KSÜ Mühendislik Bilimleri Dergisi, 28(2), 2025
Araştırma Makalesi

KSU J Eng Sci, 28(2), 2025
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 24.07.2024
Kabul Tarihi :23.01.2025

Received Date : 24.07.2024
Accepted Date : 23.01.2025

COMPARATIVE ANALYSIS OF VISION TRANSFORMERS AND CONVOLUTIONAL NEURAL NETWORKS IN DIABETIC RETINOPATHY DIAGNOSIS

GÖRÜNTÜ TRANSFORMATÖRLERİ VE EVRİŞİMLİ SİNİR AĞLARININ DİYABETİK RETİNOPATİ TEŞHİSİNDE KARŞILAŞTIRMALI ANALİZİ

Esra YÜZGEÇ ÖZDEMİR[1] (ORCID: 0000-0003-2914-2603)
Canan KOÇ[1] (ORCID: 0000-0002-2651-9471)
Fatih ÖZYURT[1](ORCID: 0000-0002-8154-6691)

[1]Firat University, Engineering Faculty, Software Engineering

*Sorumlu Yazar / Corresponding Author: Esra YÜZGEÇ ÖZDEMİR, eyuzgec@firat.edu.tr

ABSTRACT

Diabetic retinopathy can lead to significant visual complications and significantly affects individuals' quality of life. This study focuses on comparing the performance of Vision Transformer (ViT) models and Convolutional Neural Networks (CNN) methods in diabetic retinopathy diagnosis and aims to evaluate their potential as an alternative to traditional diagnostic methods. In this study, the performance of four different ViT model architectures and four different convolutional neural network (CNN) models in training and testing phases were comparatively analyzed. ViT models achieved accuracy rates of 97.83%, 98.41%, 95.2%, and 98.26% for "tiny," "base," "small," and "large," respectively. Additionally, models trained with VGG13, ResNet18, ResNet50, and SqueezeNet architectures from CNN techniques achieved accuracy rates of 96.1%, 97.83%, 90.9%, and 93.93%, respectively. ViT architectures achieved higher accuracy rates than CNN architectures. When the results were evaluated, it was concluded that ViT methods were more successful in the diagnosis of diabetic retinopathy.

Keywords: Diabetic retinopathy, vision transformers, deep learning, convolutional neural networks

ÖZET

Diyabetik retinopati, önemli görsel komplikasyonlara yol açabilen ve bireylerin yaşam kalitesini önemli ölçüde etkileyen bir hastalıktır. Bu çalışma, diyabetik retinopatinin erken evrelerde teşhis edilmesinin önemini vurgulamakta, mevcut teşhis yöntemlerinin sınırlılıklarına dikkat çekmekte ve geleneksel yöntemlere alternatif olarak Görüntü Dönüştürücüsü (ViT) modellerinin potansiyelini ele almaktadır. Bu çalışmada, dört farklı ViT model mimarisinin yanı sıra döt farklı evrişimli sinir ağı (CNN) modellerinin eğitim ve test aşamalarındaki performansları karşılaştırmalı olarak analiz edilmiştir. ViT modelleri 'tiny', 'base', 'small' ve 'large' sırasıyla %97,83, %98,41, %95,2 ve %98,26 doğruluk oranlarına ulaşmıştır. Ayrıca CNN tekniklerinden VGG13, ResNet18, ResNet50 ve SqueezeNet mimarileri ile eğitilen modeller sırasıyla %96,1, %97,83, %90,9 ve %93,93 doğruluk oranlarına ulaşmıştır. Çalışma sonucunda ViT mimarileri CNN mimarilerine göre daha yüksek doğruluk oranlarına ulaşmıştır. Sonuçlar değerlendirildiğinde ViT yöntemlerinin diyabetik retinopati teşhisinde daha başarılı olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Diyabetik retinopati, görüntü dönüştürücüleri, derin öğrenme, evrişimli sinir ağları

ToCite: YÜZGEÇ ÖZDEMİR, E., KOÇ, C., & ÖZYURT, F., (2025). COMPARATIVE ANALYSIS OF VISION TRANSFORMERS AND CONVOLUTIONAL NEURAL NETWORKS IN DIABETIC RETINOPATHY DIAGNOSIS. Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi, 28(2), 592-600.

INTRODUCTION

Diabetic retinopathy is a serious health condition, particularly in people with diabetes, that causes damage to the thin blood vessels behind the eye's retina. This condition occurs when high blood sugar (glucose) levels caused by diabetes damage the microvessels in the inner part of the eye. Diabetes is a chronic metabolic disease characterized by the body's inability to properly use insulin or produce sufficient amounts of insulin, while insulin is a vital hormone that enables the absorption of blood sugar by cells (Uğurlu et.al., 2018) Thus, dysfunction or deficiency of insulin results in persistently high blood sugar levels. These persistently high blood sugar levels can lead to many complications in the body and cause serious damage, especially to the eyes. The retina is a layer at the back of the eye that detects the light that allows us to see. Chronic high glycemia caused by diabetes can disrupt the fine vasculature of the retina, causing vessels to leak, block, or bleed. This damage builds up over the duration of diabetes and can severely impair the function of the retina, leading to vision loss and even blindness.

Diabetic retinopathy occurs in people with diabetes and, without intervention, can lead to visual impairment and even blindness in advanced stages. Diagnosis in the early stages of the disease can be difficult because the symptoms are minimal and nonspecific (Özçelik & Altan, 2021). If the disease remains undiagnosed, abnormalities such as decreased vision, visual field deterioration, floaters, and lightning flashes can be seen in the later stages. At the same time, impaired color perception, difficulties in night vision, increased pressure inside the eye, and pain are prominent symptoms that may occur in the later stages. The doctor diagnoses diabetic retinopathy through comprehensive eye examinations and performs this procedure by evaluating changes in the retina (Rahmanlar et al., 2019). When necessary, laser treatment, intraocular drug infections, and surgery are the treatment methods for the disease diagnosed at an early stage. The aim of these methods is to slow or stop the progression of retinal damage. The main goal is to keep the quality of the patient's vision as high as possible. Experts recommend that people with diabetes have regular eye examinations.

The innovative studies and developments in artificial intelligence and deep learning technologies, especially in the field of healthcare, have led to significant advances in the diagnosis and treatment of diabetic retinopathy. Sunkari et al. developed a model for automatic classification of diabetic retinopathy (Sunkari et al., 2024). In this model, the evaluation was made with the patient's real-time data set. They achieved a 93.51% success rate using ResNet18 and Swish. In another study, In another study, Özçelik and Altan proposed a robust AI-based model utilizing chaotic swarm intelligence optimization and RNN-LSTM architecture to classify diabetic retinopathy into five classes, achieving a success rate of 98.7% and demonstrating the model's ability to handle nonlinear dynamics in fundus images (Özçelik & Altan, 2023). Karthika et al. used a diabetic retinopathy dataset consisting of five classes and performed classification with a model named SE-ResCA-GTNet. In addition to the developed model, they achieved a 99.8% success rate using the Gazelle Optimization algorithm (Karthika & Durgadevi, 2024). These studies reveal the achievements in this field. In another classification study, Fang et al. developed a new DAG network model to classify diabetic retinopathy data (Fang & Qiao, 2022). In this study, three important data features were extracted and combined for classification. The study used the DIARETDB1 dataset and real data from Dalian NO.3 People's Hospital. As a result of the study, 98.7% and 98.5% success rates were obtained. Patil et al. developed an automatic diagnosis method for diabetic retinopathy using transfer learning methods (Patil et al., 2023). ResNet-50 DL model was trained using the APTOS dataset and tested with the EyePACS dataset. Both preprocessing and augmentation were used in the study. Thus, the problem of data imbalance was avoided. As a result, a 98.5% success rate was achieved.

In addition to all these studies, especially ViTs, have led to important steps in the field of eye health (Chen et al., 2021; Wang et al., 2022). In these studies, high-resolution eye scans are analyzed to prevent problems that may occur in human eye work. ViTs and similar image-processing technologies in healthcare offer up-to-date solutions in many studies, from disease classification to treatment planning. The use of these technologies in diagnosing diabetic retinopathy facilitates accurate diagnosis at an early stage and contributes to the preservation of patients' vision with timely intervention.

Huang et al. developed a transducer-based framework for automating retinopathy diagnosis (Huang et al., 2023). In this study, the success of the transducers was demonstrated by improving classification accuracy to 85.49% and report generation performance to a BlEU-1 score of 0.422. Although not much work has been done in ViT in diabetic retinopathy, there have been studies on medical images. Manzari et al. proposed a CNN-Transducer hybrid model for medical image classification using MedMNIST-2D datasets in another study (Manzari et al., 2023). This proposed

model achieved an AUC improvement of 2.3% and 1.1% over AutoML methods on RetinaMNIST and TissueMNIST, respectively. This approach showed high generalizability and outperformed the existing best methods. Lian et al. proposed a model for detecting and classifying white blood cells. The model was innovated by combining YOLO and ViT technologies (Lian & Li, 2024). In addition to the accuracy rate achieved with the model, at the end of the study, 96.449% was obtained from two types of images: white, blood cells and nuclei, and 16 cell classes.

Chintamreddy and Seshasayee conducted a study highlighting the importance of fundus photographs for automatic classification of diabetic retinopathy (DR) severity. In this study, a model is developed that can identify various severity levels of DR. The dataset utilized in the study was sourced from the APTOS-2019 blindness detection dataset and underwent preprocessing through image processing techniques. Furthermore, the Conv-ViT model was taken as a basis and a hybrid structure was developed to increase the effectiveness of this model. This model utilized the feature extraction of Inception-V3, ResNet-50, and Vision Transformer (ViT) models and outperformed other methods in the literature with an accuracy of 93.75% (Chintamreddy & Seshasayee, 2024).

Wu et. al. conducted a study focusing on the application of Vision Transformer (ViT) in the recognition of diabetic retinopathy (DR) severity. In this study, a model that can classify five different severity levels of DR was designed. Fundus images are segmented to be used as input to the ViT model and these segments are processed for classification by adding positional information. The study reported that the 384_L_32 ViT model achieved high performance metrics such as 91.4% accuracy, 92.6% sensitivity, 97.7% specificity and 0.988 AUC. It was reported that this model outperformed traditional Convolutional Neural Network (CNN) based approaches and emphasized the importance of the attention mechanism in DR classification tasks (Wu et al., 2021).

ViT models offer significant advantages compared to traditional Convolutional Neural Networks (CNN) architectures, particularly due to their ability to model long-term dependencies. Despite their ability to recognize local features, CNNs have limited capabilities in modeling global relationships. ViT models, on the other hand, divide images into small patches, process these patches as a sequence, and learn the global relationships of these patches using the transformer mechanism. *The vit model receives image patches, the patch size is fixed for all image parts, it does not vary.* In cases such as diabetic retinopathy, the analysis of retina images requires accurately modeling both local features and structural connections. The global attention mechanism of ViT enables a more in-depth analysis of retinal images, thereby achieving higher accuracy rates. For complex medical imaging problems like diabetic retinopathy, ViT is more suitable due to its parametric flexibility and strong overall performance on large datasets. This study demonstrates that ViT models are a good option for medical image analysis due to their high accuracy rates compared to CNNs.

This study emphasizes diabetic retinopathy using deep learning techniques. In particular, eye images obtained using image processing technologies and Vit methods were used to classify diabetic retinopathy diseases. A two-class process was used to diagnose this disease, and discrimination between images was provided. The "Materials and Methods" section of the study includes the applied dataset, ViT, and detailed descriptions of the deep learning methods. The " Findings and Discussion" section includes the conclusions and discussions of the findings. Finally, the "Conclusion" section summarizes the results obtained and highlights the importance of the study's classification of diabetic retinopathy. Considering that early diagnosis of diabetic retinopathy is essential, this study makes an important contribution to the field.

MATERIALS AND METHODS

Dataset

Diabetic retinopathy is an essential disease to consider for people with diabetes. Early diagnosis and treatment can prevent vision loss and improve outcomes. Moreover, the automatic classification of retinal images for diabetic retinopathy screening has a significant role in this field, as the interpretation of retinal images by the human eye is prone to some errors. Most of the work is done manually, which can lead to inefficiency and inconsistencies in diagnosis. Therefore, there is a need to develop a reliable and robust automatic classifier.

The dataset used in this study is a two-class dataset shared by Parisa Karimi Darabi on the Kaggle platform (Darabi, n.d.). This dataset has two classes, Diabetic Retinopathy and Healthy class, as shown in Figure 1. In the Train dataset, there is a balanced data distribution with 1050 Diabetic Retinopathy and 1026 healthy.

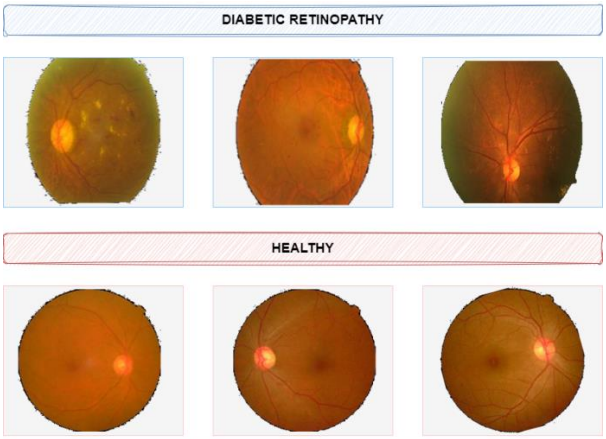

Figure 1. Dataset sample images

This study was conducted using training and test folders from these folders. The dataset consists of a large collection of high-resolution retinal images taken under various imaging conditions. A medical professional assessed the presence of Diabetic Retinopathy in each image, and categories were assigned a rating on a scale from 0 to 1. The obtained images were reduced to 224x224 size and subjected to various pre-processing such as flipping, rotating and zooming in order to increase model success during model training.

Table 1. Numerical Distribution of Data in Dataset

Class Name	Number of Training Images	Number of Test Images
Diabetic Retinopathy	1050	113
Healthy	1026	118

The study applied an extensive hyperparameter tuning to train ViT models. The AdamW optimization algorithm was chosen and the learning rate was 0.001. The models were trained for 100 epochs and a batch size of 16 was used. For each model, images were resized and normalized to 224x224 pixels.
ViT models were initialized using weights pre-trained on the ImageNet-1k dataset. This pre-training helped the models to learn general visual features and helped to achieve better results on limited medical data. Data augmentation techniques such as randomization, rotation, contrast enhancement and zooming were used during the training process.

Vision Transformers (ViTs)
It is based on the fact that for the first time, in 2021, the Image-Based Transformers model was put forward and immediately became widely used, bringing astounding results (Dosovitskiy et al., 2020). In 2022, the model's implementation of ViT showed improved performance compared to the outdated traditional (Zhou et al., 2021). Since then, the disciplines have made a great leap forward, and now they have moved beyond the traditional processing of images since it has been broken. In other words, a vision transformer that rivals CNNs has also gradually been adopted as the new standard.

The specified model ViT has been significantly successful in image classification studies in the last two years and the 2022 ImageNet competition (Wu et al., 2023; Beyer et al., 2022). This competition declared ViT the best image output processing model, proving it can become the leading choice in this field. This result demonstrates that the ViT model is at a remarkable point in image processing and provides many benefits to users with this project. The ViT architecture is an image processing network that uses sequential transformation of images and no convolutional layers. In the first processing step, the images in the dataset are also divided into pieces of different sizes. After this step, low-dimensional linear embeddings are created from these image fragments. A state-of-the-art transform encoder is used for the input sequence. The basic structure of the ViT method is as shown in Figure 2 (Alhawas & Tüfekçi, 2022).

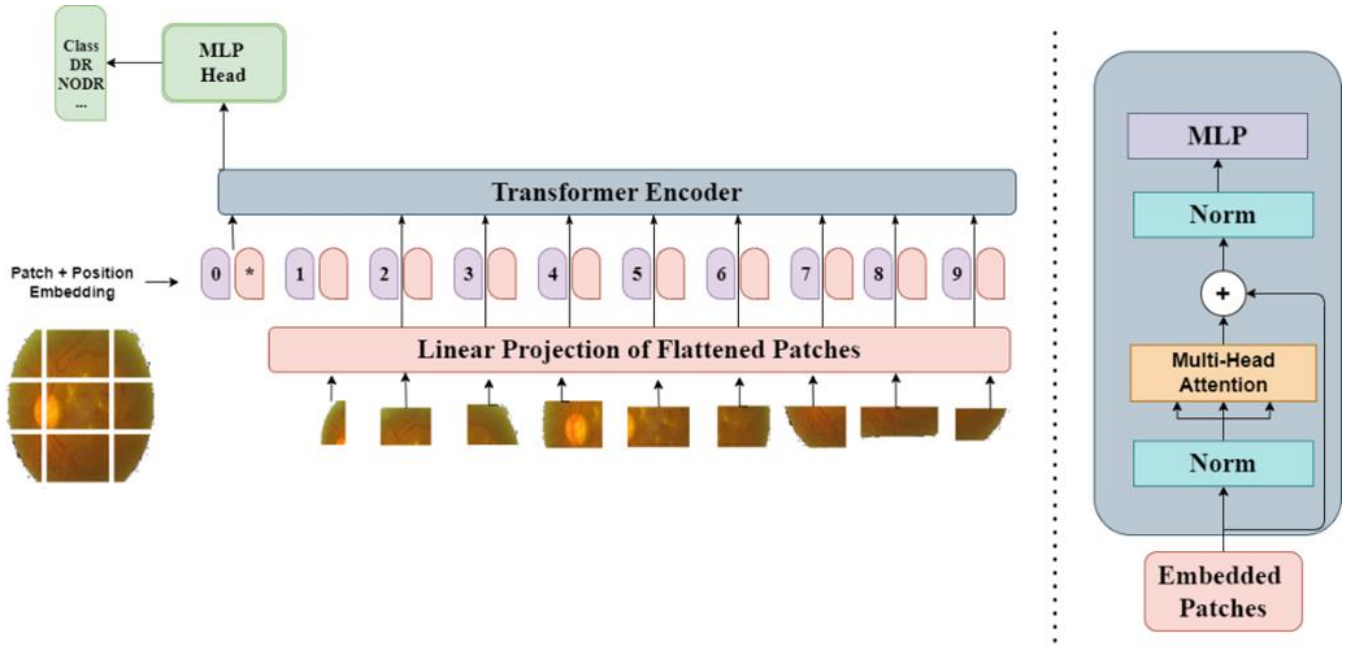

Figure 2. Architectural structure of ViT method

In this study, four different Vision Transformer (ViT) models, namely Tiny, Small, Base and Large, were used for diabetic retinopathy classification. The Tiny ViT model consists of 12 transformer encoder layers, each with 6 attention heads and 192 hidden sizes, and has a total of approximately 5.7 million parameters. The Small ViT model consists of 12 transformer encoder layers, 12 attention heads, 384 hidden sizes and 22 million parameters. Similarly, the Base ViT model consists of 12 transformer encoder layers, 12 attention heads, 768 hidden dimensions and has about 86 million parameters. The Large ViT model is the most complex model, with 24 transformer encoder layers, 16 attention heads, 1024 hidden dimensions and 307 million parameters in total. All of these models have the features given in Table 2 and accept input images resized to 224x224 pixels. These configurations are based on the ViT framework and adapted to the specific requirements of diabetic retinopathy classification (Touvron et al., 2021).

Table 2. Comparison of architectural content of ViT models

Model	Layers	Attention Heads	Hidden Size	Parameters	Input Size
Tiny	12	6	192	5.7M	224x224
Small	12	12	384	22M	224x224
Base	12	12	768	86M	224x224
Large	24	16	1024	307M	224x224

Furthermore, the dataset used in this study contains two folders: training and test. The training data is used to learn the model parameters, while the test data is kept completely independent from the training process as it is in a separate folder. This makes it suitable for evaluating the generalization ability of the model. This separation is important to avoid data leakage and to ensure the reliability of the test results. Keeping the test data set independent allowed an objective measurement of the model's performance on unknown data.

Furthermore, various data augmentation techniques were applied during the training process to increase the diversity of the data set and to support the generalization ability of the models. These techniques included randomization, rotation, contrast adjustments and scaling. This approach allowed the model to be sensitive not only to the training data but also to the variety of possible real-world data.

Findings And Discussion

In this study, the diagnosis and classification of Diabetic Retinopathy have been conducted using Vision Transformer (ViT) models at "tiny," "base," "small," and "large" scales alongside deep learning architectures such as VGG13, ResNet18, ResNet50, and SqueezeNet. During the training phase of the models, the "tiny" model achieved an accuracy rate of 97.83%, the "base" model reached 98.55%, the "large" model attained 99.03%, and the "small" model secured 98.07%. For the deep learning models, during their training phase, the ResNet18 architecture reached

an accuracy rate of 98.07%, ResNet50 achieved 97.83%, SqueezeNet also reached 98.07%, and the VGG13 architecture attained an accuracy rate of 97.59%.

Besides these accuracy rates, confusion matrices representing the models' false predictions on the validation data are given in Figure 3. This detailed evaluation reveals the effectiveness of each model and architecture in accurately diagnosing and classifying Diabetic Retinopathy. It also highlights their strengths and improvement areas based on the performance measurements and error analyses performed.

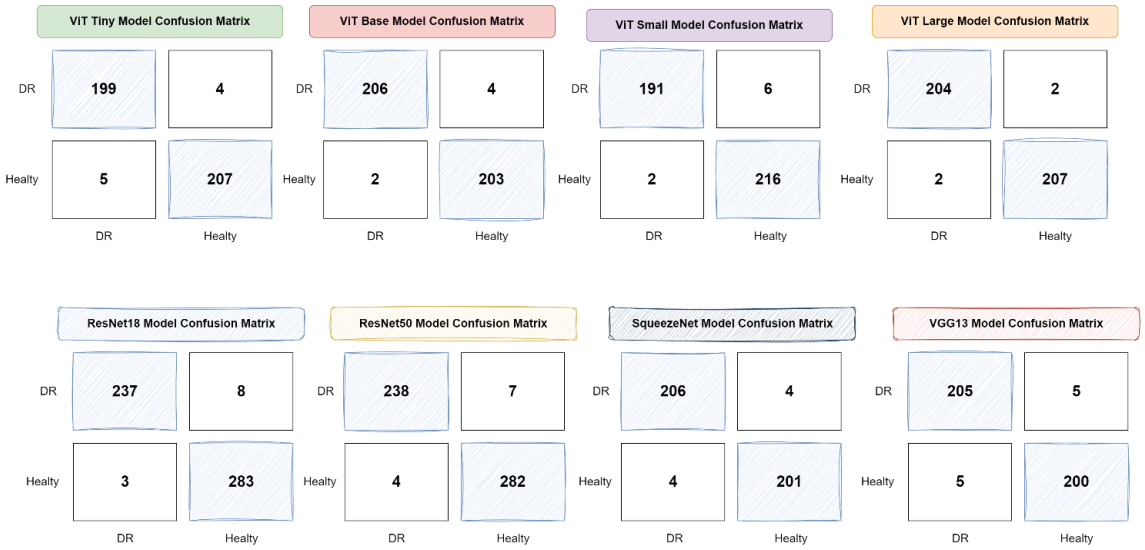

Figure 3. Confusion matrices obtained on validation data of ViT models

In addition, training and validation losses are obtained by estimating the validation data during training. These loss rates are visualized as loss graphs as in Figure 4 and when the loss graphs are examined; it is seen that the training and validation losses are reduced in a balanced manner and no over-fitting or under-fitting problems are encountered during model training.

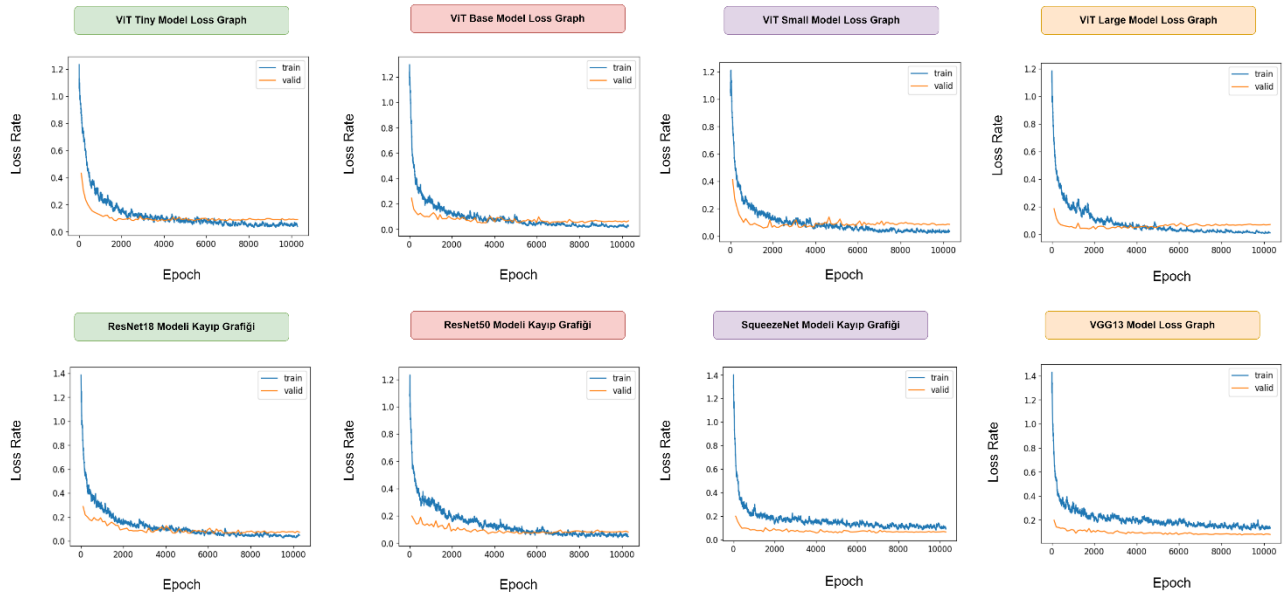

Figure 4. Loss graphs obtained from training ViT models

The confusion matrices and loss graphs given in Figure 3 and Figure 4 above were obtained from the estimations from the validation data used during training. These results were obtained by considering 10% of the data as validation data. The dataset also includes a test folder the model has not encountered before and no examples in the

training data. This test data was used for the classification success of the models after the model training. Test accuracies of 97.83%, 98.41%, 95.2%, and 98.26% were achieved for "tiny", "base", "small," and "large" ViT architectures, respectively. In addition, VGG13, ResNet18, ResNet50, SqueezeNet architectures achieved 96.1%, 97.83%, 90.9%, 93.93% test accuracy respectively. The complexity matrices for the prediction of these test data are given in Figure 5.

Figure 5. Confusion matrices obtained by ViT models on test data

The performance of the models used in our study was examined using evaluation metrics, including accuracy, precision, recall and F1-score (Powers, 2011). These metrics were used to assess both the overall accuracy of the models and the distribution of classification errors. Details of these metrics for each model are shown in Table 3 below.

Table 3. Comparison of Classification Performance Metrics for ViT and CNN Models

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ViT Tiny	97.83	98.03	97.55	97.79
ViT Base	98.55	98.10	98.07	98.56
ViT Small	98.07	96.95	98.96	97.95
ViT Large	99.03	99.03	99.04	99.03
ResNet18	98.07	96.73	97.25	97.73
ResNet50	97.83	98.58	97.14	97.74
SqueezeNet	98.07	98.05	98.10	98.10
VGG13	97.59	97.62	97.56	97.62

The values shown in the table indicate that the ViT Large model shows the best performance with an accuracy of 99.03%, while the ViT Base and ViT Small models show strong performance with accuracy rates above 98%. ViT models can be considered to perform better, especially in metrics such as sensitivity and F1-score. Although CNN-based models offer competitive results, they are limited compared to ViT's advantage of global relationship modeling due to their focus on learning local features. These results show that ViT models have great potential for medical imaging. It offers a viable alternative for complex imaging problems such as diabetic retinopathy.

CONCLUSION

Diabetic retinopathy is an eye disease that significantly affects the quality of life of patients. Diabetic retinopathy, which severely impairs vision, is of great importance for early diagnosis. If the disease can be detected and classified especially in its early stages, it is possible to stop its progression and develop effective treatment methods. This is of great importance for improving the patient's living standards and developing effective treatment methods.

This study investigates the classification of diabetic retinopathy using deep learning techniques. As an alternative to traditional diagnostic methods and CNN architectures, ViT models are reported to offer promising results in this field. In this study, we have done a comparative study of the ViT models and convolutional neural network models. The findings indicated that the ViT models, particularly the “base” model, achieved high performance when tested

on the allocated test datasets. With these results, it can be inferred that the ViT models, including the "base" model, have great possibility in the diagnosis of diabetic retinopathy. Further research based on our study can be performed. It is also predicted that the performance of the model will be more good if applied to distinct variations of ViT model and actual patient's unique datasets. In conclusion, the use of ViT models offers new and efficient ways to diagnose diabetic retinopathy and promises a promising future for the integration of this technology into clinical applications.

REFERENCES

Alhawas, N., & Tüfekçi, Z. (2022). The Identification of Red-Meat Types using The Fine-Tuned Vision Transformer and MobileNet Models. European Journal of Science and Technology. https://doi.org/10.31590/ejosat.1112892

Beyer, L., Zhai, X., & Kolesnikov, A. I. (2022). Better plain ViT baselines for ImageNet-1k. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2205.01580

Chen, J., He, Y., Frey, E. C., Li, Y., & Du, Y. (2021). VIT-V-Net: Vision Transformer for unsupervised Volumetric Medical Image Registration. arXiv.org. https://arxiv.org/abs/2104.06468

Chintamreddy, D., & Seshasayee, U. R. (2024, June). Detection of Diabetic Retinopathy (DR) Severity from Fundus Photographs using Conv-ViT. In 2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI) (pp. 1-6). IEEE.

Darabi, P. K. (n.d.). Competitions Contributor. Kaggle. https://www.kaggle.com/pkdarabi/competitions. Accessed [24.07.2024].

Dosovitskiy, A., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv.org. https://arxiv.org/abs/2010.11929

Fang, L., & Qiao, H. (2022). Diabetic retinopathy classification using a novel DAG network based on multi-feature of fundus images. Biomedical Signal Processing and Control, 77, 103810. https://doi.org/10.1016/j.bspc.2022.103810

Huang, Y.-H., et al. (2023). Model long-range dependencies for multi-modality and multi-view retinopathy diagnosis through transformers. Knowledge-Based Systems, 271, 110544. https://doi.org/10.1016/j.knosys.2023.110544

Karthika, S., & Durgadevi, M. (2024). Improved ResNet_101 assisted attentional global transformer network for automated detection and classification of diabetic retinopathy disease. Biomedical Signal Processing and Control, 88, 105674. https://doi.org/10.1016/j.bspc.2023.105674

Lian, J., & Li, T. (2024). Lesion identification in fundus images via convolutional neural network-vision transformer. Biomedical Signal Processing and Control, 88, 105607. https://doi.org/10.1016/j.bspc.2023.105607

Manzari, O. N., Ahmadabadi, H., Kashiani, H., Shokouhi, S. B., & Ayatollahi, A. (2023). MedViT: A robust vision transformer for generalized medical image classification. Computers in Biology and Medicine, 157, 106791. https://doi.org/10.1016/j.compbiomed.2023.106791

Özçelik, Y. B., & Altan, A. (2021). Diyabetik Retinopati Teşhisi için Fundus Görüntülerinin Derin Öğrenme Tabanlı Sınıflandırılması. European Journal of Science and Technology. December 2021. https://doi.org/10.31590/ejosat.1011806

Özçelik, Y. B., & Altan, A. (2023). Overcoming nonlinear dynamics in diabetic retinopathy classification: a robust AI-based model with chaotic swarm intelligence optimization and recurrent long short-term memory. Fractal and Fractional, 7(8), 598.

Patil, M. S., Chickerur, S., Abhimalya, C., Naik, A., Kumari, N., & Maurya, S. K. (2023). Effective deep learning data augmentation techniques for diabetic retinopathy classification. Procedia Computer Science, 218, 1156-1165. https://doi.org/10.1016/j.procs.2023.01.094

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

Rahmanlar, H., Atılgan, C. Ü., Çıtırık, M., Yaradilmiş, İ. M., & Gürsöz, H. (2019). Türkiye'de diyabetik retinopati tanısında endikasyon dışı ilaç kullanımı. Sakarya Medical Journal, 9(3), 499-505. https://doi.org/10.31832/smj.543998

Sunkari, S., et al. (2024). A refined ResNet18 architecture with Swish activation function for Diabetic Retinopathy

classification. Biomedical Signal Processing and Control, 88, 105630. https://doi.org/10.1016/j.bspc.2023.105630

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. arXiv preprint arXiv:2012.12877.

Uğurlu, N., Taşlıpınar, A. G., Yülek, F., Özdemir, D., Ersoy, R., & Çakır, B. (2018). Evaluation of Retinal Microvascular Structures in Type 1 Diabetic Patients without Diabetic Retinopathy. Ankara Medical Journal. December 2018. https://doi.org/10.17098/amj.501136

Wang, Z., Dong, N., & Voiculescu, I. (2022). Computationally-Efficient Vision transformer for medical image semantic segmentation via dual Pseudo-Label supervision. 2022 IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/icip46576.2022.9897482

Wu, J., Hu, R., Xiao, Z., Chen, J., & Liu, J. (2021). Vision Transformer-based recognition of diabetic retinopathy grade. Medical Physics, 48(12), 7850-7863.

Wu, K., et al. (2023). TinyCLIP: CLIP distillation via affinity mimicking and weight inheritance. arXiv.org. https://arxiv.org/abs/2309.12314

Zhou, D., et al. (2021). DeepVIT: Towards Deeper Vision Transformer. arXiv.org. https://arxiv.org/abs/2103.11886