



A hybrid approach based on deep learning for gender recognition using human ear images

Bahadır Karasulu^{1*} , Fatih Yücalar² , Emin Borandağ²

¹Department of Computer Engineering, Çanakkale Onsekiz Mart University, Çanakkale, 17100, Turkey

²Department of Software Engineering, Manisa Celal Bayar University, Manisa, 45400, Turkey

Highlights:

- Gender recognition was made using images of human ears
- Gender recognition based on the hybrid deep learning architectural models is presented to the literature
- The success of the proposed approach has been shown using various human ear image datasets

Keywords:

- Human Ear
- Gender Recognition
- Deep Learning
- Convolutional Neural Network
- Recurrent Neural Network

Graphical/Tabular Abstract

A hybrid deep learning approach is presented in our study to recognize more accurately the human gender with the use of ear images obtained unconditionally (i.e., *in-the-wild*) with various scale, orientation and size. On the contrary of classical approaches, it is seen by our hybrid approach that the requirements expected from any biometric system can be met by bridging the semantic gaps between lower-level and higher-level features thanks to the learned ear image attributes obtained with the use of deep learning. The main contribution of this study to the literature is to present the experimental results of both the convolutional neural network (CNN) model (as standalone) and a model based on the novel hybridization approach. Hybrid model was constructed with CNN component and the component with recurrent neural network (RNN)-type layers. Consequently, it was tested for classification (i.e., female or male recognition).

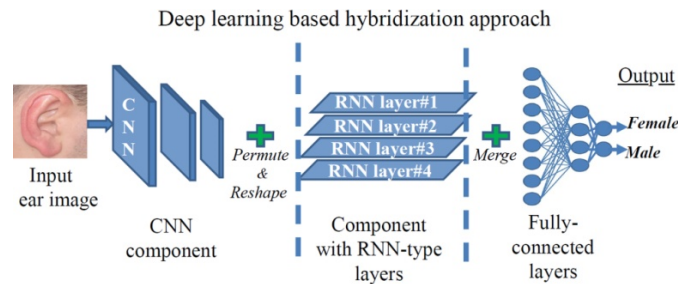


Figure A. Schematic for gender recognition based on hybrid deep learning with human ear images.

Article Info:

Research Article
Received: 30.05.2021
Accepted: 16.10.2021

DOI:

10.17341/gazimmfd.945188

Acknowledgement:

We thank to respective owners of human ear image datasets in the study.

Correspondence:

Author: Bahadır Karasulu
e-mail:
bahadirkarasulu@comu.edu.tr
phone: +90 286 218 0018 /
20034

Purpose: This study aims to proof the necessity and capabilities of proposed our novel hybrid deep learning approach for gender recognition using human ear images. The study presents the performance benchmarking results of our gender recognition approach with the use of two different datasets with various scale, orientation and image sizes that these images were tagged with gender mark came from their datasets.

Theory and Methods:

To obtain hybridization, proposed approach uses convolutional neural network (CNN) model and recurrent neural network (RNN) models such as gated recurrent unit (GRU) and long-short term memory (LSTM). Our hybridization approach was tested by using the combination of these abovementioned deep learning architectural models to proof the human gender recognition performance with ear images.

Results:

In this study, the human gender classification and recognition performance achieved by the hybridization of deep learning architectural models varies with parameter configuration of these models in given experiments. By considering performance benchmark on ear images from two different datasets, the best test accuracy rates with various combination of these models were obtained 85.16% for EarVN (i.e., Vietnam) dataset and 87.61% for WPUT (i.e., Poland) dataset as well.

Conclusion:

Gender recognition was made on higher-level abstract features obtained throughout the representational learning of the hybrid deep neural network model working on the various images. Alongside the CNN component, thanks to components with RNN-type (e.g., LSTM and GRU) layers, a better understanding of the relational dependencies between pixel regions in the ear images for hybrid neural network model has been provided. The two-component hybrid deep neural network model consisting of CNN and GRU type layers synergy achieved quite higher accuracy rates in the experimental test results in terms of gender classification (i.e., female or male recognition) performance compared to the standalone CNN model.



İnsan kulağı görüntüleri kullanarak cinsiyet tanıma için derin öğrenme tabanlı melez bir yaklaşım

Bahadır Karasulu^{1*}, Fatih Yücalar², Emin Borandag²

¹Çanakkale Onsekiz Mart Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 17100, Çanakkale, Türkiye

²Manisa Celal Bayar Üniversitesi, Hasan Ferdi Turgutlu Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, 45400, Manisa, Türkiye

Ö N E Ç İ K A N L A R

- İnsan kulağı görüntüleri kullanılarak cinsiyet tanıma yapılmıştır
- Melez derin öğrenme mimari modellerini temel alan cinsiyet tanıma literatüre sunulmuştur
- Önerilen derin öğrenme yaklaşımının başarısı çeşitli insan kulağı görüntü veri kümeleri kullanılarak gösterilmiştir

Makale Bilgileri

Araştırma Makalesi
Geliş: 30.05.2021
Kabul: 16.10.2021

DOI:

10.17341/gazimmfd.945188

Anahtar Kelimeler:

İnsan kulağı,
cinsiyet tanıma,
derin öğrenme,
evrişimli sinir ağı,
tekrarlayıcı sinir ağı

ÖZ

Günümüzde insan kulak görüntülerinin kullanımı, biyometrik yetkilendirme ve gözetleme sistemlerinin sürdürülebilirliği adına önem kazanmaktadır. Güncel çalışmalar, böyle işlemlerin el ile yapılması yerine yarı otomatik veya tam otomatik olarak yapılabileceğini göstermektedir. Derin öğrenme soyut öznitelikleri (temsili öğrenme) kullanması nedeniyle klasik yöntemlere göre oldukça yüksek başarımlar elde etmektedir. Çalışmamızda insanların cinsiyetlerine göre tam otomatik olarak sınıflandırılmasında insan kulağı görüntülerinin kullanımına dayanan melez derin öğrenme tabanlı sinerjik bir cinsiyet tanıma yaklaşımı oluşturulmuştur. Melezleme yoluyla hem evrişimli sinir ağı bileşeni hem de tekrarlayan sinir ağı tipli bileşenlerini bir arada içeren melez derin sinir ağı mimari modelleri kullanılmıştır. Bu modellerde tekrarlayan sinir ağı tipi bileşenler olarak uzun kısa süreli bellek ve kapalı tekrarlayan birim alınmıştır. Bu bileşenler sayesinde melez model görüntü içerisindeki piksel bölgeleri arasındaki ilişkisel bağımlılıkları oldukça iyi elde etmektedir. Bu sinerjik yaklaşım sayesinde çalışmamızdaki tek başına evrişimli sinir ağı modeline göre melez modellerin cinsiyet sınıflandırma doğruluğu daha yüksek olmaktadır. Cinsiyet işaretlemesine sahip iki farklı görüntü veri kümesi deneylerimizde kullanılmıştır. Deneysel sonuçların güvenilirliği nesnel ölçütlerle kanıtlanmıştır. Yapılan deneylerde melez modellerle yapılan cinsiyet tanımda en yüksek değerler sırasıyla, EarVN veri kümesi için test doğruluğu %85,16 ve WPUT veri kümesi için test doğruluğu %87,61 oranlarında elde edilmiştir. Çalışmamızın son bölümünde tartışma ve sonuçlara yer verilmektedir.

A hybrid approach based on deep learning for gender recognition using human ear images

H I G H L I G H T S

- Gender recognition was made using images of human ears
- Gender recognition based on the hybrid deep learning architectural models is presented to the literature
- The success of the proposed deep learning approach has been shown by using various human ear image datasets

Article Info

Research Article
Received: 30.05.2021
Accepted: 16.10.2021

DOI:

10.17341/gazimmfd.945188

Keywords:

Human ear,
gender recognition,
deep learning,
convolutional neural
network,
recurrent neural network

ABSTRACT

Nowadays, the use of the human ear images gains importance for the sustainability of biometric authorization and surveillance systems. Contemporary studies show that such processes can be done semi-automatically or fully automatically, instead of being done manually. Due to the fact that deep learning uses abstract features (i.e., representation learning), it reaches quite high performance values compared to classical methods. In our study, a synergistic gender recognition approach based on hybrid deep learning was created based on the use of human ear images in classifying people fully automatically according to their gender. By means of hybridization, hybrid deep neural network architectural models are used, which include both convolutional neural network component and recurrent neural network type components together. In these models, long-short term memory and gated recurrent unit are taken as recurrent neural network type components. Thanks to these components, the hybrid model extracts the relational dependencies between the pixel regions in the image very well. On account of this synergistic approach, the gender classification accuracy of hybrid models is higher than the standalone convolutional neural network model in our study. Two different image datasets with gender marking were used in our experiments. The reliability of the experimental results has been proven by objective metrics. In the conducted experiments, the highest values in gender recognition with hybrid models were obtained with the test accuracy of 85.16% for the EarVN dataset and 87.61% for the WPUT dataset, respectively. Discussion and conclusions are included in the last section of our study.

1. GİRİŞ (INTRODUCTION)

Günümüzde biyometrik yetkilendirme ve tanıma (recognition) popülerlik kazanmıştır. Biyometrik tanımda yüz bölgesi, kulak, parmak izi ve göz iris deseni gibi bireyler arasındaki ayrıştırmayı kolaylaştıracak olgu ve öğeler kullanılmaktadır [1, 2]. Biyometrik teknolojiler, insanların fizyolojik ve davranışsal karakteristiklerini temel alarak bu insanların kimliğinin tanınması ve doğrulanması için otomatikleştirilmiş yöntemlerdir [1, 3]. Biyometrik tanıma temel oluşturan karar aşamasındaki sınıflandırma ve tahminlemede, makine öğrenmesi ve derin öğrenme yöntemleri kullanılmaktadır. Özellikle bilgi elde etme kuramındaki ölçütlerle, bahsi geçen sınıflandırma ve tahminleme başarımları değerlendirilmektedir [4]. Kulağın fizyolojik yapısının biyometrik değerlendirmeye alınabilecek hale gelmesi insanın 8 yaşından itibaren olmaktadır [2, 5]. 70 yaşına kadar kulakta dramatik bir değişiklik yaşanmadığı sürece kulak bölgesi biyometrik açıdan güvenilir ve kararlı bilgilerin elde edilmesinde kullanılabilir [2, 6]. Görüntü işleme yöntemleri ile kulak görüntülerinden birey tespiti ve tanıma yapılmakta, bu işlemlerde poz veya yüz ifadesindeki değişimler kulak bölgesi tabanlı işlemleri etkilememektedir. Kulak görüntüleri üzerinden yerel öznitelikler elde edilerek, biyometrik ayrıştırmaya dayalı tespit veya tanıma işlemleri uygulanmaktadır [6, 7]. Otomatik kulak tanıma ve benzeri sistemlerde istatistiksel bilgilere dayanan; temel bileşen analizi [8], ölçekle değişmez öznitelik dönüşümü [9] ve dalgacık dönüşümü [10] gibi yöntemleri de içeren çeşitli çalışmalar literatürde yer almaktadır [7]. Çoğunlukla görüntüden yerel öznitelik elde etme ve bağımsız betimleyiciler kullanımıyla kontrol edilemeyen görüntüleme koşulları olarak; ışıklandırma değişimleri, döndürme, ölçeklendirme, başka nesnelere kulağı örtmesi ve görüntüdeki gürültü mevcut olsa bile kulak görüntüsünden oldukça iyi biyometrik sonuçlar elde edilebilmektedir [11-13]. Kulak tanıma sistemleri çoğunlukla el ile belirlenmiş kulak bölgelerini temel alarak işlem yaparlar. Otomatik veya yarı otomatik kulak bölgesi tespitinin görüntüye dair çeşitli bilgilerle yapılmasında ise; nesne kenar tespiti, çevrit tabanlı yaklaşımlar, matematiksel morfoloji işlemleri veya insan cildinin tonları tabanlı yaklaşımlar kullanılabilir [2]. El ile öznitelik elde etmeye (handcrafted features) ve yarı otomatik yöntemlere göre derin öğrenme yöntemleri daha gürbüz (robust) olmaktadır. Bu nedenle nesnelere, görüntü içerisindeki görsel ve biçimsel değişikliklere rağmen tam otomatik bir biçimde derin öğrenme yöntemleriyle tanınabilmektedir. Klasik makine öğrenmesi ile derin öğrenme arasındaki temel farklardan biri; klasik makine öğrenmesi için el ile oluşturulan öznitelikler yerine derin öğrenmede özniteliklerin soyutlanmasıdır. Böylece katman sayısı arttıkça derinlik artmakta ve filtreleme yoluyla daha rafine bir öznitelik kümesi oluşturularak temsili öğrenme (representation learning) gerçekleştirilmektedir [14]. Belirli kategorilerin (sınıfların) önceden tanımlı etiketlerle Yapay Sinir Ağı'na (YSA) girdi olarak verildiği danışmanlı öğrenme alanında biyometrik bilgilere dayanan birey tespiti

ve tanıma için literatürde çalışmalar vardır [13, 14]. Bu alanda en çok kullanılan derin öğrenme yöntemleri arasında Evrişimli Sinir Ağı (ESA) ve Tekrarlayan Sinir Ağı (TSA) yer almaktadır. ESA kullanılarak ağın girdi boyutlarına denk gelen bir adet görüntü bloğundaki özniteliklerin tanınması gerçekleştirilmektedir. Veri kümesindeki görüntülerin girdi olarak kullanılmasıyla eğitilmektedir. İlgili etiketlerle bu eğitimde verilen sınıf etiketlerinin karşılaştırılması sayesinde sınıflandırma ve tahminleme yapılabilmektedir [13]. TSA ağları ise girdi ile çıktı arasındaki ilişkinin TSA ağının kendi çıktısını işlem sırasında tekrar girdi olarak kullanabildiği, böylece girdi ile çıktı arasındaki gizli durumları katlanmış ağ düğümleri olarak ortaya çıkarabilen ve işleyen bir sinir ağı mimari modeli olarak literatürde yer almaktadır. Derin öğrenme kapsamında çok katmanlı yapıda TSA kullanılmasıyla ham verideki veya verilen öznitelik vektöründeki örüntü temsili sayesinde ilişkiler daha anlamlı hale getirilerek işlenmektedir. Literatürde en sık kullanılan TSA tipli mimari modeller olarak Uzun Kısa Süreli Bellek (UKSB) ve Kapılı Tekrarlayan Birim (KTB) modelleri birçok çalışmada kullanılmıştır [15]. Çalışmamızdaki ana motivasyon, melezleme yoluyla hem ESA hem de TSA tipli (UKSB ve KTB) derin sinir ağı mimari modellerini bir arada kullanan sinerjik bir yaklaşım ile insan kulak görüntülerinden cinsiyet sınıflandırma (tanıma) işleminin başarıyla gerçekleştirilmesidir. Bahsedilen melezleme yaklaşımına ihtiyaç duyulmasının temel nedeni, derin öğrenmenin otomatikleştirdiği öznitelik haritalarının elde edilmesine dayanan sınıflandırmanın, sinir ağının öğrenme çabasını anlamsal olarak zenginleştirebilecek bir sinerjiyi oluşturmasıdır. Bu çalışma kapsamında iki farklı görüntü veri kümesi ve bunların bileşkesi olan veri kümesi kullanılmıştır. Yapılan deneyler birkaç farklı kombinasyon ile tekrarlanmış, sonuçların güvenilirliği bilgi elde etme kuramındaki nesnel ölçütlerle kanıtlanmıştır. Bu çalışmanın literatüre ana katkısı, kulak görüntülerinden cinsiyet tanıma için hem tek başına derin öğrenme mimari modelinin (ESA) hem de melezleme yoluyla birden fazla modeli bir arada kullanan (ESA ve TSA bileşenleriyle oluşturulan) bir sinerjik yaklaşımın sonuçlarının nesnel bir kıyaslama yoluyla literatüre sunulmasıdır.

Çalışmanın geri kalan bölümleri şu şekilde düzenlenmiştir. İkinci bölümde, materyal ve metot kısmı ile insan kulak fizyolojisi, görüntüden kulak tespiti ve tanıma yapılması yanı sıra çalışmamıza temel oluşturan derin öğrenme mimari model alt yapıları detaylandırılmıştır. Üçüncü bölümde insan kulağı görüntülerini kullanarak cinsiyet tanıma için çalışmamızda önerilen melez derin öğrenme yaklaşımının detayları sunulmuştur. Dördüncü bölümde, ilgili görüntü veri kümeleri ile yapılan deneysel çalışmalara yer verilmektedir. Son bölümde ise elde edilen sonuçlar üzerine yapılan tartışma sunulmaktadır.

2. MATERYAL VE METOT (MATERIAL AND METHOD)

Bu bölümde kulak tanıma, tespit ve biyometrisi, derin öğrenme, literatürde geniş bir kullanıma sahip sinir ağı

modellerinden ESA, TSA ile TSA'nın özelleşmiş halleri olan UKSB ve KTB modelleri incelenmiştir.

2.1. Kulak tanıma ve kulak tespiti (Ear recognition and ear detection)

Zaman içerisinde değişmeyen kulak yapısı, insan tanımlamada kullanılan mükemmel veri kaynaklarından biridir [16-18]. Kulak tanıma ve kulak tespiti birbirinden farklı kavramlardır. Kulak tanıma, kulakları tespit etmek ve tam bir eşleşme bulmak için bir veri kümesi üzerinde arama işleminin gerçekleştirilmesidir. Kulak tespiti ise herhangi bir eşleşme olup olmadığının aranması ve bir eşleşme bulunduğu anda arama işleminin durdurulmasıdır [19]. Literatürde, kulak tespiti ve tanıma üzerine yapılan birçok klasik makine öğrenmesi ve derin öğrenme çalışmaları vardır [20-22]. Alshazly vd. [23] tarafından yapılan çalışmada, dört adet ESA tabanlı model üç farklı kulak veri kümesi üzerinde denemiştir. Emerşiç vd. [24] çalışmasında AlexNet [25], görsel tabanlı VGG-16 [26] ve SqueezeNet [27] mimarilerini kullanmıştır. Tanıma sistemlerinin başarımı, görüntü içeriğini temsil edebilmenin gürbüzlüğüne ve ayırt ediciliğine bağlıdır. Ağ modellerinin eğitilmesinde, üstün parametre (hyperparameter) uzayının eniyenmesi için büyük miktarda etiketli eğitim verisi ve güçlü hesaplama kaynakları gerekmektedir [28]. Alshazly vd. [29] çalışmasında ESA tabanlı temsiller ve özellikle kulak tanıma sistemlerinin oluşturulması için farklı derinlikteki VGG tabanlı ağlar ele alınmıştır.

2.1.1. Kulak biyometrisi (Ear biometry)

Kulak büyümesi orantılı olsa da tıbbi literatüre göre yerçekimi kulağın dikey yönde gerilmesine neden olabilmektedir [22]. Kulak yapısındaki oranlar yaş ilerledikçe asla değişmemektedir. Kulak anatomisi iç, orta ve dış kulak olmak üzere üç kısımdan oluşmaktadır [30, 31]. Dış kulak yapısının biyometri alanında kullanımı çalışmamız içerisinde önerilen derin öğrenme yaklaşımı için temel alınmıştır.

2.2. Derin öğrenme (Deep learning)

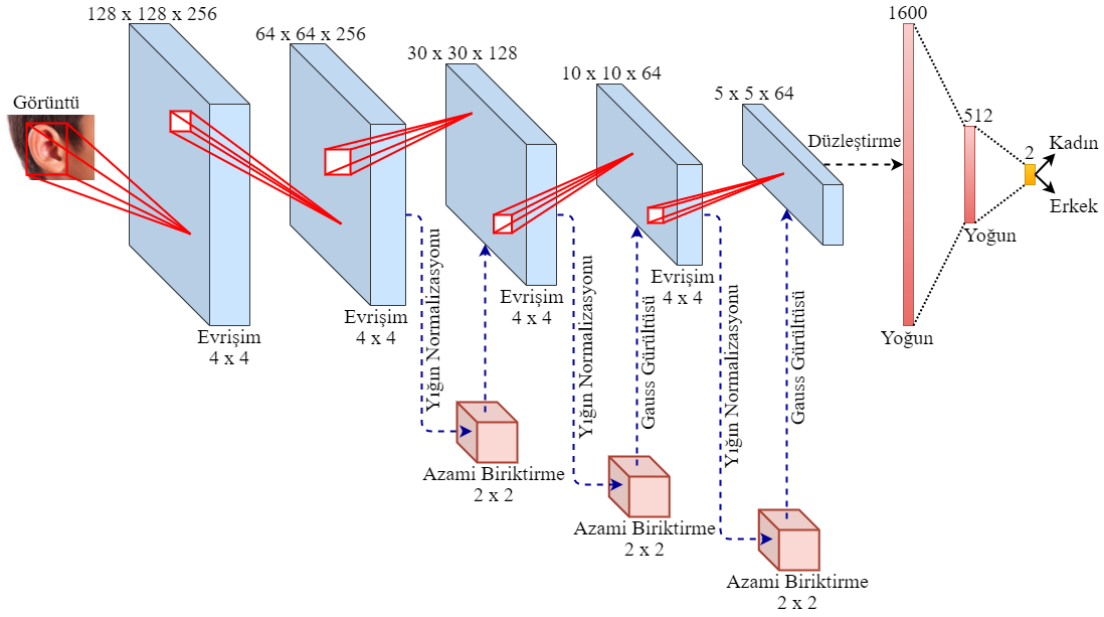
Derin öğrenme, yapay zekâ araştırma alanında yer alan makine öğrenmesinin bir alt alanıdır [32]. Klasik çok katmanlı algılayıcının (ÇKA) farklılaştırılmasıyla oluşturulmuş bir sinir ağ mimari modeli olarak derin öğrenme birden çok gizli katman içermektedir. Derin sinir ağlarının eğitiminde, eğitim yineleme sayısı (epoch) boyunca birden çok işleme katmanı üzerinden temsili öğrenmeye göre soyut özellikler elde edilir. Yapılandırılmamış veriler üzerinde çok daha iyi başarımlar göstermektedir. Derin öğrenme; danışmanlı, yarı danışmanlı veya danışmansız olarak gerçekleştirilebilir [32]. ESA ve TSA, literatürde yer alan en temel derin öğrenme sinir ağ mimari modellerinden ikisidir [33] ve görüntü işleme, biyomedikal sinyal işleme, yüz tanıma, sağlık uygulamaları ve tavsiye sistemleri gibi pek çok alanda kullanılmaktadır [34, 35].

2.2.1. Evrişimli sinir ağı (Convolutional neural network)

ESA modelleri (ConvNets) girdi olarak bir görüntüyü alıp, bu görüntü içerisindeki çeşitli nesnelerin birbirinden ayırt edilebilmesini sağlayan bir derin öğrenme mimari modelidir [36]. Çok sayıda YSA katmanından oluşan ve ileri yönlü bir sinir ağı olan ESA'nın sinir düğümleri çeşitli görsel özellikleri seçer [34, 37]. ESA'na bir görüntü verildiğinde, katmanlarının her biri farklı özellik haritası (feature map) oluşturur. Özellik haritaları, görüntünün ilgili özelliklerini vurgular. Sinir düğümlerinin her biri, girdi olarak bir piksel bölgesi (yama) alır, renk değerlerini ağırlıklarıyla çarpıp, elde edilen değerleri toplar ve bunları aktivasyon fonksiyonu aracılığıyla işleme alır [37]. ESA, evrişim katmanı (convolution layer), biriktirme katmanı (pooling layer), düzleştirme katmanı (flattening layer) ve tam bağlantılı katman (fully-connected layer) olmak üzere başlıca dört katmandan oluşur [38]. Evrişim katmanı, filtreler yardımıyla görüntü üzerindeki düşük ve yüksek seviyeli özelliklerin saptanması amacıyla kullanılır [39]. Biriktirme katmanı, özellik haritaları üzerinden elde edilen bilgilerin öznetlenmesini sağlayarak ağırlık sayısını azaltır [39]. Ayrıca, verinin uzamsal boyutunu, parametre sayısını ve modelin aşırı uyum (overfitting) göstermesi olarak tarif edilen ağırlık ezberlemesini azaltarak sinir ağının doğru karar vermesine yardımcı olur. Düzleştirme katmanı, tam bağlantılı katman girişindeki verileri hazırlayarak kendinden önce gelen katmanlardaki matris biçimindeki verilerin tek boyutlu diziye indirgenmesini gerçekleştirir. Tam bağlantılı katman ise sınıflandırma işlemini gerçekleştirmek adına kendinden önceki katmanlardan gelen verileri ağırlıklandırarak birleştirilmektedir. Hata miktarını da gösteren kayıp fonksiyonu yardımıyla eğitim boyunca kullanılacak uygun ağırlıklar bu katmanda belirlenmektedir. ESA için kullanılan tam bağlantılı katmanlarının sayısı ve bu katmanlar içerisindeki sinir düğümlerinin sayısı değişiklik göstermektedir [40]. Soyutlama, derin öğrenmede sınıflandırma kalitesini artıracak gerekli özelliklere odaklanmaktadır. ESA ağ modellerinde temsili öğrenme soyutlama amacıyla yapılmaktadır. Görüntü içerisindeki düşük ve yüksek seviyeli özellikler arasındaki anlamsal boşluk (semantic gap) problemini soyut özelliklerin kullanımıyla olabildiğince ortadan kaldırarak, özellik zenginleşmesine katkı sağlayarak sınıflandırma başarımını arttırmaktadır [41]. Şekil 1'de iki boyutlu ESA (2B-ESA) mimarisi görülmektedir.

2.2.2. Tekrarlayan sinir ağı (Recurrent neural network)

TSA, sıralı verileri işlemek için kullanılan derin öğrenme mimarilerinden biridir [42, 43]. TSA mimarisinin tekrarlayan olarak ifade edilmesinin nedeni, bir dizinin her elemanı için aynı görevi önceki çıktıları dikkate olarak icra etmesidir [15, 44]. Herhangi bir uzunluktaki girdinin işlenmesine olanak sağlayan bu mimari, hesaplama sırasında geçmiş bilgileri ele alır. Tekrarlanan aşamalarla bir önceki girdi durumu saklanır ve elde edilen yeni girdi değeriyle birleştirilir. Bu sayede yeni elde edilen girdinin bir önceki girdiyle olan ilişkisi sağlanarak hafızada tutulmuş olur. Buna



Şekil 1. 2B-ESA mimarisi (2D CNN architecture)

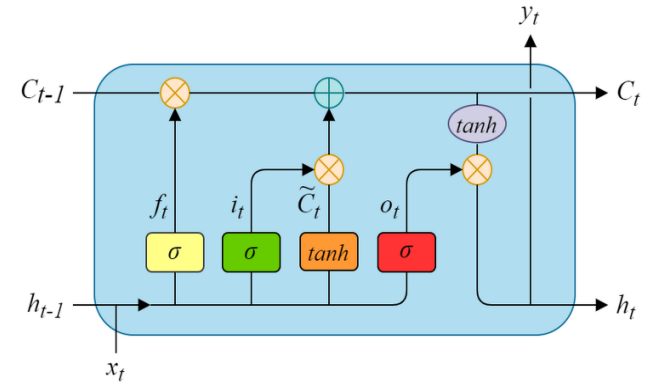
istinaden yapılan işlemlerle, Eş. 1 ve Eş. 2’de her bir t zamanında; $x^{<t>}$ girdiyi, $a^{<t>}$ aktivasyonu ve $y^{<t>}$ ise çıktıyı ifade eder.

$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a) \quad (1)$$

$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y) \quad (2)$$

Eş. 1 ve Eş. 2 üzerindeki W_{aa} , W_{ax} ve W_{ya} geçici olarak paylaşılan ağırlıkları; b_a ve b_y ise sapma (bias) miktarlarını; g_1 ve g_2 ise aktivasyon fonksiyonunu ifade eder. Sinir ağları üzerinde ileri yöndeki işlemde sonra eğitimin gerçekleştirilebilmesi için bir de geriye dönülerek zaman boyunca geri yayılım işleminin yapılması gerekmektedir. Bu işlem eniyileme amaçlı gerçekleştirilmektedir. İlgili araştırmalarda diğer derin öğrenme modellerinde (ESA gibi) farklı eniyileme yöntemleri de kullanılabilir [45]. Eniyileme işlemiyle ağ üzerindeki ağırlıkların güncellenmesi [46] için literatürde çokça tercih edilen eğimli iniş (gradient descent) [47] tekniği kullanılır. Eğitim yineleme sayısına [48] göre bu işlemler tamamlandığında model eğitilmiş olur. TSA’da geri yayılım işlemi esnasında patlayan gradyan (exploding gradient) veya kaybolan gradyan (vanishing gradient) problemleri oluşur [48]. Bunlara çözüm olarak TSA’nın özelleşmiş halleri olan UKSB ve KTB modelleri geliştirilmiştir [49]. UKSB ağırları bir TSA mimarisi olarak ortaya çıkmıştır [50]. UKSB’ler, geleneksel ileri beslemeli sinir ağları ve TSA’ya göre birçok yönden üstünlüğe sahiptir. Tipik bir UKSB ağı, hücre adı verilen farklı bellek bloklarından oluşur. Hücre durumu ve gizli durum olmak üzere bir sonraki hücreye aktarılan iki durum vardır. Hücre, bellek bloklarındaki değerleri hatırlamaktan sorumludur. Şekil 2’de UKSB yapısı görülmektedir. Burada hücreye giren ve çıkan bilgi akışını düzenleyen, giriş kapısı (input gate), çıkış kapısı (output gate) ve unutma kapısı (forget gate) olmak üzere üç ana

mekanizma bulunmaktadır [51]. Bunlar Şekil 2’de i , o ve f terimleriyle; ağır işlemsel hücre bellek durumları ise C ve \tilde{C} terimleri ile gösterilmektedir. Tüm terimlerde indis olarak t harfi, ilgili terimin geçerli zaman adımını göstermektedir.



Şekil 2. UKSB yapısı (Structure of LSTM)

Hücre durumundan bilgilerin kaldırılmasını unutma kapısı sağlar [51]. Unutma kapısı; h_{t-1} ve x_t olmak üzere iki girdi alır. Buna göre; h_{t-1} , önceki hücrenin gizli durumu veya bu hücrenin çıktısıdır. Burada, x_t ise belirli zaman adımındaki girdidir. Girdiler ağırlık matrisleri (W_f) ile çarpılır ve bir sapma (b_f) miktarı eklenir. Bunu takiben sigmoid (σ) işlevi bu değere uygulanır [52]. KTB’ler, hücre durumundan kurtulması ve bilgi aktarması için gizli durumu kullanması dışında UKSB’ye oldukça benzer yapıdadırlar [53]. KTB’ler güncelleme kapısı (update gate) ve sıfırlama kapısı (reset gate) olmak üzere iki kapıya sahiptir. UKSB’deki unutma ve giriş kapıları, KTB’de güncelleme kapısı içinde birleştirilmiştir. Aynı zamanda, KTB’de hücre durumu ve gizli durumu birleştirilmiş, bazı değişikliklerin yapılarak basitleştirilmiştir. Uzun dönemli bağımlılıkların

çözümlemesinde açıklayıcılık özelliği olarak KTB modeli diğer modele göre daha iyidir [53, 54]. Bu modellerin dışında, bir dizi işleme modeli olarak birinin girdiyi ileri yönde, diğerinin ise geri yönde ele aldığı iki UKSB modelinin birleşimini ifade eden çift yönlü UKSB modeli vardır. Çift yönlü UKSB'ler, ağda mevcut olan bilgi miktarını etkin bir şekilde artırarak, algoritma yardımıyla mevcut içeriği iyileştirir. Literatürde, önceki ($t-1$) ve sonraki ($t+1$) zaman adımlarını işleyebilen çift yönlü UKSB benzeri bir bakış açısına sahip çift yönlü KTB yapıları ile ilgili çalışmalar da yer almaktadır.

3. DERİN ÖĞRENME YAKLAŞIMI KULLANILARAK KULAK GÖRÜNTÜLERİNDEN CİNSİYET TANIMA (GENDER RECOGNITION FROM EAR IMAGES BY USING DEEP LEARNING APPROACH)

Bu bölümde, ESA, UKSB, KTB derin ağ mimari modellerinin ilgili ağ katman yapıları ve parametrelerinin iki farklı görüntü veri kümesi ile bunların bileşkesi olan bir veri kümesi üzerinden yapılan deneylerde sınıflandırma sayesinde cinsiyet tanımada nasıl kullanıldıklarına ilişkin detaylar yer almaktadır.

3.1. Derin öğrenme modellerini oluşturma (Construction of deep learning models)

İnsan kulağından cinsiyet sınıflandırılması için oluşturulan iki boyutlu ESA (2B-ESA) ağ modelinin detayları Tablo 1'de görülmektedir. Temsili öğrenmenin ana fikrine uygun olarak Tablo 1'de yer alan 2B-ESA ağ modeli dört evrişim katmanından oluşmaktadır. Ağdaki katman sayısının artışıyla derinliğin artmasından dolayı filtreleme sayesinde daha rafine ve soyut öznitelik kümesinin oluşturulması adına temsili öğrenme tercih edilmektedir. Bu sayede söz konusu sinir ağının öğrenme başarımının yüksek olması hedeflenmiştir. Bu tasarımda her bir katmanda yer alan

parametre sayıları birbirinden farklıdır. Tablo 1 ve Tablo 2'de görülen # işareti, işlevin kaçınıcı sırada olduğunu veya ağdaki aynı tipli katmanlardan kaçınıcı katman olduğunu gösterir.

Tablo 1'de görülen evrişim katmanları arasına sıklıkla eklenen azami biriktirme (max-pooling) katmanı, evrişim katmanlarının oluşturduğu öznitelik haritasının belirli bir boyuta indirgenmesi amacıyla kullanılmaktadır. Evrişim katmanlarının daha düzenli hale getirilmesi amacıyla kullanılan yığın normalizasyon (batch normalization) yöntemi, iki boyutlu ESA mimarisi içerisinde yer alan herhangi bir evrişim katmanına ortalaması sıfır olan ve $[0, 1]$ arasında değerlere sahip verilerin girdi olarak verilmesini sağlar. Eğitim esnasında karşılaşılan aşırı uyum gösterme şeklindeki ağın ezberlemesini azaltmak için Gauss gürültü katmanında rastgele veri artırma yöntemi olarak o katmandaki girdi değerlerine beyaz gürültü (Gaussian noise) uygulanmaktadır [55]. Kulak görüntüsü önce ilk evrişim katmanına 128×128 büyüklüğündeki bir matris halinde girdi olarak verilmektedir. İlk katmanda, x ve y eksenleri için çekirdek büyüklüğünün (kernel size) 4 olarak alınıp işlenmesi sonucu $(125, 125)$ büyüklüğündeki matris ile bu katmanın öznitelik haritası oluşturularak evrişim işlemi gerçekleştirilmiş olur. İlk evrişim katmanı için filtre büyüklüğüne bağlı olarak düğüm sayısı 256 olarak ayarlanmıştır. İkinci evrişim katmanı da ilk evrişim katmanına benzer şekilde 64×64 büyüklüğündeki bir matris haline getirilen öznitelik haritasını alır ve evrişim işlemi, her iki eksen için çekirdek büyüklüğünün 4 olarak alınıp işlenmesi sonucu $(61, 61)$ büyüklüğündeki matris ikinci evrişim katmanının öznitelik haritasını oluşturmaktadır. İkinci evrişim katmanı için de filtre büyüklüğüne bağlı olarak düğüm sayısı 256 olarak ayarlanmıştır. Öznitelikler üzerinde belirli bir filtreleme yapılabilmesi için ara katmanlarda her bir katmanın aktivasyon işlemlerinde, doğrultulmuş doğrusal birim (*ReLU*) fonksiyonu kullanılmıştır. Ağ içerisindeki sonraki katmanlarda uygun

Tablo 1. 2B-ESA tipli ağ modeli şablonu (Network model template of 2D CNN type)

Katmanlar	Çıktı Birim Biçimi	Parametreler*
2B_ESA_#1	(Yok, 125, 125, 256)	4352
2B_ESA_#2	(Yok, 61, 61, 256)	1048832
Yığın_normalizasyon_#1	(Yok, 61, 61, 256)	1024
Azami_biriktirme_#1	(Yok, 30, 30, 256)	0
2B_ESA_#3	(Yok, 27, 27, 128)	524416
Yığın_normalizasyon_#2	(Yok, 27, 27, 128)	512
Azami_biriktirme_#2	(Yok, 13, 13, 128)	0
Gauss_gürültü_#1	(Yok, 13, 13, 128)	0
2B_ESA_#4	(Yok, 10, 10, 64)	131136
Yığın_normalizasyon_#3	(Yok, 10, 10, 64)	256
Azami_biriktirme_#3	(Yok, 5, 5, 64)	0
Gauss_gürültü_#2	(Yok, 5, 5, 64)	0
Düzleştirme	(Yok, 1600)	0
Yoğun_#1	(Yok, 512)	819712
İletim_sönümü	(Yok, 512)	0
Yoğun_#2	(Yok, 2)	1026

*Toplam parametreler: 2,531,266

Tablo 2. 2B-ESA ile TSA tipli bileşen içeren melez derin sinir ağ modeli
(Hybrid deep neural network model with 2D CNN and RNN type component)

Katmanlar	Çıktı Birim Biçimi	Parametreler*
2B_ESA_#1	(Yok, 125, 125, 256)	4352
2B_ESA_#2	(Yok, 61, 61, 256)	1048832
Yığın_normalizasyon_#1	(Yok, 61, 61, 256)	1024
Azami_birikirme_#1	(Yok, 30, 30, 256)	0
2B_ESA_#3	(Yok, 27, 27, 128)	524416
Yığın_normalizasyon_#2	(Yok, 27, 27, 128)	512
Azami_birikirme_#2	(Yok, 13, 13, 128)	0
Gauss_gürültü_#1	(Yok, 13, 13, 128)	0
2B_ESA_#4	(Yok, 10, 10, 128)	262272
Yığın_normalizasyon_#3	(Yok, 10, 10, 128)	512
Azami_birikirme_#3	(Yok, 5, 5, 128)	0
Gauss_gürültü_#2	(Yok, 5, 5, 128)	0
Yeniden_şekillendirme	(Yok, 128, 25)	0
Permütasyon	(Yok, 25, 128)	0
TSA_#1	(Yok, 25, 512)	984576
TSA_#2	(Yok, 25, 256)	590592
TSA_#3	(Yok, 25, 128)	147840
TSA_#4	(Yok, 64)	37056
Yoğun_#1	(Yok, 4096)	266240
İletim_sönümü_#1	(Yok, 4096)	0
Yoğun_#2	(Yok, 512)	2097664
İletim_sönümü_#2	(Yok, 512)	0
Yoğun_#3	(Yok, 2)	1026

TSA_tip#: UKSB, KTB veya çift yönlü işlevsel yapı.

*Toplam parametreler: 5,966,914

evrişim, yığın normalizasyon, azami biriktirme ve Gauss gürültüsü işlemlerinin kullanımı ile işlenen matris büyüklüğü katmanlar arasında ilerledikçe indirgenmektedir. Buna istinaden yığın normalizasyon işlemi ve azami biriktirme katmanının kullanılmasıyla birlikte üçüncü evrişim katmanında öznetelik haritasının boyutu (30, 30) büyüklüğündeki bir matris olarak yarıya indirgenmiş ve düğüm sayısı da 128 olarak ayarlanmıştır. Üçüncü evrişim katmanından sonra yığın normalizasyon işlemi ve azami biriktirme katmanının kullanılmasına ek olarak Gauss gürültü katmanı uygulanarak dördüncü evrişim katmanında öznetelik haritasının boyutu (10, 10) büyüklüğündeki bir matris olarak yarıya indirgenmiş ve düğüm sayısı 64 olarak belirlenmiştir. Takip eden katmanlardan sonra sırasıyla yığın normalizasyon işlemi ve azami biriktirme katmanının kullanılmasıyla birlikte Gauss gürültü katmanının da etkisini içerir şekilde öznetelik haritasının boyutu (5, 5) büyüklüğündeki bir matris olarak yarıya indirgenmiştir. Önceki katmanlardan matris biçimindeki veri, sinir ağının düzleştirme katmanına gönderilerek öznetelik vektörü biçimine dönüştürülmüştür. Bu sayede, tıpkı ÇKA gibi tam bağlantılı katman uygun aktivasyon fonksiyonu ile çalıştırılarak çıktı katmanının cinsiyet için erkek veya kadın sınıfını gösteren sonuçlar üretmesi sağlanmıştır. Tek başına ele alınan 2B-ESA modelinin cinsiyet sınıflandırma yeteneklerinin iyileştirilmesi amacıyla sinerjik bir yaklaşım çalışmamızda oluşturularak Tablo 2’de görülen bir melez derin sinir ağ mimari modeli sunulmuştur. TSA tipli sinir ağlarından verilen görüntüdeki düşük seviyeli özneteliklerin elde edilmesi çalışmamızda ana amaç olarak hedeflenmemiş,

daha çok bu ağların bağlamsal ilişkileri elde edebilme yeteneği öncelikli olarak dikkate alınmıştır. ESA’nın soyutlama ve temsili öğrenme yapabilme yeteneğini, TSA tipi ağların ise örüntü temsiline dayanan öznetelikler arasındaki bağlamsal ilişkileri elde edebilme yeteneğini kullanabilmek için uygun bir melez derin sinir ağ mimari modeli oluşturulmuştur. Melez model, UKSB ve KTB gibi TSA tipli bileşenler aracılığıyla görüntü içerisindeki piksel bölgelerini temsil eden öznetelikler arasındaki ilişkisel bağımlılıkların elde edilmesini sağlamıştır.

Tablo 1’deki mimari modelden farklı olarak Tablo 2’de dördüncü evrişim katmanından sonra takip eden katmanlarda sırasıyla yığın normalizasyon işlemi, azami biriktirme katmanının kullanımı ve Gauss gürültü katmanının da etkisi ile öznetelik haritasının boyutu (5, 5) büyüklüğündeki bir matris olarak yarıya indirgenmiş ve düğüm sayısı da 128 olarak ayarlanmıştır. TSA tipli ağlar ESA modelindeki öznetelik vektöründen farklı biçimdeki bir veri düzenleşimi ile çalıştığından mevcut öznetelik haritasını gösteren matristeki eksenlerin yerlerini değiş tokuş etmek ve düzenleme için yeniden şekillendirme katmanı [56] ve hemen ardından permütasyon katmanı [57] uygulanmıştır. Tablo 2’de görüldüğü üzere ESA ağ bileşeninin arkasına TSA tipinde uygun bir ağ bileşeni eklenmesi tercih edilecekse bu ağ modeline uygun olarak dört adet TSA katmanı (TSA_#1, TSA_#2, TSA_#3 ve TSA_#4) ile melez derin sinir ağ modeli oluşturulmuştur. Tablo 2’de yer alan TSA tipli bileşenler UKSB veya KTB olabilmektedir. Tablo 2’de TSA tipli katmanlarla kurulmuş örnek bir melez derin

sinir ağ modeli için son sütunda parametre sayıları gösterilmiştir. Bu tablodaki toplam parametre sayısı, ESA bileşeni ve KTB katmanlarından oluşan melez model için verilmiştir. Tam bağlantılı katman biçimindeki sırasıyla 4096 ve 512 düğümünden oluşan yoğun katmanlar uygun aktivasyon fonksiyonu ile çalıştırılarak, bu katmanlar arasında belirlenerek çıktı katmanının cinsiyet için erkek veya kadın sınıfını gösteren sonuçlar üretmesi sağlanmıştır. Yoğun katmanda aktivasyon fonksiyonu tipi olarak *softmax* fonksiyonu seçilmiştir. Bu fonksiyon normalize edilmiş üstel bir fonksiyondur. Böylece verilen her bir girdi değerinin belirli bir sınıfla ilişkili olma ihtimalini gösteren [0, 1] arası çıktı değerleri üretilmektedir [58]. Sinir ağlarının eğitiminde literatürde yer alan stokastik eğimli iniş yaklaşımına benzer olan ADAM (Adaptive Moments Estimation) eniyileme yöntemi kullanılmıştır [59, 60]. Modelin eğitimi esnasında erken durdurma (early stopping) işlemi uygulanarak, modelin sınıflandırma kalitesini geçermek (validation) adına veri kümelerinin her biri için eğitim amacıyla kullanılan kısmının %20 kadarlık belirli bir parçası üzerinden geçermek kayıp fonksiyonu değeri takip edilmiştir. Buradaki kayıp fonksiyonunun tipi için seyrek kategorik çapraz entropi kullanılmıştır. Eğitim süreci boyunca izlenen geçermek kayıp fonksiyonunun minimum değerinin elde edilmesi için delta miktarı olarak 0,0001'lik bir değişim oluşmuyorsa eğitim sırasındaki sınıflandırma kalitesinde olumlu yönde bir gelişme olmadığına dair programatik olarak eğitimin durdurulmasına karar verilir. Bir derin öğrenme uygulama programlama arayüzü olan Keras'ın sağladığı "ReduceLRonPlateau" ve "EarlyStopping" fonksiyonları kullanılarak eğitim kalitesinin iyileştirilmesinin takibinde yararlanılmıştır [61]. EarVN [62] veya WPUT [63] veri kümeleri için gerçekleştirilen deneylerde ADAM eniyileme yöntemi Keras üzerindeki "optimizer" fonksiyonunun sağladığı "Adam" sınıfı kullanılarak öğrenme oranı için ilk değer olarak 0,001 ile eğitime başlayıp ilgili adımlarda 0,3162 katı kadar öğrenme oranı küçültülerek alt sınır 5×10^{-7} olacak şekilde işlemler yapılarak gerekli ağ parametreleri erken durdurma için takip edilmiştir [60].

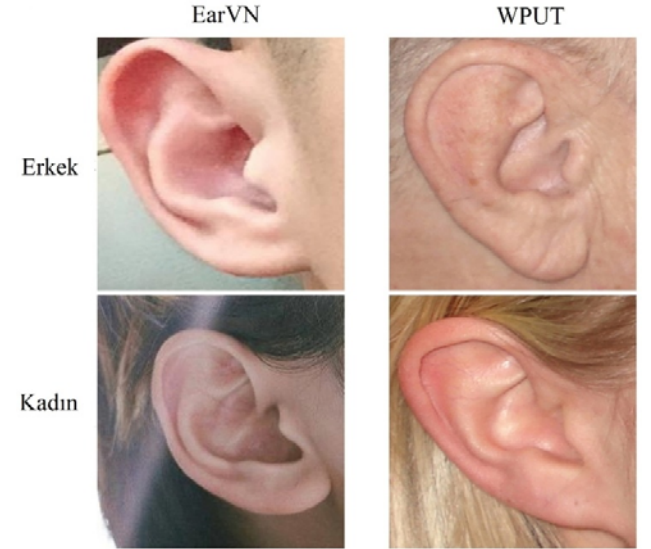
4. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

Bu bölümde, kullanılan veri kümeleri, uygulanan deneysel prosedür, dikkate alınan nesnel başarımlar değerlendirme kriterleri ve elde edilen deneysel sonuçlar hakkında detaylı bilgiler yer almaktadır.

4.1. Veri kümesi (Dataset)

İnsan kulağından cinsiyet tanıma için gerçekleştirilen çalışma kapsamında EarVN [62] ve WPUT [63] olmak üzere iki farklı görüntü veri kümesi kullanılmıştır. EarVN veri kümesi, 98'i erkek ve 66'sı kadın olmak üzere toplam 164 Asyalı insandan toplanan kulak görüntülerinden oluşmaktadır [64]. Bu veri kümesi içerisinde 28412 renkli kulak görüntüsü bulunmakta ve kısıtlanmamış koşullar

altında her bir kişinin iki kulağına ait görüntüleri içermektedir. Kişilerin orijinal yüz görüntüleri, kamera sistemleri ve ışık koşulları dahil olmak üzere kısıtlanmamış ortamda elde edilmiştir. Kulak görüntüleri daha sonra büyük poz, ölçek ve aydınlatma varyasyonları üzerinden yüz görüntülerinden kırılmıştır. WPUT veri kümesi ise mevcut kulak veri kümelerindeki bazı kısıtlamalardan kaçınmak ve daha fazla veri temsili elde etmek amacıyla toplanmış bir kulak görüntüsü veri kümesidir [65]. Bu veri kümesi, 254'ü kadın ve 247'si erkek olmak üzere her yaşta 501 bireyin toplamda 2071 kulak görüntüsünü içermektedir. Bunun yanı sıra çeşitli çalışmalarda belirtildiği gibi 27 bireyin görüntülerinin veri kümesinin Internet'ten indirilen dosyalarında eksik olmasından dolayı 474 birey içerecek şekilde toplamda 3348 görüntüye sahiptir. Her kişi için farklı günlerde ve aydınlatma koşullarında kulak başına profil ve yarı profil fotoğrafları çekilmiştir [6]. Şekil 3'te sırasıyla EarVN [62, 64] ve WPUT [63, 65] veri kümeleri için erkek ve kadın sınıf etiketli görüntülerden birkaç örnek görüntü verilmiştir.



Şekil 3. EarVN ve WPUT veri kümelerinden alınan birkaç örnek görüntü

(A few sample images taken from EarVN and WPUT datasets)

4.2. Başarımlar değerlendirme (Performance evaluation)

Sınıflandırma modellerinin değerlendirilmesinde literatürde yer alan Çapraz Tahmin Matrisi'ne (Confusion Matrix) göre; Doğru Pozitif (DP), Yanlış Negatif (YN), Yanlış Pozitif (YP), Doğru Negatif (DN) değerleri kullanılmaktadır. Doğruluk (accuracy), sınıflandırıcının sınıf ayrıştırıcılığı yeteneğini belirlemek amacıyla yaygın olarak kullanılan ölçütlerden biridir. Kullanılan modelde doğru olarak sınıflandırılan test örneklerinin değeri olarak tanımlanmaktadır [45]. Bu değer, işlemsel gösterimlerde yüzdelik oran olarak ifade edilebilmektedir. Doğruluk oranı ölçütü, Eş. 3'te verilen biçimiyle hesaplanmaktadır.

$$\text{Doğruluk} = (DP + DN) / (DP + YP + DN + YN) \quad (3)$$

ROC (Receiver Operating Characteristic) olarak literatürde geçen Alıcı İşletim Karakteristik (AİK) eğrisi, modelin ayırt etme gücünün belirlenmesini sağlar [45, 48]. Yaygın olarak kullanılan bir diğer ölçüt ise, AİK eğrisinin altında kalan alanı ifade eden, Eğri Altında Kalan Alan (EAKA)'dır. EAKA oranı ne kadar büyük ise sınıflandırıcının başarı oranı da o kadar iyidir. EAKA'nın olası değerleri 0,5 ile 1,0 arasında değişim gösterir. EAKA oranı Eş. 4'te verilen biçimiyle hesaplanmaktadır.

$$EAKA = \frac{1}{2} \left(\frac{DP}{DP + YN} + \frac{DN}{DN + YP} \right) \quad (4)$$

Duyarlılık (precision) ölçütü, Eş. 5'teki gibi; tespit olasılığı ölçütü ise Eş. 6'da görüldüğü gibi hesaplanır. Tespit olasılığı ölçütü, anma (recall) olarak da adlandırılır. F₁ skoru ölçütü, özellikle dengesiz (unbalanced) olan veri kümelerinde yapılan tahminin doğruluğunun sınanması için Eş. 7'de görüldüğü gibi hesaplanarak kullanılır [48]. Eş. 5, Eş. 6 ve Eş. 7'de verilen başarı ölçütleri her bir sınıf için ayrı olarak göz önüne alınır, çoklu sınıflar söz konusu olduğunda ölçütlerin aritmetik ortalaması ile hesaplama yapılmaktadır.

$$Duyarlılık = DP / (DP + YP) \quad (5)$$

$$Tespit Olasılığı = DP / (YN + DP) \quad (6)$$

$$F_1 = 2 * \left(\frac{Tespit Olasılığı * Duyarlılık}{Tespit Olasılığı + Duyarlılık} \right) \quad (7)$$

4.3. Deneysel sonuçlar ve tartışma (Experimental results and discussion)

Çalışmada başarı değerlendirme bölümünde belirtilen ölçütler kullanılmıştır. Deneylerdeki başarı sonuçları bu ölçütlere göre nesnel olarak değerlendirildiğinde melezleme sayesinde daha yüksek oranlara ulaşıldığı görülmektedir. EarVN ve WPUT isimli iki farklı veri kümesinin yanı sıra bu iki veri kümesinin bileşkesi olan bir veri kümesi üzerinden hem tek başına ESA hem de ESA bileşenli ve TSA tipi katmanlarla oluşturulan melez mimari model test edilmiştir. Çalışmada veriler işlenirken yığın büyüklüğü (batch size) olarak 16 seçilmiştir. Her iki görüntü veri kümesi için de yaklaşık %80'lik kısmı eğitim ve kalan kısmı ise test için ayrılmıştır. Eğitime ayrılan kısmın %20'si de geçiş için kullanılmıştır. Bu çalışmada, Python programlama dili v3.7 versiyonu kullanılarak derin öğrenme için yazılımın alt yapısı oluşturulmuştur [66, 67]. Derin öğrenme kütüphanesi olarak Tensorflow [68] v2.10, Keras [61] için v2.31

versiyonları kullanılmıştır. Çizilen grafikler için Matplotlib [69] kütüphanesinin v3.32 versiyonu tercih edilmiştir. Görüntülerin işlenmesinde OpenCV kütüphanesi [70] v3.4.2 versiyonu kullanılmıştır. Deneyler, 16 GB yerel bellekli, Intel i7 64 bit işlemciye sahip ve 4 GB belleği bulunan 896 CUDA çekirdekli Nvidia GTX 1650 ekran kartı olan bilgisayar üzerinde gerçekleştirilmiştir. Tablo 3 ve Tablo 5 içerisinde; kullanılan mimari model, erken durdurma, eğitim yinleme sayısı (EYS), sınıflandırma hata değeri olarak eğitim kayıp değeri (EKD) ve yüzdelik oran olarak verilmekte olan eğitim doğruluğu (ED), test doğruluğu (TD) değerleri yanı sıra tespit olasılığı (TO), duyarlılık ve F₁ skoru değerleri yer almaktadır. İki farklı sınıf içeren deneylerde başarı değerlendirme için duyarlılık, TO, F₁ skoru Tablo 3 ve Tablo 5'te makro ortalama değerleri olarak verilmiştir. Makro ortalama değeri hesaplanırken her iki sınıfı eşit ağırlıklı olarak değerlendirebilmek adına Keras altyapısı aracılığıyla her bir sınıf için ilgili başarı ölçütleri bağımsız olarak hesaplandıktan sonra bu ölçütlerin aritmetik ortalaması alınmıştır. Testler sonucunda elde edilen en büyük değerler, tablolar içerisinde italik yazı tipinde gösterilmektedir.

Yapılan deneyler göstermiştir ki 2B-ESA+KTB modeli için WPUT veri kümesi üzerinde %87,61 test doğruluk oranı elde edilmiştir. 2B-ESA modeli tek başına kullanıldığında %82,08 test doğruluk oranı değerine ulaşırken, 2B-ESA+UKSB melez modeli kullanıldığında elde edilen en yüksek test doğruluk oranı değerinin %82,98 olarak kaldığı görülmektedir. Veri kümesinin ölçek büyüklüğüne bağlı olarak, modelin eğitimi esnasında erken durdurma yöntemi kullanıldığında bazı deneylerde derin sinir ağının veriyi daha az öğrenmesi (underfitting) durumu gerçekleşebilmektedir. Erken durdurmanın mevcut olmadığı durumlarda, ağ tarafından verinin ezberlenmesi ile bazı deneylerde karşılaşılabilir. Deney sonuçlarına dair başarının değerlendirilmesinde bu tarz sorunların etkisinin olmaması adına önerilen yaklaşım hem erken durdurma varken hem de erken durdurmanın mevcut olmadığı deneylerle sınanmıştır. Tablo 3'te yer alan deney sonuçlarına bakıldığında, 2B-ESA+KTB melez modelinin test doğruluğu oranı erken durdurmanın mevcut olmadığı deneyde, erken durdurmanın mevcut olduğu deneye göre yaklaşık %2 daha yüksek çıkmıştır. Melez modeller birbirleri ile kıyaslandığında 2B-ESA+KTB modelinin, 2B-ESA+UKSB modeline göre erken durdurma mevcutken %4,5 kadar daha iyi bir test doğruluğu başarımı elde ettiği görülmektedir. Tablo 4'te görüldüğü üzere WPUT veri kümesinde yapılan deneylerde erken durdurma varken, en yüksek başarıya sahip 2B-ESA+KTB

Tablo 3. WPUT veri kümesi için eğitim ve test sonuçları (Training and testing results for WPUT dataset)

Kullanılan Mimari Model	Erken Durdurma	EYS	EKD	ED (%)	TD (%)	Duyarlılık	TO	F ₁ skoru
2B-ESA	Açık	11	0,14400	94,5	62,23	0,742	0,604	0,549
2B-ESA	Kapalı	50	0,00100	100	82,08	0,839	0,815	0,816
2B-ESA+UKSB	Açık	13	0,03900	98,5	81,04	0,810	0,809	0,809
2B-ESA+UKSB	Kapalı	50	0,00001	100	82,98	0,830	0,828	0,829
2B-ESA+KTB	Açık	18	0,00001	100	85,67	0,859	0,854	0,855
<i>2B-ESA+KTB</i>	<i>Kapalı</i>	<i>50</i>	<i>0,00001</i>	<i>100</i>	<i>87,61</i>	<i>0,882</i>	<i>0,873</i>	<i>0,874</i>

Tablo 4. WPUT veri kümesi için çapraz tahmin matrisi (The confusion matrix for WPUT dataset)

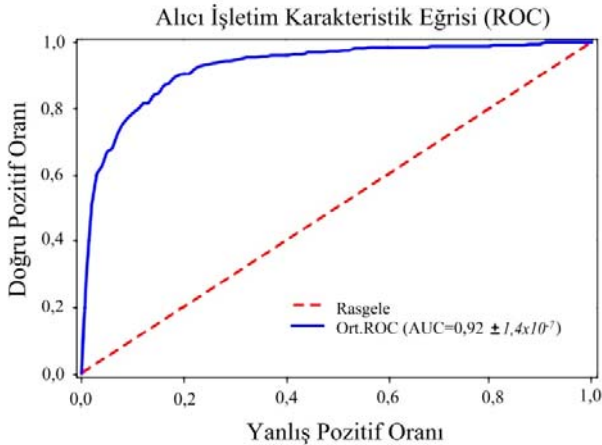
		2B-ESA		2B-ESA + UKSB		2B-ESA+ KTB		
		Tahminlenen Sınıf		Tahminlenen Sınıf		Tahminlenen Sınıf		
		Erkek	Kadın	Erkek	Kadın	Erkek	Kadın	
Geçerli Sınıf	Erkek	74	245	255	64	257	62	<i>Erken Durdurma Mevcut</i>
	Kadın	8	343	63	288	34	317	
	Erkek	221	98	255	64	258	61	<i>Erken Durdurma Mevcut Değil</i>
	Kadın	22	329	50	301	22	329	

Tablo 5. EarVN veri kümesi için eğitim ve test sonuçları (Training and testing results for EarVN dataset)

Kullanılan Mimari Model	Erken Durdurma	EYS	EKD	ED (%)	TD (%)	Duyarlılık TO	F ₁ skoru
2B-ESA	Açık	17	0,041	98,62	83,90	0,866	0,837
2B-ESA	Kapalı	50	0,003	99,90	84,45	0,866	0,843
2B-ESA+UKSB	Açık	10	0,110	95,79	79,75	0,828	0,794
2B-ESA+UKSB	Kapalı	50	0,002	99,92	82,56	0,855	0,823
2B-ESA+KTB	<i>Açık</i>	<i>10</i>	<i>0,039</i>	<i>98,52</i>	<i>85,16</i>	<i>0,866</i>	<i>0,851</i>
2B-ESA+KTB	Kapalı	50	0,002	99,92	81,68	0,842	0,823

ile 319 adet erkek etiketli görüntüden 257 adedi doğru şekilde erkek cinsiyeti etiketiyle, 351 adet kadın etiketli görüntüden ise 317 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmıştır. Ayrıca, WPUT veri kümesinde gerçekleştirilen erken durdurmanın mevcut olmadığı çalışmada, en yüksek başarıma sahip 2B-ESA+KTB modeli, 319 adet erkek etiketli görüntüden 258 adedi doğru şekilde erkek cinsiyeti etiketiyle, 351 adet kadın etiketli görüntüden ise 329 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmıştır.

AİK eğrisi grafiğinde koyu kalın eğri ile ifade edilen ortalama değerler doğru pozitif ve yanlış pozitif oranlarından elde edilerek, mevcut standart sapma da göz önüne alındığında oluşan sınıflandırma başarıma Şekil 4'te görülmektedir.



Şekil 4. WPUT üzerinde çalışılan 2B-ESA bileşenli ve KTB katmanları olan melez model için AİK
(The ROC plot for hybrid model with 2D CNN component and GRU layers running on WPUT dataset)

2B-ESA+KTB melez modeli WPUT veri kümesi için Şekil 4'te görüldüğü gibi oldukça iyi bir sınıflandırma başarıma ile EAKA değeri olarak 0,92 oranı ve standart sapma değeri

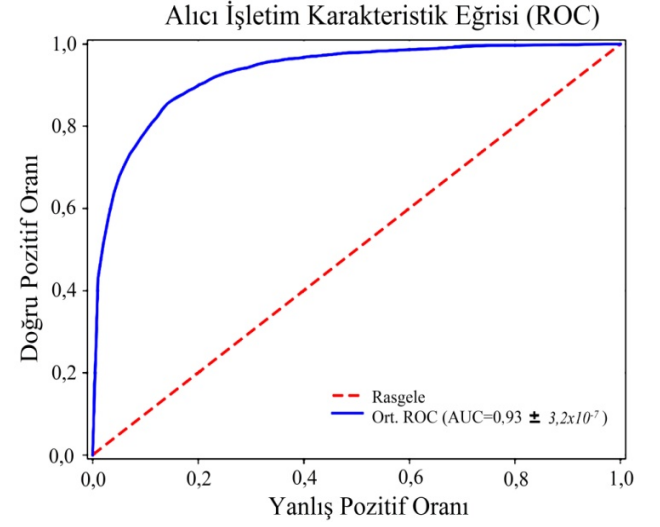
olarak $1,4 \times 10^{-7}$ elde etmiştir. AİK eğrilerinde en iyi sınıflandırma başarıma sınıflandırıcının ayrıştırıcılık özelliğini gösterecek şekilde eğrinin sol üst köşedeki 1 değerine yaklaştığı halinde elde edilmektedir. Tablo 5'te görüldüğü üzere, yapılan deneylerde 2B-ESA+KTB mimari modeli EarVN veri kümesi için %85,16'lık bir test doğruluğu oranı elde edilmiştir. 2B-ESA modeli tek başına kullanıldığında %84,45 test doğruluk oranı değerine ulaşırken, 2B-ESA+UKSB melez modeli kullanıldığında elde edilen en yüksek test doğruluk oranı değerinin %82,56 olarak kaldığı görülmektedir. Tablo 5'te yer alan deney sonuçlarına bakıldığında, 2B-ESA+KTB melez modelinin test doğruluğu oranı erken durdurmanın mevcut olduğu deneyde, erken durdurmanın mevcut olmadığı deneyde göre yaklaşık %3,5 daha yüksek çıkmıştır. Melez modeller birbirleri ile kıyaslandığında 2B-ESA+KTB modelinin, 2B-ESA+UKSB modeline göre erken durdurma mevcutken %5 kadar daha iyi bir test doğruluğu başarıma elde ettiği görülmektedir.

Tablo 6'da görüldüğü üzere EarVN veri kümesinde yapılan deneylerde erken durdurma varken, 2B-ESA+KTB ile 2532 adet erkek etiketli görüntüden 2398 adedi doğru şekilde erkek cinsiyeti etiketiyle, 2774 adet kadın etiketli görüntüden ise 2121 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmıştır. Ayrıca, EarVN veri kümesi için erken durdurmanın mevcut olmadığı çalışmada, 2B-ESA+KTB modeli gerçekleştirilen deneyde 2532 adet erkek etiketli görüntüden 2428 adedi doğru şekilde erkek cinsiyeti etiketiyle, 2774 adet kadın etiketli görüntüden ise 1906 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmıştır.

Şekil 5'te görüldüğü üzere, koyu kalın eğri ile ifade edilen ortalama değerler doğru pozitif ve yanlış pozitif oranlarından elde edilerek, AİK eğrisi grafiğinde mevcut standart sapma da hesaba katılarak sınıflandırma başarıma oranı verilmektedir. 2B-ESA+KTB melez modeli EarVN veri kümesi için Şekil 5'te görüldüğü gibi oldukça iyi bir sınıflandırma başarıma ile EAKA değeri olarak 0,93 oranı ve

standart sapma değeri olarak $3,2 \times 10^{-7}$ elde etmiştir. Yapılan deneyleri kapsam olarak daha fazla kulak görüntüsü ile genişletmek için EarVN ve WPUT veri kümelerini bir arada kullanarak, bileşke (composite) adı verilen veri kümesi oluşturulmuştur. Bu bileşke veri kümesi (BVK) için diğer iki görüntü veri kümesinde de olduğu gibi yaklaşık %80'lik kısmı eğitim ve yaklaşık %20'lik kısmı ise test için ayrılmıştır. Eğitime ayrılan kısmın %20'si de değerlendirme için kullanılmıştır. Oluşturulan BVK veri kümesi için yapılan deneyler erken durdurma mevcutken ve erken durdurma mevcut değilken durumlarında gerçekleştirilmiştir. Deneysel sonuçlar sayesinde deneylerde kullanılan veri kümelerindeki görüntü sayısı ve çeşitliliğinin ilgili modelin başarımına etkisinin olup olmadığı elde edilen deney sonuçlarının değerlendirildiği nesnel ölçütler ile ortaya konulmuştur. Yapılan deneylerde bileşke veri kümesi, EarVN ve WPUT veri kümelerinin bir araya getirilmesiyle oluşturulduğundan görüntü sayısındaki artış nedeniyle veri ölçeceğindeki değişimin ve veri çeşitliliğinin çalışmada önerilen derin sinir ağı modellerinin başarımlarını nasıl etkilediği incelenmiştir. Bu deneylerde, 2B-ESA+UKSB mimari modeli ile bileşke veri kümesi (BVK) için %85,64'lük bir test doğruluğu oranının en yüksek değer olarak deneylerde elde edildiği Tablo 7'de görülmektedir. 2B-ESA modeli tek başına kullanıldığında %85,30 test doğruluk oranı değerine ulaşırken, 2B-ESA+KTB melez modeli kullanıldığında elde edilen en yüksek test doğruluk oranı değerinin %85,05 olarak kaldığı görülmektedir. Tablo 7'de yer almakta olan deney sonuçları incelendiğinde, 2B-ESA+UKSB melez modelinin test doğruluğu oranı erken durdurmanın mevcut olduğu deneyde, erken durdurmanın mevcut olmadığı deneye göre yaklaşık %1,68 daha yüksek çıktığı görülmektedir. Melez modeller birbirleri ile kıyaslandığında 2B-ESA+UKSB modelinin erken durdurma mevcut değilken, 2B-ESA+KTB modelinin erken durdurma mevcut olan deney sonucuna göre %0,59 kadar daha iyi bir test doğruluğu başarımı elde ettiği görülmektedir. İki melez modelin başarımları arası farkın %1 değerinin altında olması

nedeniyle her iki melez modelin BVK veri kümesi için birbirine denk başarımda olduğu anlaşılmaktadır.



Şekil 5. EarVN üzerinde çalışan 2B-ESA bileşenli ve KTB katmanları olan melez model için AİK
(The ROC plot for hybrid model with 2D CNN component and GRU layers running on EarVN dataset)

Tablo 8'de görüldüğü üzere BVK veri kümesinde yapılan deneylerde erken durdurma varken, 2B-ESA+UKSB ile 2851 adet erkek etiketli görüntüden 2628 adedi doğru şekilde erkek cinsiyeti etiketiyle, 3125 adet kadın etiketli görüntüden ise 2390 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmıştır. Ayrıca, BVK veri kümesi için erken durdurmanın mevcut olmadığı çalışmada, 2B-ESA+UKSB modeli gerçekleştirilen deneyde 2851 adet erkek etiketli görüntüden 2705 adedi doğru şekilde erkek cinsiyeti etiketiyle, 3125 adet kadın etiketli görüntüden ise 2413 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmıştır.

Tablo 6. EarVN veri kümesi için çapraz tahmin matrisi (The confusion matrix for EarVN dataset)

	2B-ESA		2B-ESA + UKSB		2B-ESA+ KTB			
	Tahminlenen Sınıf		Tahminlenen Sınıf		Tahminlenen Sınıf			
	Erkek	Kadın	Erkek	Kadın	Erkek	Kadın		
Geçerli Sınıf	Erkek	2461	71	2416	116	2398	134	<i>Erken Durdurma Mevcut</i>
	Kadın	783	1991	958	1816	653	2121	
Sınıf	Erkek	2463	69	2474	58	2428	104	<i>Erken Durdurma Mevcut Değil</i>
	Kadın	756	2018	867	1907	868	1906	

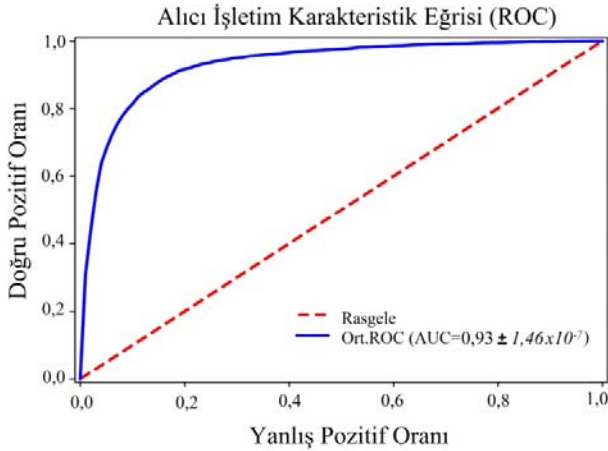
Tablo 7. BVK veri kümesi için eğitim ve test sonuçları (Training and testing results for BVK dataset)

Kullanılan Mimari Model	Erken Durdurma	EYS	EKD	ED (%)	TD (%)	Duyarlılık	TO	F ₁ skoru
2B-ESA	Açık	13	0,08100	96,99	82,63	0,852	0,832	0,824
2B-ESA	Kapalı	50	0,00180	99,98	85,30	0,866	0,857	0,852
2B-ESA+UKSB	Açık	13	0,02470	99,16	83,96	0,848	0,843	0,839
2B-ESA+UKSB	Kapalı	50	0,00062	99,99	85,64	0,867	0,860	0,856
2B-ESA+KTB	Açık	11	0,03610	98,76	85,05	0,857	0,853	0,850
2B-ESA+KTB	Kapalı	50	0,00043	100	84,32	0,855	0,847	0,842

Tablo 8. BVK veri kümesi için çapraz tahmin matrisi (The confusion matrix for BVK dataset)

	2B-ESA		2B-ESA + UKSB		2B-ESA+ KTB			
	Tahminlenen Sınıf		Tahminlenen Sınıf		Tahminlenen Sınıf			
	Erkek	Kadın	Erkek	Kadın	Erkek	Kadın		
Geçerli Sınıf	Erkek	2763	88	2628	223	2641	210	<i>Erken Durdurma Mevcut</i>
	Kadın	950	2175	735	2390	683	2442	
Sınıf	Erkek	2721	130	2705	146	2689	162	<i>Erken Durdurma Mevcut Değil</i>
	Kadın	748	2377	712	2413	775	2350	

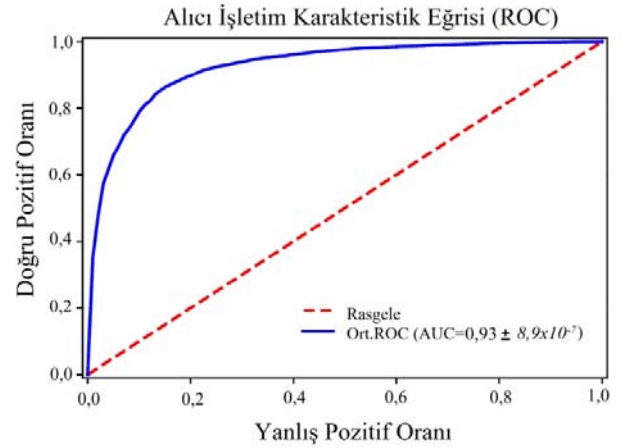
BVK veri kümesinde yapılan diğer deneylerde erken durdurma varken, 2B-ESA+KTB ile 2851 adet erkek etiketli görüntüden 2641 adedi doğru şekilde erkek cinsiyeti etiketiyle, 3125 adet kadın etiketli görüntüden ise 2442 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmış olduğu Tablo 8'den görülmektedir. Ayrıca, BVK veri kümesi için erken durdurmanın mevcut olmadığı çalışmada, 2B-ESA+KTB modeli gerçekleştirilen deneyde 2851 adet erkek etiketli görüntüden 2689 adedi doğru şekilde erkek cinsiyeti etiketiyle, 3125 adet kadın etiketli görüntüden ise 2350 adedi doğru şekilde kadın cinsiyeti etiketi ile tanınarak sınıflandırılmış olduğu da yine Tablo 8'den görülmektedir. Şekil 6 ve Şekil 7'de görüldüğü üzere, koyu kalın eğri ile ifade edilen ortalama değerler doğru pozitif ve yanlış pozitif oranlarından elde edilerek, AİK eğrisi grafiğinde mevcut standart sapma da hesaba katılarak sınıflandırma başarımları verilmektedir.



Şekil 6. BVK üzerinde çalışan 2B-ESA bileşenli ve UKSB katmanları olan melez model için AİK
(The ROC plot for hybrid model with 2D CNN component and LSTM layers running on BVK dataset)

2B-ESA+UKSB melez modeli BVK veri kümesi için erken durdurma mevcut değilken yapılan deneyde Şekil 6'da görüldüğü gibi oldukça iyi bir sınıflandırma başarımları ile EAKA değeri olarak $0,93$ oranı ve standart sapma değeri olarak $1,46 \times 10^{-7}$ elde etmiştir. Ayrıca, Şekil 7'de görüldüğü gibi BVK veri kümesi için erken durdurma mevcutken yapılan deneyde 2B-ESA+KTB melez modeli oldukça iyi bir sınıflandırma başarımları ile EAKA değeri olarak $0,93$ oranı ve standart sapma değeri olarak $8,9 \times 10^{-7}$ elde etmiştir. Şekil 6 ve Şekil 7'deki AİK eğrileri incelendiğinde, 2B-ESA+KTB

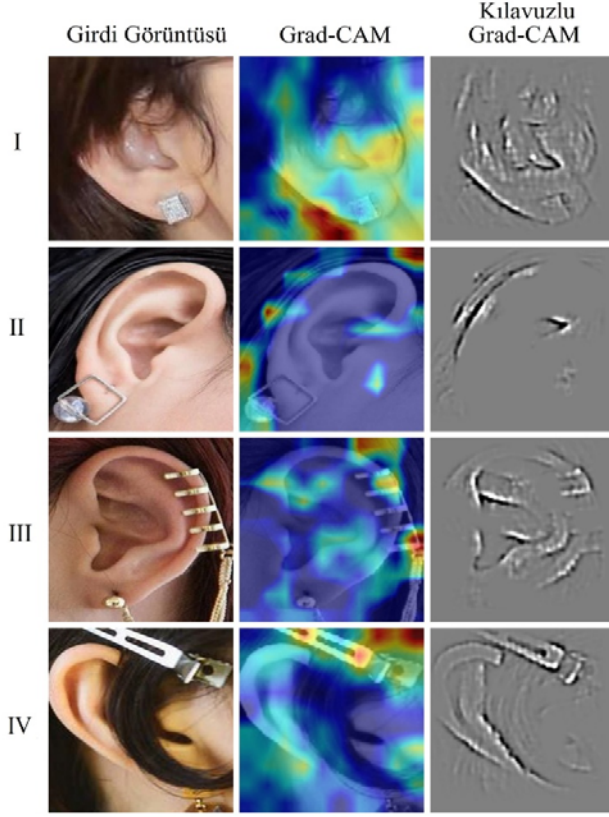
melez modelinin standart sapma değerinin 2B-ESA+UKSB melez modelinkine göre farklı oluşu başarımları arasındaki farkı da işaret etmektedir. Diğer veri kümelerinin (EarVN ve WPUT) her biri ile bağımsız yapılan önceki deneylerde 2B-ESA+KTB melez modeli daha yüksek başarımlı sonuçlar elde ederken, BVK veri kümesi için 2B-ESA+UKSB melez modeli çok daha yüksek başarımlı sonuçlar elde etmiştir. Bunun temel nedeni, UKSB modelinin veri çeşitliliği ve veri kümesi ölçeği göz önüne alındığında KTB modeline göre daha uzun dizileri (sequence) anlamsallık adına incelerken daha başarılı olmasından kaynaklanmaktadır. BVK veri kümesinin EarVN ve WPUT veri kümelerinin bileşkesi şeklinde ele alınarak oluşturulan bu yeni biçim verideki ölçek değişikliğine neden olmuştur. Yapılan bu ölçek değişikliğinin, KTB modelinin işleyebildiğinden daha uzun erimli korelasyonların UKSB tarafından ortaya çıkartılması yoluyla öğrenmenin anlamsal olarak zenginleştirilmesinde oldukça yararlı olduğu deneysel sonuçlardan anlaşılmaktadır.



Şekil 7. BVK üzerinde çalışan 2B-ESA bileşenli ve KTB katmanları olan melez model için AİK
(The ROC plot for hybrid model with 2D CNN component and GRU layers running on BVK dataset)

Sınıf aktivasyon haritası görselleştirmesinde, Grad-CAM [59, 71, 72] yöntemi kullanılarak ilgili renk kanalının sınıfına göre gradyan (eğim) değerleri üzerinden ağırlıklandırma yapılır. Buradaki yoğunluk haritası etkinliğin görüntünün hangi bölümlerinde yoğunlaştığını gösterir. Literatürde genellikle bu yoğunluk haritası ısı haritası (heatmap) biçiminde verilmektedir. Sonrasında ilgili ısı haritası girdi görüntüsünün üzerine bindirilerek

(superimpose) uygulanmaktadır [71-73]. Çalışmamızdaki melezleme yaklaşımının cinsiyet ayırtma başarımına dair sınıf aktivasyon haritaları görselleştirilmiştir. Şekil 8’de deneylerimizdeki veri kümelerinden [62-65] rastgele seçilmiş olan bazı görüntülerin üzerine bindirilen ısı haritaları tabanlı Grad-CAM sonuçları görülmektedir.



Şekil 8. Cinsiyet tanımda melez modelin gradyan ağırlıklandırılmış sınıf aktivasyonu ısı haritalarına dayanan bindirilmiş görüntüler

(The superimposed images based on the heatmaps of gradient-weighted class activation of hybrid model for gender recognition)

Deneylerdeki her iki veri kümesiyle (EarVN [62, 64] ve WPUT [65]) en yüksek başarıyı elde eden 2B-ESA+KTB melez derin sinir ağı modeli Grad-CAM yöntemi kullanılarak sınıflandırmada verilen test görüntüleri üzerinden Şekil 8’de görülebileceği gibi sınanmıştır. Şekildeki ilgili sütunlarda sırasıyla; orijinal girdi görüntüsü, ısı haritası bindirilmiş Grad-CAM ve kılavuzlu Grad-CAM görüntülerine yer verilmektedir. Ayrıca melez modelin sınıflandırmadaki deneysel sonuçları olarak, Şekil 8’de ilk satırdaki (I) görüntülerde “küpe takmış ve doğru sınıflandırılmış erkek” örneği, ikinci satırdaki (II) görüntülerde “küpe takmış ve yanlış sınıflandırılmış kadın” örneği, “farklı nesnelere kulağın bir kısmını örttüğü” örnekler olarak üçüncü satırdaki (III) görüntülerde erkek sınıfına ait örneğin melez model tarafından doğru sınıflandırıldığı ve dördüncü satırdaki (IV) görüntülerde ise kadın sınıfına ait örneğin melez model tarafından yanlış sınıflandırıldığı deneysel Grad-CAM sonuçları

verilmektedir. Kulağın bir kısmını örten nesnenin cinsiyet ayırt etmede etkin belirleyici olmadığı anlaşılmıştır.

5. SONUÇLAR VE TARTIŞMALAR (RESULTS AND DISCUSSIONS)

Bu çalışmada insan kulağı görüntüleriyle insanların cinsiyetlerine göre sınıflandırılmasına dayanan derin öğrenme tabanlı melezlemeyi kullanan sinerjik bir cinsiyet tanıma yaklaşımı oluşturulmuştur. Derin ağı mimari modelleriyle öğrenilmiş öznelikler sayesinde mevcut anlamsal boşluk çalışmamızdaki melezleme tabanlı sinerji sayesinde doldurulmaktadır. TSA tipli bileşenler (UKSB ve KTB) sayesinde melez sinir ağı modelinin görüntü içerisindeki piksel bölgeleri arası ilişkisel bağımlılıkların daha iyi anlaması sağlanmıştır. Bu çalışmanın ana katkısı, literatürdeki kulak bölgesi ile tespit ve tanıma çalışmalarına farklı bir açıdan bakarak yeni bir çözüm önermesi, cinsiyet tanıma alanında derin öğrenme mimari modellerinin kullanımını yüksek başarımla sonuçlarıyla sunmasıdır. Çalışmamızdaki bir diğer katkı ise melezleme yoluyla mevcut tek başına modele (ESA) göre daha yüksek doğrulukta sınıflandırmayı ve daha az tanıma hatasını içerecek bir derin sinir ağını oluşturacak mimari yapının TSA tipli bileşen içeren yeni melez modellerin kullanımıyla önermiş olmasıdır. Melezleme yaklaşımına dayanan melez sinir ağı modellerinin veri kümesindeki veri çeşitliliğine neden olan ölçeklendirme, gürültü ve çeşitli çevresel etkilerden çok daha az etkilendiği, böylece cinsiyet tanımda daha gürbüz olduğu görülmektedir. Kulaktan birey tanıma veya cinsiyet tanıma kısıtlanmamış koşullarda (*in-the-wild*) oldukça zor çözülebilen bir problemdir. Çözüm olarak, öznelik işleme daha başarılı olan derin öğrenmenin bağlamsal bilgiyi işleyebilme yeteneği sayesinde çeşitli etkilerden oldukça az etkilenen ve sınıflandırma için daha iyi ayrıştırıcılık sağlayan yüksek seviyeli öznelikler çalışmamızda oluşturulabilmiştir. İleriki çalışmalarımızda çeşitli derin öğrenme sinir ağı mimari modellerinin farklı ölçekte ve hacimdeki görüntü veri kümeleri üzerinden sınanması planlanmaktadır.

TEŞEKKÜR (ACKNOWLEDGEMENT)

Bu çalışmada kullanılan halka açık insan kulağı görüntüleri içeren veri kümelerinin sahiplerine teşekkür ederiz.

KAYNAKLAR (REFERENCES)

- Wayman, J. L., Jain, A. K., Maltoni, D., Maio, D., Biometric Systems: Technology, Design and Performance Evaluation, Springer-Verlag London, XIV-370, 2005.
- Resmi, K. R., Raju, G., Automatic 2D Ear Detection: A Survey, International Journal of Scientific & Technology Research, 8 (11), 3643-3647, 2019.
- Phadke, S., The Importance of a Biometric Authentication System, The SIJ Transactions on Computer Science Engineering & its Applications CSEA, 1 (4), 2013.

4. Watne, K. S., Thermal Imaging of Ear Biometrics for Authentication Purposes, Master's Thesis, Gjøvik University College, Department of Computer Science and Media Technologies, Norway, 2008.
5. Iannarelli, A.V., Ear Identification, Forensic Identification Series, Paramount Publishing Company, 1989.
6. Hassaballah, M., Alshazly, H. A., Ali, A. A., Ear Recognition using Local Binary Patterns: A Comparative Experimental Study, *Expert Systems with Applications*, 128, 182-200, 2019.
7. Emeršič, Z., Štruc, V., Peer, P., Ear Recognition: More Than A Survey, *Neurocomputing*, 255, 26-39, 2017.
8. Victor, B., Bowyer, K., Sarkar, S., An Evaluation of Face and Ear Biometrics, 16th International Conference on Pattern Recognition, IEEE, 1, 429-432, 2002.
9. Hurley, D. T., Nixon, M. S., Carter, J. N., Ear Biometrics by Force Field Convergence, In *Proceedings of the Audio- and Video-Based Biometric Person Authentication*, Springer, 386-394, 2005.
10. Nosrati, M. S., Faez, K., Faradji, F., Using 2D Wavelet and Principal Component Analysis for Personal Identification based on 2D Ear Structure, *International Conference on Intelligent and Advanced Systems*, IEEE, 616-620, 2007.
11. Annapurani, K., Sadiq, M. A. K., Malathy, C., Fusion of Shape of the Ear and Tragus - A Unique Feature Extraction Method for Ear Authentication System, *Expert System with Applications*, 42 (1), 649-656, 2015.
12. Anwar, A. S., Ghany, K. K. A., ElMahdy, H., Human Ear Recognition using SIFT Features, *Third World Conference on Complex Systems (WCCS)*, 1-6, 2015.
13. Galdamez, P. L., Raveane, W., Arrieta, A. G., A Brief Review of the Ear Recognition Process using Deep Neural Networks, *Journal of Applied Logic*, 24 (A), 62-70, 2017.
14. Lia, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M., Deep Learning for Generic Object Detection: A Survey, *International Journal of Computer Vision*, 128, 261-218, 2020.
15. Metin, İ. A., Karasulu, B., İnsan Aktivitelerinin Sınıflandırılmasında Tekrarlayan Sinir Ağı Kullanan Derin Öğrenme Tabanlı Yaklaşım, *Veri Bilimi*, 2 (2), 1-10, 2019.
16. Fırtına, N., Silahtaroglu, G., Kulaktan Kişi ve Kimlik Tespiti için Örnek Bir Uygulama, *Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 7 (2), 21-46, 2014.
17. Choras, M., Ear Biometrics Based on Geometrical Feature Extraction, *Electronic Letters on Computer Vision and Image Analysis*, 5 (3), 84-95, 2005.
18. Hurley, D. J., Nixon, M. S., Carter, J. N., Forced Field Feature Extraction for Ear Biometrics, *Computer Vision and Image Understanding*, 98, 491-512, 2005.
19. Kocaman, B., Ear Biometrics In Personal Identification, M.Sc.Thesis, Istanbul Technical University, Institute of Science and Technology, Turkey, 2008.
20. Singh, D., Singh, S. K., A Survey on Human Ear Recognition System Based on 2D and 3D Ear Images, *Open Journal of Information Security and Applications*, 1 (2), 21-30, 2014.
21. Abaza, A., Harrison, M. A. F., Ear recognition: a complete system, *Proc. SPIE 8712, Biometric and Surveillance Technology for Human and Activity Identification X*, 87120N, Baltimore, Maryland, United States, 2013.
22. Srivastava P., Agrawal D., Bansal A., Ear Detection and Recognition Techniques: A Comparative Review. In: Kolhe M., Tiwari S., Trivedi M., Mishra K. (eds) *Advances in Data and Information Sciences. Lecture Notes in Networks and Systems*, 94. Springer, Singapore, 2020.
23. Alshazly, H., Linse, C., Barth, E., Martinetz, T., Handcrafted versus CNN Features for Ear Recognition, *Symmetry*. 11 (12), 1493, 2019.
24. Emeršič, Z., Štepec, D., Štruc, V., Peer, P., Training convolutional neural networks with limited training data for ear recognition in the wild, *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, IEEE, Washington, DC, USA, 987-994, 2017.
25. Krizhevsky, A., Sutskever, I., Hinton, G.E., ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems (NIPS)*, 1097-1105, 2012.
26. Simonyan, K., Zisserman, A., Very deep convolutional networks for large-scale image recognition, *International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 7-9, 2015.
27. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size, arXiv:1602.07360, 2016.
28. Yosinski, J., Clune, J., Bengio, Y., Lipson, H., How transferable are features in deep neural networks? *27th International Conference on Neural Information Processing Systems*, Montreal, QC, Canada, 8-13 December, 3320-3328, 2014.
29. Alshazly, H., Linse, C., Barth, E., Martinetz, T., Ensembles of Deep Learning Models and Transfer Learning for Ear Recognition, *Sensors*. 19 (19), 4139, 2019.
30. Salah, A. A., Gevers, T., Sebe, N., Vinciarelli, A., Computer Vision for Ambient Intelligence, *Journal of Ambient Intelligence and Smart Environments*, 3 (3), 187-191, 2011.
31. Yuan, L., Mu, Z. C., Yang, F., A Review of Recent Advances in Ear Recognition, *Proceedings of 6th Chinese Conference on Biometric Recognition (CCBR)*, Beijing, China, 2011.
32. Bengio, Y., LeCun, Y., Hinton, G., Deep Learning, *Nature*, 521 (7553), 436-444, 2015.
33. Brownlee, J., What is Deep Learning? *Machine Learning Mastery*. Yayınlanma Tarihi: Ağustos 16, 2019. Çevrimiçi: <https://machinelearningmastery.com/what-is-deep-learning/>, Erişim Tarihi: Mart 23, 2021.

34. Şeker, A., Diri, B., Balık, H. H., Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme, Gazi Mühendislik Bilimleri Dergisi, 3 (3), 47-64, 2017.
35. Gündüz, G., Cedimoğlu, İ. H., Derin Öğrenme Algoritmalarını Kullanarak Görüntüden Cinsiyet Tahmini. Sakarya University Journal of Computer and Information Sciences, 2 (1), 9-17, 2019.
36. Takhtardeshir, S., Mahdipour, M., Ghaderi, R., Azimi, P., How Can Deep Learning Track Brain Metastasis Using Convolutional Neural Network?, 7th Iranian Human Brain Mapping Congress (IHBM 2020), November, 9-12. Iran, 2020.
37. Dickson, B., What are convolutional neural networks (CNN)?, TechTalks, Yayınlanma Tarihi: Haziran 6, 2020, Çevrimiçi: <https://bdtechtalks.com/2020/01/06/convolutional-neural-networks-cnn-convnets/>, Erişim Tarihi: Mart 24, 2021.
38. Aydın, E., Yüksek, S. E., Buried Target Detection with Ground Penetrating Radar Using Deep Learning Method, 25th Signal Processing and Communications Applications Conference (SIU 2017), 1-4, May, 15-18. Antalya, Turkey, 2017.
39. Namatëvs, I., Deep Convolutional Neural Networks: Structure, Feature Extraction and Training, Information Technology and Management Science, De Gruyter, 20, 40-47, 2017.
40. Cengil, E., Çınar, A., New Approach for Image Classification: Convolutional Neural Network, European Journal of Technic, 6 (2), 96-103, 2016.
41. Karasulu, B., Çoklu Ortam Sistemleri için Siber Güvenlik Kapsamında Derin Öğrenme Kullanarak Ses Sahne ve Olaylarının Tespiti, Acta Infologica, 3 (2), 60-82, 2019.
42. Lipton, Z. C., Berkowitz, J., Elkan, C., A Critical Review of Recurrent Neural Networks for Sequence Learning, arXiv e-prints arXiv:1506.00019, 2015.
43. Onan, A., Evrişimli Sinir Ağı Mimarilerine Dayalı Türkçe Duygu Analizi, Avrupa Bilim ve Teknoloji Dergisi, Özel Sayı, 374-380, 2020.
44. Amidi, A., Amidi, S., Recurrent Neural Networks. CS 230 - Deep Learning Course Materials. Stanford University. Yayınlanma Tarihi: Kasım 26, 2018, Çevrimiçi: <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks>, Erişim Tarihi: Mart 10, 2021.
45. Yıldız O., Melanoma detection from dermoscopy images with deep learning methods: A comprehensive study, Journal of the Faculty of Engineering and Architecture of Gazi University, 34 (4), 2241-2260, 2019.
46. Kaytan, M., Yeroğlu, C., Aydılek, İ. B., Yapay Sinir Ağları Eğitiminde Kullanılan Optimizasyon Yöntemlerinin İncelenmesi ve Kan Nakli Hizmet Merkezi Veri Seti Üzerinden Değerlendirilmesi. Anatolian Journal of Computer Sciences. 5 (2), 99-113, 2020.
47. Gülcü, A., Kuş, Z. A Survey of Hyper-parameter Optimization Methods in Convolutional Neural Networks. Gazi Üniversitesi Fen Bilimleri Dergisi, 7 (2), 503-522, 2019.
48. Metin İ. A., Karasulu B., A novel dataset of human daily activities: Its benchmarking results for classification performance via using deep learning techniques, Journal of the Faculty of Engineering and Architecture of Gazi University, 36 (2), 759-777, 2021.
49. Kilimci Z.H., Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (2), 635-650, 2020.
50. Hochreiter, S., Schmidhuber, J., Long short-term memory, Neural Computation, 9 (8), 1735-1780, 1997.
51. Srivastava, P., Essentials of Deep Learning: Introduction to Long Short Term Memory, Analytics Vidhya, Çevrimiçi: <https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to- lstm/>, Erişim Tarihi: Nisan 4, 2020.
52. Aytaç Altan, A., Karasu, S., Ayırıştırma Yöntemlerinin Derin Öğrenme Algoritması ile Tanımlanan Rüzgâr Hızı Tahmin Modeli Başarımına Etkisinin İncelenmesi, Avrupa Bilim ve Teknoloji Dergisi, 20, 844-853. 2020.
53. Chung, J., Gulcehre, C., Cho, K. H., Bengio, Y., Empirical evaluation of gated recurrent neural networks on sequence modeling. In NIPS 2014 Workshop on Deep Learning and Representation Learning, arXiv:1412.3555, December 11, 2014.
54. Görgel, P., Kavlak, E., Uzun Kısa Süreli Hafıza ve Evrişimsel Sinir Ağları ile Rüzgâr Enerjisi Üretim Tahmini, DÜMF Mühendislik Dergisi, 11 (1), 69-80, 2020.
55. GaussianNoise layer, Keras, Çevrimiçi: https://keras.io/api/layers/regularization_layers/gaussian_noise/, Erişim Tarihi: Mayıs 3, 2021.
56. Reshape layer, Keras, Çevrimiçi: https://keras.io/api/layers/reshaping_layers/reshape/, Erişim tarihi Mayıs 4, 2021.
57. Permute layer, Keras, Çevrimiçi: https://keras.io/api/layers/reshaping_layers/permute/, Erişim tarihi Mayıs 4, 2021.
58. Buduma, N., Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms, O'Reilly UK Ltd., 2017.
59. Chollet, F., Deep Learning with Python, Manning Publications, 2017.
60. Kingma, D. P., Ba, J., Adam: A Method for Stochastic Optimization, ICLR 2015, arXiv e-prints arXiv:1412.6980, <https://arxiv.org/abs/1412.6980>, 2014.
61. Keras, About Keras, <https://keras.io/about/>, Erişim tarihi Mayıs 8, 2021.
62. Hoang, V. T., EarVN1.0 dataset, Çevrimiçi: <https://data.mendeley.com/datasets/yws3v3mwx3/4>, Erişim tarihi Nisan 8, 2020.
63. Zhou, Y., Zaferiou, S., Deformable Models of Ears in-the-Wild for Alignment and Recognition, 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 626-633, 2017.

64. Hoang, V. T., EarVN1.0: A new large-scale ear images dataset in the wild, *Data in Brief*. Elsevier, 27, 104630, 2019.
65. Frejlichowski D, Tyszkiewicz N., The West Pomeranian University of Technology Ear Database A Tool for Testing Biometric Algorithms. In: Campilho A., Kamel M., (Editors). *Image Analysis and Recognition*. vol. 6112 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 227–234, 2010.
66. Bingöl K., Akan A., Örmecioglu H., Er A., Artificial intelligence applications in earthquake resistant architectural design: Determination of irregular structural systems with deep learning and Image AI method, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (4), 2971-2210, 2020.
67. Python programming language website, <https://www.python.org/>, Erişim tarihi Mayıs 20, 2021.
68. Tensorflow websitesi, <https://www.tensorflow.org/>, Erişim tarihi Mayıs 20, 2021.
69. Matplotlib: Visualization with Python web sitesi, <https://matplotlib.org/>, Erişim tarihi Mayıs 20, 2021.
70. OpenCV web sitesi, <https://opencv.org/>, Erişim tarihi Mayıs 20, 2021.
71. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra D., Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, 128 (2), 336–359, 2020.
72. Chattopadhyay, A., Sarkar, A., Howlader, P., Balasubramanian, V. N., Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks, *Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 839-884, March, 12-15, Lake Tahoe, NV, USA, 2018.
73. Islam, M. Z., Islam, M. M., Asraf, A., A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images, *Informatics in Medicine Unlocked*, 20 (100412), 1-11, 2020.