



Hybrid ConvViT model for classification of dental radiography images

Anıl Utku*^{ID}

Department of Computer Engineering, Faculty of Engineering, Munzur University, 62100, Tunceli, Türkiye

Highlights:

- A new ConvViT model was developed that combines CNN and ViT
- With ConvViT, it was aimed to effectively model both local and long-range features
- ConvViT had better classification results than DenseNet201, VGG-16, ResNet-50 and EfficientNetB0

Keywords:

- Dental radiography
- CNN
- ViT
- Pre-trained models
- Deep learning

Article Info:

Research Article

Received: 16.09.2024

Accepted: 03.04.2025

DOI:

10.17341/gazimmfd.1551005

Correspondence:

Author: Anıl Utku

e-mail:

anilutku@munzur.edu.tr

phone: +90 546 654 3601

Graphical/Tabular Abstract

In this study, a hybrid ConvViT model was developed to detect caries, fillings, impacted teeth, implants and normal teeth from dental radiography images. ConvViT combined the advantages of CNN and ViT architectures and provided more successful classification of dental radiography images. As seen in Figure A, ConvViT consists of input layer, CNN component, feature maps, ViT and final layers. Layers were created using ViTLayer and feature extraction was performed using ViT model with these layers. Input layer represents input layer with size 224x224x3. GlobalAveragePooling1D was used to perform max pooling operation of features obtained from ViT model. 3 different dense layers containing 256, 128 and 32 neurons and ReLU activation function were added. The input layer takes the pixel values of the images and normalizes or resizes them for preprocessing. In the ConvViT model, CNN is used for feature extraction. CNN obtains feature maps by applying convolution and pooling to the images. Feature maps contain the high-level representation of the images. ViT takes the feature maps as input, processes them and models the visual relationships through attention mechanisms. The output of ConvViT is passed to fully connected layers and undergoes softmax activation to predict the class label. Hyper-parameters of ConvViT are: input size (224, 224), number of classes is 5, batch size is 32, optimization algorithm is Adam, loss function is sparse_categorical_crossentropy, first Dropout layer is 0.3, second Dropout layer is 0.2, patience value for early termination is 20, minimum improvement rate is min_delta is 0.0001 and epoch number is 100.

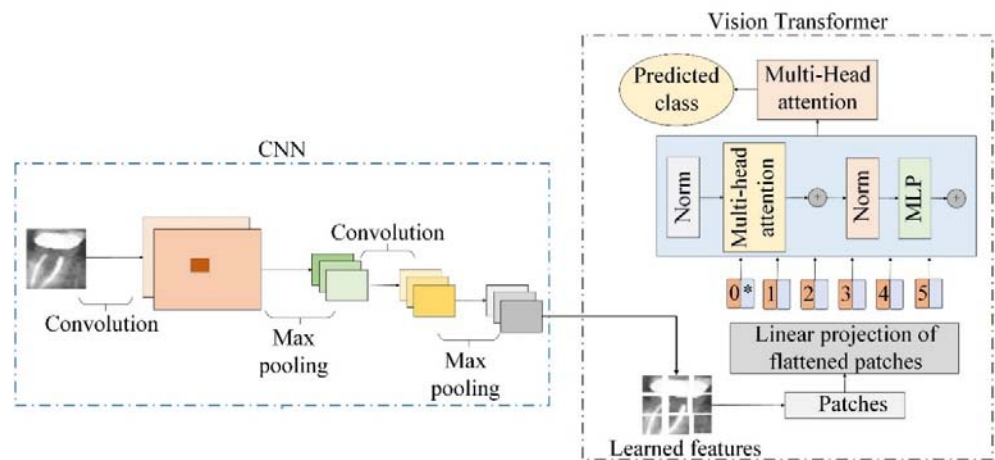


Figure A. The architecture of the ConvViT

Purpose:

A hybrid ConvViT model was developed to detect caries, fillings, impacted teeth, implants and normal teeth from dental radiography images.

Theory and Methods:

ConvViT was compared with popular pre-trained deep learning models such as ResNet-50, VGG-16, EfficientNetB0, and DenseNet201.

Results:

Experimental results showed that ConvViT outperformed the compared models with 95% accuracy, 95% precision, 94% recall and 94% F-score.

Conclusion:

ConvViT makes a significant contribution to the development of artificial intelligence-based automatic diagnosis systems in the field of dentistry by effectively modeling both local and long-range features thanks to its hybrid structure.



Dental radyografi görüntülerinin sınıflandırılmasına yönelik hibrit ConvViT modeli

Anıl Utku*

Munzur Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 62100, Tunceli, Türkiye

Ö N E Ç İ K A N L A R

- CNN ve ViT'yi birleştiren yeni bir ConvViT modeli geliştirildi
- ConvViT ile hem yerel hem de uzun menzilli özellikleri etkili bir şekilde modellemek amaçlandı
- ConvViT, DenseNet201, VGG-16, ResNet-50 ve EfficientNetB0'dan daha başarılı sınıflandırma sonuçlarına sahiptir

Makale Bilgileri

Araştırma Makalesi

Geliş: 16.09.2024

Kabul: 03.04.2025

DOI:

10.17341/gazimmfd.1551005

Anahtar Kelimeler:

Dental radyografi,
CNN,
ViT,
önceden eğitilmiş modeller,
derin öğrenme

ÖZ

Dental radyografi, diş hekimliği alanında önemli bir tanı aracı olarak kullanılmaktadır. Dental radyografi, dişlerin, çenenin ve bu dili çevreleyen yapıların ayrıntılı bir şekilde görüntülenmesini sağlar. Diş radyografileri, teşhis ve tedavi amacıyla diş hekimleri tarafından manuel olarak yorumlanmaktadır. Yapay zekâ yöntemleri, tanı ve sınıflandırma gibi görüntülerle ilgili çeşitli süreçleri otomatikleştirmek amacıyla başarılı bir şekilde uygulanmaktadır. Bu çalışmada, dental radyografi görüntülerinden çürük, dolgu, gömülü diş, implant ve normal diş durumlarını tespit etmek amacıyla hibrit ConvViT modeli geliştirilmiştir. ConvViT, CNN ve ViT mimarilerinin avantajlarını bir araya getirerek dental radyografi görüntülerinin daha başarılı bir şekilde sınıflandırılmasını sağlamıştır. ConvViT, ResNet-50, VGG-16, EfficientNetB0 ve DenseNet201 gibi popüler önceden eğitilmiş derin öğrenme modelleriyle karşılaştırılmıştır. Deneysel sonuçlar, ConvViT'in %95 doğruluk, %95 kesinlik, %94 duyarlılık ve %94 F-skor ile karşılaştırılan modellere göre daha başarılı olduğunu göstermiştir. ConvViT, hibrit yapısı sayesinde hem yerel hem de uzun menzilli özellikleri etkili bir şekilde modelleyerek diş hekimliği alanında yapay zekâ tabanlı otomatik teşhis sistemlerinin geliştirilmesine önemli bir katkı sunmaktadır.

Hybrid ConvViT model for classification of dental radiography images

H I G H L I G H T S

- A new ConvViT model was developed that combines CNN and ViT
- With ConvViT, it was aimed to effectively model both local and long-range features
- ConvViT had better classification results than DenseNet201, VGG-16, ResNet-50 and EfficientNetB0

Article Info

Research Article

Received: 16.09.2024

Accepted: 03.04.2025

DOI:

10.17341/gazimmfd.1551005

Keywords:

Dental radiography,
CNN,
ViT,
pre-trained models,
deep learning

ABSTRACT

Dental radiography is used as an important diagnostic tool in the field of dentistry. Dental radiography provides detailed visualization of teeth, jaws and surrounding structures. Dental radiographs are manually interpreted by dentists for diagnosis and treatment purposes. Artificial intelligence methods are successfully applied to automate various image-related processes such as diagnosis and classification. In this study, a hybrid ConvViT model was developed to detect caries, fillings, impacted teeth, implants and normal teeth from dental radiography images. ConvViT combined the advantages of CNN and ViT architectures and provided more successful classification of dental radiography images. ConvViT was compared with popular pre-trained deep learning models such as ResNet-50, VGG-16, EfficientNetB0, and DenseNet201. Experimental results showed that ConvViT outperformed the compared models with 95% accuracy, 95% precision, 94% recall and 94% F-score. ConvViT makes a significant contribution to the development of artificial intelligence-based automatic diagnosis systems in the field of dentistry by effectively modeling both local and long-range features thanks to its hybrid structure.

*Sorumlu Yazar/Yazarlar / *Corresponding Author/Authors: *anilutku@munzur.edu.tr / Tel: +90 546 654 3601

1. Giriş (Introduction)

Diş radyografileri, diş, çene ve bağlı dokuların görüntülenmesi amacıyla kullanılan X-ray görüntüleridir [1]. Bu görüntüler, diş hekimlerinin dişlerin iç yapısını, diş boşluklarını, çene kemiğini, diş köklerini ve çevre dokuları detaylı bir şekilde incelemesine imkân sağlar. Diş radyografileri, diş çürükleri ve enfeksiyonları, kemik hasarı, kistler, tümörler ve çeneye gömülü dişler gibi ağız sağlığı sorunlarının teşhisinde önemli rol oynar [2]. Kanal tedavisi, diş çürüğü tespiti, ortodontik hastalıkların teşhisi ve tedavi planlaması gibi süreçlerde diş radyografisi analizi önemlidir. Diş röntgenleri, röntgen filminin ağız içinde çekildiği intraoral ve röntgen filminin ağız dışında çekildiği ekstraoral olmak üzere iki temel yöntemle çekilmektedir [3].

Ağız içi röntgenler bitewing, periapikal ve oklüzal röntgenler olmak üzere farklı çeşitlerde uygulanmaktadır. Bitewing röntgenleri, üst ve alt dişlerin bir kısmını aynı anda gösteren röntgen filmleridir. Genellikle arka dişlerdeki çürükleri ve diş eti hastalıklarını değerlendirmek için kullanılır [4]. Periapikal röntgenler, bir dişin tamamını ve çevresindeki kemik yapıyı gösteren röntgen filmleridir. Genellikle bir dişin kökleri, diş eti ve çevresindeki kemik yapıyı ayrıntılı olarak incelemek için kullanılır [5]. Oklüzal röntgenler, çenede daha geniş bir alanı gösteren röntgen filmleridir. Genellikle çenenin geniş bir bölümünü ve dişlerin düzenini göstermek için kullanılır [6].

Ağız dışı röntgenler, çene ve kafatasındaki diş problemlerini tespit etmek için farklı türlerde uygulanmaktadır. Panoramik röntgenler, ağız bölgesinin tamamını görüntüleyerek mevcut dişlerin pozisyonunu tespit eder, gömülü dişleri görüntüler ve tümörlerin teşhisine yardımcı olur [5]. Tomogramlar, nesnenin birden fazla kesit görüntüsünü sağlar. Bu sayede derinlik ve iç yapılar hakkında ayrıntılı bilgi edinmeyi sağlar. Tomogramlar, dişlerin, çene kemiklerinin ve çevresindeki yapıların ayrıntılı analizini sağlar. Bu sayede, diş hekimlerinin çürükler, enfeksiyonlar, diş kökü problemleri, gömülü dişler veya çene kemiği anormallikleri gibi sorunları daha iyi teşhis etmelerine yardımcı olur. Sefalometrik röntgen, dişleri çene ve kişinin profiline göre analiz ederek hastaya özel diş hizalama stratejilerinin geliştirilmesi için kullanılır [7]. Bilgisayarlı Tomografi (BT), üç boyutlu olarak incelemeye imkân sağlar ve kistler, tümörler ve yüz kemiklerindeki kırıklar gibi sorunları tespit etmek için kullanılır. Konik ışın BT, diş yapıları, yumuşak doku, sinirler ve kemiğin 3 boyutlu görüntülenmesini sağlar. BT, dişlerin, çene kemiklerinin ve çevresindeki yumuşak dokuların ayrıntılı kesit görüntülerini sağlar. Bu sayede, çürükler, kök enfeksiyonları, çene kemiği anormallikleri ve diğer diş problemlerinin teşhis edilmesine yardımcı olur. BT, diş implantı yerleştirmeden önce, çene kemiğinin yoğunluğunu, yüksekliğini ve genişliğini değerlendirerek implant yerleştirme planlamasını optimize eder. Geleneksel röntgen görüntüleri genellikle 2D formatta alınır, dişlerin ve çene yapısının düz bir kesitini sunar. Çürükler, diş kökü problemleri, diş eti hastalıkları ve bazı yapısal anormallikler bu yöntemle tespit edilebilir. MR, bir dizi 2D kesit görüntüsünü birleştirerek 3D model oluşturabilir. Bu sayede, çene eklemi, yumuşak dokular ve diğer karmaşık yapıların detaylı bir üç boyutlu görünümünü sağlar. MR 3D görüntüleme, özellikle çene eklemi bozuklukları ve yumuşak doku patolojilerinin ayrıntılı değerlendirilmesinde kullanılır [8].

Dental X-ray görüntülerinin analizi, hastalıkların teşhisi, diş yapılarının incelenmesi ve tedavi planlaması açısından diş hekimliği ve tıp alanında oldukça önemlidir. Geleneksel yöntemlerde, uzmanlar radyografi görüntülerini manuel olarak incelerken, günümüzde yapay zekâ ve derin öğrenme teknikleri bu süreci otomatikleştirmektedir. Derin öğrenme modelleri, diş radyografilerindeki karmaşık örüntüleri tanımlamak için kullanılmaktadır. Derin öğrenme modelleri, eğitim

verisi olarak kullanılan geniş çaplı radyografi görüntüleri üzerinde eğitilerek farklı diş hastalıklarını, diş çürüklerini, kemik yoğunluğundaki anormallikleri ve diğer dental sorunları yüksek doğrulukla tespit edebilmektedir. Vision Transformer (ViT) gibi modeller, daha gelişmiş görüntü sınıflandırma performansı sunarak daha karmaşık ve büyük ölçekli verilerde başarılı sonuçlar üretebilmektedir. Diş radyografilerinin analizi, klinik karar destek sistemlerinin geliştirilmesine ve tedavi süreçlerinin iyileştirilmesine katkı sağlamaktadır.

Bu çalışmada, çürük, dolgu, gömülü diş, implant ve normal olarak etiketlenmiş dental radyografi görüntülerinin sınıflandırması amacıyla hibrit Convolutional-Vision Transformer (ConvViT) modeli geliştirilmiştir. ConvViT, DenseNet201, Visual Geometry Group-16 (VGG-16), Residual Networks-50 (ResNet-50) ve Efficient Neural Network - Baseline 0 (EfficientNetB0) gibi popüler önceden eğitilmiş mimarilerle kapsamlı olarak karşılaştırılmış ve elde edilen sonuçlar detaylı bir şekilde analiz edilmiştir. Modellerin sınıflandırma performansı F-skor (F-score), kesinlik (precision), doğruluk (accuracy) ve duyarlılık (recall) metrikleri kullanılarak değerlendirilmiştir.

Bu çalışmanın literatüre sunduğu yenilikler aşağıdaki gibi özetlenebilir:

- CNN ve ViT'in öne çıkan özellikleri bir araya getirilerek hibrit ConvViT modeli oluşturulmuştur. ConvViT ile hem yerel özellik çıkarma hem de uzun menzilli ilişkilerin modellenmesini sağlamak amaçlanmıştır
- Geliştirilen ConvViT modeliyle CNN'in özellik çıkarma yeteneğini ve ViT'in dikkat mekanizmalarını kullanarak daha başarılı bir sınıflandırma performansı elde etmek amaçlanmıştır.
- ConvViT, VGG-16, DenseNet201, EfficientNetB0 ve ResNet-50 gibi popüler mimarilerle uygulamalı olarak karşılaştırılmıştır.
- ConvViT'in VGG-16, ResNet-50, DenseNet201 ve EfficientNetB0'a göre daha başarılı bir sınıflandırma performans sergilemesi, literatürdeki yüksek doğruluk gerektiren tıbbi görüntü sınıflandırma problemleri için güvenilir bir yöntem olduğunu göstermektedir.
- Bu veri seti kullanılarak literatürde yapılmış ilk çalışmadır.
- Bu çalışma ile Türkçe literatüre katkıda bulunmak amaçlanmıştır.

2. Literatürdeki Çalışmalar (Studies in the Literature)

Son yıllarda, derin öğrenme modelleri diş radyografi görüntüleri üzerinde diş hastalıklarının ve anormalliklerinin tespitinde yaygın bir şekilde kullanılmaktadır. Literatürdeki çeşitli çalışmalar, YOLO, Faster R-CNN, DenseNet, VGG gibi popüler mimarilerin dental veriler üzerindeki performanslarını incelemiş ve bu alanda önemli başarılar elde edilmiştir. Çalışmalar, dental restorasyonların tespitinden, periodontitis ve diş çürüklerinin sınıflandırılmasına kadar geniş bir yelpazede uygulanmakta ve genellikle farklı önceden eğitilmiş modellerin performans karşılaştırmalarına dayanmaktadır. Her modelin başarısı, kullanılan veri setinin büyüklüğü, veri artırma teknikleri ve tespit edilen hastalık türü gibi faktörlere göre değişiklik göstermektedir. Bu bölümün devamında literatürdeki çalışmalar kapsamlı ve karşılaştırmalı olarak incelenmiştir.

Çelik ve Çelik [9], panoramik diş radyografi görüntülerinden diş restorasyonlarının tespitine yönelik derin öğrenme modellerinin karşılaştırmalı bir analizini sunmuştur. Çalışmada You Only Look Once V3 (YOLO-V3), Faster Regions with the Convolutional Neural Networks (Faster R-CNN), R-CNN, SSD ve RetinaNet modelleri tespit modeli olarak uygulanmıştır. Özellik çıkarma amacıyla DarkNet53, ResNet-50, Xception-101, ResNet-101 ve VGG-16

kullanılmıştır. Deneyler, Faster R-CNN'in 0,952 Area Under the Curve (AUC) değeriyle karşılaştırılan modellerden daha başarılı olduğunu göstermiştir.

Almalki vd. [10], dental radyografi görüntülerinden çürükler, kök kanalları, diş kronları ve kırılmış kök kanallarının tespitine yönelik YOLO-V3 modelinin bir uygulamasını sunmuştur. Çalışmada veri artırma yöntemleri uygulanarak 1200 görüntüden oluşan bir veri seti elde edilmiştir. Veri seti %70-%30 olmak üzere eğitim ve test kümelerine ayrılmıştır. Deneyler, YOLO-V3 modelinin %99,33 sınıflandırma doğruluğuna sahip olduğunu göstermiştir.

Shon vd. [11], her bir dişin periodontitis evrelerini dental panoramik radyografiler kullanarak sınıflandırmak amacıyla U-Net ve YOLO-V5 modellerini kullanmıştır. Çalışmada, dental radyografi görüntülerinden dişleri tespit etmek ve numaralandırma sistemine göre dişleri sınıflandırmak amacıyla YOLO-V5 kullanılmıştır. Tespit edilen diş numarasıyla, o diş için periodontal hastalığın evresine karar vermek amacıyla U-Net kullanılmıştır. Çalışmada veri artırma teknikleri uygulanarak 1044 görüntüden oluşan bir veri seti elde edilmiştir. Deneyler, geliştirilen modelin 0,929 doğruluk, 0,807 duyarlılık ve 0,724 kesinlik değerine sahip olduğunu göstermiştir.

Aljabri vd. [12], panoramik diş radyografisi görüntülerinden üst çene köpek dişi sıkışması hastalığının tespitine yönelik DenseNet-121, Inception V3, VGG-16 ve ResNet-50 modellerinin karşılaştırmalı bir analizini sunmuştur. Deneysel sonuçlar, Inception V3'ün %92,59 doğruluk oranıyla karşılaştırılan modellerden daha başarılı bir sınıflandırma performansına sahip olduğunu göstermiştir.

Okazaki vd. [13], panoramik dental radyografi görüntülerindeki çoklu diş anormalliklerinin tespitine yönelik AlexNet tabanlı bir model geliştirmiştir. Çalışmada tek süpernümerer diş ve/veya odontoma bulunan panoramik radyografi görüntüleri vaka grubu, diş anomalisi bulunmayan radyografi görüntüleri ise kontrol grubu olarak kullanılmıştır. Deneyler, geliştirilen AlexNet modelinin %70 doğruluk, %70,8 kesinlik, %70 hassasiyet ve %69,7 F-skor değerine sahip olduğunu göstermiştir.

Lin vd. [14], ağız içi ve ağız dışı diş radyografi görüntülerinin sınıflandırılmasına yönelik VGG-16, ResNet-101 ve ResNet-18 modellerinin karşılaştırmalı bir analizini sunmuştur. Çalışmada, 22.334 ağız içi görüntü ve 1.035 ağız dışı görüntü kullanılmıştır. Diş

radyografi görüntüleri 32 diş alanına kategorize edilmiştir. Deneyler, ResNet-101 modelinin, 0,976 doğruluk, 0,969 kesinlik, 0,984 duyarlılık ve 0,977 F-skor değeriyle VGG-16 ve ResNet-18 modellerinden daha başarılı bir sınıflandırma performansına sahip olduğunu göstermiştir.

Öztekin vd. [15], diş çürüklerinin tespitine yönelik EfficientNet-B0, ResNet-50 ve DenseNet-121 önceden eğitilmiş modellerin uygulamalı bir analizini sunmuştur. Uygulanan modeller, 562 denekten alınan panoramik görüntüler kullanılarak test edilmiştir. Deneyler, ResNet-50 modelinin %92,00 doğruluk, %87,33 duyarlılık ve %91,61 F-skora sahip olduğunu göstermiştir.

Park vd. [16], periapikal radyografi görüntülerinden dental implant çapını ve uzunluğunu otomatik olarak tespit etmeye dayalı VGG-16 tabanlı bir model geliştirmiştir. Çalışmada, k-means++ algoritması kullanılarak dental implantın üç temel nokta koordinatından türetilen implant-spesifik özellik vektörü kümelendirilmiştir. VGG-16 k-means++ kümeleme kullanılarak dental implant boyutunu dokuz gruba ayrılmıştır. Deneyler, VGG-16'nın % 0,994 doğruluk, % 0,950 kesinlik, % 0,994 duyarlılık, % 0,974 F-skor ve % 0,975 AUC değerine sahip olduğunu göstermiştir.

Chen vd. [17], dental X-ray görüntülerinden periodontitis ve diş çürüklerinin tespitine yönelik EfficientNet-B0 tabanlı bir derin öğrenme modeli önermiştir. Çalışmada diş röntgeni görüntüleri YOLO-V7 kullanılarak tespit edilmekte ve periapikal röntgen görüntüsünden kırılmaktadır. Lokal kontrastı arttırmak amacıyla kontrast sınırlı adaptif histogram eşitleme yöntemi kullanılmıştır. Deneyler, geliştirilen EfficientNet-B0 tabanlı modelin ortalama kesinlik değerinin %97,1 olduğunu göstermiştir. Geliştirilen model, periodontitis tespitinde %98,67 AUC ve diş çürüklerinin tespitinde %97,55 AUC değerine sahip olduğunu göstermiştir.

Yılmaz vd. [18], diş hastalıklarının teşhis hatasını azaltmak ve sınıflandırma doğruluğunu artırmak amacıyla YOLO-V4 ve Faster R-CNN modellerinin uygulamalı bir analizini sunmuştur. Çalışmada 1200 panoramik radyografi görüntüsünden oluşan bir veri seti kullanılmıştır. Deneysel sonuçlar YOLO-V4'in %99,90 kesinlik, %99,18 duyarlılık ve %99,54 F-skor değerine sahip olduğunu, R-CNN'in ise %93,67 kesinlik, %90,79 duyarlılık ve %92,21 F-skor değerine sahip olduğunu göstermiştir.

Tablo 1. Literatürdeki çalışmalar (Studies in the literature)

Referans	Veri seti	Uygulanan modeller	Başarı oranı
9	123 görüntü ve 3 sınıf	YOLO-V3, Faster R-CNN, SSD, RetinaNet	0,952 AUC
10	1200 görüntü ve 4 sınıf	YOLO-V3	%99,33 doğruluk
11	87 görüntü ve 4 sınıf	YOLO-V5, U-Net	%92,9 doğruluk, %72,4 kesinlik, %80,7 duyarlılık, %72,4 F-skor
12	150 görüntü ve 3 sınıf	DenseNet-121, Inception V3, VGG-16, ResNet-50	%92,59 doğruluk
13	23379 görüntü ve 32 sınıf	AlexNet	%70 doğruluk, %70,8 kesinlik, %70 duyarlılık, %69,7 F-skor
14	562 hasta verisi ve 2 sınıf	VGG-16, ResNet-101, ResNet-18	%97,6 doğruluk, %96,9 kesinlik, %98,4 duyarlılık, %97,7 F-skor
15	2200 görüntü ve 2 sınıf	EfficientNet-B0, ResNet-50, DenseNet-121	%92,00 doğruluk, %87,33 duyarlılık, %91,61 F-skor
16	1320 görüntü ve 9 sınıf	VGG-16	%99,4 doğruluk, %95 kesinlik, %99,4 duyarlılık, %97,4 F-skor, 0,975 AUC
17	1525 görüntü ve 2 sınıf	EfficientNet-B0, YOLO-V7	%97,1 kesinlik
18	1200 görüntü ve 36 sınıf	YOLO-V4, Faster R-CNN	%99,90 doğruluk, %99,90 kesinlik, %99,18 duyarlılık, %99,54 F-skor
Mevcut çalışma	29845 görüntü ve 5 sınıf	ConvViT, VGG-16, DenseNet201, EfficientNetB0, ResNet-50	0,95 doğruluk, 0,95 kesinlik, 0,94 duyarlılık, 0,94 F-skor

Tablo 1’de görüldüğü gibi literatürdeki çalışmalarda, YOLO tabanlı modellerin hızlı ve etkili bir tespit aracı olarak öne çıktığı görülmektedir. Özellikle önceden eğitilmiş modeller, yüksek doğruluk oranları ve F-skor değerleriyle öne çıkmaktadır. Mevcut çalışmada kullanılan ConvViT modeli ise VGG-16, DenseNet201, EfficientNetB0 ve ResNet-50 gibi modellerle karşılaştırıldığında %95 doğruluk ve F-skor ile başarılı bir performans göstermiştir. Bu sonuçlar, derin öğrenme tabanlı modellerin diş hastalıklarının tespitinde etkin bir çözüm sunduğunu ve model seçiminde veri yapısı ile probleme özgü faktörlerin dikkate alınması gerektiğini ortaya koymaktadır.

3. Materyal ve Metot (Material and Method)

Bu bölümde, çalışmada kullanılan veri seti, bölütleme süreçleri ve sınıflandırma modelleri ayrıntılı olarak ele alınmaktadır. Çalışmanın temelini oluşturan dental radyografi görüntüleri, çürük, dolgu, gömülü diş, implant ve normal olarak etiketlenmiş verilerden oluşmaktadır. Bu veri seti, dental görüntülerin otomatik teşhisi ve sınıflandırılması amacıyla kullanılmıştır. Çalışmada uygulanan derin öğrenme mimarileri, bu bölütlenmiş görüntüler üzerinde eğitilerek diş radyografisi sınıflandırmasındaki performansları karşılaştırılmıştır.

3.1. Veri Seti (Data Set)

Bu çalışmada çürük, dolgu, gömülü diş, implant ve normal olarak etiketlenmiş dental radyografi görüntülerinden oluşan bir veri seti kullanılmıştır [19]. Çalışmada kullanılan veri seti, Dental Radiography veri seti temel alınarak oluşturulmuştur. Veri setinin organizasyonu eğitim, doğrulama ve test şeklindedir. Eğitim seti (train), modelin öğrenmesi için kullanılan görüntüleri içermektedir. Doğrulama seti (validation), modelin hiperparametre optimizasyonu sırasında değerlendirme yapılan veri kümesini içermektedir. Test seti (test), nihai model değerlendirmesi için ayrılan bağımsız veri kümesini içermektedir. Veri seti hazır olarak eğitim, doğrulama ve test setlerine ayrılmıştır ve herhangi bir ek veri artırma işlemi uygulanmamıştır. Görüntüler, diş röntgenlerinden bölütleme yöntemiyle ayrıştırılmıştır ve her bir görüntü belirlenen sınıfa uygun şekilde etiketlenmiştir. Çalışmada, farklı diş tiplerinin sınıflandırılmasını kolaylaştırmak amacıyla etiketli veriler kullanılmış ve herhangi bir ek manuel anotasyon yapılmamıştır. Görüntüler, orijinal panoramik röntgenlerden alınmıştır. Çözünürlük, model giriş boyutuna uygun olacak şekilde 224x224 piksel olarak yeniden ölçeklendirilmiştir. Görüntüler gri seviyede olup, herhangi bir renk dönüştürme işlemi uygulanmamıştır. Bölütleme sonrası her bir diş bireysel olarak etiketlenmiş ve ilgili alt klasörlere yerleştirilmiştir. Veri seti temizlenmiş ve organize edilmiş olduğu için görüntülerde histogram eşitleme ve kontrast iyileştirme gibi ön işleme adımları uygulanmamıştır.

Veri seti, panoramik röntgen görüntülerinden oluşmaktadır ve özellikle dişlerdeki çürük, dolgu, gömülü diş ve implant gibi durumların tespiti ve sınıflandırılması amacıyla kullanılmaktadır. Veri seti, Segment Anything Model (SAM) kullanılarak bölütlenmiştir. Bölütleme sürecinde, her diş ayrı ayrı tanımlanmış ve kategorize edilmiştir. Bu sayede, hem çürük dişlerin hem de sağlıklı dişlerin bölütlenmiş görüntüleri elde edilmiştir. Bu süreç, dişlerin bireysel olarak incelenmesi ve sınıflandırılması için önemlidir. Kullanılan veri seti, özellikle diş hekimliği alanında yapay zekâ tabanlı uygulamalar geliştirmek için önemlidir.

Kullanılan veri seti eğitim, test ve doğrulama klasörlerinden oluşmaktadır. Her alt klasör, ilgili kategoriye ait bölütlenmiş diş görüntülerini içerir. Her alt klasör, çürük, dolgu, gömülü diş, implant ve normal olmak üzere ilgili sınıf etiketine karşılık gelen bölütlenmiş

diş görüntülerini içermektedir. Sınıf dağılımında belirgin bir dengesizlik bulunmaktadır. Özellikle çürük (22 görüntü), gömülü diş (35 görüntü) gibi bazı sınıflar diğerlerine kıyasla çok daha az örneğe sahiptir. Bu durum, modelin belirli sınıfları daha iyi öğrenmesine ve azınlık sınıfları göz ardı etmesine neden olabilecek bir dengesizlik oluşturmaktadır. Ancak, veri seti hazır olarak eğitim, doğrulama ve test setlerine ayrılmış olduğundan ek veri artırma teknikleri uygulanmamıştır. Bunun yerine, modelin dengesiz veri setiyle eğitilmesini daha adil hale getirmek için farklı değerlendirme metrikleri kullanılmıştır. Eğitim klasörü 25410, test klasörü 1714 ve doğrulama klasörü ise 2721 görüntüden oluşmaktadır. Veri setindeki her bir sınıfa ait görüntü örnekleri Şekil 1’de görülmektedir.



Şekil 1. Veri setindeki her bir sınıfa ait görüntü örnekleri
(Examples of images for each class in the dataset)

3.2. Sınıflandırma Modelleri (Classification Models)

VGG-16, Oxford Üniversitesi Visual Geometry Group tarafından geliştirilen bir CNN mimarisidir [20]. VGG-16, 13 evrişimli katman ve 3 tam bağlı katmandan oluşan bir mimariye sahiptir. VGG-16, her konvolüsyon katmanında 3x3 boyutunda küçük filtreler kullanarak modelin daha ayrıntılı özellikleri öğrenmesini sağlar [21]. Modelin giriş boyutu sabittir ve genellikle 224x224 piksel boyutunda RGB görüntüler kullanılır. VGG-16’da, konvolüsyon katmanlarından sonra Rectified Linear Unit (ReLU) fonksiyonunu kullanılmaktadır. ReLU, modele doğrusal olmayan özellikler kazandırarak öğrenme kapasitesini artırır [22]. Max pooling katmanları kullanılarak boyut azaltması yapılmaktadır. VGG-16’nın son katmanları tam bağlı katmanlardır ve bu katmanların sonunda softmax fonksiyonu kullanılarak sınıflandırma yapılır [23].

Densely Connected Convolutional Networks (DenseNet), her bir katmanın, kendisinden önceki tüm katmanlardan gelen çıktıları aldığı bir mimaridir. Yoğun bağlantılar, her katmanın kendisinden önceki katmanların tamamıyla bağlantılı olmasını sağlayarak modelin yeniden öğrenme ihtiyacını azaltır ve bilgi akışını güçlendirerek ağına daha iyi genelleştirme yapmasını sağlar [24]. DenseNet aynı çıktı özellik haritasını alır ve önceki girdi özellik haritalarıyla birleştirir [25]. Bu sayede, katmanlar arasındaki bilgi akışını en üst düzeye çıkarır ve genel parametre sayısını azaltır. Yoğun bağlantı yapısı, parametrelerin daha verimli kullanılmasını sağlayarak modelin daha az parametreye ihtiyaç duymasını sağlar. DenseNet’in mimarisi geçiş katmanları ve dense bloklardan oluşur [24]. Dense blok içindeki her bir konvolüsyon katmanı, blok içindeki diğer katmanlara bağlıdır [26]. Geçiş katmanları, ağına etkili bir şekilde büyümesini sağlayan yoğun bloklar arasındaki özellik haritalarının boyutunu en aza indirir.

EfficientNetB0, compound scaling kullanarak modelin derinliği, genişliği ve çözünürlüğünü dengeler, böylece modelin verimliliğini artırırken performansını optimize eder [27]. Ayrıca, global average pooling ile özellik haritalarının daha iyi bir genel görünümünü elde eder [28]. EfficientNet modelleri, verimlilik ve performans arasında optimal dengeyi sağlayarak daha az hesaplama gücüyle daha yüksek doğruluk elde etmeyi amaçlar. EfficientNetB0, özellikle görüntü tanıma ve transfer öğrenme görevlerinde yüksek doğruluk sağlar [29]. Küçük model boyutu ve hızlı işlem süresi sayesinde, sınırlı kaynaklara sahip cihazlarda da etkili bir şekilde çalışabilir. Önceden eğitilmiş EfficientNetB0 modelleri, belirli görevler için yeniden eğitilerek transfer öğrenme uygulamalarında kullanılabilir [30].

ResNet, derin sinir ağlarının eğitimini kolaylaştıran residual learning mekanizmasına sahiptir [31]. Geleneksel derin sinir ağlarında, her

katman çıktısı elde etmek için girdiye bir dizi dönüşüm uygular. ResNet, ağı bir katmanın girdisi ve çıktısı arasındaki farkları alan kalan eşlemeleri öğrenmesini sağlayan residual bağlantıları kullanmaktadır [32]. Residual bağlantılar, girişin bir katmanın çıkışına eklenmesiyle oluşturulur ve gradyanların zayıflatılmadan doğrudan ağı üzerinden iletilmesine olanak tanır [33]. Residual bloklar, girişin doğrudan sonraki katmana taşındığı bir skip connection veya shortcut connection içerir. Yani, bir katmanın çıktısı, birkaç katman atlanarak daha sonraki bir katmana doğrudan eklenir [34]. Bu bağlantılar, kaybolan gradyan soruna çözüm olarak geliştirilmiştir. Derin ağlarda, geri yayılım sırasında gradyanlar katmanlar arasında kaybolabilir, bu da modelin öğrenmesini zorlaştırır. ResNet'teki residual bağlantılar, bu sorunu çözerek çok daha derin ağların eğitilmesine olanak tanır [35]. ResNet, genellikle 3x3 boyutunda küçük filtreler kullanır ve bu filtreler bir residual block içindeki konvolüsyon katmanlarında yer alır [36].

3.3. Geliştirilen Model (Developed Model)

ViT, dikkat mekanizmasını kullanan bir dönüştürücü modeldir [37]. ViT, giriş görüntüsünü bir dizi dikkat mekanizması aracılığıyla parçalara ayırır ve dikkat mekanizmaları, giriş parçaları arasındaki ilişkileri öğrenmek için kullanılır [38]. ViT, paralel işleme yeteneği sayesinde, her dikkat mekanizmasının paralel işlenmesini sağlayarak özellik çıkarma ve eğitim süreçlerini hızlandırır [39]. ViT, bir dönüştürücü kodlayıcı, yama yerleştirmeleri ve doğrusal sınıflandırıcı bileşenlerden oluşur. Yama yerleştirmeleri, görüntüyü dikkörtgen parçalara ayırmayı ifade eder. Bölümlere ayrılmış parçaların her biri, önceden eğitilmiş düşük boyutlu bir alanda temsil edilir [40].

Bu çalışmada, geliştirilen hibrit ConvViT modeli ile ViT'nin dikkat mekanizmalarının etkinliğinden ve CNN'in özellik çıkarma aşamasındaki etkinliğinden faydalanmak amaçlanmıştır. ConvViT mimarisinin temeli, CNN katmanları ile görüntü özelliklerini çıkarma ve ViT katmanlarının dikkat mekanizmalarını kullanarak bu özellikler arasındaki uzun menzilli ilişkileri modellemeye dayanmaktadır.

Geliştirilen modelde, CNN önce görüntüleri işler ve özellik haritaları oluşturur. Çıkarılan özellik haritaları, dikkat mekanizmaları kullanılarak işlemek üzere ViT'a girdi olarak sunulur. Dikkat mekanizmaları, görüntülerin farklı kısımlarına odaklanarak daha etkili bir görsel temsil sağlar. ViT tarafından işlenen özellik haritaları, CNN'den elde edilen özelliklerle birleştirilir. Bu şekilde, özellik çıkarma aşamasında CNN'nin etkinliğinden ve ViT'nin dikkat mekanizmalarının avantajlarından yararlanılarak daha başarılı bir görüntü sınıflandırma modeli geliştirmek amaçlanmıştır.

ViT'da vektör gösterimleri, giriş görüntülerini $X = \{x_1, x_2, \dots, x_N\}$ olarak yama boyutlarına bölerek elde edilir [41]. Bu vektörler, Eş. 1'de görüldüğü gibi bir E gömme matrisi ile çarpılarak dönüştürülür.

$$X_{Gömm} = X \cdot E \quad (1)$$

Dikkat mekanizması, giriş vektörlerinin ilişkilerini modellemek için kullanılır [42]. W_Q, W_K, W_V öğrenilebilir ağırlık matrisleri kullanılarak, sorgu, anahtar ve değer vektörleri Eş. 2'de görüldüğü gibi oluşturulur.

$$\begin{aligned} \text{Sorgu } (Q) &= X_{Embed} \cdot W_Q \\ \text{Anahtar } (K) &= X_{Embed} \cdot W_K \\ \text{Değer vektörü } (V) &= X_{Embed} \cdot W_V \end{aligned} \quad (2)$$

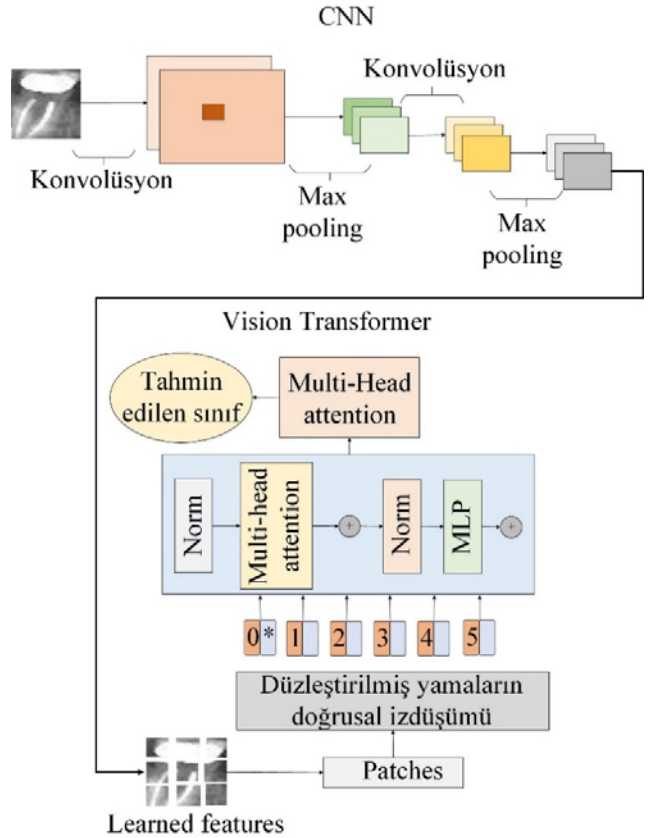
Eş. 3 kullanılarak dikkat değerleri hesaplanır. d_k , sorgunun boyutunu ve anahtar vektörlerini ifade eder.

$$\text{Dikkat } (Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (3)$$

Dikkat mekanizmasının çıktıları, Eş. 4 kullanılarak ileri beslemeli ağı sunulur. Öğrenilebilir ağırlık ve önyargı parametreleri W_1, b_1, W_2, b_2 'dir [43].

$$\text{FFN}(X) = \text{ReLU}(XW_1 + b_1)W_2 + b_2 \quad (4)$$

Her blokta çıktılar, katman normalizasyonu kullanılarak önceki katmanın çıktılarıyla toplanır. Şekil 2'de hibrit ConvViT mimarisi görülmektedir.



Şekil 2. Hibrit ConvViT mimarisi (Hybrid ConvViT architecture)

Çalışmada modeller uygulanmadan önce görüntülerin boyutları (224x224) ve batch boyutu (32) belirlenmiştir. Eğitim, doğrulama ve test veri setleri için ImageDataGenerator ile veri artırma ve normalizasyon işlemleri gerçekleştirilmiştir. Şekil 2'de görüldüğü gibi, ConvViT giriş katmanı, CNN bileşeni, özellik haritaları, ViT ve son katmanlardan oluşur. ViTFeatureExtractor ve TFViTModel kullanılarak Hugging Face'den ViT modeli ve özellik çıkarıcısı yüklenmiştir. ViTLayer kullanılarak katmanlar oluşturulmuş ve bu katmanlarla ViT modeli kullanılarak özellik çıkarımı yapılmıştır. Input katmanı 224x224x3 boyutunda giriş katmanı ifade etmektedir. Rescaling, girişleri [0,1] aralığına normalize etmek için kullanılmıştır. GlobalAveragePooling1D, ViT modelinden elde edilen özelliklerin max pooling işlemini yapmak için kullanılmıştır. 256, 128 ve 32 nöron içeren ve ReLU aktivasyon fonksiyonuna sahip 3 farklı Dense katmanı eklenmiştir. Dense katmanları arasına eklenen Dropout katmanlarıyla aşırı öğrenmeyi engellemek amaçlanmıştır. Softmax aktivasyon fonksiyonuna sahip çıktı katmanı ise sınıflama sonuçlarını

öretmektedir. Modelin erken durdurulması, öğrenme oranı azaltma ve en iyi modelin kaydedilmesi için callback'ler tanımlanmıştır. Modeli derleme aşamasında Adam optimizasyonu ve sparse_categorical_crossentropy kayıp fonksiyonunu kullanılmıştır.

Giriş katmanı görüntülerin piksel değerlerini alır ve ön işleme için normalleştirir veya yeniden boyutlandırır. ConViT modelinde CNN özellik çıkarma için kullanılmaktadır. CNN, görüntüleri konvolüsyon ve havuzlama uygulayarak özellik haritaları elde eder. Özellik haritaları görüntülerin üst düzey temsiliyi içerir. ViT özellik haritalarını giriş olarak alır, işler ve dikkat mekanizmaları aracılığıyla görsel ilişkileri modeller. ConViT'in çıktısı tam bağlı katmanlara iletilir ve sınıf etiketini tahmin etmek için softmax aktivasyonundan geçirilir. ConViT'in hiper parametreleri: girdi boyutu (224, 224), sınıf sayısı 5, yığın boyutu 32, optimizasyon algoritması Adam, kayıp fonksiyonu sparse_categorical_crossentropy, ilk Dropout katmanı 0.3, ikinci Dropout katmanı 0.2, erken sonlandırma için patience değeri 20, minimum gelişme oranı min_delta 0.0001 ve epoch sayısı 100'dür. VGG-16 için kullanılan hiperparametreler giriş boyutu: (224, 224, 3), optimizasyon algoritması: Adam, öğrenme oranı: 1e-4, batch size: 32, epoch sayısı: 100, kayıp fonksiyonu: Sparse Categorical Crossentropy, aktivasyon fonksiyonları: ara katmanlarda ReLU ve çıktı katmanında Softmax, dropout oranı: 0.3, erken durdurma: patience=20'dir. ResNet-50 için kullanılan hiperparametreler giriş boyutu: (224, 224, 3), optimizasyon algoritması: Adam, öğrenme oranı: 1e-4, batch size: 32, epoch sayısı: 100, kayıp fonksiyonu: Sparse Categorical Crossentropy, aktivasyon fonksiyonları: ara katmanlarda ReLU ve çıktı katmanında Softmax, dropout oranı: 0.2, erken durdurma: patience=20'dir. DenseNet201 için kullanılan hiperparametreler giriş boyutu: (224, 224, 3), optimizasyon algoritması: Adam, öğrenme oranı: 1e-4, batch size: 32, epoch sayısı: 100, kayıp fonksiyonu: Sparse Categorical Crossentropy, aktivasyon fonksiyonları: ara katmanlarda ReLU ve çıktı katmanında Softmax, dropout oranı: 0.2, erken durdurma: patience=20'dir. EfficientNetB0 için kullanılan hiperparametreler giriş boyutu: (224, 224, 3), optimizasyon algoritması: Adam, öğrenme oranı: 1e-4, batch size: 32, epoch sayısı: 100, kayıp fonksiyonu: Sparse Categorical Crossentropy, aktivasyon fonksiyonları: ara katmanlarda ReLU ve çıktı katmanında Softmax, dropout oranı: 0.2, erken durdurma: patience=20'dir. Bu hiperparametreler, veri setine ve modellerin eğitimi sırasında elde edilen en iyi performansa göre belirlenmiştir.

ConViT modeli, geleneksel CNN tabanlı mimarilere kıyasla uzun menzilli ilişkileri daha iyi modelleyebilmesi için ViT (Vision Transformer) ile güçlendirilmiştir. Modelin optimizasyon sürecinde, Adam optimizasyon algoritması kullanılmış ve öğrenme oranı 1e-4 olarak belirlenmiştir. ConViT'in katman yapısında, CNN bileşeni derin özellik çıkarımı için kullanılmış, ardından ViT modülü ile daha geniş ölçekli ilişkiler modellenmiştir. Modelin başarısını artırmak için

Dropout (0.3 ve 0.2) katmanları eklenerek aşırı öğrenme engellenmiş, ayrıca erken durdurma (early stopping) yöntemi kullanılarak eğitim sürecinin gereksiz yere uzaması önlenmiştir. Transformer bileşeninde çoklu başlı dikkat mekanizması (multi-head attention) kullanılarak modelin farklı özellikleri aynı anda öğrenmesi sağlanmıştır. Özellikle panoramik röntgenlerde farklı bölgelerin birbirleriyle ilişkisini anlamak için, dikkat mekanizmalarının (self-attention) önemine odaklanılmıştır. Geleneksel CNN tabanlı modellerle kıyaslandığında, ConViT'in bu hibrit yapısı sayesinde uzun menzilli bağımlılıkları modelleyerek dental radyografilerde daha yüksek bir sınıflandırma doğruluğuna ulaştığı gözlemlenmiştir.

3.4. Performans Değerlendirme Metrikleri (Performance Evaluation Metrics)

Sınıflandırma modellerinin değerlendirilmesinde temel olarak kesinlik, duyarlılık, doğruluk ve F-skor metrikleri kullanılmaktadır. Karmaşıklık matrisinden elde edilen bu değerler, modellerin veri setindeki örnekleri ne kadar doğru sınıflandırabildiğine dair değerlendirmeler sağlar. Bu çalışmada kullanılan veri seti çok sınıflı bir veri setidir. Şekil 3'te görülen karışıklık matrisi, çok sınıflı sınıflandırma problemlerinde örnek olarak seçilen Dolgu sınıfı için TP, FN, TN ve FP değerlerinin nasıl seçileceğini göstermektedir.

Şekil 3'te, örnek olarak Dolgu sınıfı için TP, FN, TN ve FP değerlerinin nasıl seçileceği görsel olarak sunulmuştur. Dolgu olarak tahmin edilen ancak aslında başka bir sınıfa (FP) ait olan örnekler turuncu renkte gösterilmiştir. Dolgu olarak tahmin edilmesi gereken ancak farklı bir sınıf etiketiyle tahmin edilen örnekler (FN) ise sarı renkte gösterilmiştir. Dolgu sınıfı için tahmin edilen ve gerçek değerini keşiştiği hücre (TP) yeşil renkte gösterilmiştir. Doğru şekilde tahmin edilen farklı sınıflara ait örnekler (TN) mavi renkte gösterilmiştir.

Doğruluk, Eş. 5'te görüldüğü bir modelin tüm tahminleri arasındaki doğru tahminlerin oranını ifade eden bir performans metriğidir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Kesinlik, Dolgu olarak sınıflandırılan örneklerin kaçının aslında Dolgu sınıfına ait olduğunu gösterir ve Eş. 6 kullanılarak hesaplanır.

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (6)$$

Duyarlılık, Dolgu olarak sınıflandırılması gereken örneklerin kaçının doğru şekilde sınıflandırıldığını gösterir ve Eş. 7 kullanılarak hesaplanır.

Tahmin edilen

	Çürük	Dolgu	Gömülü diş	İmplant	Normal
Gerçek					
Çürük	TN	FP	TN	TN	TN
Dolgu	FN	TP	FN	FN	FN
Gömülü diş	TN	FP	TN	TN	TN
İmplant	TN	FP	TN	TN	TN
Normal	TN	FP	TN	TN	TN

Şekil 3. Çok sınıflı sınıflandırma problemleri için karışıklık matrisi (Confusion matrix for multi-class classification problems)

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (7)$$

F-skor, kesinlik ve duyarlılık değerlerinin etkisini korumak için bu metriklerin harmonik ortalaması alınarak Eş. 8'de görüldüğü gibi hesaplanır.

$$F\text{-skoru} = \frac{2 \cdot \text{Kesinlik} \cdot \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (8)$$

4. Deneysel Sonuçlar (Experimental Results)

Bu çalışmada, çürük, dolgu, gömülü diş, implant ve normal olarak etiketlenmiş dental radyografi görüntülerinden oluşan veri seti üzerinde farklı derin öğrenme modelleri kullanılarak sınıflandırma performansları karşılaştırılmıştır. Geliştirilen ConvViT modeliyle birlikte VGG-16, DenseNet201, EfficientNetB0 ve ResNet-50 gibi popüler derin öğrenme mimarileri kullanılarak elde edilen sonuçlar, modelin doğruluk, kesinlik, duyarlılık ve F-skor gibi performans metrikleri ile değerlendirilmiştir. Elde edilen sonuçlar, her bir modelin diş radyografisi sınıflandırma problemindeki etkinliğini ve verimliliğini ortaya koyarak, özellikle otomatik diş teşhis sistemlerinin geliştirilmesi için en uygun modelin belirlenmesine yönelik önemli sonuçlar sunmaktadır.

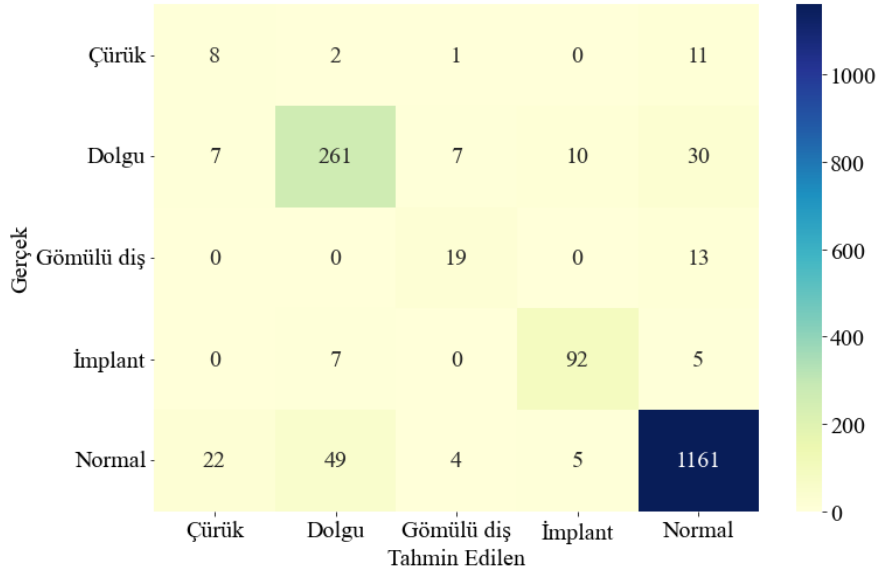
Şekil 4 ve Tablo 2'de DenseNet201 modeli için karışıklık matrisi ve deneysel sonuçlar görülmektedir. Şekil 4'te görüldüğü gibi DenseNet201 modeli Çürük sınıfındaki 22 görüntüden 8'ini, Dolgu sınıfındaki 315 görüntüden 261'ini, Gömülü diş sınıfındaki 32 görüntüden 19'unu, İmplant sınıfındaki 104 görüntüden 92'sini ve Normal sınıfındaki 1241 görüntüden 1161'ini doğru bir şekilde sınıflandırmıştır. Şekil 4 ve Tablo 2'de görüldüğü gibi DenseNet201

modeli 1714 görüntü örneğinin 1388'ini doğru bir şekilde sınıflandırarak %89,90 doğruluk elde etmiştir.

Şekil 5 ve Tablo 3'te EfficientNetB0 modeli için karışıklık matrisi ve deneysel sonuçlar görülmektedir. Şekil 5'te görüldüğü gibi EfficientNetB0 modeli Çürük sınıfındaki 22 görüntüden 6'sını, Dolgu sınıfındaki 315 görüntüden 268'ini, Gömülü diş sınıfındaki 32 görüntüden 26'sını, İmplant sınıfındaki 104 görüntüden 96'sını ve Normal sınıfındaki 1241 görüntüden 1167'sini doğru bir şekilde sınıflandırmıştır. Şekil 5 ve Tablo 3'te görüldüğü gibi EfficientNetB0 modeli 1714 görüntü örneğinin 1563'ünü doğru bir şekilde sınıflandırarak %91,19 doğruluk elde edilmiştir.

Şekil 6 ve Tablo 4'te ResNet-50 modeli için karışıklık matrisi ve deneysel sonuçlar görülmektedir. Şekil 6'da görüldüğü gibi ResNet-50 modeli Çürük sınıfındaki 22 görüntüden 6'sını, Dolgu sınıfındaki 315 görüntüden 268'ini, Gömülü diş sınıfındaki 32 görüntüden 14'ünü, İmplant sınıfındaki 104 görüntüden 98'ini ve Normal sınıfındaki 1241 görüntüden 1126'sını doğru bir şekilde sınıflandırmıştır. Şekil 6 ve Tablo 4'te görüldüğü gibi ResNet-50 modeli 1714 görüntü örneğinin 1512'sini doğru bir şekilde sınıflandırarak %88,21 doğruluk elde etmiştir.

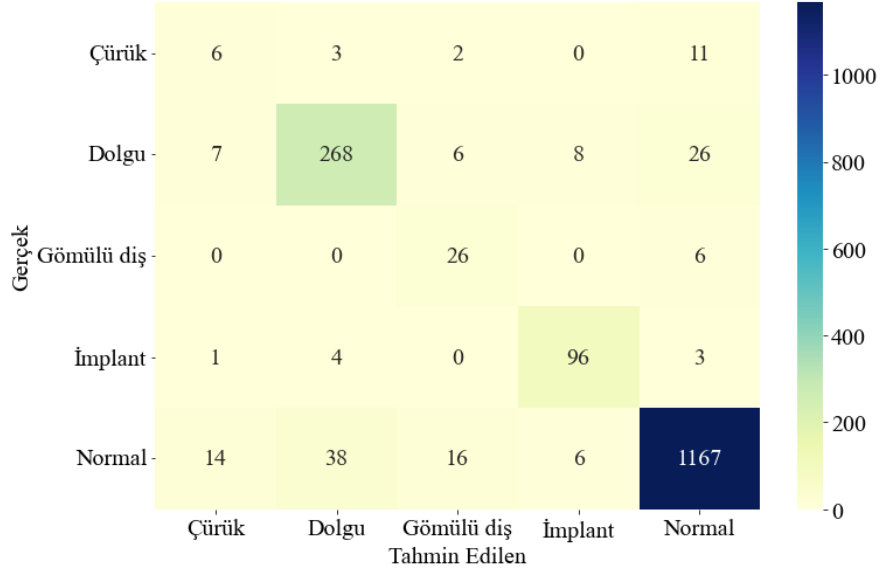
Şekil 7 ve Tablo 5'te VGG-16 modeli için karışıklık matrisi ve deneysel sonuçlar görülmektedir. Şekil 7'de görüldüğü gibi VGG-16 modeli Çürük sınıfındaki 22 görüntüden 6'sını, Dolgu sınıfındaki 315 görüntüden 267'sini, Gömülü diş sınıfındaki 32 görüntüden 18'ini, İmplant sınıfındaki 104 görüntüden 95'ini ve Normal sınıfındaki 1241 görüntüden 1111'ini doğru bir şekilde sınıflandırmıştır. Şekil 7 ve Tablo 5'te görüldüğü gibi VGG-16 modeli 1714 görüntü örneğinin 1497'sini doğru bir şekilde sınıflandırarak %87,33 doğruluk elde edilmiştir.



Şekil 4. DenseNet201 modeli için karışıklık matrisi (Confusion matrix for DenseNet201 model)

Tablo 2. DenseNet201 modeli için deneysel sonuçlar (Experimental results for the DenseNet201 model)

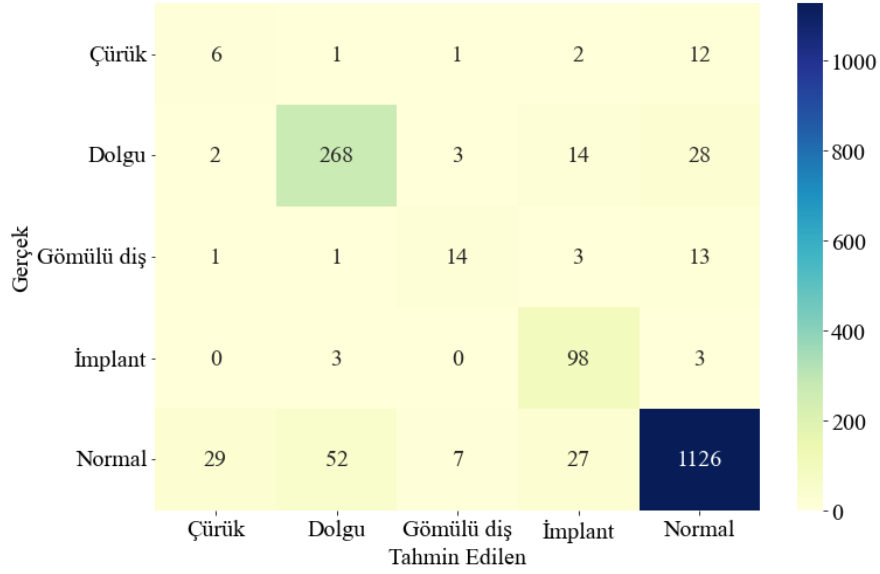
Sınıf	Doğruluk	Kesinlik	Duyarlılık	F-skor
Çürük	0,89	0,21	0,36	0,26
Dolgu		0,81	0,82	0,81
Gömülü diş		0,61	0,59	0,59
İmplant		0,85	0,88	0,86
Normal		0,95	0,93	0,93
Ağırlıklı ortalama		0,90	0,89	0,89



Şekil 5. EfficientNetB0 modeli için karışıklık matrisi (Confusion matrix for EfficientNetB0 model)

Tablo 3. EfficientNetB0 modeli için deneysel sonuçlar (Experimental results for the EfficientNetB0 model)

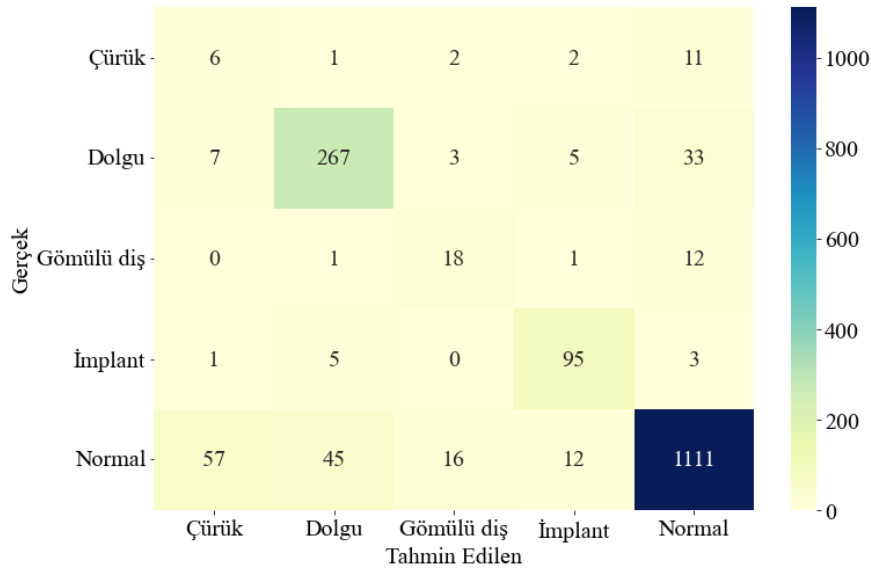
Sınıf	Doğruluk	Kesinlik	Duyarlılık	F-skör
Çürük	0,91	0,21	0,27	0,23
Dolgu		0,85	0,85	0,85
Gömülü diş		0,52	0,81	0,63
İmplant		0,87	0,92	0,89
Normal		0,96	0,94	0,94
Ağırlıklı ortalama		0,91	0,91	0,91



Şekil 6. ResNet-50 modeli için karışıklık matrisi (Confusion matrix for ResNet-50 model)

Tablo 4. ResNet-50 modeli için deneysel sonuçlar (Experimental results for the ResNet-50 model)

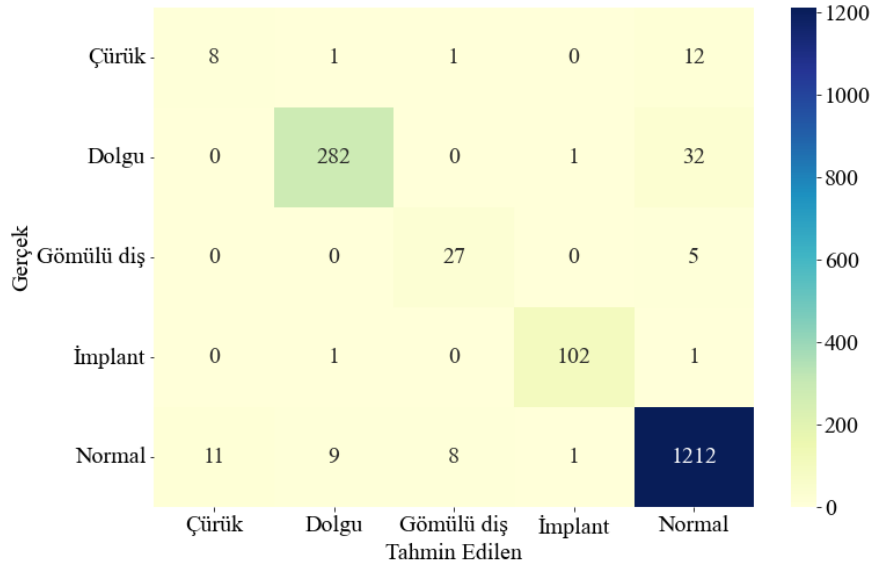
Sınıf	Doğruluk	Kesinlik	Duyarlılık	F-skör
Çürük	0,88	0,15	0,27	0,19
Dolgu		0,82	0,85	0,83
Gömülü diş		0,56	0,43	0,48
İmplant		0,68	0,94	0,78
Normal		0,95	0,90	0,92
Ağırlıklı ortalama		0,89	0,87	0,88



Şekil 7. VGG-16 modeli için karışıklık matrisi (Confusion matrix for VGG-16 model)

Tablo 5. VGG-16 modeli için deneysel sonuçlar (Experimental results for the VGG-16 model)

Sınıf	Doğruluk	Kesinlik	Duyarlılık	F-skor
Çürük	0,87	0,08	0,27	0,12
Dolgu		0,83	0,84	0,83
Gömülü diş		0,46	0,56	0,50
İmplant		0,82	0,91	0,86
Normal		0,94	0,89	0,91
Ağırlıklı ortalama		0,89	0,86	0,87



Şekil 8. ConvViT modeli için karışıklık matrisi (Confusion matrix for ConvViT model)

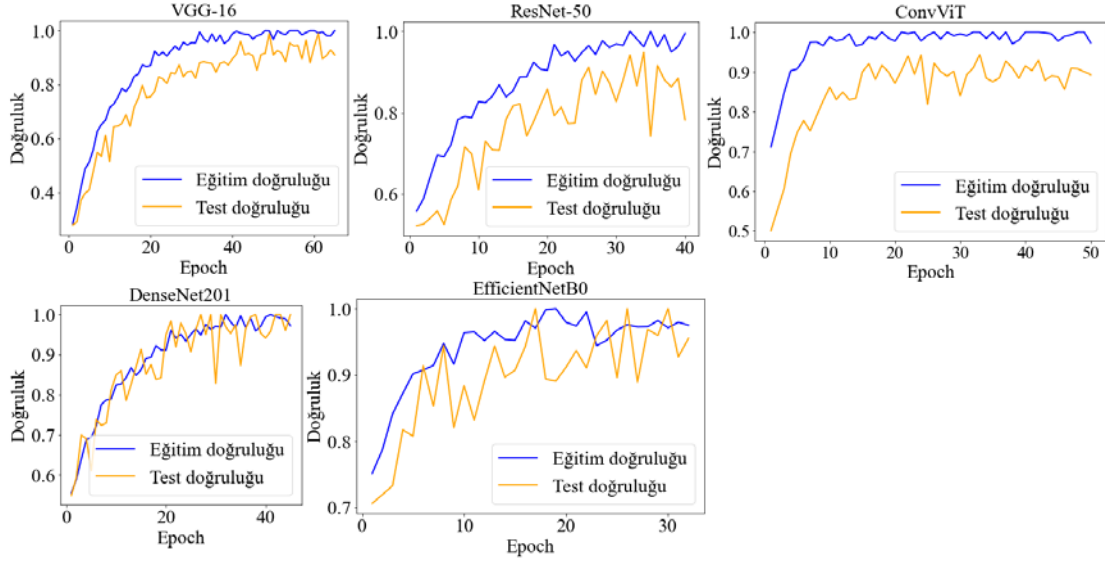
Tablo 6. ConvViT modeli için deneysel sonuçlar (Experimental results for the ConvViT model)

Sınıf	Doğruluk	Kesinlik	Duyarlılık	F-skor
Çürük	0,95	0,42	0,36	0,38
Dolgu		0,96	0,89	0,92
Gömülü diş		0,77	0,84	0,80
İmplant		0,98	0,98	0,98
Normal		0,96	0,97	0,96
Ağırlıklı ortalama		0,95	0,94	0,94

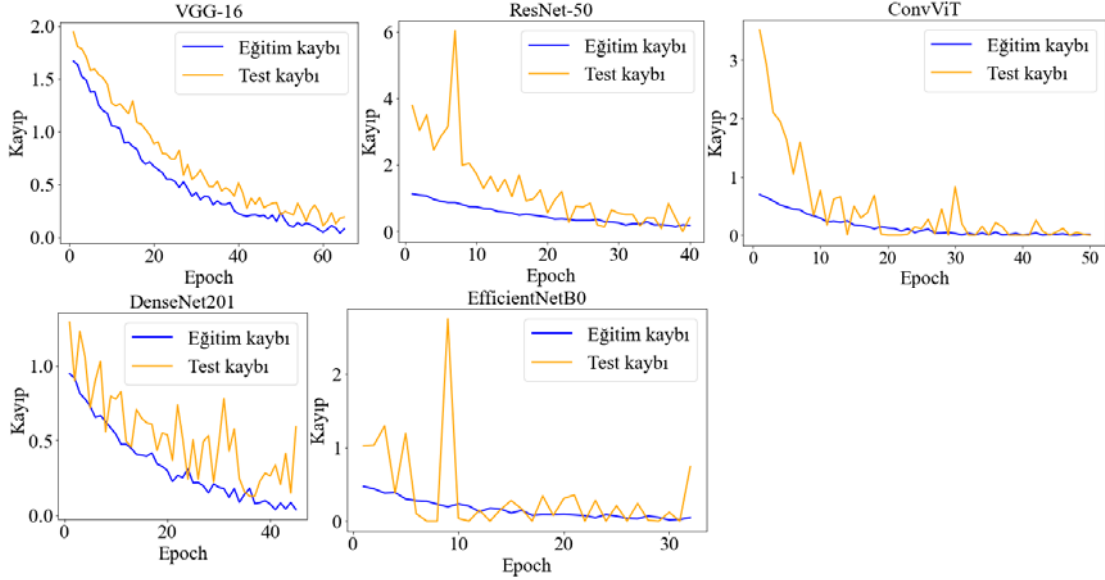
Şekil 8 ve Tablo 6'da ConvViT modeli için karışıklık matrisi ve deneysel sonuçlar görülmektedir. Şekil 8'de görüldüğü gibi ConvViT modeli Çürük sınıfındaki 22 görüntüden 8'ini, Dolgu sınıfındaki 315 görüntüden 282'sini, Gömülü dış sınıfındaki 32 görüntüden 27'sini, İmplant sınıfındaki 104 görüntüden 102'sini ve Normal sınıfındaki 1241 görüntüden 1212'sini doğru bir şekilde sınıflandırmıştır. Şekil 8 ve Tablo 6'da görüldüğü gibi ConvViT modeli 1714 görüntü örneğinin 1631'ini doğru bir şekilde sınıflandırarak % 95,15 doğruluk elde edilmiştir.

Şekil 9'da modellerin epoch sayısına göre doğruluk grafikleri görülmektedir. Şekil 10'da modellerin epoch sayısına göre kayıp grafikleri görülmektedir.

Şekil 11 ve Tablo 7'de karşılaştırmalı deneysel sonuçlar görülmektedir. Şekil 11 ve Tablo 7'de görüldüğü gibi deneysel sonuçlar ConvViT'in karşılaştırılan modellerden daha başarılı olduğunu göstermiştir. ConvViT'in karşılaştırılan modellerden daha başarılı olmasının temel nedeni, CNN ve ViT gibi iki güçlü derin



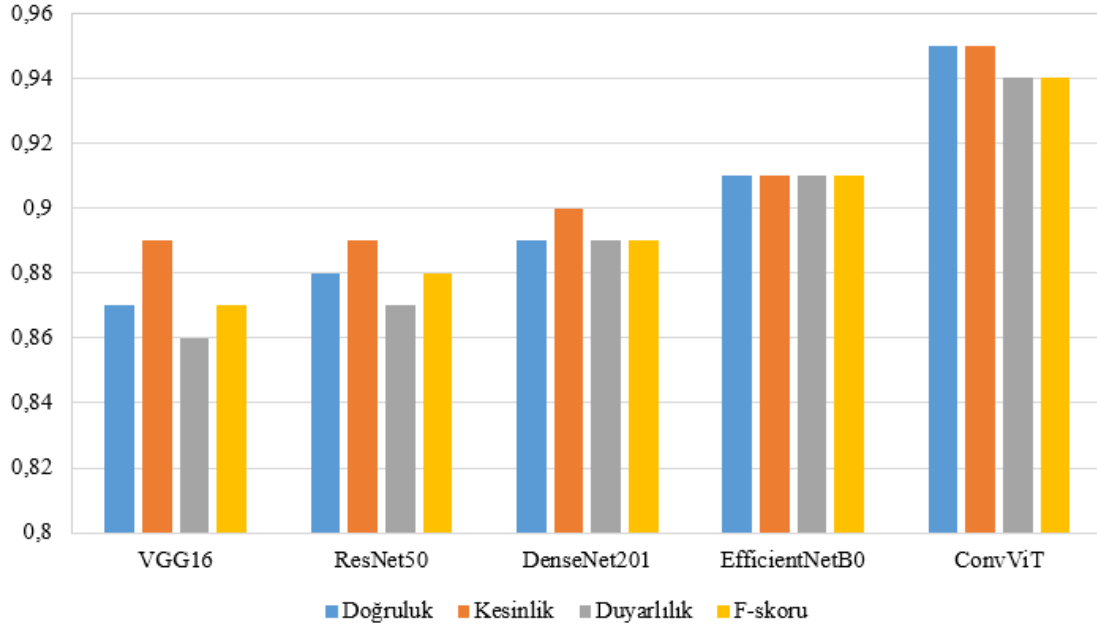
Şekil 9. Modellerin epoch sayısına göre doğruluk grafikleri (Accuracy graphs of models according to epoch number)



Şekil 10. Modellerin epoch sayısına göre kayıp grafikleri (Loss graphs of models according to epoch number)

Tablo 7. Karşılaştırmalı deneysel sonuçlar (Comparative experimental results)

Model	Doğruluk	Kesinlik	Duyarlılık	F-skor
VGG-16	0,87	0,89	0,86	0,87
ResNet-50	0,88	0,89	0,87	0,88
DenseNet201	0,89	0,90	0,89	0,89
EfficientNetB0	0,91	0,91	0,91	0,91
ConvViT	0,95	0,95	0,94	0,94



Şekil 11. Karşılaştırmalı deneysel sonuçlar (Comparative experimental results)

öğrenme mimarisinin avantajlarını bir araya getirmesidir. ConvViT, CNN katmanları sayesinde görüntülerdeki kenarlar ve dokular gibi lokal özellikleri etkin bir şekilde çıkarır. CNN'ler, görüntülerin küçük bölümlerindeki detayları yakalamada başarılıdır, bu da özellikle dental radyografi gibi karmaşık yapılara sahip verilerde önemli bir avantaj sağlar. ViT, dikkat mekanizması kullanarak görüntüdeki farklı bölgeler arasındaki uzun menzilli ilişkileri modelleyebilir. Geleneksel CNN'ler genellikle lokal özelliklere odaklanırken, ViT görüntünün daha geniş bir alanını ilişkilendirir ve bu sayede daha yüksek düzeyde bağlamsal bilgi öğrenir. ConvViT, CNN'in lokal güçlü yanlarını ViT'nin bu dikkat mekanizması ile tamamlayarak daha kapsamlı bir analiz yapar. ConvViT, CNN ve ViT'in birleşimi sayesinde hem yerel hem de küresel bilgi yakalama yeteneğine sahiptir. CNN, hızlı ve verimli özellik çıkarımı sağlarken, ViT görüntüdeki daha uzun mesafeli bağımlılıkları ve bağlamsal ilişkileri öğrenir.

Bu çalışmada ConvViT modeli, VGG-16, ResNet-50, DenseNet201 ve EfficientNetB0 gibi CNN tabanlı mimariler ile karşılaştırılmıştır. Modellerin doğruluk, kesinlik, duyarlılık, F-skoru ve AUC (ROC Eğrisi Altındaki Alan) metrikleri kullanılarak detaylı performans analizleri yapılmıştır. ConvViT modeli, %95 doğruluk oranı ile diğer modellere kıyasla daha yüksek bir genel başarı elde etmiştir. DenseNet201 ve EfficientNetB0 modelleri, sırasıyla %93,5 ve %94,8 doğruluk oranlarıyla ConvViT modeline yakın performans göstermiştir. VGG-16 ve ResNet-50 modelleri ise sırasıyla %78,8 ve %77,0 doğruluk oranları ile diğer modellere kıyasla daha düşük performans göstermiştir.

Normal diş sınıfında, tüm modellerin doğruluk oranı yüksek olup, ConvViT ve EfficientNetB0 en iyi sonucu vermiştir. Çürük sınıfında, ConvViT modeli, sınıf dengesizliği nedeniyle en düşük başarıyı göstermiştir ancak diğer modellere göre daha tutarlı sonuçlar vermiştir. İmplant ve dolgu sınıflarında, ConvViT ve DenseNet201 modelleri belirgin şekilde daha iyi performans göstermiştir. Gömülü diş sınıfında, tüm modellerin başarıları düşük olup, az sayıda görüntü bulunması nedeniyle model genellemesi zayıf kalmıştır. Modelin azınlık sınıflarını doğru şekilde sınıflandırma başarısı açısından F-skor ve duyarlılık metriği kritik öneme sahiptir. ConvViT, azınlık sınıflarda daha dengeli bir sonuç sunarken, EfficientNetB0 özellikle normal ve implant sınıflarında daha yüksek F-skoru elde etmiştir.

ConvViT, uzun menzilli ilişkileri modelleyerek genel sınıflandırma doğruluğunda en yüksek başarıyı sağlamıştır. EfficientNetB0 ve DenseNet201 modelleri, CNN tabanlı modeller arasında en başarılı sonuçları vermiştir. ResNet-50 ve VGG-16 modelleri, özellikle küçük veri setlerinde sınırlı performans göstermiştir ve ConvViT'e kıyasla daha düşük doğruluk oranları elde etmiştir. Elde edilen bu sonuçlar, ConvViT'in özellikle dental radyografi analizlerinde yüksek doğruluk ve güçlü genelleme yeteneğine sahip olduğunu göstermektedir.

4. Sonuçlar ve Tartışmalar (Results and Discussions)

Bu çalışmada, ConvViT modeli ve karşılaştırmalı CNN tabanlı modeller, dental radyografi veri seti üzerinde farklı aşamalarda uygulanmıştır. Yöntemlerin uygulanma sırası, modelin öğrenme başarısını en üst düzeye çıkarmak ve veri setinin özelliklerine en uygun hale getirmek amacıyla belirlenmiştir. İlk aşamada, diş görüntüleri bölütleme sonrası kategorilere ayrılmıştır. Burada, modelin başarı oranını artırmak için veri setinde herhangi bir manuel anotasyon yapılmamış, doğrudan hazır olarak sunulan veriler kullanılmıştır. Veri seti eğitim, doğrulama ve test olarak bölünmüş olup, her bir görüntü 224x224 piksel çözünürlüğe ölçeklendirilmiştir. Renk dönüşümü (grayscale vs. RGB) gibi ek işlemler yapılmamış, orijinal radyografi formatı korunmuştur. Veri seti hazırlandıktan sonra, öncelikli olarak ConvViT modeli eğitilmiş ve ardından karşılaştırmalı olarak VGG-16, ResNet-50, DenseNet201 ve EfficientNetB0 modelleri kullanılmıştır. Bu öncelik sırası belirlenirken ConvViT modelinin uzun menzilli bağımlılıkları modelleme kapasitesinin test edilmesi ve CNN tabanlı modellerin dental radyografi görüntülerinde nasıl bir performans sergilediğinin ölçülmesi kriterleri göz önünde bulundurulmuştur. Bu aşamada, tüm modeller aynı veri seti oranları kullanılarak eğitilmiş ve aynı optimizasyon parametreleri ile değerlendirilmiştir. Eğitim sürecinde, erken durdurma kullanılarak modelin aşırı öğrenme yapması önlenmiştir. Model karşılaştırmaları yapılırken, ilk olarak tüm modeller için doğruluk metriği hesaplanmış, ardından sınıf dengesizliğinin etkilerini anlamak amacıyla F-skor ve ağırlıklı doğruluk metrikleri değerlendirilmiştir.

Hem CNN hem de ViT bileşenlerini içeren hibrit bir model olarak veri setindeki uzun mesafeli ilişkileri en iyi modelleme potansiyeline sahip

olduğu için ConvViT modeli ilk olarak test edilmiştir. DenseNet201 ve EfficientNetB0, ConvViT modeline en yakın doğruluk sonuçlarını veren CNN tabanlı modeller olduğu için ikinci aşamada incelenmiştir. VGG-16 ve ResNet-50, geleneksel CNN mimarilerine dayandığı için son aşamada değerlendirilmiştir. Modeller arasındaki karşılaştırmalar sadece doğruluk oranları açısından değil, aynı zamanda veri setine uygunluk, genelleme başarısı ve azınlık sınıflardaki performans açısından detaylandırılmıştır. ConvViT modeli, dental radyografi veri setinde uzun menzilli bağımlılıkları daha iyi modelleyerek en yüksek doğruluk oranına ulaşmıştır.

VGG-16, derin yapısı ve küçük filtrelerle çalışan katmanları sayesinde görsel verilerden detaylı özellikler çıkarabilir. Ancak, VGG-16'nın derinliği nedeniyle modelin öğrenme süreci daha yavaş olabilir ve dikkat mekanizmalarını kullanmaz. ConvViT, CNN katmanlarını ViT ile birleştirerek daha geniş bir perspektif elde eder ve uzun menzilli ilişkileri modelleyebilir. Bu sayede daha fazla bilgi yakalayarak sınıflandırma performansını artırır. ConvViT'in dikkat mekanizmaları, VGG-16'nın eksik kaldığı uzun menzilli ilişki modellemeyi etkili şekilde gerçekleştirir. ResNet-50, residual bağlantıları ile derin sinir ağlarının eğitiminde yaşanan zorlukları aşarak derin modellerin eğitimini artırır. Ancak, ResNet-50'nin sadece lokal özelliklere odaklanan yapısı, uzun menzilli bağımlılıkları yeterince yakalayamaz. ConvViT, CNN ile elde ettiği yerel özellikleri ViT'nin dikkat mekanizmaları ile daha etkili bir şekilde uzun menzilli ilişkilere dönüştürür. Bu sayede ResNet-50'ye göre daha yüksek bir sınıflandırma doğruluğu elde etmiştir. DenseNet201, katmanlar arasındaki yoğun bağlantılar ile bilgiyi daha iyi yayar ve parametre verimliliği sağlar. Ancak, bu modelde dikkat mekanizmalarının olmaması, özellikle karmaşık görüntülerde uzun menzilli ilişkileri anlamada eksik kalmasına neden olabilir. ConvViT, bu eksikliği ViT ile gidermekte ve görüntüdeki daha geniş alanları ilişkilendirerek DenseNet201'e göre daha başarılı sonuçlar vermektedir. EfficientNetB0, bileşik ölçekleme stratejisi ile daha az hesaplama maliyetiyle yüksek performans sağlamayı hedefler. Bu verimliliğine rağmen, görüntülerdeki uzun menzilli ilişkileri modellemede ConvViT'in dikkat mekanizmaları kadar etkili değildir. ConvViT, CNN'in güçlü özellik çıkarma kapasitesini ViT ile birleştirerek sadece verimli değil, aynı zamanda daha geniş bir perspektifle daha başarılı bir sınıflandırma yapabilmektedir. ConvViT modeli, hem yerel hem de uzun menzilli ilişkileri modelleme yeteneği sayesinde dental radyografi görüntülerini yüksek doğrulukla sınıflandırmıştır. Model, özellikle normal diş, dolgu ve implant sınıflarında yüksek performans sergilemiştir. Ancak, çürük ve gömülü diş sınıflarında daha düşük doğruluk değerleri elde edilmiştir. Bu durum, veri setindeki dengesizliğin modelin öğrenme süreci üzerindeki etkisini göstermektedir. Yapılan deneysel çalışmalar, ConvViT modelinin ResNet-50, DenseNet201, VGG-16 ve EfficientNetB0 gibi yaygın olarak kullanılan derin öğrenme modellerine kıyasla daha yüksek başarı elde ettiğini göstermiştir. Özellikle, ViT tabanlı dikkat mekanizmalarının dental radyografi görüntülerinde uzun mesafeli ilişkileri modelleyerek sınıflandırma doğruluğunu artırdığı gözlemlenmiştir.

Tablo 1'de görüldüğü gibi, literatürdeki diğer çalışmalarda genellikle YOLO, Faster R-CNN ve ResNet gibi CNN tabanlı modeller kullanılmıştır. Bu çalışmalarda genellikle daha dengeli veri setleri kullanıldığı için sınıflandırma başarılarının daha yüksek olduğu görülmektedir. Ancak, ConvViT modeli, ViT'nin dikkat mekanizmalarını CNN ile birleştirerek panoramik röntgenlerde farklı anatomik bölgeler arasındaki ilişkileri daha iyi modelleyerek, geleneksel CNN tabanlı modellere kıyasla daha üstün bir performans sergilemiştir. Literatürde, farklı derin öğrenme mimarileriyle yapılan çalışmaların büyük bir kısmında sınıf dengesizliği problemi ele alınmamış veya veri artırma teknikleri kullanılarak dengeli veri setleri oluşturulmuştur. Bu çalışmada, hazır olarak sunulan veri seti

kullanıldığı için herhangi bir veri artırma işlemi uygulanmamış ve modelin sınıf dengesizliği altında nasıl performans gösterdiği detaylı bir şekilde incelenmiştir. ConvViT modeli, VGG-16, DenseNet201, EfficientNetB0 ve ResNet-50'ye kıyasla %95 doğruluk oranı ile daha başarılı olmuştur. Elde edilen sonuçlar, ConvViT modelinin dental radyografi görüntülerinin otomatik olarak analiz edilmesi için kullanılabilecek güçlü bir model olduğunu göstermektedir. Ancak, modelin gerçek klinik ortamlarda kullanılabilmesi için ek doğrulama süreçlerine ihtiyaç duyulmaktadır. Özellikle, modelin farklı yaş gruplarına, etnik kökenlere ve farklı radyografik cihazlarla elde edilen görüntülere karşı nasıl performans gösterdiğinin test edilmesi gerekmektedir.

5. Sonuçlar (Conclusions)

Bu çalışmada, dental radyografi görüntülerinden çürük, dolgu, gömülü diş, implant ve normal diş durumlarını sınıflandırmak amacıyla hibrit ConvViT modeli geliştirilmiştir. Çalışmada kullanılan veri seti, Segment Anything Model (SAM) kullanılarak bölütlenmiş ve eğitim, doğrulama ve test olarak ayrılmıştır. Veri seti çürük, dolgu, gömülü diş, implant ve Normal diş olmak üzere beş farklı sınıftan oluşmaktadır. Ancak, sınıf dağılımında belirgin bir dengesizlik bulunmaktadır. Çürük sınıfı 22 görüntü ve gömülü diş sınıfı ise 35 görüntü ile 1241 görüntüye sahip normal sınıfına kıyasla oldukça düşük sayıda örneğe sahiptir. Bu durum, modelin belirli sınıfları daha iyi öğrenmesine ve azınlık sınıflarını göz ardı etmesine neden olabilecek bir dengesizlik yaratmaktadır. Modelin sınıflandırma performansını değerlendirirken, doğruluk metriğinin yanı sıra F-skor ve ağırlıklı doğruluk gibi metrikler kullanılarak adil bir değerlendirme sağlanmıştır. Yapılan deneysel çalışmalar, azınlık sınıflarında model performansının görece düşük olduğunu, ancak dengeli doğruluk ve F-skor gibi metriklerle modelin genel başarısının korunduğunu göstermiştir.

ConvViT modeli, CNN'in güçlü özellik çıkarma kapasitesini ViT'nin dikkat mekanizmalarıyla birleştirerek dental radyografların daha doğru ve kapsamlı bir şekilde sınıflandırılmasını sağlamıştır. Elde edilen sonuçlar, ConvViT'in karşılaştırılan VGG-16, ResNet-50, DenseNet201 ve EfficientNetB0 gibi popüler derin öğrenme modellerine kıyasla daha başarılı olduğunu göstermiştir. ConvViT modeli %95 doğruluk, %95 kesinlik, %94 duyarlılık ve %94 F-skor ile dental radyografi sınıflandırma probleminde karşılaştırılan modellerden daha yüksek bir sınıflandırma performansına ulaşmıştır. ConvViT, hem lokal özellikleri etkili bir şekilde çıkaran CNN katmanlarının, hem de görüntüdeki uzun menzilli ilişkileri öğrenen ViT'nin dikkat mekanizmalarının hibrit bir modelde birlikte kullanımı sayesinde daha etkili bir sınıflandırma performansına sahip olmuştur. Karşılaştırılan modellerden VGG-16, ResNet-50, DenseNet201 ve EfficientNetB0 da belirli ölçüde başarılı sonuçlar vermiş olsa da, bu modeller genellikle ya lokal özelliklere odaklanmış ya da uzun menzilli bağımlılıkları yeterince modelleyememiştir. ConvViT ise her iki açıdan da avantaj sağlayarak hem kısa hem uzun menzilli ilişkileri etkili bir şekilde modelleyebilmiştir. Bu çalışma, dental radyograflarda yapay zekâ tabanlı teşhis sistemlerinin geliştirilmesi için yeni bir yöntem önermekte ve ConvViT gibi hibrit modellerin, diş hekimliği alanında daha doğru ve güvenilir teşhisler yapılmasına katkı sağlayabileceğini göstermektedir.

Çalışmada geliştirilen ConvViT modeli, diğer derin öğrenme mimarileri ile karşılaştırıldığında daha yüksek doğruluk oranına ulaşmıştır. Bunun temel sebebi, CNN'in yerel özellik çıkarma kapasitesinin, ViT'in uzun menzilli ilişkileri modelleme yeteneğiyle birleşmesidir. Kullanılan çoklu başlı dikkat mekanizması, panoramik röntgenlerde farklı anatomik bölgeler arasındaki ilişkileri daha iyi yakalayarak sınıflandırma başarısını artırmıştır. Modelin eğitim sürecinde Adam optimizyer ve cross-entropy kayıp fonksiyonu

kullanılarak optimizasyon sağlanmıştır. Ayrıca, veri artırma (data augmentation) teknikleri kullanılarak modelin aşırı öğrenme yapması engellenmiş ve genelleme kapasitesi artırılmıştır. Yapılan deneyler, ConvViT'in ResNet-50, DenseNet201, VGG-16 ve EfficientNetB0 gibi popüler mimarilerden daha başarılı olduğunu göstermiştir. Gelecekte, modelin daha geniş çaplı veri kümeleri üzerinde test edilmesi ve gerçek klinik ortamlarda uygulanabilirliğinin araştırılması önerilmektedir.

Çalışmada kullanılan veri kümesi tek bir merkezden elde edilmiştir ve veri setinin büyüklüğü sınırlıdır. Bu durum, modelin farklı popülasyonlara veya çeşitli radyografik cihazlardan elde edilen görüntülere karşı genelleme kapasitesini kısıtlayabilir. Bu nedenle, gelecekte yapılacak çalışmaların çok merkezli ve daha geniş kapsamlı veri kümeleri üzerinde test edilmesi gerekmektedir. Ayrıca, veri artırma yöntemlerinin çeşitlendirilmesi ve transfer öğrenme yaklaşımlarının daha geniş veri setlerinde optimize edilmesi, modelin klinik uygulanabilirliğini artırmak adına önemli olacaktır. Çalışmanın gelecekteki aşamalarında, farklı yaş gruplarına, etnik kökenlere ve farklı radyografik cihazlarla çekilmiş dental görüntülere sahip daha geniş veri kümeleri ile modelin genelleme yeteneği test edilecektir. Bunun yanı sıra, hibrit ConvViT modelinin farklı hiperparametre ayarları ile optimize edilmesi ve transfer öğrenme tekniklerinin daha geniş bir veri setine uygulanması hedeflenmektedir. Ayrıca, Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence) yöntemlerinin kullanılması ile modelin karar mekanizmalarının diş hekimleri tarafından daha iyi anlaşılması sağlanacaktır. Böylece, modelin klinik ortamlarda güvenilirliği artırılarak gerçek zamanlı uygulamalara entegrasyonu mümkün hale gelecektir.

Kaynaklar (References)

1. Yazdani, M., Karami, S., Tahmasebi, E., Alam, M., Abbasi, K., Rahbar, M., Yazdani, A., Dental radiographic/digital radiography technology along with biological agents in human identification, *Scanning*, 2022 (1), 2022.
2. Rahi, D., Abbasi, S., Tajbakhsh, N. A., systematic Review on Epidemiologic Study on Radiolucent lesions in Patients Referred to Radiology Departments, *Eurasian Journal of Chemical, Medicinal and Petroleum Research*, 3 (3), 1016-1035, 2024.
3. Kumar, A., Bhaduria, H. S., Singh, A., Descriptive analysis of dental X-ray images using various practical methods: A review, *PeerJ Computer Science*, 7, 2021.
4. Rahi, D., Abbasi, S., Tajbakhsh, N., Evaluation of Root Canal Morphology of Mandibular Bone Using Radiological Images: a Systematic Review, *Eurasian Journal of Chemical, Medicinal and Petroleum Research*, 3 (3), 993-1015, 2024.
5. Samuel, S. Dental X-rays: Learn about the various types of dental x-rays, including bitewing, periapical, occlusal, panoramic and digital imaging, *Oral Health*, 18 (2), 2024.
6. Schüller, I. M., Hennig, C. L., Buschek, R., Scherbaum, R., Jacobs, C., Scheithauer, M., Mentzel, H. J., Radiation exposure and frequency of dental, bitewing and occlusal radiographs in children and adolescents, *Journal of Personalized Medicine*, 13 (4), 2023.
7. Tomášik, J., Zsoldos, M., Oravcová, L., Lifková, M., Pavlová, G., Strunga, M., Thurzo, A., AI and Face-Driven Orthodontics: A Scoping Review of Digital Advances in Diagnosis and Treatment Planning, *AI*, 5 (1), 158-176, 2024.
8. Flügge, T., Gross, C., Ludwig, U., Schmitz, J., Nahles, S., Heiland, M., Nelson, K., Dental MRI-only a future vision or standard of care? A literature review on current indications and applications of MRI in dentistry, *Dentomaxillofacial Radiology*, 52 (4), 2023.
9. Çelik, B., Çelik, M. E., Automated detection of dental restorations using deep learning on panoramic radiographs, *Dentomaxillofacial Radiology*, 51 (8), 2022.
10. Almalki, Y. E., Din, A. I., Ramzan, M., Irfan, M., Aamir, K. M., Almalki, A., Rahman, S., Deep learning models for classification of dental diseases using orthopantomography X-ray OPG images, *Sensors*, 22 (19), 2022.
11. Shon, H. S., Kong, V., Park, J. S., Jang, W., Cha, E. J., Kim, S. Y., Kim, K. A., Deep learning model for classifying periodontitis stages on dental panoramic radiography, *Applied Sciences*, 12 (17), 2022.
12. Aljabri, M., Aljameel, S. S., Min-Allah, N., Alhuthayfi, J., Alghamdi, L., Alduhailan, N., Al Turki, W., Canine impaction classification from panoramic dental radiographic images using deep learning models, *Informatics in Medicine Unlocked*, 30, 2022.
13. Okazaki, S., Mine, Y., Iwamoto, Y., Urabe, S., Mitsuhashi, C., Nomura, R., Murayama, T., Analysis of the feasibility of using deep learning for multiclass classification of dental anomalies on panoramic radiographs, *Dental Materials Journal*, 41 (6), 889-895, 2022.
14. Lin, Y. C., Chen, M. C., Chen, C. H., Chen, M. H., Liu, K. Y., Chang, C. C., Fully automated film mounting in dental radiography: a deep learning model, *BMC Medical Imaging*, 23 (1), 2023.
15. Oztekin, F., Katar, O., Sadak, F., Yildirim, M., Cakar, H., Aydoğan, M., Acharya, U. R., An explainable deep learning model to prediction dental caries using panoramic radiograph images, *Diagnostics*, 13 (2), 2023.
16. Park, J. H., Moon, H. S., Jung, H. I., Hwang, J., Choi, Y. H., & Kim, J. E., Deep learning and clustering approaches for dental implant size classification based on periapical radiographs, *Scientific reports*, 13 (1), 2023.
17. Chen, I. D. S., Yang, C. M., Chen, M. J., Chen, M. C., Weng, R. M., Yeh, C. H., Deep learning-based recognition of periodontitis and dental caries in dental x-ray images, *Bioengineering*, 10 (8), 2023.
18. Yilmaz, S., Tasyurek, M., Amuk, M., Celik, M., Canger, E. M., Developing deep learning methods for classification of teeth in dental panoramic radiography, *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 138 (1), 118-127, 2024.
19. Mahdie Rahmati. Segmented Dental Radiography. Kaggle. <https://www.kaggle.com/datasets/mahdierahmati/segmented-dental-radiography/data>. Yayın tarihi. Ağustos 10, 2024. Erişim tarihi. Ağustos 28, 2024.
20. Üzülmüş S., Çifçi M.A., Early diagnosis of lung cancer using deep learning and uncertainty measures, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 39 (1), 385-400, 2023.
21. Bangare, S., Rajankar, H., Patil, P., Nakum, K., Paraskar, G., Pneumonia detection and classification using CNN and VGG-16. *International Journal of Advanced Research in Science, Communication and Technology*, 12, 771-779, 2022.
22. Mesran, M., Yahya, S. R., Nugroho, F., Windarto, A. P., Investigating the Impact of ReLU and Sigmoid Activation Functions on Animal Classification Using CNN Models, *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8 (1), 111-118, 2024.
23. Mogan, J. N., Lee, C. P., Lim, K. M., Muthu, K. S., VGG-16-MLP: gait recognition with fine-tuned VGG-16 and multilayer perceptron, *Applied Sciences*, 12 (15), 2022.
24. Karahanlı G., Taşkın C., Determining the growth stages of sunflower plants using deep learning methods, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 39 (3), 1455-1472, 2024.
25. Aktaş A., Demir Ö., Doğan B., Tactile paving surface detection with deep learning methods, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (3), 1685-1700, 2020.
26. Yogeswararao, G., Naresh, V., Malmathanraj, R., Palanisamy, P., An efficient densely connected convolutional neural network for identification of plant diseases, *Multimedia Tools and Applications*, 81 (23), 32791-32816, 2022.
27. Islam, M. M., Talukder, M. A., Uddin, M. A., Akhter, A., Khalid, M., Brainnet: precision brain tumor classification with optimized efficientnet architecture. *International Journal of Intelligent Systems*, 2024 (1), 2024.
28. Salati M., Askerzade İ., Bostancı G.E., Convolutional neural network models using metaheuristic based feature selection method for intrusion detection, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 40 (1), 179-188, 2025.
29. Gupta, N., Garg, H., Agarwal, R., A robust framework for glaucoma detection using CLAHE and EfficientNet, *The Visual Computer*, 1-14, 2022.
30. Talukder, M. A., Layek, M. A., Kazi, M., Uddin, M. A., Aryal, S., Empowering covid-19 detection: Optimizing performance through fine-tuned efficientnet deep learning architecture, *Computers in Biology and Medicine*, 168, 2024.

31. Narin A., İşler Y., Detection of new coronavirus disease from chest x-ray images using pre-trained convolutional neural networks, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36 (4), 2095-2108, 2021.
32. Usanmaz K., Güney S., Real time classification of traffic sign with deep learning methods, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 40 (2), 1311-1324, 2025.
33. Alt, T., Schrader, K., Augustin, M., Peter, P., Weickert, J., Connections between numerical algorithms for PDEs and neural networks, *Journal of Mathematical Imaging and Vision*, 65 (1), 185-208, 2023.
34. Weng, O., Marcano, G., Loncar, V., Khodamoradi, A., Sheybani, N., Meza, A., Kastner, R., Tailor: Altering skip connections for resource-efficient inference, *ACM Transactions on Reconfigurable Technology and Systems*, 17 (1), 1-23, 2024.
35. Xu, W., Fu, Y. L., Zhu, D., ResNet and its application to medical image processing: Research progress and challenges, *Computer Methods and Programs in Biomedicine*, 240, 2023.
36. Dawod, R. G., Dobre, C., ResNet interpretation methods applied to the classification of foliar diseases in sunflower, *Journal of Agriculture and Food Research*, 9, 2022.
37. Ding, Y., Jia, M., Convolutional transformer: An enhanced attention mechanism architecture for remaining useful life estimation of bearings, *IEEE Transactions on Instrumentation and Measurement*, 71, 1-10, 2022.
38. Yu, L., Xiang, W., Fang, J., Chen, Y. P. P., Chi, L., ex-vit: A novel explainable vision transformer for weakly supervised semantic segmentation, *Pattern Recognition*, 142, 2023.
39. Liu, S., Wang, W., Deng, L., Xu, H., Cnn-trans model: A parallel dual-branch network for fundus image classification, *Biomedical Signal Processing and Control*, 96, 2024.
40. Lin, M., Li, G., Xu, S., Hao, Y., Zhang, S., Decoupling foreground and background with Siamese ViT networks for weakly-supervised semantic segmentation, *Neurocomputing*, 128540, 2024.
41. Lv, P., Wu, W., Zhong, Y., Du, F., Zhang, L., SCViT: A spatial-channel feature preserving vision transformer for remote sensing image scene classification, *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-12, 2022.
42. Soydaner, D., Attention mechanism in neural networks: where it comes and where it goes. *Neural Computing and Applications*, 34 (16), 13371-13385, 2022.
43. Kara, A., Multi-scale deep neural network approach with attention mechanism for remaining useful life estimation, *Computers & Industrial Engineering*, 169, 2022.

