T.C.

BURSA ULUDAĞ ÜNİVERSİTESİ

Mühendislik Fakültesi

Bilgisayar Mühendisliği

Bölümü BMB3008-SUNUM

YÖNTEMLERİ

DÖNEM PROJESİ - FAZ II

SES TANIMA GÜVENLİĞİ

Takım Üyeleri

Berat Sercan Çelik- 032090095 **(Takım Lideri)**

Şevket Binali- 032090116 **(Raportör)**

Takım No: 32

25 Mayıs 2022

T.C.

BURSA ULUDAĞ ÜNİVERSİTESİ

Mühendislik Fakültesi

Bilgisayar Mühendisliği

Bölümü BMB3008-SUNUM

YÖNTEMLERİ

DÖNEM PROJESİNİN ÖZGÜNLÜK TAAHHÜTNAMESİ

Ekteki projenin özgün çalışmamızı içerdiğini taahhüt ederiz. Bu projede sunulan bilgiler, daha önce başkaları tarafından yayınlanmamış ve yazılmamıştır; proje metninde bahsigeçip, başkaları tarafından yapılmış olan çalışmalara ise uygun atıflar yapılmıştır.

Takım Lideri: Berat Sercan Çelik İmza:

Raportör: Şevket Binali İmza:

İletim Mektubu

Tarih: 25 Mayıs, 2022

Alıcı: Prof. Dr. Kemal Fidanboylu

Bursa Uludağ Üniversitesi

Gönderici: Berat Sercan Çelik, 32. Grup Lideri

Şevket Binali, 32. Grup Raportörü

Konu: Ses Tanıma Güvenliği için Proje Raporu

"Ses Tanıma Güvenliği" adlı çalışmamızın raporu ektedir. 25 Mayıs 2022 tarihli teklifimizde açıklanan görevleri tamamladık. Bu görevler dünya genelinde ses tanıma güvenliğinin önemini vurgulamak, medya platformlarının ve çağrı merkezleri üzerinden işlem yapan kuruluşların bu konudaki tutumlarını incelemek, uygulama için farklı modellerin incelenmesi, veri setlerinin toparlanması, toplanan verilerin Evrişimsel Sinir Ağı'nda (CNN) sınıflandırılması ve doğruluk oranının en üst düzeyde tutulmasıdır. Bu görevleri yerine getirmek için birincil ve ikincil araştırmalar yaptık. Ses tanıma güvenliği ile ilgili literatürü inceledik, ilgili kuruluşlardan veriler topladık. Konunun öneminin anlaşılması için gerçek video ve ses kayıtlarının aynısının sahtesini yaptık. Farklı modelleri derinine inceledik, uygun veri setlerini toplayıp ESA'da analiz ettik ve raporu yazdık. Çalışmalarımız sonucunda yöntemimiz %85 gibi yüksek doğruluk oranına sahiptir. Bu oran ciddi manada tatmin edicidir. Yapay sinir ağları ile bu denli doğruluk oranına sahip olduğumuz projemize gerekli destek verilirse daha fazla ilerleme kaydedeceğimizden şüphemiz yoktur. Herhangi bir ek bilgiye ihtiyaç duyulursa veya herhangi bir soruyu yanıtlayabilirsek lütfen bize bildirin. Yardımcı olmaktan mutluluk duyarız. Bize 032090095@ogr.uludag.edu.tr veya 032090116@ogr.uludag.edu.tr adresinden ulaşabilirsiniz. İlginiz için şimdiden teşekkür ederiz.

Saygılarımızla,

Şevket Binali

Berat Sercan Çelik

Ek: Ses Tanıma Güvenliği Raporu

Öz

Yapay zekanın gelişmesiyle birçok alanda kullanılması teknolojinin daha da ilerlemesini sağlamıştır. Bu alanlardan biri de kötü amaçla kullanıma açık olan deepfake (derin sahte) uygulamalarıdır. Sosyal medyada etkin bir şekilde kullanılan manipüle edilerek değiştirilmiş bu videolar insanlar tarafından ilgi görmüş ve eğlence amaçlı kullanılmıştır. Her ne kadar eğlence içeriği olarak kullanılsa da bu husus hafife alınmamalıdır. Siyasi, politik manipülasyon, algı operasyonu, yalan haberler önlenmesi zor durumlara sebep olabilir. Bu yüzden konu iyice irdelenmelidir. Google, Microsoft, Facebook, AWS gibi sosyal medya ve teknoloji geliştiricileri derin sahte uygulamalarının tespit ve analiz araştırmalarına destek sağlamakta ve açık kaynaklar sunmaktadır. Yaptığımız literatür araştırmasında çalışmaların genellikle derin sahte videolar üzerine olduğunu ve çözüm yöntemi olarak RNN (Yinelemeli Sinir Ağı) modellerini kullandıklarını gördük. Bu yüzden araştırmamızı derin sahte seslerin analizi konusunda yapmaya karar verdik. Oluşturulmuş sahte seslerin analizini yapay zeka ile tespit etmeye çalıştığımız bu projede kullandığımız yöntem CNN (Evrişimsel Sinir Ağı) modelini ASVspoof 2019 veri tabanında bulunan etiketlenmiş ses dosyalarını ses spektogramlarına çevirerek eğitmek ve yüksek doğruluk oranı elde etmektir.

Anahtar Kelimeler: Yapay Zeka, Sosyal Medya, Derin sahte, Manipülasyon, Spektogram, RNN, CNN, ASVspoof, Sahte ses analizi.

İçindekiler

Hızla ilerleyen konuşmacı tanıma teknolojisi, güvenlik için birçok sektörde kullanılmaktadır. Ses biyometrisi kullanılarak kimlik doğrulaması yapılmaktadır. Her bir insanın ağız yapısı, ses telleri ve konuşma tarzı kendine özgüdür, bu eşsiz durumlar insanın ses biyometrisini oluşturur. Bu hızla gelişen teknoloji, beraberinde güvenlik sorunlarınıda getirmiştir. Kayıttan oynatılan seslerin yanı sıra, bilgisayar yardımıyla oluşturulan ses gerçeğeoldukça yakınlaşmıştır.

Bilgisayar ile oluşturulan, sahte sesler ile gerçekleştirilen saldırıları etkili bir şekilde tespit etmek oldukça önemlidir. Yeni tür konuşma sentezi teknolojileri ve ses benzetim teknolojileri geliştikçe, sahte sesi tespit etmek oldukça zor olmaktadır. Bu araştırmada, gerçeksesleri ve sahte sesleri içeren veri seti kullanılmaktadır.

Bir sesin analizi için sesi spektrogramlara dönüştürüyor ve zamanla oluşan farklılıklar için frekans spektrumlarını inceliyoruz. Spektrogramlar, ses dosyalarının görsel temsilleridir. Sesin baskınlığını ve sıklığını tespit etmemizi sağlar. Bu verileri, sahte sesleri tanımlamak içinEvrimsel Sinir Ağı'nda (ESA) girdi olarak kullanıyor ve veriler burada sınıflandırıyoruz. Bu araştırmada ESA kategorik şekilde çaprazlama yaparak sınıflandırma için entropi değerlerini kullanır. Uygulamada yanlış bir sınıflandırma büyük zararlara yol açabileceğinden yüksek doğruluk oranı vazgeçilmezdir. Yüksek doğruluk elde ederken de sesi spektrogramlara dönüştürmenin büyük zaman gerektirdiğini farkettik. Bu araştırma, yapay zeka ile spektrogram tabanlı bir ses sınıflandırıcısının araştırmasıdır.

Yönetici Özeti

Konuşmacı tanıma programları medya ve çağrı merkezi ile işlem yaptıran tüm kuruluşlar için bir güvenlik mekanizmasıdır. Çalışmamız, hızla gelişen bilişim teknolojileri karşısında kişilerin, toplumların ve kuruluşların güvenlik problemleri sebebiyle ortaya çıkmıştır. Problemin kaynağı ise hızla gelişen teknolojinin kötü amaçlı kişiler tarafından kurnaz bir şekilde kullanılmasıdır. Çalışmamız, medya aracılığı ile yayılan sahte videolar, ses imzası kullanan kuruluşlar (banka, telekomünikasyon vb.) ve devletler gibi büyük oluşumların güvenliklerini sağlamayı amaçlamıştır.Bizim yöntemimiz derin öğrenme algoritmaları ile üretilen ve kişilerin algılayamayacağı sahteliği Evrimsel Sinir Ağları'nı (ESA) kullanarak tespit etmektir.

Bunu yaparken, biyolojik çeşitliliğimizin yapısını araştırıp formülüze ettik ve yapay sinir ağları ile dijitalleştirdik. Her insanın kendine ait eşsiz bir sesi olması bunu başarabilmemizin en önemli nedenlerindendir.Bizim çözümümüzün muhatabları global çapta, bankalar, telekomünikasyon şirketleri, medya kuruluşları, devletlerin istihbarat servisleri, ses imzası kullanan büyük şirketler ve daha niceleridir.Bizim çözümümüzün mevcut çözümlere göre en önemli artısı, derin öğrenme algoritmaları ile üretilmiş sahte verileri tespit edebilmesidir. Ayrıca tasarladığımız sistem kendi kendini yeni çıkan sahteciliklere karşıda eğitebilmektedir.Araştırmalarımız sonucunda ortaya koyduğumuz yöntem şüphesiz gelişmeye devam edecektir. Araştırmalarımız ve yöntemimizin en büyük dayanağı bilim ve teknolojidir. Ömür boyu sürecek bilimsel araştırmalara verdiğimiz önem ve geliştirdiğimiz, gelişmekte olan sistemimizi dünya çapında rekabete sokmak, bunu yaparken korsancılığın önüne geçmek yegane amaçlarımızdandır.

Giriş

İnsanlar arasında ve yazılımsal arayüzlerde hızla büyüyen ses oluşturma teknolojileri büyük oranda doğruluk ölçen ses biyometrisi stratejilerine ihtiyaç duyulmasına yol açmıştır. Ses doğrulama teknolojisinin doğruluğu, son yıllarda derin öğrenmenin yardımıyla büyük sıçramalar yaşamıştır. Bu kapsamda derin öğrenme tabanlı konuşma sentezi kullanarak sesleritaklit ederek sosyal manipülasyonlara, kimlik doğrulama sahteciliğine aynı şekilde güvenlik oluşturmaktır.

Derin öğrenme sahtekarlığının sonuçları, tüm hükümet sistemini istikrarsızlaştıracak kadar önemli değil; ancak, derin sahtekarlıklar, bireysel varlıklara muazzam derecede zarar verme yeteneğine sahiptir. Bunun nedeni, deepfake'lerin genellikle bir kişiyi ve/veya diğerleriyle ilişkilerini, kamuoyunu veya inançları etkileyecek kadar güçlü bir anlatı yaratma umuduyla hedef almasıdır. Bu, sahte telefon görüşmeleri veya konuşmalar oluşturmak için sesi değiştiren derin sahte sesli kimlik avı yoluyla yapılabilir. Deepfake kullanımının bir başka yöntemi de, zarar verici yorumları dile getiren bireyleri iletmek için medyayı manipüle eden uydurma açıklamalardır. [1].

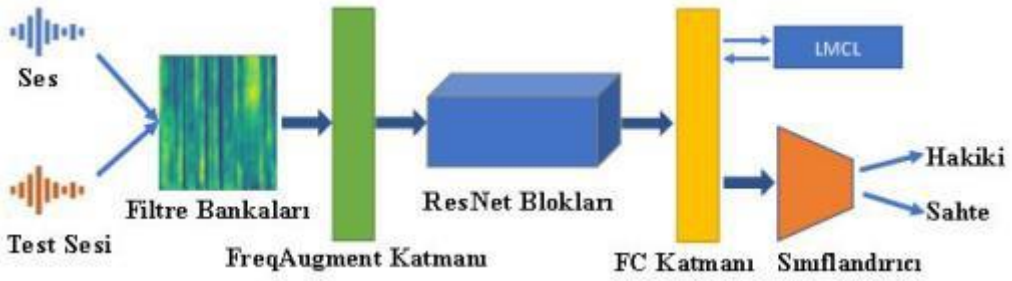
Geçtiğimiz yıllarda Kuzey Kore lideri Kim Jong-un ve Rusya Devlet Başkanı Vladimir Putin'in Deepfake'leri RepresentUS adlı grup tarafından oluşturuldu. Derin öğrenme ile oluşturulan bu sahte sesli videolar, bu liderlerin ABD seçimlerine müdahelesinin ABD demokrasisine zarar vereceği düşüncesini iletmek için yayınlanmak üzereydi; bu reklamaynı zamanda Amerikalıların medya yoluyla yayımlanan bu sahte ses ve videoların güvenilirliğini ve gerçekliğini incelemeden ülkenin ne ölçüde etkileneceğini görmek istiyordu. ABD seçimlerinden sonra bu reklamların yapay zeka ile oluşturulduğunu ve Amerikalıların karşılaşabilecekleri korku ve hassasiyet nedeniyle reklamların yayınlanmadığını duyurdu. Örnek olarak verilebilecek bir diğer deepfake ise NBC'nin The Night Show programında gerçekleşen Donald Trump'ın Jimmy Fallon skeciydi. Bu skeç derinöğrenme sahteciliği ile Youtube'da komedi içeriği olarak paylaşıldı. Bir diğer örnek ise Barrack Obama'nın Donald Trump'a üretilen deepfake ile küfrediyormuş gibi yansıtılmasıdır. Bu videonun amacı deepfake'lerin ne kadar tehlikeli olduğu ve ses ile videoların gerçekliğe çok yakın olmasıydı.[2]

Günümüzde mobil cihazlara kadar ulaşan deepfake benzeri uygulamalar gerçeğe çok yakın olmasa da üretilebilmektedir. Genelde eğlence içeriği olarak kullanılsa da gelişen teknoloji ile gerçekliğe çok yakın yaratılan sahte ses ve videolar belirtilen örneklerde olduğu gibi manipüle ve dezenformasyon amaçlı kullanılabilir. Üst düzey bir yöneticinin sesi taklit edilerek yanlış yönlendirmeler, siyasi bir kişinin üretilmiş sahte sesi ile şahsa yönelik karalama kampanyaları yapılabilir. Üretilen bu sahte seslerin analizi ve sahte olup olmadığının tespiti bu yüzden önem taşımaktadır.

Literatür Araştırması

Bu nitelikteki geçmiş çalışmalar; Mel Frekans Cepstral Katsayıları, Sabit Q değerli Cepstral Katsayıları, Elektrik Şebeke Frekansı Yardımı gibi yöntemlerdir. Bu yöntemler ses verilerinin sayısal temsilleri Destekli Vektör Makinesi'ni(DVM) ve Gizli Markov Modelini eğitmek için kullanılabilmektedir.

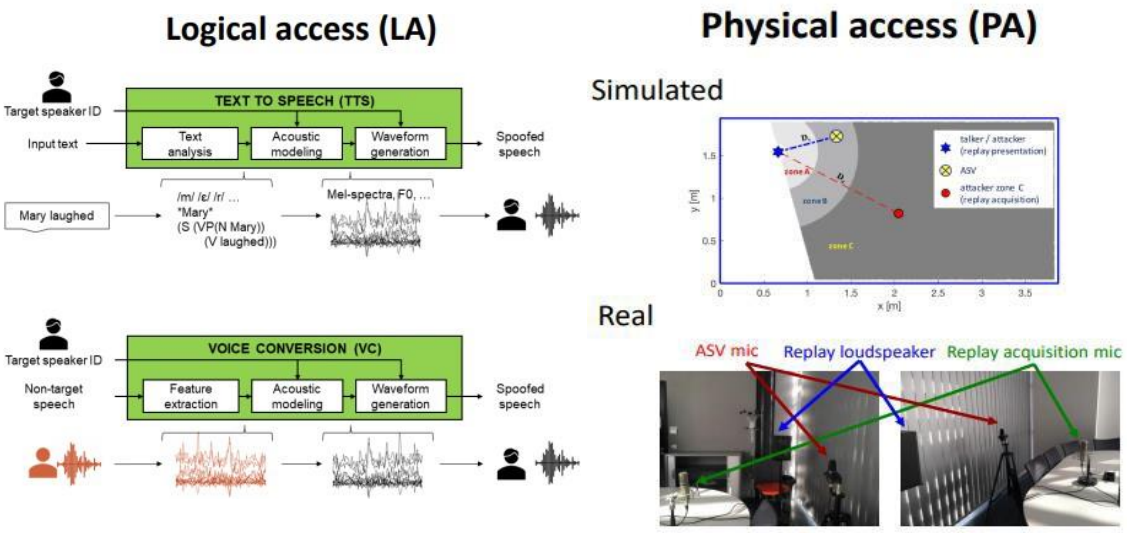
Sahte ses tespiti için güncel kullanılan yöntemlerde öğrenmeyi daha etkili kılmak için Büyük Marj Kosinüs Kaybı Fonksiyonu ve çevrimiçi frekans artırma işlemleri tanınmıştır. Derin sinir ağı modellemelerinin genelleme özelliğini artırmak için derin sinir ağı eğitiminde birleşik frekans kanallarını rastgele maskeleyen katman olan FreqAugment katmanı kullanılmıştır. [3]



Şekil 1: Sahte ses tespit yöntemi

Sahte seslerin üretilmesinde iki ana metot olan Mantıksal Erişim ve Fiziksel Erişim bulunmaktadır. Mantıksal Erişim durumunda üretilen sahte ses, metinden konuşmaya veya sesdönüştürme yazılımı kullanılarak üretilir. Yazılmış metinden konuşmaya sentez işlemi, yazılan metinlerin girişi, hedeflenen konuşmacının sesine benzer bir tona dönüştürülebilir.

Fiziksel Erişim yönteminde ise güvenlik algısını hedef konuşmacının gerçekten konuştuğunun kanaatine varması için konuşmacının önceden kaydedilmiş bir konuşması tekrar oynatılır. Sesin sahte olup olmadığının tespiti için Derin Sinir Ağları ile birlikte Sabit Q DeğerliKepstral Katsayıları, Mel Frekansı Kepstral Katsayıları ve Gauss Karışım Modeli gibikonuşma özellikleri kullanılır.[4]



Şekil 2: Mantıksal ve Fiziksel Erişim

Bankacılık gibi sektörlerde ses imzası oluşturmada kullanılan yöntemler Sıfır Geçiş Sayısı (SGS), Kök Ortalama Karesel Enerji (KOKE), Kepstral Akı (KA), İzgesel Akı (İA), Mel Frekansı Kepstrum Sayıları gibi öznitelik çıkarma modelleri kullanılmaktadır. Destek Vektör Makinesi kullanılarak ses imzasını öğrenen ve sınıflandıran bir sistem mevcuttur.

Alternatif Çözüm Önerileri

Ses verilerinden hem zaman hem de frekans alanlarında çeşitli özellikleri bulunmaktadır. Spektrogram, Mel-Spektrogramı, Mel-Frekans Katsayıları ve Sabit Q KepstralSayıları ve RMS enerjisi, bir zaman alanı özelliğidir. Sinyaldeki gücü hesaplar. Bu, boyutsallığı azalttığı için temel zaman alanı özelliği olarak kullanışlıdır. Ses verilerinin yönetilebilir bir düzeye getirilebilir. Bir ses sinyalinin spektrogramı, zaman alanı temsiline karşı bir frekans alanıdır. Hesaplamak için, bir zaman etki alanı sinyali, zaman içinde belirli bir boyuta sahip bir dizi pencereye bölünür.

Her bir pencereyi frekans alanına taşımak için zamana karşı bir frekans grafiği oluşturarak Fourier dönüşümü için hesaplanır. Mel-Spektrogramı Frekans alanı için doğrusal bir ölçek kullanarak çizim yapmak yerine Mel ölçeğinin kullanıldığı spektrogramın varyantı, dinleyiciler tarafından eşit mesafede olduğuna karar verilen, ampirik olarak türetilmiş doğrusal olmayan bir frekans ölçeğidir.

Mel frekansı kepstral katsayıları, ses analizi için yaygın olarak kullanılan özelliklerdir. Genelde bu yöntem tarafından hesaplanır. Bir ses sinyalinin pencereli Fourier dönüşümünü almak, onu Mel frekans ölçeğine eşlemek mümkündür. Logaritmayı almak Mel frekanslarının her birinde sinyal gücünün ve elde edilen güçlerin Ayrık Kosinüs Dönüşümünün hesaplanması onları yeni bir sinyal olarak ele alınabilir hale getirir. Mel Frekans Katsayıları daha sonra elde edilen DCT katsayılarının büyüklüğü olarak alınır ve bir spektrogramla aynı şekilde zamana karşı çizilir. CQCC'ler ilk olarak [6]'da, sahtekarlık tespiti için önceki tüm özellik çıkarma yöntemlerine göre bir gelişme olarak sunulmuştur. görevler. MFCC'ler tarafından kullanılan Fourier dönüşümü yerine, bu yöntem sabit Q dönüşümünü (CQT) kullanır. İnsan algısına daha yakın olması amaçlanan geometrik olarak aralıklı frekans kullanırlar.

3.1 Lojistik Regreasyon

Lojistik regresyon, istatistiksel analizde, önceki gözlemlere dayanarak bir veri değerini tahmin etmeye çalışan bir tekniktir. Lojistik regresyon algoritması, bağımlı bir değişken ile bir veya daha fazla bağımlı değişken arasındaki ilişkiye bakar. Çok basit ama çok güçlüdür veen popüler klasik makine öğreniminden biridir. Sınıflandırma görevleri için kullanılan modeller ve girdiye bağlı bir ikili değişkeni modellemek için lojistik işlevi kullanır.

Lojistik regresyonun makine öğreniminde birçok uygulaması vardır. Bir lojistik regresyon algoritması, tüm seçim sonuçlarının ortalamasını alarak bir seçimde hangi adayın kazanacağını tahmin etmeye çalışabilir. Daha sofistike bir algoritma ekonomik verileri ve geçmiş seçimleri modeline dahil edebilir. Başka bir algoritma, bir web sitesinin hangi kullanıcılarının belirli reklamları tıklayacağını belirlemeye çalışabilir. Ayrıca veritabanı hazırlamada, çıkarma, dönüştürme ve yükleme (ETL) işlemleri için verileri sınıflandırmak için yaygın olarak kullanılır.[6]

3.2 Gauss Karışım Modeli

Gauss Karışım Modeli yaklaşımı da değerlendirilebilir. Bir üretkenlik biçimi olarak GKM'lerin esnekliği sınıflandırmanın yanı sıra ilgili temellerde kullanımı el verişlidir. Biri sahte kliplerde olmak üzere iki GKM eğitilir. Sahte ses orijinal ses üzerinde ve ardından yeni bir örnek sesin olma olasılığı puanlanarak tahminler yapılır. İki GKM'nin her biri tarafından oluşturulur.

Bir Gauss karışım modeli (GKM), üretilen bütün veri noktalarının bilinen bir parametresi olmayan sonlu Gauss dağılımlarının bir karışımından türetildiğini belirten olasılıklı bir model kategorisidir. Gauss karışım modelleri için parametreler ya iyi bir posteriori kestirimden ya da iyi eğitilmiş önceki bir modelden yinelemeli bir beklenti maksimizasyon algoritmasından türetilir. Gauss karışım modelleri, modelleme verileri,özellikle de çeşitli gruplardan gelen veriler söz konusu olduğunda çok yararlıdır. [7]

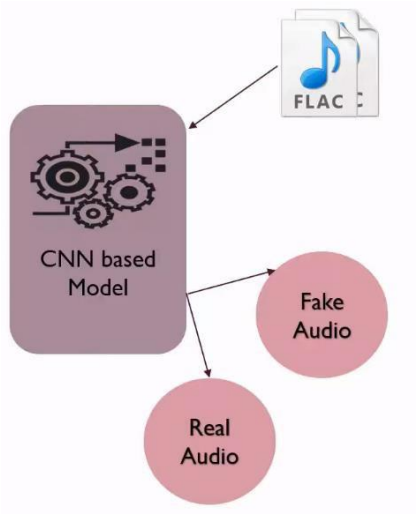
Bir Gauss karışım modeli (GMM), veri noktalarının bilinen bir parametresi olmayan sonlu Gauss dağılımlarının bir karışımından türetildiğini belirten olasılıklı bir model kategorisidir. Gauss karışım modelleri için parametreler iyi bir kestirimden ya da iyi eğitilmiş önceki modelden yinelemeli bir maksimizasyon algoritmasından türetilir. Gauss karışım modelleri, modelleme verileri, özellikle de çeşitli gruplardan gelen veriler söz konusu olduğunda çok yararlıdır.[8]

Gauss karışım modelleri, parametrik modelin vokal kanalı spektral özellikleri gibi özelliklerle veya ölçümleri yardımcı olduğu biyometrik sistemlerde kullanılır. Gauss karışım modelleri de yoğunluk tahmini için kullanılır ve kümelenme için istatistiksel olarak en olgun teknikler olarak kabul edilir.[8]

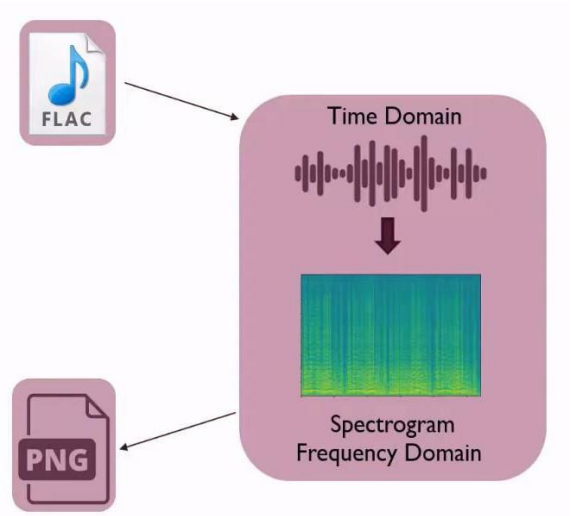
Gauss karışım modelleri, parametrik modelin vokal kanalı spektral özellikleri gibi özelliklerle veya ölçümleri anlamada yardımcı olduğu biyometrik sistemlerde kullanılır. Gauss karışım modelleri de yoğunluk tahmini için kullanılır ve kümelenme için en istatistikselolarak en olgun teknikler olarak kabul edilir.[9]

Çözüm Yöntemi

Derin öğrenme algoritmalarıyla üretilen ve insan kulağının ayırt edemeyeceği sesleri analiz etme konusunda Evrişimsel Sinir Ağları (Convolutional Neural Network) kullanarak yüksek başarı oranına sahip çıktı almayı planladık. Girdi olarak alınan raw ses dosyasını Şekil 4.'teki Evrişimsel Sinir Ağı Modeli ile ses spektogramına dönüştürmeyi ve çıktı olarak spektogramın görselini almayı başardık.



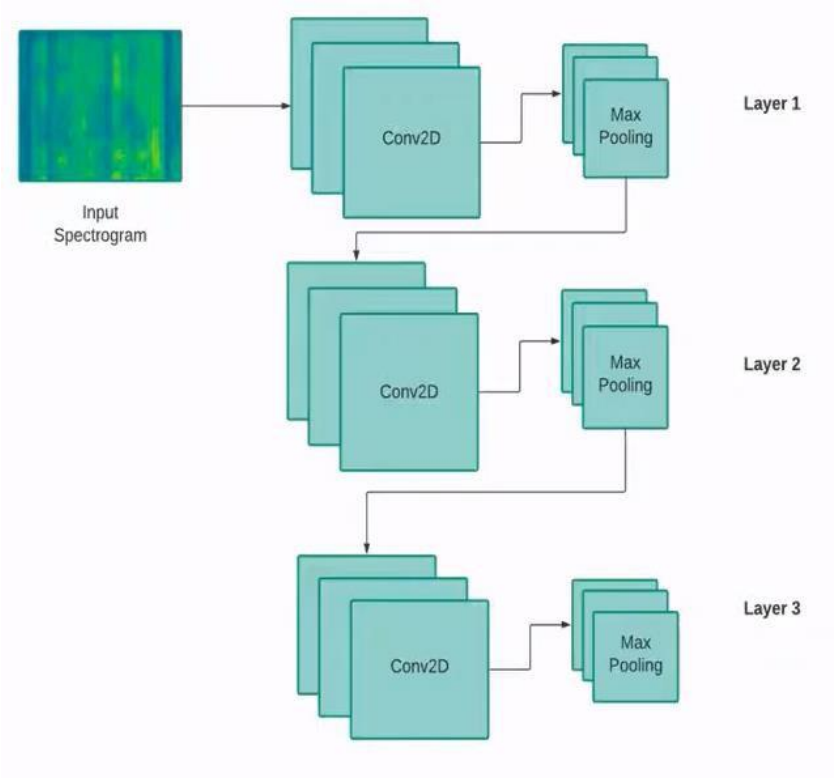
Şekil 3: Model girdi ve çıktıları



Şekil 4: Spektogram dönüşümü

Evrişimsel sinir ağlarında her bir katmanda matris üzerinde işlem yapan kerneller kullanılmaktadır. Bunlar basit Gabor filtreleri gibidir. Bu kernellerin boyutu öğrenme üzerinde hayli etkilidir. Çünkü kernel boyutu ile ne kadar genişlikte verinin birbirini etkileyeceğine karar verilmektedir. Genelde 3x3, 5x5, 7x7 kerneller kullanılmaktadır.

Büyük boyutlu kernel kullanılması evrişim uygulandıktan sonra oluşacak verinin küçük olmasına neden olacaktır. Bu durumda bilgi kaybına yol açtığından 3x3 gibi küçük boyutlu kernel kullanılacaktır. Kenar bulma işleminde işlem yapılan pikselin sağına-solunaüstüne-altına bakılabilmesi için tek sayılardan oluşan merkezi olabilecek filtreler kullanılmıştır.



Şekil 5: 3x3 Evrişimsel Sinir Ağı Modeli (CNN)

Araştırma metodolojileri ve ses verisi ön işleme adımından sonra dosyaların mevcut spektrogram görüntüleri alınarak Tensorflow frameworküne girdi olarak yüklemiştir. Tenserflow, neredeyse her tür derin öğrenme modelini tanımlamak ve eğitmek için uygun bir yol sağlayanPython için bir derin öğrenme kütüphanesidir. Keras, Tensorflow , Theano ve CNTK üzerindeçalışabilen Python ile yazılmış bir üst düzey sinir ağları API'sıdır.[8] Keras'a yüklenen spektogram örnekleriyle sıralı bir model oluşturup Evrişimli Sinir Ağı ile 32, 64 ve 128 katmanları ile kategorik durum sınıflandırması ve çapraz entropi kullanılacaktır.

CNN yapısı çok katmanlı bir sinir ağı modelidir. Bu modelde diğer modellerden farklı olarak çoklu evrişimsel ve havuzlama katmanları bulunmaktadır. Evrişimsel katmanının temel görevi verilen görüntüden özelliklerin çıkarılmasıdır. Bu özelliklerin çıkarılması işleminde filtre denilen n x n boyutunda matrisler kullanılır. Bu filtreler ile görüntü üzerinde piksel pixel ilerleyerek filtreler uygulanır. Pooling katmanında ise alt örnekleme yapılarak bir önceki katmanda bulunan özelliklerin içerdiği en önemli bilgileri koruyarak boyutunu azaltır.[10]

Modelin Evrişimli Sinir Ağları ile eğitilmesi için veri tabanı gereklidir. Çözümde kullanacağımız veri tabanı 2015 yılında düzenlenen konuşmacı doğrulama yarışmasında oluşturulan ASVspoof adlı veri tabanı olacaktır. Veri tabanı 16 kHz'de örneklenmiş, 16 bit çözünürlüğünde standart .wav formatında kaydedilmiş ses dosyalarından oluşmaktadır. Bu sesdosyaları eğitim, geliştirme ve değerlendirme olmak üzere birbiri ile örtüşmeyen üç alt kümeye ayrılmıştır. Şekil 6'da gösterildiği gibi; 3014 adet eğitim, 1710 adet geliştirme ve 13306 adet değerlendirme alt kümelerine ait ses kayıtları mevcuttur.

Alt Küme	Konuşmacı Sayısı	Gerçek Kayıt Sayısı	Sahte Kayıt Sayısı
Eğitim	10	1507	1507
Geliştirme	8	760	950
Değerlendirme	24	1298	12008

Şekil 6: ASVspoof 2017 veri tabanı

Bulgular ve Sonuçlar

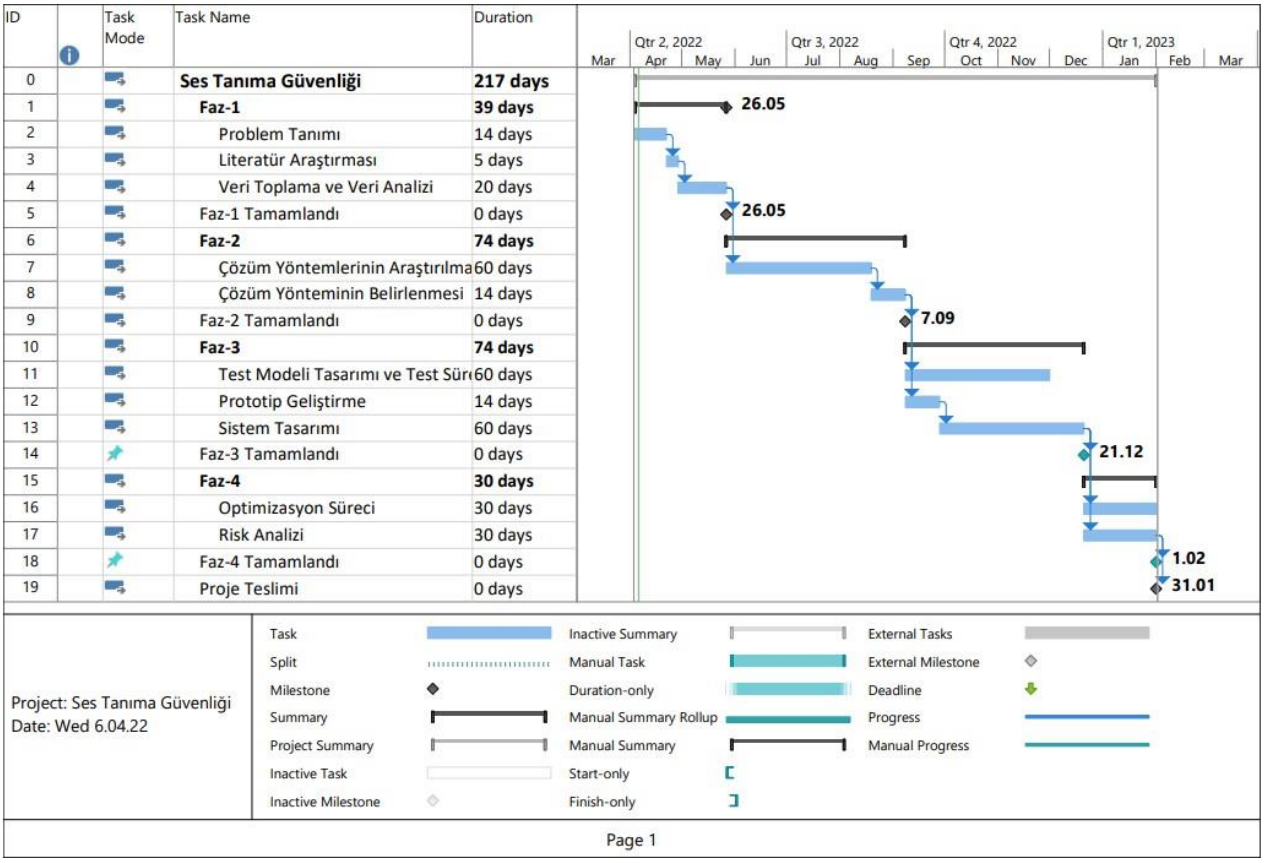
Bu çalışmada güncel derin öğrenme ve makine öğrenmesi yöntemlerinden olan CNN kullanılmıştır. ASVSpoof 2017-2019 veri tabanında etiketli halde bulunan ses dosyaları Fourier Dönüşümü ile ses spektogramlarına dönüştürülmüştür. Dönüştürülen ses spektogramları bilgi kaybı olmaması için 3x3 boyutlu kerneller kullanılmıştır. Veri tabanında eğitim, geliştirme ve değerlendirme için ayrılan ses kayıtları belirtilen amaçlarlar doğrultusunda modeli eğitmek ve değerlendirme amacıyla kullanılmıştır. Bu bilgiler doğrultusunda eğitilen CNN (Evrişimsel Sinir Ağı) modeli %85 gibi yüksek doğrulukta çıktı vermesi beklenmiştir.

Gelişen yapay zeka teknolojisi ile tehdit haline gelen deepfake ile manipüle edilmiş içeriklerin oluşturulması hızla gelişen sorundur. Deepfake için yapay zeka ile geliştirilen sahte uygulamalar, sahte tespiti için geliştirilen algoritmalardan daha hızlı gelişmektedir. Sosyal medyada kullanımı popüler olan deepfake içeriklerinin hızla paylaşılarak yayılması da kritik bir sorun oluşturmaktadır. Buna karşın çözüm araştırmalarının ölçeklenebilir, genelleştirebilir yöntemler sunmaya odaklanması gerekmektedir. Sahte içeriklerin yayılmasını ve farkındalığın yaratılması için eğitimler verilmelidir. Deepfake karşıtı algoritmaların geliştirilmesi için yarışmalar düzenlenmeli, yüksek çıktı oranına sahip projeler desteklenmelidir. Teknolojinin doğal bir sorunu olarak ortaya çıkan bu sorun güncelliğini her zaman koruyacaktır.

Proje Bütçesi

Kaynaklar	Harcama Periyodu	Tutar
Proje Personeli		
Yazılım Geliştirici	15.000 ₺ / ay * 6 ay	90.000 ₺
Sistem Tasarımcısı	17.000 ₺ / ay * 6 ay	102.000 ₺
Teknik Destek		
Yapay Zeka Uzmanı	30.000 ₺ / ay * 3 ay	90.000 ₺
Teknik Doküman Uzmanı	500 ₺ / gün * 1 hafta	3.500 ₺
Donanım		
Tensor İşleme Birimi (TPU) v3-8	8 $ / saat * 200	22.400 ₺
Bilgisayar	10.000 ₺	10.000 ₺
Diğer Giderler		
Ofis Kira Giderleri	2000 ₺ 6 ay	12.000 ₺
Personel Yeme, içme	50 ₺ / gün * 2 * 120	12.000 ₺
TOPLAM GİDER		341.900 ₺

Gantt Şeması



Projeyi tamamlamak için gerekli olan yazılım ürünleri Python programlama dilininkütüphaneleri olan numpy, pandas, scipy, mathplotlib ve özellikle derin öğrenmede kullanılan keras,tesnerflow ile sklearn modülleridir. Veri tabanı ise ASVspoof 2017-2019 kaynağından sağlanacaktır.

-Yazılım Ürünleri

-Python

-Tensorflow

-Numpy, Pandas, Scipy, Mathplotlib, Sklearn

-ASVspoof 2017-2019

-Donanım Ürünleri

-İşlemci: Intel® Core™ i5-10500H Processor or higher.

-Bellek: 16.00 GB

-Tensor İşleme Birimi (TPU) v3-8

Özet

Bu proje önerisinde, KD sistemlerinde yaygın olarak kullanılan Mel-Frekansı Kepstrum katsayıları, Otomatik Konuşmacı Doğrulama Yanıltma Saldırıları ve Karşı Önlemler (ASVspoof 2017) yarışmasında başarısı kanıtlanan sabit Q kepstral katsayıları ve Uzun dönem Ortalama Spektrum öznitelikleri mevcut literatür taraması ile bahsedilmiştir.[2],[3],[4],[5] Alternatif Çözüm olarak irdelenen Lojistik Regreasyon ve Gauss Karışım Modeli yöntemi araştırılmış ve çıktıları hakkında bilgi verilmiştir.[6],[7],[8] Çözüm yöntemimizde ise ASVspoof veri tabanından elde ettiğimiz sesleri ses spektrogramına dönüştürerek Evrişimsel Sinir Ağları ile modelimizi eğitip yapay zeka öğrenimine dayalı sonuç elde edilmiştir.

Referanslar

[1] Bateman, Jon (2020). "Özet" . Finansal Sistemde Deepfakes ve Sentetik Medya 1-2. 20 Nisan 2021 tarihinde kaynağından arşivlendi . Erişim tarihi: 28 Ekim 2020

[2] "Deepfake'nin Yükselişi ve Demokrasiye Tehdit" . Gardiyan . 1 Kasım 2020 tarihinde kaynağından arşivlendi . Erişim tarihi: 3 Kasım 2020 .

[3] Fagan, Kaylee. "Obama'nın Trump'a 'dips-- ' dediğini gösteren viral bir video, 'deepfakes 'olarak adlandırılan rahatsız edici yeni bir trendi gösteriyor " İş İçeriden . 22 Eylül 2020 tarihinde kaynağından arşivlendi . Erişim tarihi: 3 Kasım 2020 .

[4] T. Chen, A. Kumar, P. Nagarsheth, G. Sivaraman ve E. Khoury, Generalization of audio deepfake detection, Proceedings of the Odyssey Speaker and Language Recognition Workshop, Tokyo, Japan, 2020.

[5] A. Chintha, B. Thai, S. J. Sohrawardi, K. Bhatt, A. Hickerson, M. Wright ve R. Ptucha,Recurrent convolutional structures for audio spoof and video deepfake detection, IEEE Journal of Selected Topics in Signal Processing, cilt 14, p. 1024–1037, 2020.

[6] URL 1 https://theastrologypage.com/logistic-regression (27.03.2022)

[7] Bhattacharyya, S. T. Srikanthan, P. Krishnamurthy, (2001) Ideal GMM. parameters & Posterior Log Likelihood for Speaker Verification, Proceedings of the IEEE Signal Processing Society Workshop, USA. ISBN: 0-7803-7196-8, p. 471-480.

[8] https://tr.theastrologypage.com/gaussian-mixture-model

[9] URL 2. https://tr.wikipedia.org/wiki/Tensorflow (30.03.2022)

[10] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

Ekler;

Ek- A: Raportör Toplantı Tutanakları

1.HAFTA

Takım Üyelerinin tanışması ve rollerin belirlenmesi.

Lider: Berat Sercan Çelik
Raportör: Şevket Binali

2.HAFTA

Proje konusu hakkında fikir tartışmaları ve beyin fırtınası. Belirlenen konu: Ses tanıma sistemlerinde yapay zeka ile sahte ses analizi

3.HAFTA

Seçilen proje konusunun hocaya sunumu ve onay alınması.

4.HAFTA

Haftalık Online olarak Discord programı üzerinden görüşerek toplantıların ve içeriklerin belirlenmesi.

5.HAFTA

Problem tanımının ekipçe yapılması, Literatür araştırması ve problemin çözümü hakkında beyin fırtınası

6.HAFTA

Konuların ekip üyelerine dağıtılması, belirlenen amaç ve hedef kitle doğrultusunda yazı bütünlüğü korunarak ilgili içeriklerin yazılması.

7.HAFTA

Toplanan bilgilerin ve hazırlanan taslakları bir araya getirerek yazım kuralları ve metin bütünlüğü hedef kitleye anlatım amacı korunarak birleştirilmesi ve hazırlanması.

8.HAFTA

ARA

9.HAFTA

Çözüm yönteminin genişletilmesi hakkında görüş bildirimi

10.HAFTA

Faz 1 raporunun hoca tarafından değerlendirmesi ve söylediği eksikliklerimizin tamamlanması için görev dağılımı

11.HAFTA

Faz 2 raporu için İletim Mektubu, Öz, Yönetici Özeti, Bulgular ve Sonuçlar kısımlarının görev dağılımı

12.HAFTA

Hazırlanan tüm belgelerin bir araya getirilmesi, yazım kurallarının kontrolü, yazım bütünlüğünün korunması ve düzenlenmesi

T.C.

BURSA ULUDAĞ ÜNİVERSİTESİ

Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü
BMB3008-SUNUM YÖNTEMLERİ
2021-2022 Bahar Yarıyılı

DÖNEM PROJESİ - FAZ II

PROJE RAPORU

FAZ II DEĞERLENDİRME TABLOSU

Takım No: 32

Performans Kriteri	**Puan**	**Yorumlar**
Kapak Sayfası	/2	
Raporun Uzunluğu (Min. 30 sayfa)	/5	
Biçimlendirme Kuralları	/5	
Dilbilgisi ve Cümle Yapısı	/5	
İletim Mektubu	/5	
Öz	/5	
İçindekiler	/5	
Problem Tanımı	/5	
Yönetici Özeti	/5	
Giriş	/10	
Literatür Araştırması (Geçmiş ve Mevcut Çözümler)	/5	
Alternatif Çözüm Önerileri	/5	
Önerilen Çözümün Kısıtlamaları	/5	
Çözüm Yöntemi	/15	
Bulgular ve Sonuçlar	/5	
Proje Bütçesi	/3	
Gantt Şeması	/2	
Referanslar (Bibliyografya)	/3	
Her Takım Üyesinin Özgeçmişi	/5	
Toplam Puan	**/100**	

Takım Liderinin Adı, Soyadı: Berat Sercan Çelik **Öğrenci No: 032090095**

Raportörün Adı, Soyadı: Şevket Binali **Öğrenci No: 032090116**