AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

■ RESEARCH ARTICLE

Classification of Diabetic Retinopathy Disease with Deep Learning Methods

Metin Tuncel [a†], **Murat Uçar** [a]

[a] Department of Computer Engineering, İzmir Bakırçay University, İzmir, Türkiye
[†] tuncel155@gmail.com, corresponding author

RECEIVED DECEMBER 05, 2024
ACCEPTED APRIL 15, 2025

CITATION Tuncel, M. & Uçar, M. (2025). Classification of diabetic retinopathy disease with deep learning methods. *Artificial Intelligence Theory and Applications, 5(1)*, 1-17.

Abstract

Diabetes is defined as a chronic disease caused by an increase in blood glucose levels (hyperglycemia), in which the organism is unable to make adequate use of carbohydrates, fats, and proteins due to the pancreas' inability to produce enough insulin hormone or the hormone's inability to function properly. Diabetes is the most severe chronic condition, according to a World Health Organization report. Diabetic retinopathy (DR) is a complication of type 1 diabetes. Diabetes problems can cause damage to the blood vessels in the light-sensitive tissue (retina) at the rear of the eye, resulting in DR. Diabetes is one of the top three causes of blindness, according to the International Diabetes Federation's Diabetes Atlas 10th Edition (2021). Diabetes-related blindness is mostly caused by the loss of small vessels in the retina because of chronic hyperglycemia. Approximately 25% of individuals with diabetes globally have DR of any severity. Our country has around 2 million diabetes patients, with DR accounting for 25% of the total. There are five categories of DR. These are non-proliferative diabetic retinopathy (NPDR), mild non-proliferative retinopathy, moderate non-proliferative retinopathy, severe non-proliferative retinopathy, and proliferative diabetic retinopathy (PDR), in order of severity. Using the APTOS2019 dataset, this study develops a computer-aided diagnosis system to assist doctors in making early diagnoses using convolutional deep learning (DL) models. Binary and multi-class classification was done utilizing cutting-edge models such as VGG16, InceptionResNetV2, ResNet152V2, EfficientNetB0, and MobileNetV2, which are extensively used in the literature for medical image classification. Since the amount of data in the multi-class classification in diabetic retinopathy disease images was not equal, the data were equalized utilizing data augmentation techniques in the training dataset with the Albumentations library. Among the cutting-edge models employed in binary classification, VGG16 performed best, with accuracy, precision, sensitivity, and F1-score metric values of 0.97. VGG16 was the best model employed in multi-class classification, with accuracy, precision, sensitivity, and f1-score metric values of 0.78 and 0.79, respectively.

Keywords: diabetic retinopathy, deep learning, convolutional neural networks, transfer learning

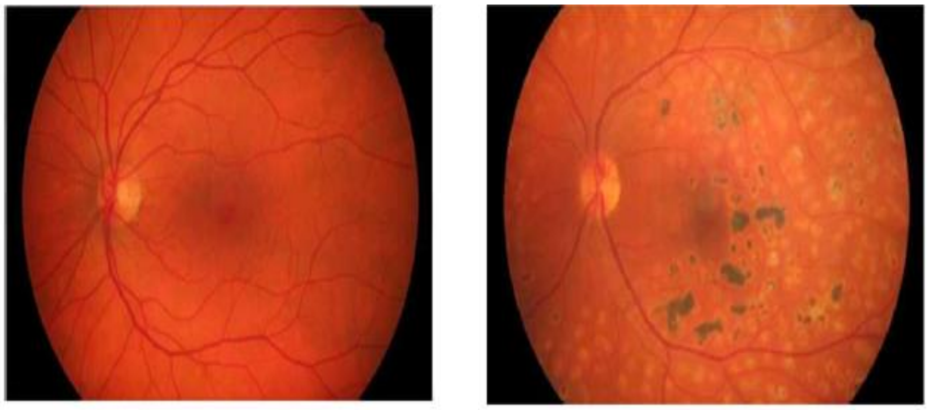
1. Introduction

Diabetes is a rapidly increasing global health problem, with the number of people with diabetes predicted to rise to 643 million (11.3%) by 2030 and 783 million (12.2%) by 2045. [2,3]. High blood glucose is a common effect of uncontrolled diabetes. Over time, it causes serious damage to numerous systems of the body, especially the eyes, heart,

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2025 İzmir Bakırçay University

kidneys, nerves, and blood vessels [2,4,5]. Diabetes is generally classified into 4 groups: Type I, Type II, gestational diabetes, and other specific types. The most common types of diabetes are Type I and Type II diabetes. The type of diabetes that occurs during pregnancy is defined as gestational diabetes, while other types are defined as high blood glucose levels that occur for many reasons and affect the pancreas [1].

Diabetic retinopathy can progress slowly or quickly, yet it can also improve on its own. However, if it worsens, it may result in partial or permanent vision loss.



a) Healthy retina image b) Retina with diabetic retinopathy

Figure 1. Sample images: a) healthy retina image, b) retina image with diabetic retinopathy.

Figure 2 shows hemorrhages, soft and hard exudates, and microaneurysms on the retina.

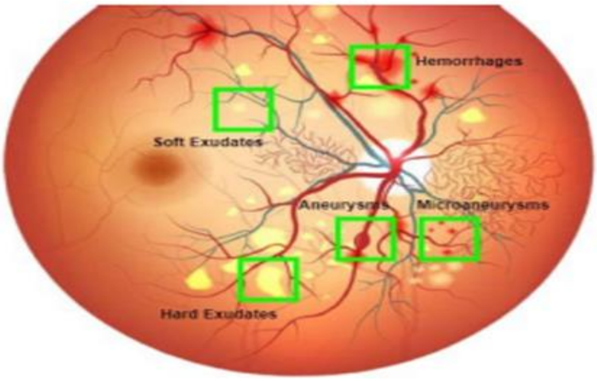


Figure 2. Fundoscopic illustration of the retina displaying microaneurysms, hemorrhages, and exudates [6]

There are 5 stages of diabetic retinopathy:[7]

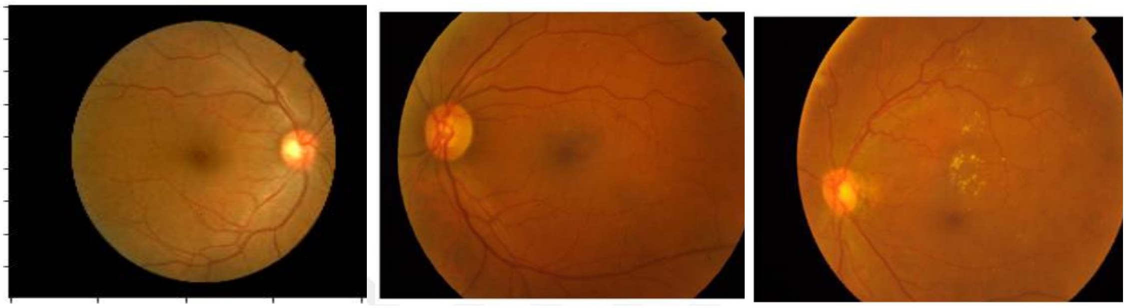
Non-proliferative diabetic retinopathy [7]: It can be seen in Figure 3a. This is the time when the sickness first appears. During this period, fluid leakage occurs in the damaged vessels, causing retinal hemorrhages. In general, the patient's vision is not affected during this period [8].

Mild non-proliferative retinopathy [7]: It can be seen in Figure 3b. It represents the first stage of diabetic retinopathy. Swellings (microaneurysms) occur in small blood vessels in the rear of the eye (retina) [9, 10].

Moderate non-proliferative retinopathy [7]: It can be seen in Figure 3c. Diabetic macular edema occurs as a result of the accumulation of blood and other fluids in the small central part of the retina (macula). Visual problems are also seen at this stage due to diabetic macular edema [9, 10].

Severe non-proliferative retinopathy [7]: It can be seen in Figure 3d. At this stage, new blood vessels and scar tissue are formed. Some or all of the blood vessels are occluded. Complete occlusion of blood vessels is called macular ischemia. As a result of this condition in the blood vessels, dark spots (flying objects) form in the visual field, and this causes blurred vision. At this stage, the possibility of visual loss is quite high [9, 10].

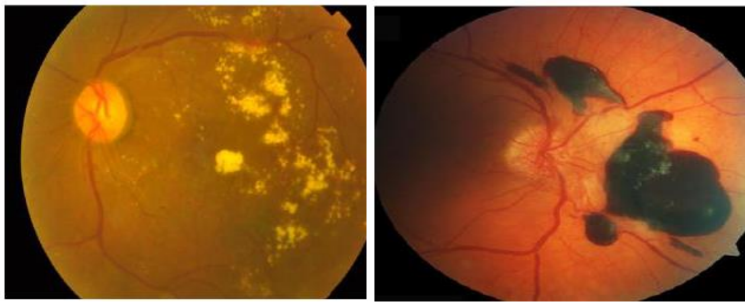
Proliferative diabetic retinopathy [7]: It can be seen in Figure 3e. This figure represents the most dangerous period of diabetic retinopathy. During this period, the blood vessels in the retinal layer are severely impaired, resulting in the formation of areas that cannot be nourished. These areas cause the development of new blood vessels. Since these vessels are very thin and fragile, they can cause sudden bleeding in the eye. The patient experiencing this process notices black spots appearing in front of the eye and moving in that direction, wherever the eye is turned. Patients with significant hemorrhage suffer visual loss and blindness [9, 10].



a) Non-proliferative diabetic retinopathy (NPDR) retinal image

b) Retinal image of mild DR [11].

c) Retinal image of moderate DR [11].



d) Retinal image of severe DR.

e) The retinal image of proliferative DR [11].

Figure 3. Sample images for the stages of diabetic retinopathy

In recent years, DL methods have been frequently used in object recognition, image classification, and segmentation in medical and ophthalmological images, and very successful results have been obtained. In particular, deep convolutional neural networks (DNNs) have been used for early detection and identification of retinal diseases such as DR, age-related macular degeneration, and glaucoma from retinal images [12,13,14,15]

In this study, image classification for DR was performed using DL methods. VGG16, InceptionResNetV2, ResNet152V2, EfficientNetB0, and MobileNetV2 transfer learning (TL) architectures from state-of-the-art models were used. Binary and multi-class classification was performed for DR. Since the amount of data in the multi-class classification was not equal, the data was equalized utilizing data augmentation techniques using the albumentations library in the training dataset.

This study contributes to the literature by demonstrating the effectiveness of deep learning and transfer learning methods in the early diagnosis of DR. The successful use of advanced models such as VGG16, InceptionResNetV2, ResNet152V2, EfficientNetB0, and MobileNetV2 in DR classification brings a significant innovation to the existing knowledge in this field. Additionally, addressing class imbalances through data augmentation techniques leads to more accurate and reliable classification results.

The rest of this paper is organized as follows. Section 2 presents the literature review. In Section 3, the proposed methods and the dataset used in the study are described in detail. Section 4 presents the results obtained in the study. The discussion and conclusion section of the paper is presented in Section 5, where a comparison with the findings of similar studies in the literature is made.

2. Literature Review

Cao et al. suggested a DL model with three major components. First, a transducer structure was incorporated into a convolutional neural network (CNN) to effectively utilize both local and long-range information. Second, disease details were collected from multiple images before self-attention was applied to improve inter-image interactions and reduce overfitting. Finally, an attention-based image transformation approach was proposed to filter information from different stages of the feature maps and adaptively capture lesion-related details. Their experiments produced a multi-class accuracy of 85.96% on the APTOS dataset and a binary class accuracy of 95.33% on the Messidor dataset, exceeding current approaches [16].

The aim of the study by Chandra et al. in 2024 was to create a CNN-based model for detecting and categorizing diabetic retinopathy using the APTOS dataset. The APTOS dataset was a large, open-access collection of fundus images that ophthalmologists analyzed for the likelihood and severity of DR. The accuracy obtained through the CNN model and the AlexNet model was 97% and 93% in quintuple classification, respectively [17].

Ohri et al. employed vision transformer-based deep learning models to classify DR diseases. The vision transformer-based DL architecture can classify fundus images into one of the DR categories by pre-training unlabeled fundus images before undertaking supervised training on labeled data. This study used DINO[VIT], MAE[VIT], and MSN[VIT] transducer-based DL architectures. The best MAE[VIT] transducer-based design achieved kappa values of 0.8341 in the low data regime and 0.9027 in the full data regime on 50 EyePACS training data using the Masked Autocoder framework [18].

Oulhadj et al. used four datasets, including EyePACS, DDR, Messidor-2, and APTOS datasets, for their study. As a preprocessing step on fundus images, they applied contrast-limiting adaptive histogram equalization and power law transformation approaches to classify DR illness using a modified capsule network-based and a fine-tuned vision transducer hybrid DL method. On the EyePACS, DDR, Messidor-2, and APTOS datasets, their efforts produced flawless test accuracy scores of 78.64%, 80.36%, 87.78%, and 88.18%, respectively [19].

In 2023, Mondal et al. combined the updated ResNet and DenseNet101 models, an enhanced version of the ResNet model for DR classification, to propose an automated ensemble DL model for better feature extraction. Two datasets, APTOS19 and DIARETDB1, were used in experiments for binary and multi-class classification. They preprocessed the images using the CLAHE approach for histogram equalization. For data augmentation, they used a GAN-based boosting strategy because of the dataset's severe class imbalance. The accuracy of the suggested approach is 86.08% for multi-class classification and 96.98% for binary classification [20].

Vijayan et al. proposed an early detection approach for automatic detection of DR that uses regression rather than multiclass classification using the convolution-based EfficientNetB0 model, one of the most advanced TL models. Better generalization was the initial advantage of the regression problem approach, and finer-grained predictions were made possible by the model's ability to give a value that falls between conventionally distinct labels. The APTOS and DDR datasets were used to test the idea. The DDR dataset had an accuracy of 84.80, whereas the APTOS dataset had an accuracy of 86.20 [21].

By merging the pyramid hierarchy of the discrete wavelet transform of the retinal fundus image with a modified capsule network and a proposed modified starting block, Oulhadj et al. discovered a novel deep hybrid model for determining the severity level of DR. Their proposed method's performance was assessed using the APTOS dataset, yielding training accuracy ratings of 97.71% and test accuracy scores of 86.54% [22].

In another study, Oulhadj et al. proposed comparing ensemble voting and five cutting-edge TL CNN models (ResNet50, DenseNet121, VGG16, InceptionV3, and Xception) for the automatic assessment of the severity of diabetic retinal disease. The results of five cutting-edge models were used by the community voting to make its choice. Using the APTOS dataset for training and testing, the suggested effort produced an accuracy of 83.63 [23].

A novel two-step convolutional DL-based approach for automatic DR detection was presented by Oulhadj et al. In the first step, known as preprocessing, the background influence was eliminated from the classification process by applying the deformable registration that covers the entire image to the retina. To identify the stage of DR, they trained four cutting-edge TL CNN models (ResNet50, InceptionV3, Xception, and DenseNet121) in the second step. The APTOS 2019 dataset was used to evaluate the proposed architecture's performance, and their model's accuracy was 85.28% [24].

Islam et al. used the APTOS 2019 dataset to automatically predict DR severity from fundus images. The state-of-the-art TL CNN model Xception was chosen as the encoder, and CLAHE was used to enhance image quality. For the binary classification of DR, the suggested model produced an AUC score of 98.50% and a test accuracy of 98.36%; for the multi-class classification using the APTOS 2019 dataset, the AUC score was 93.819% and the test accuracy was 84.364% [25].

3. Materials and Methods

This section provides detailed information about the dataset, methods, and evaluation criteria used in the study.

3.1. Dataset

APTOS2019 [11] was used in this study. Tables 1 and 2 indicate the quantity of data used for binary and multi-class classification. There are 3662 data points in this dataset. In binary and multi-class classification, 72% of the total data was used for training, 20% for testing, and 8% for validation. As a result of employing Albumentations library data augmentation methods, there are 1300 data points in each class, for a total of 6500 data points.

Table 1. Data Amounts for Binary Classification in Training, Testing, and Validation

	Train %72	Test %20	Validation %8	Total
0-NPDR	1300	361	144	1805
1-DR	1336	372	149	1857
Total	2636	733	293	3662

Table 2. Data Amounts for Multi-Class Classification in Training, Testing, and Validation

	Train (DataAugmented)	Train %72	Test %20	Validation %8	Total
0-NPDR	1300	1300	361	144	1805
1-MildNPDR	1300	266	74	30	370
2-ModerateNPDR	1300	719	200	80	999
3-SevereNPDR	1300	139	39	15	193
4-ProliferateDR	1300	212	59	24	295
Total	-	2636	733	293	3662
Total(DataAugmented)	6500	-	-	-	-

3.2. Method

In this study, VGG16, MobileNetV2, ResNet152V2, InceptionResNetV2, and EfficientNetB0 state-of-the-art DL models, which are frequently preferred for image classification in the field of healthcare, are used.

The Visual Geometry Group proposed the VGG16 CNN architecture. Figure 4 shows the VGG16 network architecture, which consists of 13 convolutional layers, three fully connected layers, and five pooling layers [26]. The step size is two, while the kernel size in the pooling layers is 2 × 2. In Step 1, the convolution kernel size is set to 3×3 in the convolutional layers. The rectified linear unit (ReLU) is the activation function for convolutional layers. The VGG-16 network receives an image with dimensions of 224 by 224 pixels and three channels. In the initial portion, two convolutional layers are followed by a pooling layer. Each of these convolutional layers has 64 cores and measures 224 x 224 pixels. Each of these convolutional layers has 64 cores with 224 x 224 pixels.

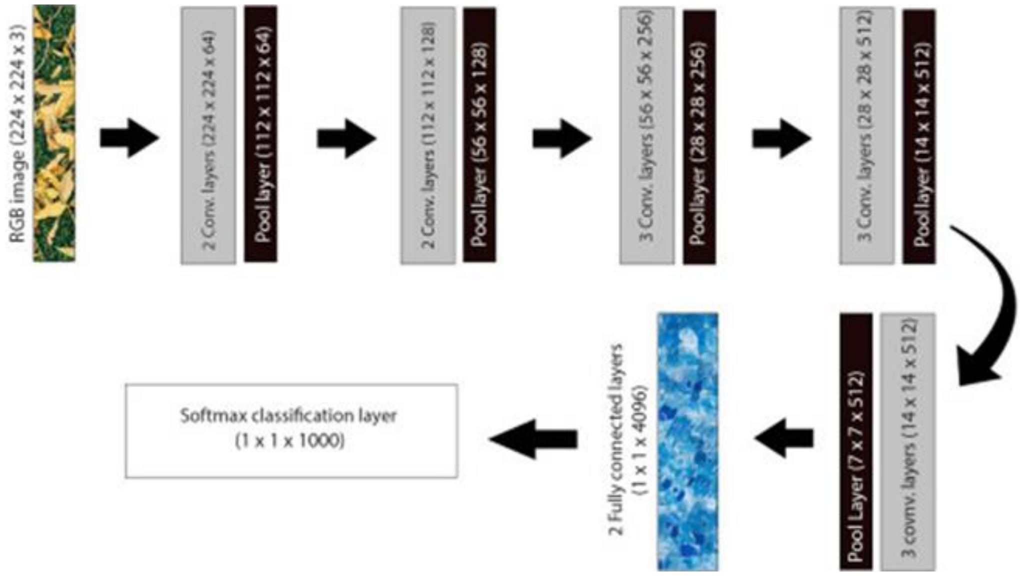


Figure 4. VGG16 Architecture Illustration [27]

The MobileNetV2 architecture shown in Figure 5 is a CNN architecture designed to work effectively on mobile devices. It is based on the inverted residual structure and enhanced with bottleneck features. This architecture improves the performance of mobile models, enabling more efficient results with lower computational requirements [28].

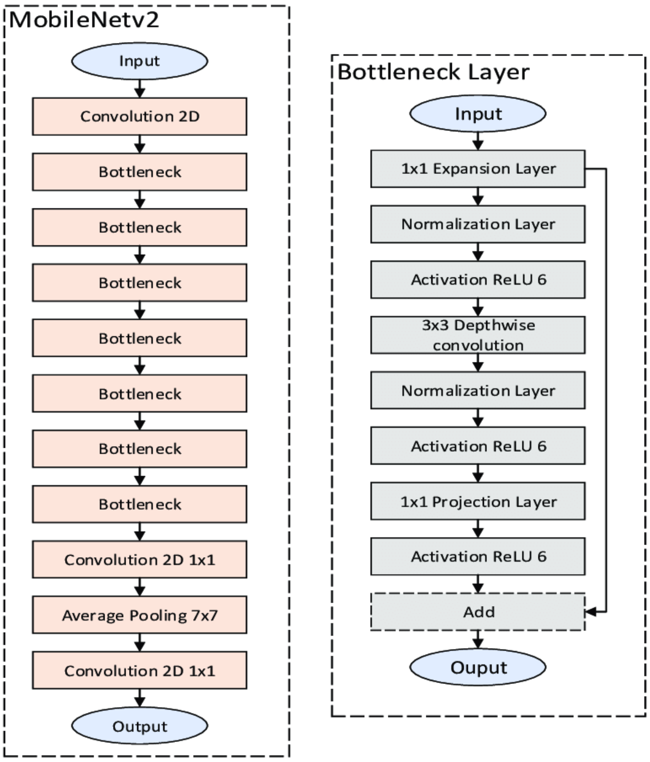


Figure 5. MobilNetV2 Architecture Illustration [29]

The ResNet152V2 architecture shown in Figure 6 is a CNN architecture used in DL. It is known for providing high accuracy, especially in image recognition tasks. The architecture includes 152 levels, improves on prior ResNet architectures, and supports the training of deeper layers utilizing residual connections [30].

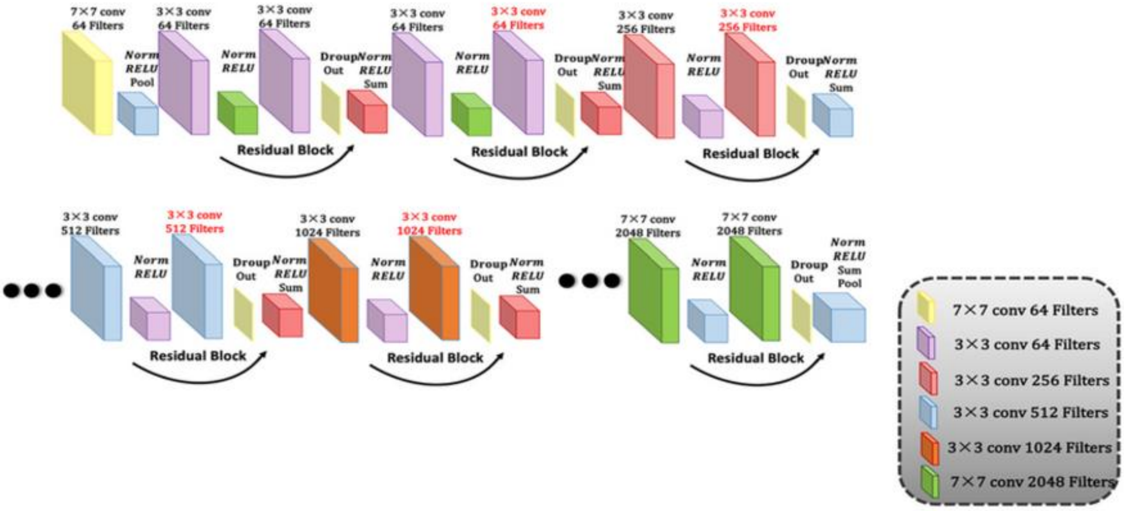


Figure 6. ResNet152V2 Architecture Illustration [31]

Figure 7 shows the Inception-ResNet-v2 architecture, which is based on the Inception architectural family. It's a CNN that combines residual connections. Training with residual connections considerably accelerates the training of initial networks [32].

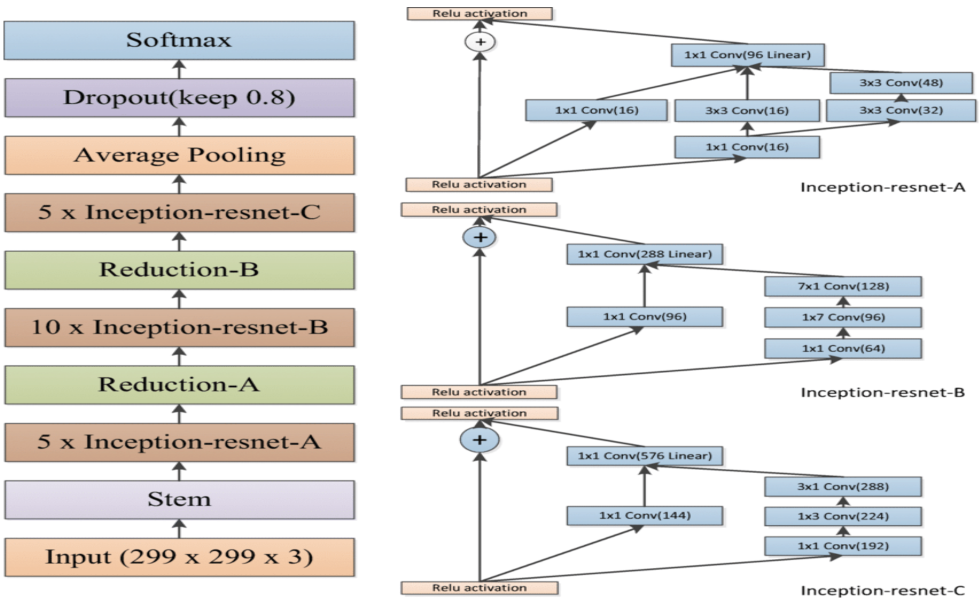


Figure 7. InceptionResNetV2 architecture [33]

The EfficientNetB0 architecture, shown in Figure 8, has a balanced design. It improves computational performance by extracting significant information from the input. EfficientNet uses a composite scaling approach that scales its layers equally in depth, width, and resolution [34].

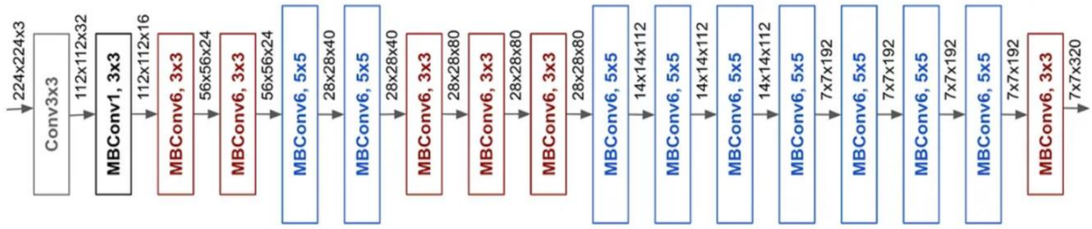


Figure 8. EfficientNetB0 architecture [35]

All models were trained by transfer learning (TL). As the initial weights of the model parameters, the weights resulting from the training on the ImageNet dataset were used. The stochastic gradient descent (SGD) method was used as the optimizer. Categorical cross-entropy was used as the loss function. The learning rate value was set as 0.00005. To reduce overfitting, the ‘patience’ parameter in EarlyStopping was set to 10, and training was stopped if there was no improvement in the verification loss for 10 epochs. Although the epochs were set to 1000 due to early stopping, the EfficientNetB0 model completed the binary and multi-class classification in 620 steps at most.

For data augmentation, the images were rotated horizontally and symmetrically using methods from the albumentations library. Random brightness and contrast transformation was applied in the range [-1, 1]. Random gamma correction was performed on the images. Grid distortion was applied to the image. Optical distortion in the range [-2, 2] and shift in the range [- 0.5, 0.5] were applied to the image with 50% probability. Shifting, scaling, and rotation operations were combined in the image. The images were resized to 224×224 dimensions.

3.3. Evaluation Criteria

In this study, binary and multi-class classifications were made, and the performances of the models used were evaluated using different metrics such as accuracy, precision, recall, and F1 score. These metrics given in Equations 1-4 are calculated using values such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) obtained in the confusion matrix.

$$Accuracy = \frac{(Tp + Tn)}{(Tp + Tn + Fp + Fn)} \tag{1}$$

$$Precision = \frac{Tp}{(Tp + Fp)} \tag{2}$$

$$Recall = \frac{Tp}{(Tp + Tn)} \tag{3}$$

$$F1 = 2 * \frac{precision * recall}{(precision + recall)} \tag{4}$$

4. Findings

4.1 Binary Classification Findings

The obtained results for the binary classification were listed in Table 3. The confusion matrices displaying the binary classification results were given in Figure 9.

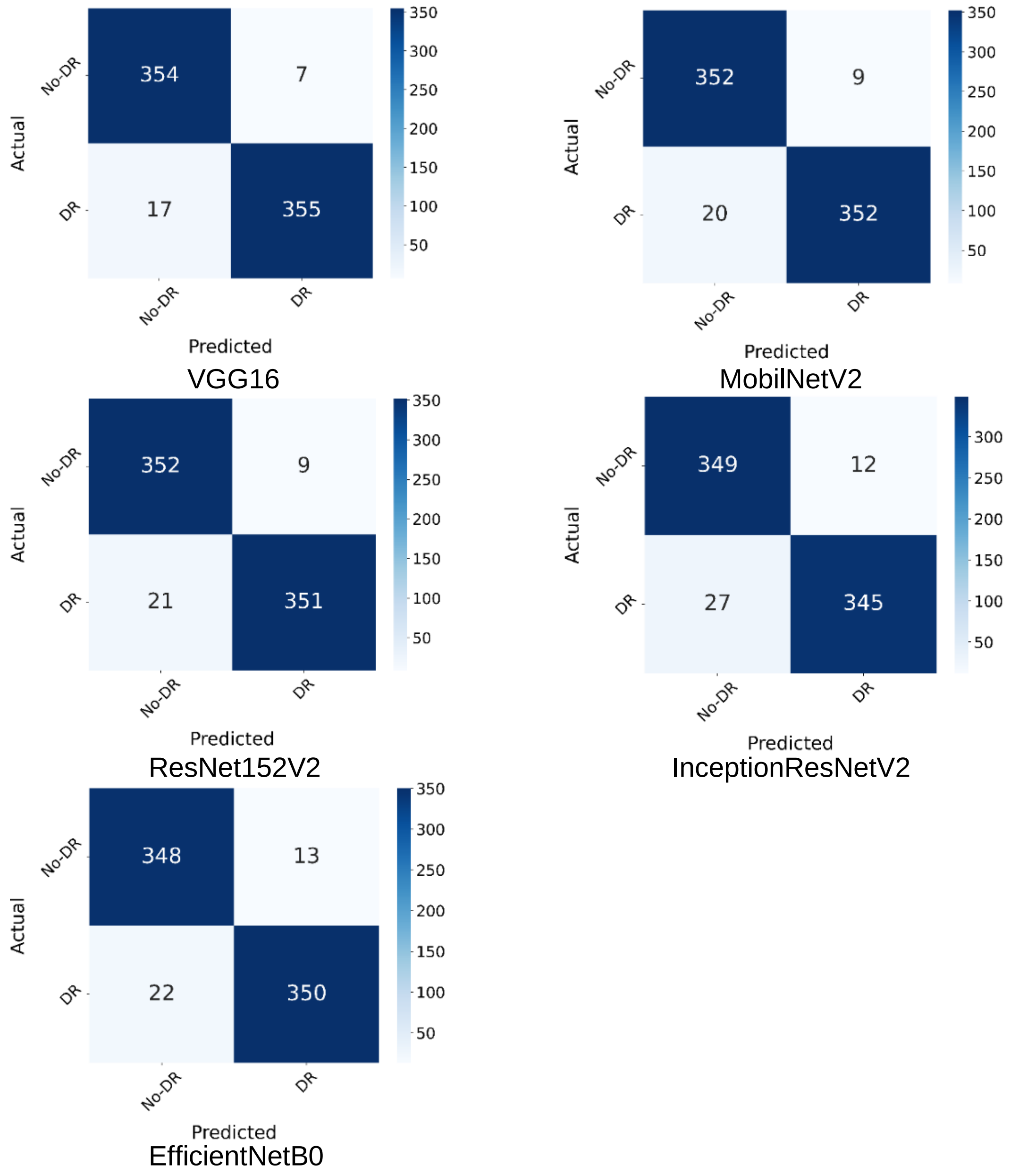


Figure 9. Confusion Matrices of Binary Classification

Model	Accuracy	Precision	Recall	F1-score	Support
VGG16	**0.97**	**0.97**	**0.97**	**0.97**	**733**
MobileNetV2	0.96	0.96	0.96	0.96	733
ResNet152V2	0.96	0.96	0.96	0.96	733
InceptionResNetV2	0.95	0.95	0.95	0.95	733
EfficientNetB0	0.95	0.95	0.95	0.95	733

Table 3. Comparison of Test Metrics for State-of-the-Art Models in Binary Classification

The best results in binary classification were obtained by the VGG16 model with an accuracy of 0.97. In the VGG16 model, according to the confusion matrix in Figure 9, 709 images were correctly classified, and 24 images were wrongly classified in the binary classification in the test data. In the No-DR class, 354 images were correctly classified and 7 incorrectly classified, while the DR class had 355 correctly classified images and 17 incorrectly classified images. The VGG16 model was followed by the MobileNetV2 and ResNet152V2 models, with an accuracy of 0.96, while the InceptionResNetV2 and EfficientNetB0 models came last with an accuracy of 0.95.

4.2 Multi-class classification findings (without data augmentation)

The multi-class classification results obtained for the InceptionResNetV2 model without data augmentation were presented in Table 4. The confusion matrix displaying the multi-class classification results of the InceptionResNetV2 model was given in Figure 10.

Table 4. Performance Metrics for Multi-Class Classification Using InceptionResNetV2 Model Without Data Augmentation

Model	Accuracy	Precision	Recall	F1-Score	Support
InceptionResNetV2 Without Data Augmentation	0.73	0.68	0.73	0.67	733

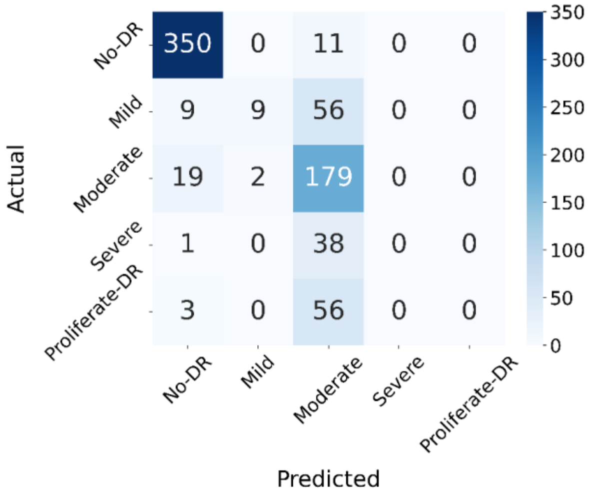


Figure 10. Confusion Matrix for Multi-Class Classification of InceptionResNetV2 Without Data Augmentation

Without data augmentation, the InceptionResNetV2 model yielded multi-class accuracy and sensitivity scores of 0.73, precision of 0.68, and an F1 score of 0.67. In the test data, 538 photos were correctly classified and 195 were wrongly classified, as shown by the confusion matrix in Figure 10. The network was unable to learn sufficiently due to the lack of classification in the Severe and Proliferate DR classes; hence, Albumentations library methods were used to enrich the data.

4.3 Multi-class classification findings (with data augmentation)

The obtained results for the multi-class classification were listed in Table 5. The confusion matrices displaying the multi-class classification results were given in Figure 11.

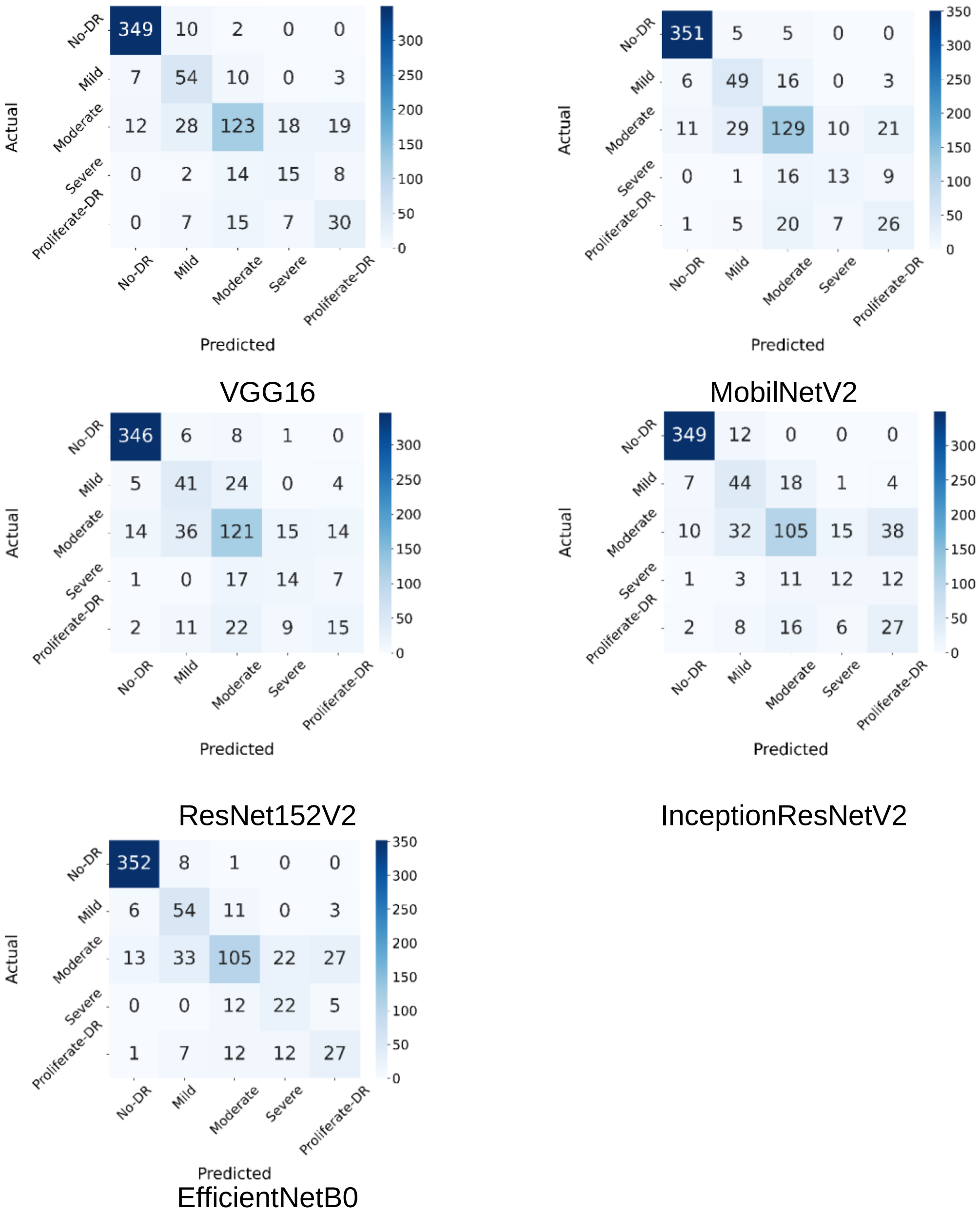


Figure 11. Confusion Matrices for Multi-Class Classification

Table 5. Comparison of Performance Metrics for State-of-the-Art Models in Multi-Class Classification

Model	Accuracy	Precision	Recall	F1-Score	Support
VGG16	**0.78**	**0.79**	**0.78**	**0.78**	**733**
MobileNetV2	0.77	0.77	0.77	0.77	733
ResNet152V2	0.73	0.73	0.73	0.73	733
InceptionResNetV2	0.73	0.75	0.73	0.73	733
EfficientNetB0	0.76	0.78	0.76	0.76	733

In multi-class classification, the VGG16 model had the best accuracy value with 0.78. As can be seen from the confusion matrix in Figure 11, the VGG16 model correctly classified 571 images and incorrectly classified 162 images on the test dataset in multi-class classification. There were 349 correct and 12 incorrect images classified in the No-DR class, 54 correct and 20 incorrect images in the Mild class, 123 correct and 77 incorrect images in the Moderate class, 15 correct and 24 incorrect images in the Severe class, and 30 correct and 29 incorrect images in the Proliferate-DR class. In terms of accuracy rankings, MobileNetV2 ranks second with a value of 0.77 on the data-augmented dataset, followed by EfficientNetB0 in third place with an accuracy value of 0.76, and the ResNet152V2 and InceptionResNetV2 models in last place with an accuracy value of 0.73.

5. Discussion and Conclusion

Studies on diabetic retinopathy using the APTOS2019 dataset are displayed in Table 6. In the best model, VGG16, our suggested method for binary classification had a 97% success rate in the accuracy metric. Our method produced results that were comparable to the accuracy metric success rates of Mondal et al. (96.98%) and Kumar et al. (96.24%). Our proposed approach in multi-class classification achieved 78% success in the accuracy metric with the best model, VGG16, and remained below the success achieved in other studies. In the literature, binary and multi-class classification studies have not been conducted on ResNet152V2 and MobileNetV2 models, which are state-of-the-art models. We contributed to the literature by performing binary and multi-class classification studies on ResNet152V2 and MobileNetV2 models.

Table 6. Previous Studies on Diabetic Retinopathy Using the APTOS2019 Dataset

Authors	**Classes**	**Method/Data Set(Number of Data)**	**Approach/Algorithm**	**Metric(%)**
Proposed approach(2024)	2 5	CNN/ APTOS(3662)	VGG16**,** InceptionResNetV2, ResNet152V2, EfficientNetB0, MobileNetV2	**Accuracy,Precision,Recall, F1-Score**:**97.00**(VGG16 Binary Classification)- **Accuracy,Recall,F1-Score:78.00,Precision:79.00** (VGG16 Data Agumentation Multi Class Classificaiton)
Cao X. Et al.(2024)[16]	5	CNN,Vit / APTOS(3662)	CNN,Vit	Accuracy:85,96
Mondal et al. (2023) [36]	2 5	CNN/ APTOS(3662)	Ensemble Deep-Learning Technique(DenseNet101 ve ResNeXt)	Accuracy: 96.98-86.08 Precision: 97.00-76.00 Recall: 97.00-82.00
Oulhadj et al. (2023) [37]	5	APTOS(3662)	ViT+ CapsNet+ PLT+CLAHE	Accuracy : 88.18 Precision: 80.00 Recall : 76.00 F1-score : 78.00 Kappa score: 81.55
Vijayan et al. (2023) [21]	5	CNN/ APTOS(3662)	Efficientnet-B0	Accuracy:86.20
Oulhadj et al. (2023) [22]	5	CNN/ APTOS(3662)	CapsNet + Inception Block + DWT	Accuracy : 86.54 Kappa score : 78.77 Precision: 76.00 Recall : 70.00 F1-score: 73.00
Oulhadj et al. (2023) [23]	5	CNN/ APTOS(3662)	Transferred Learning + Voting((Xception, InceptionV3, VGG16, DenseNet121, Resnet50))	Accuracy:83.63

M. Oulhadj et al (2022) [24]	5	CNN/APTOS(3662)	Ensemble Voting(Densenet-121, Xception, Inception-v3, Resnet-50)	Accuracy:85.28 Precision:80.00 Recall:70.00 F1-Score:73.00
Islam et al. (2022)[25]	2 5	CNN/ APTOS(3662)	Supervised Contrastive Learning(Xception)	Accuracy: 98.36-84.36 Precision: 98.37-73.84 Recall: 98.36-70.51 F1-Score: 98.37-70.49 AUC: 98.50-93.82
Bodapati et al. (2022) [38]	5	Attention based CNN/ APTOS(3662)	Stacked Convolutional Auto-Encoder(VGG16, Inception, ResNet Version2 (IRV2) , Xception)	Accuracy:84.17
Zhao et al. (2022)[39]	5	CNN/ APTOS(3662)	CoTXNet	Accuracy:84.18 Kappa score:90.00
Shaik and Cherukuri (2022)[40]	5	CNN/ APTOS(3662)	Hinge Attention Networks	Accuracy:85.54
Hu et al. (2022) [41]	2 5	CNN/ APTOS(3662)	Graph Adversarial	Accuracy: 94.30-83.50
Fan et al. (2021) [42]	5	CNN/ APTOS(3662)	Multi-Scale Features (MobileNetV3)	Accuracy:85.32 Kappa score: 77.26 F1-Score:85.30 AUC:97.00
Sugeno et al. (2021) [43]	5	CNN/APTOS(3662)	Transfer Learning + EfficientNet-B3	Accuracy:84.20
Al-Antary and Arafa (2021) [44]	2 5	CNN/ APTOS(3662)	Extraction of Features + Attention	Accuracy: 98.10- 84.60 Kappa score: -89.60 AUC: 98.20 -
Kumar et al. (2021) [45]	2 5	CNN/ APTOS(3662)	VGG16+Capsule Network+Hybrid Deep Learning, DRISTI	Accuracy: 96.24-82.06

The binary classification scenario, in which the dataset's images were categorized as sick and the healthiest (npdr), yielded the best test results. Without data augmentation, the results of the multi-class classification trials, which also attempted to predict the disease's intensity, were poor. The models were biased toward classes with more examples in the learning phase since the dataset used to classify the disease by level was an unbalanced dataset. Test findings showed satisfactory success in the tests where training data sets were equalized using data augmentation approaches. In a multi-class classification, the data augmentation strategy has been demonstrated to be advantageous when treating diabetic retinal disease.

Transformer-based architectures (Vision Transformer, VitDet, Swin Transformer, etc.) can be used in future research to do classification studies for the early diagnosis of diabetic retinopathy disease. Convolution and transformation-based DL models can be used to evaluate fundus images of diabetic retinopathy patients in the hospitals, and research can be conducted to assist the ophthalmologists in making early diagnoses.

Acknowledgement

This paper is a comprehensively expanded version of the proceedings abstract presented at the IV. International Congress on Artificial Intelligence in Healthcare.

References

[1] Coşansu, G. (2015). Diyabet: Küresel bir salgın hastalık. *Okmeydanı Tıp Dergisi*, 31, 1-6.

[2] World Health Organization. (2021). Diabetes. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/diabetes

[3] International Diabetes Federation. (2021). IDF Diabetes Atlas | Tenth Edition. Retrieved from https://www.idf.org/news/240:diabetes-now-affects-one-in-10-adults-worldwide.html and https://diabetesatlas.org/atlas/tenth-edition/

[4] Grosman, B., Ilany, J., Roy, A., Kurtz, N., Wu, D., Parikh, N., Voskanyan, G., Konvalina, N., Mylonas, C., Gottlieb, R., Kaufman, F., & Cohen, O. (2016). Hybrid closed-loop insulin delivery in type 1 diabetes during supervised outpatient conditions. Journal of Diabetes Science and Technology, 10(3), 708–713. https://doi.org/10.1177/1932296816631568

[5] Swapna, G., Vinayakumar, R., & Soman, K. P. (2018). Diabetes detection using deep learning algorithms. ICT Express, 4(4), 243–246. https://doi.org/10.1016/j.icte.2018.10.005

[6] Atwany, M. Z., Sahyoun, A. H., & Yaqub, M. (2022). Deep learning techniques for diabetic retinopathy classification: A survey. IEEE Access, 10, 28642-28655. https://doi.org/10.1109/ACCESS.2022.3157632

[7] Ağca, K. (2022). Evrişimsel sinir ağları kullanarak diyabetik retinopati hastalığının tespiti [Yüksek lisans tezi, Sivas Cumhuriyet Üniversitesi]. YÖK Ulusal Tez Merkezi. https://tez.yok.gov.tr

[8] İnan, S. (2014). Diabetik retinopati ve etiyopatogenezi. Kocatepe Tıp Dergisi, 15(2), 207-217.

[9] Cunha, J. P. (2021). What are the stages of diabetic retinopathy? eMedicineHealth. Retrieved from https://www.emedicinehealth.com/what_are_the_stages_of_diabetic_retinopathy/article_em.htm

[10] Yakar, H. K. (2018). Yaşlılıkta Diyabetin Diğer Bir Yüzü: Diyabetik Retinopati Ve Düşmeler. Izmir Democracy University Health Sciences Journal, 1(2), 13-22.

[11] APTOS. (2019). Blindness detection. Kaggle. Retrieved from https://www.kaggle.com/c/aptos2019-blindness-detection/overview/aptos-2019

[12] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

[13] Lee, C. S., Baughman, D. M., & Lee, A. Y. (2017). Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. Ophthalmology Retina, 1(4), 322–327. https://doi.org/10.1016/j.oret.2016.12.009

[14] Chen, X., Xu, Y., Wong, D. W. K., Wong, T. Y., & Liu, J. (2015, August). Glaucoma detection based on deep convolutional neural network. In 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 715-718). IEEE.

[15] Uçar, M. (2021). Glokom Hastalığının Evrişimli Sinir Ağı Mimarileri ile Tespiti. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 23(68), 521-529.

[16] Cao, X., Lin, J., Gao, X., & Li, Z. (2024). Integrating convolution and transformer for enhanced diabetic retinopathy detection. International Journal of Bio-Inspired Computation (IJBIC), 23(4). https://doi.org/10.1504/IJBIC.2024.139257

[17] Chandra, R., Tiwari, S., Kumar, S. S., & Agarwal, S. (2024). Diabetic retinopathy prediction based on CNN and AlexNet model. In 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 382–387). IEEE. https://doi.org/10.1109/Confluence60223.2024.10463351

[18] Ohri, K., Kumar, M., & Sukheja, D. (2023). Self-supervised approach for diabetic retinopathy severity detection using vision transformer. Pattern Analysis and Applications. https://doi.org/10.1007/s13748-024-00325-0

[19] Oulhadj, M., Riffi, J., Khodriss, C., Mahraz, A. M., Yahyaouy, A., Abdellaoui, M., Andaloussi, I. B., & Tairi, H. (2024). Diabetic retinopathy prediction based on vision transformer and modified capsule network. Computers in Biology and Medicine, 175, 108523. https://doi.org/10.1016/j.compbiomed.2024.108523

[20] Mondal, S. S., Mandal, N., Singh, K. K., Singh, A., & Izonin, I. (2023). EDLDR: An ensemble deep learning technique for detection and classification of diabetic retinopathy. Diagnostics, 13(1), 124. https://doi.org/10.3390/diagnostics13010124

[21] Vijayan, M., & Venkatakrishnan, S. (2023). A regression-based approach to diabetic retinopathy diagnosis using EfficientNet. Diagnostics, 13(4), 774. https://doi.org/10.3390/diagnostics13040774.

[22] Oulhadj, M., Riffi, J., Khodriss, C., Mahraz, A. M., Bennis, A., Yahyaouy, A., ... & Tairi, H. (2023). Diabetic retinopathy prediction based on wavelet decomposition and modified capsule network. Journal of Digital Imaging, 36(4), 1739-1751. https://doi.org/10.1007/s10278-023-00813-0

[23] Oulhadj, M., Riffi, J., Khodriss, C., Mahraz, A. M., Bennis, A., Yahyaouy, A., ... & Tairi, H. (2023, January). Diabetic Retinopathy Prediction Based on Transfer Learning and Ensemble Voting. In International Conference on Digital Technologies and Applications (pp. 929-937). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-29857-8_92

[24] Oulhadj, M., Riffi, J., Chaimae, K., Mahraz, A. M., Ahmed, B., Yahyaouy, A., ... & Tairi, H. (2022). Diabetic retinopathy prediction based on deep learning and deformable registration. Multimedia Tools and Applications, 81(20), 28709-28727. https://doi.org/10.1007/s11042-022-12968-z

[25] Islam, M. R., Abdulrazak, L. F., Nahiduzzaman, M., Goni, M. O. F., Anower, M. S., Ahsan, M., ... & Kowalski, M. (2022). Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images. Computers in biology and medicine, 146, 105602. https://doi.org/10.1016/j.compbiomed.2022.105602

[26] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In 3rd International Conference on Learning Representations (ICLR 2015) (pp. 1–14).

[27] Sharma, P., Nayak, D. R., Balabantaray, B. K., Tanveer, M., & Nayak, R. (2024). A survey on cancer detection via convolutional neural networks: Current challenges and future directions. Neural Networks, 169, 637-659. https://doi.org/10.1016/j.neunet.2023.11.006.

[28] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[29] Tragoudaras, A., Stoikos, P., Fanaras, K., Tziouvaras, A., Floros, G., Dimitriou, G., ... & Stamoulis, G. (2022). Design space exploration of a sparse mobilenetv2 using high-level synthesis and sparse matrix techniques on FPGAs. Sensors, 22(12), 4318. https://doi.org/10.3390/s22124318

[30] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[31] Vatanpour, M., & Haddadnia, J. (2024). Brain tumour segmentation of MR images based on custom attention mechanism with transfer-learning. IET Image Processing, 18(4), 886-896. https://doi.org/10.1049/ipr2.12992

[32] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1). https://doi.org/10.1609/aaai.v31i1.11231

[33] Peng, C., Liu, Y., Yuan, X., & Chen, Q. (2022). Research of image recognition method based on enhanced inception-ResNet-V2. Multimedia Tools and Applications, 81(24), 34345-34365. https://doi.org/10.1007/s11042-022-12387-0

[34] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR. http://proceedings.mlr.press/v97/tan19a.html

[35] Montalbo, F. J. P., & Alon, A. S. (2021). Empirical analysis of a fine-tuned deep convolutional model in classifying and detecting malaria parasites from blood smears. KSII Transactions on Internet and Information Systems (TIIS), 15(1), 147-165. https://doi.org/10.3837/tiis.2021.01.009

[36] Mondal, S. S., Mandal, N., Singh, K. K., Singh, A., & Izonin, I. (2022). Edldr: An ensemble deep learning technique for detection and classification of diabetic retinopathy. Diagnostics, 13(1), 124. https://doi.org/10.3390/diagnostics13010124

[37] Oulhadj, M., Riffi, J., Khodriss, C., Mahraz, A. M., Yahyaouy, A., Abdellaoui, M., ... & Tairi, H. (2024). Diabetic retinopathy prediction based on vision transformer and modified capsule network. Computers in Biology and Medicine, 175, 108523. https://doi.org/10.1016/j.compbiomed.2024.108523.

[38] Bodapati, J. D. (2022). Stacked convolutional auto-encoder representations with spatial attention for efficient diabetic retinopathy diagnosis. Multimedia Tools and Applications, 81(22), 32033-32056. https://doi.org/10.1007/s11042-022-12811-5

[39] Zhao, S., Wu, Y., Tong, M., Yao, Y., Qian, W., & Qi, S. (2022). Cot-xnet: contextual transformer with xception network for diabetic retinopathy grading. Physics in Medicine & Biology, 67(24), 245003. https://doi.org/10.1088/1361-6560/ac9fa0

[40] Shaik, N. S., & Cherukuri, T. K. (2022). Hinge attention network: A joint model for diabetic retinopathy severity grading. Applied Intelligence, 52(13), 15105-15121. https://doi.org/10.1007/s10489-021-03043-5

[41] Hu, J., Wang, H., Wang, L., & Lu, Y. (2022). Graph adversarial transfer learning for diabetic retinopathy classification. IEEE Access, 10, 119071-119083. https://doi.org/10.1109/ACCESS.2022.3220776

[42] Fan, R., Liu, Y., & Zhang, R. (2021). Multi-scale feature fusion with adaptive weighting for diabetic retinopathy severity classification. Electronics, 10(12), 1369.

[43] Sugeno, A., Ishikawa, Y., Ohshima, T., & Muramatsu, R. (2021). Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning. Computers in Biology and Medicine, 137, 104795.

[44] Al-Antary, M. T., & Arafa, Y. (2021). Multi-scale attention network for diabetic retinopathy classification. IEEE Access, 9, 54190-54200.

[45] Kumar, G., Chatterjee, S., & Chattopadhyay, C. (2021). DRISTI: a hybrid deep neural network for diabetic retinopathy diagnosis. Signal, Image and Video Processing, 15(8), 1679-1686.

List of abbreviations

CNN : Convolutional Neural Network
DL : Deep Learning
DR : Diabetic Retinopathy
NPDR : Non-Proliferative Diabetic Retinopathy
PDR : Proliferative Diabetic Retinopathy
TL : Transfer Learning