

# Yeni bir Koronavirüs Konuşma Veri Kümesi: CoSDa

## A Novel Coronavirus Speech Dataset: CoSDa

Gökçe Yılmaz<sup>1</sup>, Recep Tektaş<sup>1</sup>, Mevlüde Sıla Kazan<sup>1</sup>, Alain Patrick Ndigande<sup>1</sup>, Meltem Kurt Pehlivanoğlu<sup>1</sup>

<sup>1</sup>Bilgisayar Mühendisliği Bölümü, Kocaeli Üniversitesi, Kocaeli, Türkiye  
{190201073, 200202122, 190201007, 160201132, meltem.kurt}@kocaeli.edu.tr

**Özetçe—** Bu çalışmada, makine öğrenmesi modellerini eğitmek için, CoSDa olarak adlandırılan, yeni bir koronavirüs (COVID-19) konuşma veri kümesi tanıtılmıştır. CoSDa, bildiğimiz kadarıyla, COVID-19 hastalarının konuşma kayıtlarını içeren halka açık ilk Türkçe konuşma veri kümesidir. Veri kümesi, sağlıklı insanların ve COVID-19 hastalarının YouTube videolarından toplanmıştır. CoSDa veri kümesi toplam 99 sağlıklı (COVID-19 negatif) ve toplam 92 sağlıklı (COVID-19 pozitif) Türkçe konuşma kaydı sağlamaktadır. EfficientNet ve ResNet (Residual Network) mimarileri iki sınıflı sınıflandırıcı olarak COVID-19 konuşmalarını sağlıklı insanlardan ayırmak için kullanılmıştır. ResNet-152 sınıflandırıcısı ile %99,64 doğruluk, %99,51 F1-Skoru, %99,50 kesinlik ve %99,59 duyarlılık ile en iyi değerler elde edilmektedir.

**Anahtar Kelimeler—**COVID-19; konuşma veri kümesi; CoSDa ; makine öğrenmesi.

**Abstract—** In this paper, a novel coronavirus (COVID-19) speech dataset is introduced, called CoSDa, to train machine learning models. According to our knowledge, CoSDa is the first publicly available Turkish speech dataset that includes the speech recordings of COVID-19 patients. The dataset is collected from YouTube videos of healthy people and COVID-19 patients. The CoSDa dataset provides a total of 99 healthy (COVID-19 negative) and a total of 92 unhealthy (COVID-19 positive) Turkish speech recordings. The EfficientNet and ResNet (Artık Ağ- Residual Network) architectures have been used as two-class classifiers to separate the COVID-19 speech from healthy people. The ResNet-152 classifier can achieve the best values with an accuracy of 99.64%, an F1-Score of 99.51, a precision of 0.9950, and a recall of 0.9959.

**Keywords—** COVID-19; speech dataset; CoSDa ; machine learning.

### I. GİRİŞ

2020 yılında ortaya çıkan COVID-19 pandemisi kısa sürede tüm dünyayı etkisi altına almıştır. COVID-19 semptomları bireye bağlı olarak önemli ölçüde değişir ve bir birey etkilendikten sonra belirtilerin ortaya çıkması yaklaşık 14 gün kadar sürebilir. Ayrıca öksürük, yorgunluk ve ateş

gibi belirtiler, soğuk algınlığı ve/veya grip ile de kolaylıkla karışmaktadır. Bunun yanı sıra virüs akciğerlere girdiğinde solunum güçlüğüne, zatürreye ve ölüme neden olabilir. Bu nedenle enfeksiyonun akciğerlere taşınmasından önce erken teşhis amaçlı, yapay zeka tabanlı sistemlerin tasarımı önemli bir çalışma alanı haline gelmiştir [1]. Bu kapsamda literatürde, özellikle ses sinyalleri üzerinden COVID-19 tespiti amaçlı çalışmalarda kullanılmak üzere, COVID-19 pozitif ve sağlıklı kişilere ait öksürük ve konuşma kayıtlarını içeren ses veri kümeleri mevcuttur.

COVID-19 pozitif kişileri içeren öksürük ve konuşma ses veri kümeleri incelendiğinde; Brown ve arkadaşlarının çalışmasında yer alan COVID-19 Sounds veri kümesinde [2], 235'i COVID-19 pozitif, 7000 farklı kullanıcıdan alınan 10000 öksürük ve nefes alma örneği yer almaktadır. Ancak bu veri kümesi erişime açık değildir. NoCoCoDa veri kümesi [1], COVID-19 hastalarıyla yapılan medya röportajlarından (YouTube üzerinden) seçilen öksürük seslerinden oluşan bir veri kümesidir. Ancak bu veri kümesinde sadece 10 farklı kişiye ait öksürük ses dosyaları yer almaktadır. Ancak bu kadar az örneğin küresel popülasyonu başarılı bir şekilde genellemesi mümkün değildir. Coswara projesinde [3] ise, öksürüğün yanı sıra nefes alma, sayı sayma, sesli harfleri okuma seslerini de içeren 1123'ü erkek, 363'ü kadın toplam 1486 ses dosyası yer almaktadır. Çalışmaya gönüllü olarak katılan katılımcıların 1329'u Hindistan, 157'si ise farklı ülkelerden (Asya, Avustralya, Avrupa, Kuzey Amerika ve Güney Amerika) katılımcıları içermektedir. Bu katılımcıların 107 tanesi COVID-19 pozitifdir. Virufy öksürük veri kümesinde [4] ise 7 farklı COVID-19 pozitif ve 9 farklı sağlıklı olmak üzere 16 katılımcıdan alınan öksürük ses verileri yer almaktadır. Bu veri kümesi, 16 farklı kişiye ait toplam 121 farklı segmente edilmiş öksürük örneği içermektedir. COUGHVID veri kümesinde [5], 1010 farklı COVID-19 pozitif ve 19062 sağlıklı öksürük verisi yer almaktadır.

Yukarıda verilen veri kümelerine ait veriler Tablo I'de özetlenmiştir.

| Ref. | Veri Kümesi     | Kategori                                                        | # COVID-19 +Sağlıklı | Erişim         |
|------|-----------------|-----------------------------------------------------------------|----------------------|----------------|
| [2]  | COVID-19 Sounds | Öksürük, Nefes alma                                             | 235/6765             | Erişime Kapalı |
| [1]  | NoCoCoDa        | Öksürük                                                         | 10/Bilinmiyor        | Erişime Kapalı |
| [3]  | Coswara         | Öksürük, Nefes alma, Sayı sayma, Sesli harfleri okuma seslerini | 107/1379             | Erişime Açık   |
| [4]  | Virufy          | Öksürük                                                         | 7/9                  | Erişime Açık   |
| [5]  | COUGHVID        | Öksürük                                                         | 1010/19062           | Erişime Açık   |

**Tablo I.** COVID-19 öksürük ve konuşma ses veri kümeleri

Yukarıda verilen veri kümelerinde ana dili Türkçe olan COVID-19 pozitif ve sağlıklı katılımcıların yer aldığı öksürük ve konuşma veri kümesine rastlanmamıştır. Ana dil bir model belirlerken önemli bir özelliktir. Çünkü farklı dilleri konuşan (ana dilleri farklı olan) kişilerin ses yollarında da morfolojik farklılıklar olmaktadır, bu farklılıklar da akustik farklara sebep olmaktadır [6, 7]. Bu nedenle, özellikle COVID-19 pozitif ve sağlıklı kişilerin yer aldığı konuşma veri kümeleri oluşturulurken Türkçe konuşma kayıtlarını içeren bir ses veri kümesinin oluşturulması önemli bir açık problemdir.

#### A. Motivasyon ve Katkı

Literatürde COVID-19 tespiti amaçlı oluşturulan öksürük ve konuşma veri kümeleri incelendiğinde, özellikle ana dili Türkçe COVID-19 pozitif hastaların konuşma ses kayıtlarını içeren bir veri kümesinin olmadığı tespit edilmiştir.

Bu çalışmada COVID-19 pozitif (Polimeraz Zincirleme Tepkimesi (Polimeraz Chain Reaction-PCR) test sonucu pozitif: 40 kadın, 52 erkek toplam 92 kişi) ve sağlıklı (51 kadın, 48 erkek toplam 99 kişi) olduğu bilinen toplam 191 farklı kişiye ait video YouTube üzerinden toplanıp gerekli dönüşümler yapılarak, CoSDa olarak adlandırılan, özgün bir Türkçe konuşma ses veri kümesi (yaklaşık 20.7 GB) oluşturulmuştur. Çalışma kapsamında literatüre kazandırılan özgün veri kümesi erişime açık olarak paylaşılacaktır.

Çalışmada ayrıca, CoSDa veri kümesi üzerinde Evrişimsel Sinir Ağları (Convolutional Neural Network-CNN) [8] tabanlı EfficientNet mimarisi [9] ve ResNet mimarileri [10] iki sınıflı sınıflandırıcı olarak COVID-19 hastalarının konuşmasını sağlıklı insanlardan ayırt etmek için kullanılmıştır. Modelin başarımını test etmek amaçlı doğruluk (accuracy), F1-skor (F1-measure), kesinlik (precision), duyarlılık (recall) metrikleri kullanılmıştır. Kullanılan sınıflandırıcılar arasından en iyi başarımlar oranları, bu metrikler için sırasıyla 0.9964, 0.9951, 0.9950, 0.9959 olarak ResNet-152 modeli ile elde edilmiştir.

#### B. Organizasyon

Çalışmanın ikinci bölümünde, çalışma kapsamında ilk defa literatüre kazandırılan CoSDa Türkçe konuşma veri

kümesine ait bilgiler verilmiştir. Üçüncü bölümde ise, bu veri kümesi üzerinde EfficientNet ve ResNet mimarileri kullanılarak COVID-19 pozitif ve sağlıklı ses kayıtları sınıflandırılmıştır. Son bölümde ise elde edilen sonuçlar tartışılarak ileriki çalışmalar verilmiştir.

## II. CoSDa COVID-19 KONUŞMA VERİ KÜMESİ

Çalışma kapsamında ana dili Türkçe olan, 39'u kadın, 52'si erkek toplam 91 COVID-19 hastası ile 51'i kadın ve 48'i erkek toplam 99 sağlıklı (COVID-19 negatif) kişiye ait 191 farklı video YouTube üzerinden toplanarak yaklaşık 20.7 GB'lık, CoSDa olarak adlandırılan, veri kümesi oluşturulmuştur.

Ses verilerinin toplanması sürecinde kişilerin YouTube kanalları üzerinden herkesin erişimine açık olarak sunduğu videolardan yararlanılmıştır. 2019 yılında başlayan COVID-19 pandemisi dikkate alınarak, sağlıklı kişilerin ses (konuşma) verileri 2018 yılı ve öncesinde paylaşılan videolardan toplanmıştır. Veri toplama sürecinde veriler ayırt edici özelliklerine göre COVID-19 ve sağlıklı olarak etiketlenmiştir. Her verinin isimlendirilmesinde öncelikle, video id (kişi id bilgisi), cinsiyet (kadın/erkek) ve kişinin kaçınıcı konuşma kesiti olduğu bilgileri sırasıyla kullanılmıştır. Videolar kesit bilgisinin eklenmesinin nedeni; özellikle COVID-19 pozitif hastalara ait videolarda hasta kişinin dışında konuşan diğer kişiler videodan temizlenmiştir, bu nedenle aynı kişiye ait bir video birkaç kesit içerebilmektedir. Temizlenen ve etiketlenen bu videolar WAV (Waveform Audio File Format) uzantılı ses dosyalarına dönüştürülerek veri kümesi oluşturulmuştur. Dönüşüm sırasında örnekleme oranı (sampling rate) 44.1 kHz (Kilohertz) seçilmiştir. Bu oran bu dönüşümler için kullanılan en iyi değerdir [1].

Veri toplama aracı olarak; videoların indirilmesi işlemi için Internet Download Manager [11] ve safefrom.net [12] uygulamaları kullanılmıştır. Video içeriklerindeki konuşma sahnelerinin ayrılıp düzenlenmesi için Windows Video Editör [13] kullanılmıştır. Düzenlenen videoların WAV uzantılı ses dosyalarına dönüştürülmesi için VLC Media Player [14] kullanılmıştır.

CoSDa COVID-19 konuşma veri kümesinde yer alan ses verilerine ilişkin ayrıntılı bilgiler Tablo II'de verilmiştir. Tabloda ses verileri, sağlıklı ve COVID-19 pozitif olmak üzere iki ana sınıf olarak verilmiştir. Her bir sınıfa ait kadın ve erkek sayıları, toplam konuşma süresi ve dosya boyut bilgileri sunulmuştur.

| Hasta Durumu | Sağlıklı (COVID-19 negatif) |                  | Hasta (COVID-19 pozitif) |                  |
|--------------|-----------------------------|------------------|--------------------------|------------------|
|              | Kadın                       | Erkek            | Kadın                    | Erkek            |
| Kişi Sayısı  | 51                          | 48               | 40                       | 52               |
| Kayıt Süresi | 8 Saat 56 Dakika            | 9 Saat 27 Dakika | 5 Saat 11 Dakika         | 4 Saat 39 Dakika |
| Dosya Boyutu | 6.56 GB                     | 6.95 GB          | 3.81 GB                  | 3.39 GB          |

**Tablo II.** CoSDa veri kümesi özellikleri

### III. CoSDa COVID-19 KONUŞMA VERİ KÜMESİ ÜZERİNDE İKİLİ SINIFLANDIRMA

Çalışma kapsamında verilen CoSDa veri kümesi üzerinde; COVID-19 pozitif ve sağlıklı olmak üzere ikili sınıflandırma gerçekleştirilmiştir. Modelde EfficientNet ve ResNet mimarileri sınıflandırıcı olarak seçilmiş ve bu yöntemler birbirleriyle doğruluk, F1-skor, kesinlik, duyarlılık metrikleri kullanılarak karşılaştırılmıştır.

CNN modeli tabanlı transfer öğrenim mimarilerinde, öğrenme aktarımı sayesinde yüksek başarımlı sonuçlar elde edilir. Bu modeller özellikle görüntü işleme alanında yapılan çalışmalarda önemli iyileştirmeler sağlamıştır.

EfficientNet, doğruluğun yanı sıra aynı anda modellerin verimliliğini de geliştirmeye odaklanan bir mimariye sahiptir. EfficientNet, B0-B7 arasında 8 farklı modelden oluşur. Model üç ayrı parametre kullanarak ölçeklendirme yapar. Bu parametreler: derinlik, genişlik ve çözünürlük. Derinlik parametresi ağların ne kadar derin olduğunu ölçerken, genişlik parametresi ise katmanlardaki nöron sayısıdır. Çözünürlük parametresi modelin eğitileceği veri kümesinin çözünürlüğünü ifade eder. Çalışma kapsamında EfficientNet B0-B1-B2-B3 mimarileri ile sınıflandırma işlemi gerçekleştirilmiştir.

CNN mimarisinde daha fazla katman eklemek (ağı daha derin hale getirmek) teoride eğitim hatasının azalmasını sağlarken, pratikte aşırı uyum (over fitting) sebebiyle hatanın artmasına neden olur. Burada optimizasyon ve bozulma problemi ortaya çıkar. ResNet mimarisinde bu probleme çözüm odaklı artık bloklar kullanılır bu nedenle artık ağ olarak adlandırılır. ResNet mimarisinde artık blokta, ağ girişi doğrudan ağına çıkışına eklenir, böylece atlama bağlantısı oluşturulur. Bu atlama bağlantısı sayesinde ağı performansı katman sayısı arttığında da düşmez (kaybolan gradyan sorunu çözülür). Eğer ağı girişi ile çıkışı aynı ise kimlik bloğu kullanılır. Temelde artık bloktaki katmanlar ağ için önemli bilgiler öğrenebilirse modelin performansı artar. ResNet mimarisinde kullanılan ağı katman sayısına göre versiyonlar geliştirilmiştir. Bu çalışmada ResNet-50, ResNet-101 ve ResNet-152 mimarileri kullanılmıştır.

CoSDa veri kümesi için ikili sınıflandırma gerçekleştirilmeden önce, veri kümesinde yer alan her bir örneklem 4'er saniyelik parçalara (chunk) bölünerek modellere girdi verilmiştir. CoSDa veri kümesinde elde edilen parça sayıları Tablo III'de verilmiştir.

| Hasta Durumu                | Sağlıklı (COVID-19 negatif)   |                               | Hasta (COVID-19 pozitif)      |                               |
|-----------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
|                             | Kadın                         | Erkek                         | Kadın                         | Erkek                         |
| Toplam Parça (Chunk) Sayısı | 8045<br>(8045 x 4 = 32180 sn) | 8509<br>(8509 x 4 = 34036 sn) | 4663<br>(4663 x 4 = 18652 sn) | 4190<br>(4191 x 4 = 16760 sn) |

Tablo III. CoSDa veri kümesi üzerinde her bir örnekleme ait 4'er saniyelik parça sayıları

4'er saniyelik parçalara ayrılan her bir ses verisi, bu seslerin temsilleri olan Mel Frekanslı Cepstrum Katsayıları (Mel Frequency Cepstral Coefficient-MFCC) olarak

adlandırılan iki boyutlu matris şeklinde katsayılarla dönüştürülmüştür. Bu frekans ölçeği temelde, insan kulaklarının algılamasını taklit ederek, ses frekanslarındaki değişimi modeller. MFCC katsayıları ile ses sinyalinin kısa zamanlı güç spektrumu Mel ölçeği üzerinde (frekanstan Mel ölçeğine dönüşüm yapılır) ifade edilir. Ses sinyalinin MFCC katsayıları dönüştürülmesi adımı; çerçeveleme, pencereleme, Hızlı Fourier Dönüşümü (Fast Fourier Transform – FFT), Mel-frekansına çevirme ve Cepstrum adımları uygulanır. Çerçeveleme adımı; ses sinyalleri, her çerçevenin N tane öksürük ve konuşma örneğini ve önceki komşu çerçevenin belirli bir M tane ( $M < N$ ) örneğini içerir. Pencereleme adımı; her bir çerçevenin başındaki ve sonundaki süresizliğin önlenmesi amaçlanır. FFT adımı; N örnekli çerçeve FFT dönüşümüyle zaman bölgesindeki gösteriminden frekans bölgesindeki gösterimine (time domain-frequency domain) dönüştürülür. Mel-frekansına çevirme adımı ise; frekans bölgesine dönüştürülen çerçeveler, Mel-frekansına çevrilir. Cepstrum katsayıları adımı ise; ses sinyallerinin cepstrum değerleri elde edilerek frekans bölgesinden frequency bölgesine geçilir. Elde edilen MFCC katsayıları kullanılarak ilgili ses sinyallerini temsil eden öznetelikler çıkarılmıştır.

CoSDa veri kümesi %70 eğitim, %30 test olarak olmak üzere Tablo IV'de verildiği gibi ayrılmıştır.

| Durum               | Eğitim   |      |       |      | Test     |      |       |      |
|---------------------|----------|------|-------|------|----------|------|-------|------|
|                     | Sağlıklı |      | Hasta |      | Sağlıklı |      | Hasta |      |
|                     | K.       | E.   | K.    | E.   | K.       | E.   | K.    | E.   |
| Kişi Sayısı         | 41       | 32   | 32    | 37   | 10       | 16   | 8     | 15   |
| Toplam Parça Sayısı | 6458     | 5061 | 3573  | 2579 | 1587     | 3448 | 1090  | 1611 |

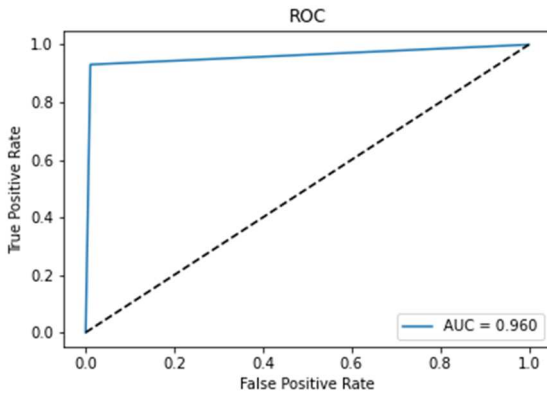
Tablo IV. CoSDa veri kümesinin eğitim ve test olarak ayrılması

Kurulan modellere ait elde edilen sonuçlar Tablo V'de karşılaştırmalı olarak verilmiştir. Tabloda her bir modele ait doğruluk, F1-skor, kesinlik, duyarlılık, iterasyon sayısı (epoch) ve ROC Eğrisinin Altındaki Alan (Area Under the ROC Curve-AUC) metrikleri verilmiştir. Modelde, EfficientNet mimarilerinden EfficientNet B0, EfficientNet B1 mimarileri, ResNet mimarilerinden ResNet-50, ResNet-101 ve ResNet-152 mimarileri kullanılmıştır. Modellere ait sonuçlar tabloya eklenmiştir. Tablodan da görüleceği gibi sınıflandırıcılar arasında en iyi başarımlı oranları, doğruluk, F1-skor, kesinlik, duyarlılık ve AUC metrikleri için sırasıyla 0.9964, 0.9951, 0.9950, 0.9959, 0.960 olarak, ResNet-152 modeli ile elde edilmiştir.

| Model           | Doğruluk | F1-skor | Kesinlik | Duyarlılık | İter. Sayısı | AUC    |
|-----------------|----------|---------|----------|------------|--------------|--------|
| EfficientNet B0 | 0.9884   | 0.9823  | 0.9815   | 0.9851     | 10           | 0.9996 |
| EfficientNet B1 | 0.9922   | 0.9874  | 0.9883   | 0.9883     | 10           | 0.9969 |
| EfficientNet B2 | 0.9935   | 0.9904  | 0.9900   | 0.9918     | 10           | 0.9975 |
| EfficientNet B3 | 0.9929   | 0.9893  | 0.9887   | 0.9910     | 10           | 0.9988 |
| ResNet-50       | 0.9839   | 0.9754  | 0.9972   | 0.9567     | 10           | 0.9780 |
| ResNet-101      | 0.9954   | 0.9929  | 0.9929   | 0.9940     | 10           | 0.945  |
| ResNet-152      | 0.9964   | 0.9951  | 0.9950   | 0.9959     | 10           | 0.960  |

**Tablo V.** CoSDa veri kümesi üzerinde farklı sınıflandırıcılar ile elde edilen başarımlar

Bunun yanı sıra ResNet-152 modeline ait Alıcı İşlem Karakteristikleri (Receiver Operating Characteristic-ROC) eğrisi Şekil 1’de verilmiştir.



**Şekil 1.** ResNet-152 modeline ait ROC eğrisi

#### IV. SONUÇ VE İLERİKİ ÇALIŞMALAR

Bu çalışmada, 40’ı kadın, 52’si erkek toplam 92 COVID-19 hastası ile 51’i kadın ve 48’i erkek toplam 99 sağlıklı kişiye ait 191 farklı video YouTube üzerinden toplanarak, WAV uzantılı ses dosyalarına (yaklaşık 20.7 GB’lık) 44.1 kHz örnekleme oranıyla dönüştürülmüştür. CoSDa olarak adlandırılan bu veri kümesi, bu çalışma kapsamında diğer çalışmaların da faydalanması amaçlı, Kaggle platformu üzerinden erişime açık olarak paylaşılmıştır [16].

Çalışmada ayrıca CoSDa veri kümesi üzerinde, COVID-19 pozitif ve sağlıklı olmak üzere ikili sınıflandırma gerçekleştirilmiştir. Modelde EfficientNet ve ResNet mimarileri sınıflandırıcı olarak seçilmiş ve bu yöntemler birbirleriyle doğruluk, F1-skor, kesinlik, duyarlılık, AUC metrikleri kullanılarak karşılaştırılmıştır. Yapılan deneyler sonucunda en iyi başarımlar, doğruluk, F1-skor, kesinlik, duyarlılık, AUC metrikleri için sırasıyla 0.9964, 0.9951, 0.9950, 0.9959, 0.960 olarak ResNet-152 modeli ile elde edilmiştir.

İleriki çalışmalarda veri kümesinde yer alan örneklem sayılarının ve sınıf sayısının artırılması planlanmaktadır. COVID-19 pozitif hastalarının yanı sıra, KOAH, astım, bronşit hastalıkları olan farklı kişilere ait videolar YouTube üzerinden toplanmaya başlanmıştır. CoSDa veri kümesi COVID-19 hastalığının tespiti amaçlı kullanılmasının yanı

sıra, ileride olası farklı hastalıkların erken teşhisinde önemli bir veri kümesi olarak kullanılabilir.

#### KAYNAKÇA

- [1] M. Cohen-McFarlane, R. Goubran and F. Knoefel, “Novel Coronavirus Cough Database: NoCoCoDa,” *IEEE Access*, vol. 8, pp. 154087-154094, 2020.
- [2] C. Brown, et al. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. *In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 11, 3474–3484, 2020.
- [3] Data repository of Project Coswara, <https://github.com/iiscleap/Coswara-Data>, Erişim Tarihi: 25 Haziran 2021.
- [4] Virufy COVID-19 Open Cough Dataset, <https://github.com/virufy/virufy-data>, Erişim Tarihi: 25 Haziran 2021.
- [5] L. Orlandic, T. Teijeiro, D. Atienza, “The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms,” *arXiv preprint arXiv:2009.11644*, 2020.
- [6] S. Xue, G. Hao, R. Mayo, “Volumetric measurements of vocal tracts for male speakers from different races,” *Clinical linguistics & phonetics*, 20. 691-702. 10.1080/02699200500297716, 2006.
- [7] J.H, Walton, R.F. Orlikoff, “Speaker race identification from acoustic cues in the vocal signal,” *Journal of speech and hearing research*, vol. 37,4, 1994.
- [8] L. Alzubaidi, J. Zhang, A.J. Humaidi, et al. “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J Big Data*, 8, 53, 2021.
- [9] EfficientNet, <https://arxiv.org/abs/1905.11946>, Erişim Tarihi: 25 Haziran 2021.
- [10] I. Amaya-Rodriguez et al., “ResNet”, In: Bernal J., Histace A. (eds) *Computer-Aided Analysis of Gastrointestinal Videos*. Springer, Cham, 2021.
- [11] A. L’Heureux, K. Grolinger, H. F. Elyamany and M. A. M. Capretz, “Machine Learning With Big Data: Challenges and Approaches,” *IEEE Access*, vol. 5, pp. 7776-7797, 2017.
- [12] Internet Download Manager, <https://www.internetdownloadmanager.com.tr/>, Erişim Tarihi: 25 Haziran 2021.
- [13] safefrom.net, <http://ww25.safefrom.net/?subid1=20210625-2350-3401-bb6c-2076f511f9bf>, Erişim Tarihi: 25 Haziran 2021.
- [14] Windows Video Editor, <https://www.howtogeek.com/355524/how-to-use-windows-10s-hidden-video-editor/>, Erişim Tarihi: 25 Haziran 2021.
- [15] VLC MediaPlayer, <https://www.videolan.org/index.tr.html>, Erişim Tarihi: 25 Haziran 2021.
- [16] CoSDa, <https://www.kaggle.com/rcptektas/a-novel-coronavirus-speech-dataset-cosda>, Erişim Tarihi: 05 Eylül 2021.