A Vision Transformer Approach to Fundus Image Classification

Danilo Leite[1], José Camara[1], João Rodrigues[3], and António Cunha[1,2(✉)]

[1] University of Trás-os-Montes and Alto Douro, Vila Real, Portugal
{danilol,acunha}@utad.pt
[2] INESC TEC—Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal
[3] LARSyS & ISE, Universidade do Algarve, 8005-226 Faro, Portugal
jrodrig@ualg.pt

Abstract. Glaucoma is a condition that affects the optic nerve, with loss of retinal nerve fibers, increased excavation of the optic nerve, and a progressive decrease in the visual field. It is the leading cause of irreversible blindness in the world. Manual classification of glaucoma is a complex and time-consuming process that requires assessing a variety of ocular features by experienced clinicians. Automated detection can assist the specialist in early diagnosis and effective treatment of glaucoma and prevent vision loss. This study developed a deep learning model based on vision transformers, called ViT-BRSET, to detect patients with increased excavation of the optic nerve automatically. ViT-BRSET is a neural network architecture that is particularly effective for computer vision tasks. The results of this study were promising, with an accuracy of 0.94, an F1-score of 0.91, and a recall of 0.94. The model was trained on a new dataset called BRSET, which consists of 16,112 fundus images of patients with increased excavation of the optic nerve. The results of this study suggest that ViT-BRSET has the potential to improve early diagnosis through early detection of optic nerve excavation, one of the main signs of glaucomatous disease. ViT-BRSET can be used to mass-screen patients, identifying those who need further examination by a doctor.

Keywords: Fundus Image · Vision transformers · BRSET

1 Introduction

Glaucoma is one of the leading causes of irreversible or disabling blindness worldwide. The World Health Organization (WHO) estimates that in 2040 there will be 112 million disease cases [1]. The earlier glaucoma is diagnosed and treated, the greater the chances of preserving vision. Manual classification of glaucoma is a complex and time-consuming process that requires assessing several ocular features [2]. Screening is essential for early diagnosis and treatment of glaucoma, as early intervention can help prevent vision loss. The cup-to-disc ratio (CDR), a widely used measure by experts to detect glaucoma, calculates the proportion between the size of the cup and the size of the disc, as well as the area of the cup and the location of the disc [3].

A. Cunha et al. (Eds.): MobiHealth 2023, LNICST 578, pp. 151–162, 2024.
https://doi.org/10.1007/978-3-031-60665-6_11

Computerized diagnostic systems (CAD) based on modern deep learning models, such as Vision Transformers (ViT), can help classify glaucoma with greater accuracy and speed than manual classification. Recent work [4–7] explores the efficacy of modern transformer architectures based on ViT in the diagnosis of retinal diseases, showing promising results. As these CADs are developed and improved, they can revolutionize the diagnosis of glaucoma, making it more accurate, faster, and accessible. However, a large amount of data is required to train Deep Learning (DL) models. Several datasets are available to help train these models and make them more efficient. Recently, a new sizeable Brazilian dataset of fundus images was published, called the Brazilian Multilabel Ophthalmological Dataset (BRSET) [8]. In addition to the pictures, BRSET includes several clinical parameters, including increased optic nerve excavations, an essential factor in diagnosing glaucoma. In Fig. 1, we see a normal eye (A) and one eye with increased excavation (B). Images A1 and B1 show the optic nerve head. Visually, we can see that the cup in B2 is enlarged compared to A1.

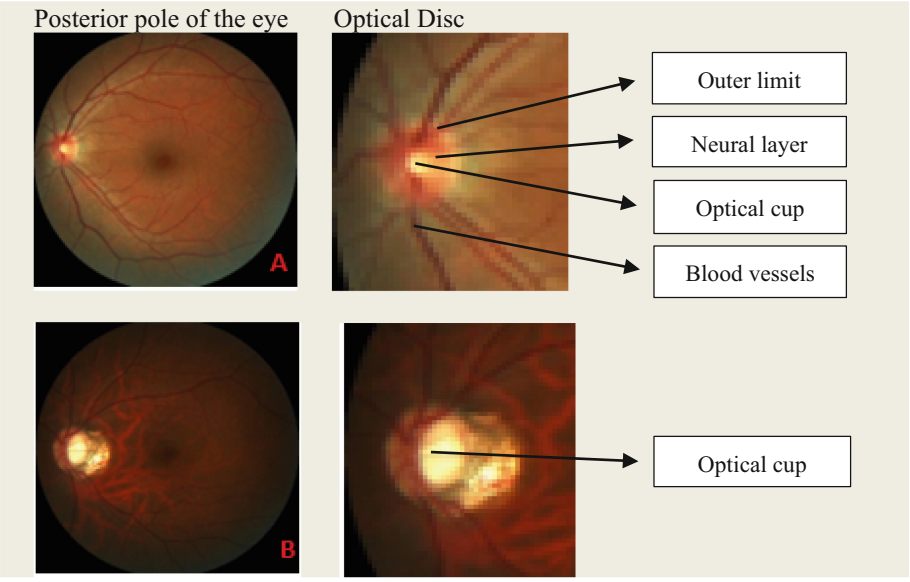

Fig. 1. Shows eye health (A) and an eye with increased excavation (B)

The optic nerve head, optic disc, or optic papilla is an oval-shaped structure, orange in color, located approximately 3 to 4 mm nasal to the center of the retina through which nerve cell fibers pass, called ganglion cell fibers, which form the nervous layer (more orange layer) and carry electrical stimuli created in the retina from visible light to the cerebral cortex, where they will be interpreted. In the central part of the optic nerve head is the emergence of retinal vessels, and the cup or excavation or optic cup is represented by a more yellowish color in the center of the optical disc due to the absence of ganglionic fibers. In the progression of the glaucomatous disease, we observe irregularities on the inner edge of the neural layer and increased excavation represented by the death of

ganglion fibers. In Fig. 1, we see a normal eye (A) with increased excavation (B). Images A1 and B1 show the optic nerve head. Visually, we can see that the cup in B2 is enlarged compared to A1.

Increased excavation in the optic nerve head is one of the main signs of the glaucomatous papilla; however, in diagnosing glaucoma, the change in the neural layer of the papilla must correspond to the difference in the visual field.

A diagnosis of glaucoma can only be made by a doctor. Therefore, this work built a model based on a promising DL approach called ViT [7]. It compared its preliminary results with other DL models in the literature for glaucoma detection. The model built was trained with the BRSET dataset. The results obtained in this study are promising and suggest that the ViT model can be a valuable tool for glaucoma detection. This study also seeks to contribute to the constant evolution of this critical ophthalmological condition's early and effective detection.

2 Literature Review

In the early stages of the disease, it can be difficult for experts to identify early changes in the optic nerve head. Recently, there have been significant advances in machine learning (ML) to diagnose glaucoma. A recent study [9] developed a convolutional neural network (CNN) for glaucoma detection. The proposed model achieved an accuracy of 0.96 with the LAG dataset and 0.82 with the RIM-ONE dataset. The algorithm uses an attention prediction subnetwork to create focused and cropped maps of fundus images, which are then used for glaucoma detection and measurement using statistical methods.

In another study [3], the authors evaluated models trained on significant public datasets to detect glaucoma in retinal images acquired by retinography and mobile devices classification methods produced model activation maps to support predictions. Segmentation methods evaluated the cup-to-disc ratio (CDR), a frequently used indicator in practice by experts to screen the optic nerve. The segmentation of the disc and cup achieved DICE 0.8 and IoU 0.7.

Another study [4] compared the DeiT (Data-efficient image Transformer) and ResNet-50 models trained on fundus photographs from the Ocular Hypertension Treatment Study (OHTS). The DeiT demonstrated performance similar to ResNet-50 on the OHTS test sets. The DeiT and ResNet-50 achieved AUROC 0.91 and 0.82, respectively. The authors highlight that image transformers can improve generalization and interpretability in ML models, detecting eye diseases and possibly other medical conditions that rely on images for clinical diagnosis and treatment.

Finally, a study [10] proposed a 13-layer CNN architecture. SoftMax and Support Vector Machine (SVM) classifiers were used to classify the images. The CNN accuracy with the SoftMax classifier was 0.93, while the CNN with an SVM classifier achieved an accuracy of 0.95. The dataset used for this investigation consisted of images collected from various public datasets and a private research centre.

The literature review in glaucoma diagnosis using ML has demonstrated significant advances, with ML being capable of detecting the disease with accuracy comparable to that of experts. However, there are still challenges to be overcome, such as the need for more robust datasets and the need to improve the interpretability of the models.

3 Datasets

3.1 Brazilian Multilabel Ophthalmological Dataset (BRSET)

The Brazilian Multilabel Ophthalmological Dataset (BRSET) [8] is a high-quality ophthalmological dataset of 16.266 images from 8.524 Brazilian patients. The images are in color and include photographs of the retinas of both eyes, along with demographic, anatomical, and clinical data. BRSET was designed to enhance the development of the scientific community and validate machine learning models. The demographic data includes age, gender, nationality, diabetes, duration of diabetes, and insulin usage. Anatomical parameters include data on the optic disc, vessels, and macula. Clinical parameters encompass diabetic retinopathy, macular edema, scars, nevi, vascular occlusion, hypertensive retinopathy, drusen, hemorrhages, retinal detachment, myopic fundus, and increased excavation.

BRSET is a valuable resource for researchers studying eye diseases. It allows machine learning models to predict demographic characteristics and classify multilabel diseases using fundus retina images.

In addition to BRSET, other publicly available datasets can be used to train machine learning models for glaucoma detection. These datasets are HRF, Drishti-GS1, RIM-ONE, sjchoi86-HRF, and ACRIMA. Table 1 provides an overview of the characteristics of these datasets [2].

Table 1. Public databases for glaucoma.

Data base	Glaucoma	Normal	Total
HRF	27	18	45
Drishti-GS1	70	31	101
RIM-UM	194	261	455
sjchoi86-HRF	101	300	401
ACRIMA	396	309	705

It is important to note that these datasets exhibit heterogeneity in their characteristics, such as lighting, field of view, and resolution. These variations can affect the performance of models trained on different datasets.

4 Fundamentals of Deep Learning

Deep learning (DL) is a subfield of machine learning that relies on the analysis of data through the representation of successive layers, inspired by the functioning of the human brain. Each layer can filter specific properties and highlight relevant features, with significant applications in medical diagnostic problems. This enables the learning of complex representations and the decomposition of these representations into intermediate spaces represented by the intermediate layers.

Deep learning has exhibited significant potential for application in the medical field, enhancing the precision of image processing and the detection pertinent diagnostic features in various examinations, including X-rays, computed tomography, ultrasounds, histological analyses of organs and tissues, as well as the scrutiny of photographic images. This methodology enables the discernment of intricate elements within extensive datasets by employing multiple intermediary layers between the input and output. Each layer is adept at refining the input signal to suit the subsequent layer, thereby unveiling progressively abstract insights.

For these methods to be successful, it is essential to have sufficient data for training and evaluation of the system. In addition, the validation of these methods requires a reference standard that can be used for comparison, which emphasizes the importance of public retinal databases that meet well-defined requirements [6, 11]. Within deep learning, convolutional neural networks (CNNs) are the most widely used architecture for image classification in computer vision.

5 Transformers

Convolutional neural networks (CNNs) have demonstrated remarkable superiority in various visual tasks over the past decade. They are particularly effective at capturing spatial features in images and are invariant to translations. CNNs have consistently proven their performance on a variety of benchmark metrics. However, they have some limitations related to how the models operate and learn from images due to the restriction on the receptive field, which results in a localized understanding of images [4, 7]. To address these limitations in visual tasks, so-called vision transformers have emerged. These architectures were developed to overcome the shortcomings of CNNs, particularly the need for a global understanding of images. After demonstrating their effectiveness in natural language processing tasks, vision transformers have also gained prominence in computer vision applications [12, 13].

In the study of[12], Dosovitskiy et al. adopt an approach of dividing the image into patches and applying self-attention calculations on each patch. When trained with sufficient data, the results achieved surpassed those obtained by CNNs regarding accuracy and computational efficiency. In addition, vision transformers can outperform CNNs in various tasks by applying multiple training techniques. Transfer learning or modifying the transformer architecture can further improve the performance of vision transformers. A notable limitation of CNNs is their restriction to the receptive field, which represents the area of the image processed by a single convolutional layer. This can make it difficult to understand complex images that contain information from different parts of the image.

6 Methodology

The dataset used was BRSET, which contains 16,112 retinal images from 8,524 Brazilian patients. The images were classified as positive for glaucoma (3,181) or negative (12,931), with an average age of 57.09 $\pm$ 18.1 years. The workflow developed is presented in Fig. 2, detailing the three main stages executed throughout this study: data collection and processing, model training and construction, and performance evaluation.

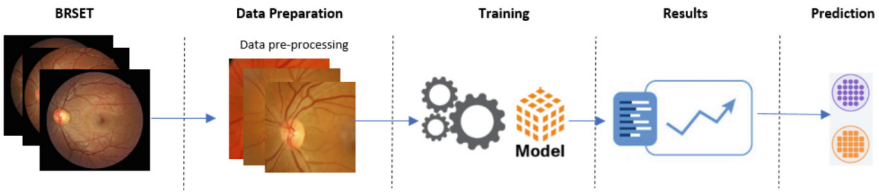

Fig. 2. The model pipeline for glaucoma screening

6.1 Data Preparation and Augmentation

Data preprocessing and data augmentation are essential steps in developing machine learning models. These processes help ensure that the data is in the correct format and that the model is trained on a representative dataset.

Data Preprocessing
This work used preprocessing to prepare images for training an object classification model. Each image were preprocessed as follows [14]:

- The region of interest (ROI) was extracted from the image to ensure that the model is trained only on the parts of the idea that are relevant to the classification task (Fig. 3).
- The image pixels were normalized to have a mean of 0 and a standard deviation of 1. This helps to ensure that the data are on the same scale and that the model is not biased towards any particular color channel.
- The color channels in the image were standardized to have the same range of values, which helps to ensure that the data are on the same scale and that the model is not biased towards any color channel.
- The noise in the image was reduced using a Gaussian filter, which helped to improve the image quality and increase the model's accuracy.
- The image was resized to a resolution of 224 $\times$ 224 pixels due to hardware limitations, which helps to ensure that the model can be run on hardware with memory or processing constraints.

To extract the ROI, we apply image thresholding to the grayscale image. This converts the image to a binary image, where bright areas are white and dark regions are black. The optic disc appears as a white area in the binary image. Finally, we extract a sub image from the original colored image that contains the optic disc. This sub image is input to a machine learning model to classify the image as glaucomatous or non-glaucomatous. Figure 2 shows a summary of this step.

Data Augmentation
After preprocessing, the data were divided into 3 folders: training, validation, and test.

- The model was trained on a set of images, and its performance was evaluated on a separate set of images. This helps to ensure that the model is balanced with the training data and that it can generalize to new data.

Data augmentation was used on the training dataset to mitigate overfitting and address potential data imbalances. Data augmentation is a powerful technique used to expand

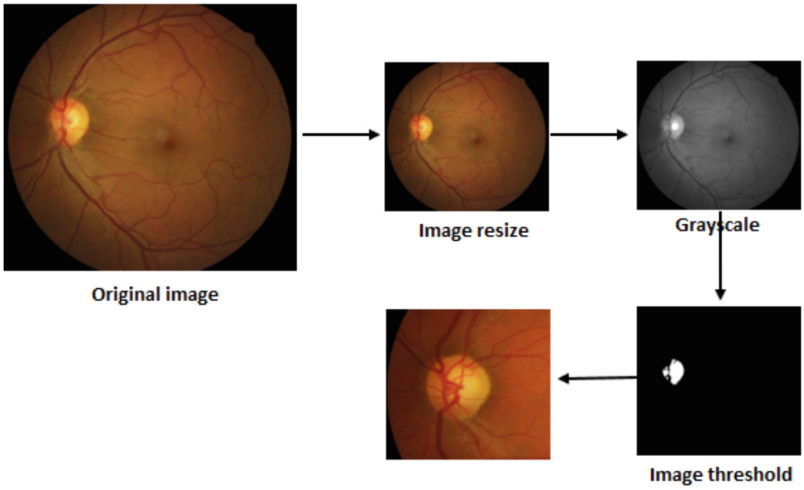

Fig. 3. Preprocessing steps

the training dataset's size and diversity, thereby enhancing the model's accuracy and its ability to generalize effectively.

6.2 Training

The dataset was randomly divided into three subsets in the following proportions: 70% for training, 15% for validation, and 15% for testing. This division of the data is essential to ensure that the model is trained on a representative dataset and evaluated on an independent dataset [15]. This division of the data into three subsets helps to ensure that the model is trained and assessed relatively. The training subset is used to train the model, and the validation and test subsets are used to evaluate the model's performance. To ensure that the class distribution was representative of the real-world data, stratified cross-validation with 10 folds was used. Stratified cross-validation is a cross-validation method that ensures that each class in the dataset is represented in each training and validation subset. After some trial and error, the best hyperparameters for the model were found, as summarized in Table 2.

6.3 Model

The Vision Transformer (ViT) is a deep learning model that uses transformers, a neural network efficient at processing sequences. The ViT can capture spatial and temporal features of images by dividing them into patches, which are then processed individually by the transformers. The transformers can then learn the relationships between the patches, allowing them to capture spatial and temporal features. The ViT was trained on the ImageNet-21k dataset, which consists of 21,841,116 images and 21.841 classes. The ImageNet-21k dataset is a large and diverse dataset created to train computer vision

Table 2. Values of hyperparameters

Hyperparameters	Values
Batch Size	32 data
Learning rate	0,001
Training, validation and testing	70%, 15%, 15%
Optimizer	Adam
Input size	224 × 224 pixels
Dropout	0.3
Epoch	20

models. The pictures from the ImageNet-21k dataset were preprocessed to have a resolution of 224 × 224 pixels. The ViT was then fine-tuned on the ImageNet dataset, consisting of 1 million images and 1.000 classes, with a resolution of 224 × 224 pixels.

This work aims to classify images into two classes including increased optic nerve excavations, an essential factor in diagnosing glaucoma. To do this, the pre-trained classification heads of each ViT model were removed. Then, a new classification head or (softmax) was added to the model with the two class labels: normal or increased optic nerve excavations. Pulling the pre-trained classification heads allowed the ViT models to be fine-tuned for classifying glaucoma images [12]. Figure 4 illustrates the architecture of the ViT.

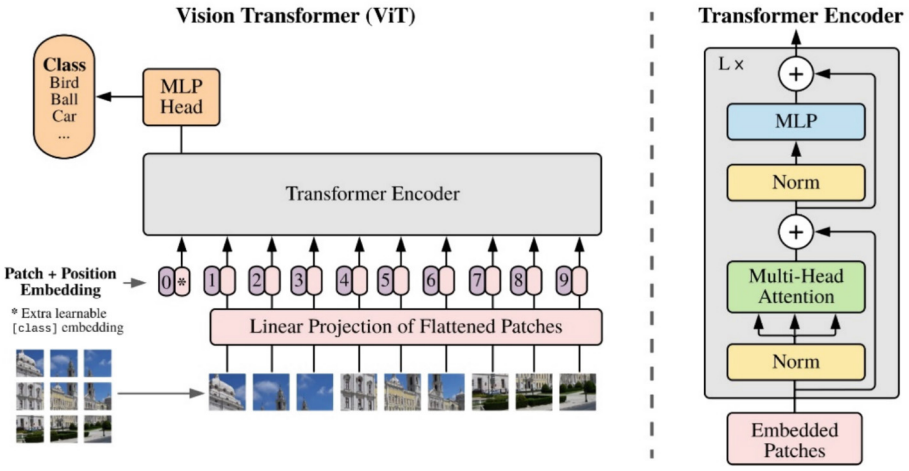

Fig. 4. Visualisation for ViT architecture

6.4 Evolution

Accuracy is one of the most critical factors in evaluating machine learning models, and it encompasses two crucial dimensions: discrimination and reliability. Discrimination measures the model's ability to distinguish between data classes, while reliability assesses its capacity to yield consistent predictions. Several techniques are available to evaluate the performance of machine learning models. However, for this study, we chose to use the accuracy (1), precision (2), recall (3), and F1-score (4) metrics. These metrics comprehensively overview the model's performance [14].

$$ACC = \frac{TP + FN}{TP + TN + FP + FN} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 * \frac{P*R}{P + R} \tag{4}$$

6.5 Results

Therefore, accuracy and sensitivity in glaucoma detection are of great importance. In this study, we used the following classification for the accuracy, f1-score, and recall metrics in the model evaluation [15]: Excellent: > 0.90; Good: 0.80–0.90; Acceptable: 0.70–0.80; Poor: 0.60–0.70; No acceptable discrimination ability: < 0.60. The results obtained in this study are presented in Table 3.

Table 3. Results

Models	Accuracy	F1-score	Recall
VIT-BRSET	**0.94**	**0.91**	**0.94**
Xception	0.91	0.89	0.90
VGG16	0.90	0.90	0.91
VGG19	0.89	0.88	0.90
MobileNetV2	0.90	**0.91**	0.91
InceptionV3	0.93	**0.91**	0.92
NASNetMobile	0.89	0.90	0.90

Table 3 presents the results of the performance of different DL models for eye disease detection trained on the BRSET. The results show that the ViT model achieved the best results in all metrics. The accuracy of 0.94 indicates that the model correctly

classified 94% of the images. Additionally, the F1-score of 0.91 and the recall of 0.94 suggest that the VIT can accurately identify most cases of glaucoma (high sensitivity) while maintaining a good balance between accuracy and sensitivity. This combination is crucial in medical applications, where accurate glaucoma identification is essential for referring patients for appropriate treatment.

The Xception model achieved the second-best performance, followed by the VGG16, VGG19, MobileNetV2, InceptionV3, and NASNetMobile models. These results suggest that the ViT is an effective neural network architecture for eye disease detection [6].

7 Discussion

Manual classification of glaucoma is a complex and time-consuming process that requires assessing a range of ocular features by experienced clinicians. Automated detection plays a crucial role in early diagnosis and effective management of glaucoma, as early intervention can prevent vision loss. DL models have been developed and deployed to identify glaucoma early, improving patient quality of life and slowing disease progression. CNNs are the most widely used DL models in developing solutions for detecting and classifying glaucoma early. They have achieved promising results in automated glaucoma detection, and their popularity is growing steadily. In addition, recent research [4–7] has explored the efficiency of other DL architecture's vision transformers. This approach is highly efficient, achieving good results in computer vision tasks. This study used a new Brazilian dataset, the Brazilian Multilabel Ophthalmological Dataset (BRSET), to train a vision transformer model to detect normal or increased optic nerve excavations, an essential factor in diagnosing glaucoma. BRSET is a high-quality dataset with well-segmented and labeled fundus images.

The results obtained in this work showed that the ViT could detect increased optic nerve excavations with high accuracy, surpassing the results of other DL models used in the literature. The model achieved an accuracy of 0.94, beating the results of the DeiT (0.91) and ResNet-50 (0.88) models of Fan et al. [4] and Souza et al. [6]. The model also achieved an accuracy of 0.916, AUROC of 0.968, and F1-score of 0.915, although it still falls below the 0.99 result obtained by He et al. [5].

In the work of Fan et al. [4], they compared the Data-efficient image Transformer (DeiT) and ResNet-50 models trained on fundus images. DeiT performed similarly to ResNet-50 on the OHTS test sets. The accuracy of DeiT and ResNet-50 were 0.91 and 0.88, respectively. However, the authors note that vision transformers can improve generalization and interpretability in machine learning models, detecting eye diseases and possibly other medical conditions that rely on images for clinical diagnosis and treatment. He et al. [5] present an interpretable transformer network for classifying retinal diseases using optical coherence tomography (OCT). The network is based on the Swin Transformer model, which is a transformer architecture that has been modified to be more interpretable. The network was trained on a dataset of OCT images from patients with various retinal diseases. The results showed that the network achieved an accuracy of 0.99 in the classification of retinal diseases. In this work, the authors obtained a better result.

The AlterNet-K model was presented in the study conducted by Souza et al. [6], a computer vision model that combines ResNets and MSAs to improve generalization.

The researchers conducted a comprehensive comparison, evaluating the performance of AlterNet-K against transformer-based models, such as ViT, DeiT-S, and the Swin Transformer, as well as against conventional deep convolutional neural network (DCNN) models, including ResNet, EfficientNet, MobileNet, and VGG. The results obtained by the authors were an accuracy of 0.916, an AUROC of 0.968, and an F1-score of 0.915. However, the results presented by Souza et al. were similar to those of the present study.

The results presented in the table indicate that Vision Transformers have a promising potential in the early increased optic nerve excavations, an essential factor in diagnosing glaucoma. The VIT-BRSET model, trained on data from fundus images of Brazilian patients, achieved an impressive accuracy of 0.94, surpassing the DL models tested in this study. Finally, we emphasize that further studies are needed to assess the efficacy of CNNs and Vision Transformers in detecting glaucoma in broader contexts. Additionally, it is crucial to develop more interpretable models, as they allow physicians to understand the decisions made by the models, increasing the confidence of healthcare professionals in using these tools to improve patient care.

8 Conclusion

This study investigated the potential of the ViT base-patch16-224 model for increased optic nerve excavations, an essential factor in diagnosing glaucoma. The ViT base patch 16-224 was trained on a dataset of fundus eye images and evaluated on an independent test dataset. The model achieved an accuracy of 0.94, an F1-score of 0.91, and a recall of 0.94, which is a promising result. The results indicate that ViT base-patch16-224 could be an effective tool to detect increased optic nerve excavations.

However, further studies are needed to assess the effectiveness of other transformers on more extensive and diverse datasets and under different clinical conditions. An important future task is to use technology to make ViT's decision-making more transparent to increase user confidence in the results. Additionally, it is essential to test the model's generalization with other datasets to ensure it can be applied to various populations and conditions.

Acknowledgements. This work was supported by the Portuguese Foundation for Science and Technology (FCT), project LARSyS - FCT Project UIDB/50009/2020 and National Funds finance this work through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020.

References

1. Tham, Y.C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., Cheng, C.Y.: Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology **121**(11), 2081–2090 (2014). https://doi.org/10.1016/j.ophtha.2014.05.013
2. Camara, J., Rezende, R., Pires, I.M., Cunha, A.: Retinal glaucoma public datasets: what do we have and what is missing? J. Clin. Med. **11**(13), 3850 (2022). https://doi.org/10.3390/JCM11133850

3. Neto, A., Camera, J., Oliveira, S., Cláudia, A., Cunha, A.: Optic disc and cup segmentations for glaucoma assessment using cup-to-disc ratio. Proc. Comput. Sci. **196**(2021), 485–492 (2021). https://doi.org/10.1016/j.procs.2021.12.040
4. Fan, R., et al.: Detecting glaucoma from fundus photographs using deep learning without convolutions transformer for improved generalization. Ophthalmol. Sci. **3**, 100233 (2023). https://doi.org/10.1016/j.xops.2022.100233
5. He, J., Wang, J., Han, Z., Ma, J., Wang, C., Qi, M.: An interpretable transformer network for the retinal disease classification using optical coherence tomography. Sci. Rep. **13**, 3637. https://doi.org/10.1038/s41598-023-30853-z. 123AD
6. D'Souza, G., Siddalingaswamy, P.C., Pandya, M.A.: AlterNet-K: a small and compact model for the detection of glaucoma **1**, 3. https://doi.org/10.1007/s13534-023-00307-6
7. Karrothu, A., Chunduru, A.: Glaucoma detection using computer vision and vision transformers (2023). https://journal.uob.edu.bh/handle/123456789/5206. Accessed 13 Sept 2023
8. Nakayama, L.F., et al.: A Brazilian multilabel ophthalmological dataset (BRSET) v1.0.0 (2023). https://physionet.org/content/brazilian-ophthalmological/1.0.0/. Accessed 13 Sept 2023
9. Li, L., et al.: A large-scale database and a CNN model for attention-based glaucoma detection. IEEE Trans. Med. Imaging **39**(2), 413–424 (2020). https://doi.org/10.1109/TMI.2019.2927226
10. Ajitha, S., Akkara, J.D., Judy, M.V.: Identification of glaucoma from fundus images using deep learning techniques. Indian J. Ophthalmol. **69**(10), 2702–2709 (2021). https://doi.org/10.4103/IJO.IJO_92_21
11. Teixeira, I., Morais, R., Sousa, J.J., Cunha, A.: Deep learning models for the classification of crops in aerial imagery: a review. Agriculture **13**(5) (2023). https://doi.org/10.3390/AGRICULTURE13050965
12. Dosovitskiy, A., et al.: An image is worth 16×16 words: transformers for image recognition at scale. In: ICLR 2021 - 9th International Conference on Learning Representations, October 2020. Accessed 14 Sept 2023. https://arxiv.org/abs/2010.11929v2
13. Wassel, M., Hamdi, A.M., Adly, N., Torki, M.: Vision transformers based classification for glaucomatous eye condition. In: Proceedings - International Conference on Pattern Recognition, vol. 2022-August, pp. 5082–5088 (2022). https://doi.org/10.1109/ICPR56361.2022.9956086
14. Leite, D., et al.: Machine Learning automatic assessment for glaucoma and myopia based on Corvis ST data. Proc. Comput. Sci. **196**(2021), 454–460 (2021). https://doi.org/10.1016/j.procs.2021.12.036
15. Leite, D.R.A., de Moraes, R.M., Lopes, L.W.: Different performances of machine learning models to classify dysphonic and non-dysphonic voices (2022). https://doi.org/10.1016/j.jvoice.2022.11.001