition with Dual-Sequence LSTM Architecture", **ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, 6474-6478, 2020.

[16] S. K. Pandey, H. S. Shekhawat, S. R. M. Prasanna, "Deep Learning Techniques for Speech Emotion Recognition: A Review", **29th International Conference Radioelektronika (RADIOELEKTRONIKA)**, Çek Cumhuriyeti, 1-6, 2019.

[17] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss “A Database of German Emotional Speech”, *Interspeech*, 1517-1520, 2005.

[18] S. Wu, T. H. Falk, W. Y. Chan, “Automatic Speech Emotion Recognition using Modulation Spectral Features”, *Speech Communication*, *53*(5), 768-785, 2011.

[19] M. Grimm, K. Kroschel, S. Narayanan, “The Vera am Mittag German Audio-Visual Emotional Speech Database”, **IEEE International Conference on Multimedia and Expo**, Almanya, 865-868, 2008.

[20] M. Yıldırım, “MFCC Yöntemi ve Önerilen Derin Model ile Çevresel Seslerin Otomatik Olarak Sınıflandırılması”, *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *34*(1), 449-457, 2022.

[21] M. Scarpiniti, D. Comminiello, A. Uncini, Y. C. Lee” Deep Recurrent Neural Networks for Audio Classification in Construction Sites” **28th European Signal Processing Conference (EUSIPCO)**, Hollanda, 810- 814, 2020.

[22] S. K. Roy, G. Krishna, S. R. Dubey, B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification", *IEEE Geoscience and Remote Sensing Letters*, 17(2), 277-281, 2019.

[23] H. Wang, J. Zhou, C. Gu, H. Lin, "Design of Activation Function in CNN for Image Classification", *Journal of Zhejiang University (Engineering Science)*, 53(7), 1363-1373, 2019.

[24] Y. Wang, Y. Li, Y. Song, X. Rong, "The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition", *Applied Sciences*, 10(5), 1897, 2020.

[25] M. A. Kizrak, B. Bolat, “Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma”, *Bilişim Teknolojileri Dergisi*, *11*(3), 263-286, 2018.