ISSN 2395-8618

Architecture for Early Detection of Age-Related Macular Degeneration Using Data Augmentation and Vision Transformers (ViT)

Augusto Javier Reyes-Delgado, Jorge Ernesto González-Díaz, José Luis Sánchez-Cervantes, Giner Alor-Hernández, José Luis Rodríguez-Loaiza, Yara Anahí Jiménez-Nieto

***Abstract*—Age-Related Macular Degeneration (AMD) is a leading cause of blindness among the geriatric population worldwide. While there is no definitive cure for this pathology, early identification of AMD allows for effective treatment administration. This paper introduces an architecture for a specialized module to identifying AMD at various evolutionary stages. An in-depth analysis was conducted examining investigations related to detecting age-related macular degeneration (AMD), with a focus on studies utilizing deep learning techniques and vision transformers. It is important to emphasize that most of these works have only addressed binary disease detection. Our initiative incorporates an architecture that emphasizes data augmentation in the training set and utilizes the ViT vision transformer for analyzing retina images. The main aim is to attain a differentiated categorization (non-AMD, mild, moderate, and advanced) that serves as a basis for diagnosing AMD. The ViT trained model has shown 96.55% accuracy in this classification. Hence, it can be inferred that the outlined module holds noteworthy value as an additional support tool for ophthalmologists in the precise detection of Age-Related Macular Degeneration.**

***Index Terms*—Age-related macular degeneration, vision transformers, data augmentation.**

I. Introduction

Age-Related Macular Degeneration (AMD) is one of the main causes of vision loss among the population over 50, particularly affecting developed countries [1]. It stands as the third cause of blindness worldwide, with a moderate or severe prevalence estimated at 10.4 million people according to the World Health Organization [2].

AMD causes the loss of central vision, implying that fine details are not perceived up close or far away in the center of the visual focus, even though peripheral vision remains normal. In addition to age, AMD is associated with overweight, smoking, hypertension, heart diseases, and high-fat consumption, conditions common in developed countries, as well as some developing countries such as Mexico.

In this context, in Mexico, over 30 million suffer from hypertension, and over 70% of the population is obese, it according to data from the Mexican government and the National Institute of Public Health [3,4]. As far as AMD is concerned, this condition is divided into two categories: 1) The most common being "dry" AMD (approximately 80% of cases), characterized by the appearance of Drusen (fat clusters in the retina, and 2) "wet" AMD, which occurs due to the growth of abnormal blood vessels under the retina [5].

Although the disease is highly treatable in its early stages, it is often asymptomatic until it reaches an advanced state, by which time the damage is irreversible. This fact underscores the urgent need to develop tools for early detection. Optical coherence tomography has become one of the predominant diagnostic methods for this disease (along with fundus images). Observing optical coherence images and fundus images are the most common methods for diagnosing the condition. Our proposal leans on computer vision techniques (Data Augmentation) and deep learning (ViT) to aid in the analysis of features in fundus images, focusing on the most common category of AMD, namely the so-called "dry" type.

Currently, artificial intelligence models, predominantly based on convolutional neural networks, have been developed to address this issue. However, in recent years, computer vision techniques have emerged as the new state-of-the-art in image analysis. These techniques are highlighted for their ability to capture the entire context of the image under analysis. This process involves segmenting the image into patches and subsequently representing them as a vector, which is processed through a transformer-type encoder [6].

Therefore, the contribution of this research, is an architecture for early AMD detection that integrates computer vision techniques (Data Augmentation) and deep learning (Vision Transformers, ViT). This is an essential part of an ongoing technological development integrating an AI module for feature analysis to detect and classify AMD in various stages (No-AMD, mild, moderate, and advanced).

This paper is organized as follows: Section 2 presents the state of the art, Section 3 introduces the design of the module's architecture, Section 4 discusses the training outcome validation, and finally, Section 5 presents the conclusions and future work.

II. State of the Art

There are several initiatives related to AMD detection using convolutional neural networks and vision transformers, such as those analyzed and briefly described below. A DCNN (Deep Convolutional Neural Network) with a 13-layer architecture to

Manuscript received on 14/04/2023, accepted for publication on 10/06/2023.

A.J. Reyes-Delgado, J. E. González-Díaz, J. L. Sánchez-Cervantes, G. Alor-Hernández, J.L. Rodríguez-Loaiza are with the Instituto de Oftalmología Conde de Valenciana, Tecnológico Nacional de México, Mexico ({m17010207, d04010291,jose.sc, giner.ah}@orizaba.tecnm.mx, jose.rodriguez@ institutode oftalmologia.org).

Y. Anahí Jiménez-Nieto is with the Universidad Veracruzana, Facultad de Negocios y Tecnologías campus Ixtaczoquitlán, Mexico (yjimenez@uv.mx).

ISSN 2395-8618

TABLE I
A COMPARISON OF THE DIFFERENT STUDIES ANALYZED IS PRESENTED, INCLUDING THE IMPLEMENTED METHODS

Author	Medical condition	Methods
Rivu Chakraborty [7]	Wet and dry AMD	DCNN
Jen Hong Tan [8]	Wet and dry AMD	CNN - Blindfold CNN - Cross-validation
Maximilian Treder [9]	AMD does not specify the type	DCNN
Joel C. De Goma [10]	AMD does not specify the type	Neural Network - SLIC Random Forest - SLIC Naive Bayes - SLIC Support Vector Machine - SLIC
Yao-Mei Chen [11]	Wet and dry AMD DME Drusen CNV	Alexnet Googlenet VGG16 VGG19 Resnet18 Resnet50 Resnet101
Zhencun Jiang [12]	AMD DME	ViT
Ordax Galindo [13]	AMD	HOG

automatically distinguish images with the presence of AMD from images with the absence of AMD was proposed in [7].

The study focused on binary classification, so only images with AMD (23 images) and control images (61 images) were used. For the IChallenge-AMD case, 89 AMD and 311 control images were taken from the set. The proposed model was trained, developed and validated in Python using TFLearn and TensorFlow.

Similarly, in [8] a CNN (Convolutional Neural Network) model for the early detection of AMD was presented. The authors evaluated it using cross-validation and blindfold strategies. As a dataset, they used 402 images with normal fundus, 583 images with evidence of AMD in different degrees up to GA (Geographic Atrophy) also known as dry AMD and 125 images with presumed wet AMD.

On the other hand, in [9] the use of deep learning for automatic AMD detection in SD-OCT (Spectral Domain Optical Coherence Tomography) images was proposed. For training, 1012 cross-sectional SD-OCT images were used, 701 of which were labeled as AMD positive, the remaining 311 images were labeled as healthy samples, 100 images were reserved for testing (50 AMD positive and 50 healthy).

A methodology for the recognition of AMD in fundus images was evaluated by de Goma et al., [10]. The authors, used image processing and 4 Machine Learning models: Naive Bayes, Neural Network, Support Vector Machine and Random Forest. Fundus images were collected. A few 13,105 unlabeled images. From the set of images, 532 are labeled as positive for AMD, plus 52 images labeled as normal. For training and classification, the image set was divided in a 70/30 ratio for training and testing respectively. The training was carried out on the 4 models mentioned above, applying 10-fold cross-validation.

Similar to the initiatives [8 and 9], the authors in [11] proposed a Convolutional Neural Network with learning transfer capability, and with hyper parameters suitable for the classification of AMD, DME (Diabetic Macular Edema) and CNV (Choroidal Neovascularization) through the analysis of OCT (Optical Coherence Tomography) images. Finally, in contrast to the initiatives of [7 to 11], in the research presented in [12] Vision Transformer was integrated with OCT to improve the diagnosis of retinopathies, mainly targeting AMD and DME. OCT images of both retinopathies as well as normal eyes were collected for training.

The dataset included 1407 OCT images of normal eyes, 723 with the AMD condition and 1101 with DME, in the research presented in [13] the author used HOG (histogram of oriented gradients) for the AMD detection, using a dataset with 400 retinography images of healthy patients and patients with AMD presence. Divided the dataset into 300 for training and 100 for testing, also applied preprocessing to the images to increase contrast. In Table 1, a summary of the analyzed studies is presented.

An exhaustive review of the existing literature shows multiple studies that concentrate on detecting AMD, with convolutional neural networks being predominantly used. However, a meticulous analysis demonstrates that these works generally tend to lean towards binary detection, meaning they determine the presence or absence of the disease, or diversify their approach by detecting multiple pathologies.

Our research is noteworthy by the use of data augmentation techniques on a comprehensive dataset, synthesized from multiple pre-existing sources, and its implementation of a multi-class classification approach to distinguish between distinct progressive phases of AMD, utilizing the Vision Transformers (ViT) framework.

III. ARCHITECTURE

In this section, we present the design architecture for the development of a module aimed at feature analysis for the early detection of age-related macular degeneration. The analysis utilized features extracted from fundus images, leveraging artificial vision techniques, specifically Data Augmentation and vision transformers.

ISSN 2395-8618

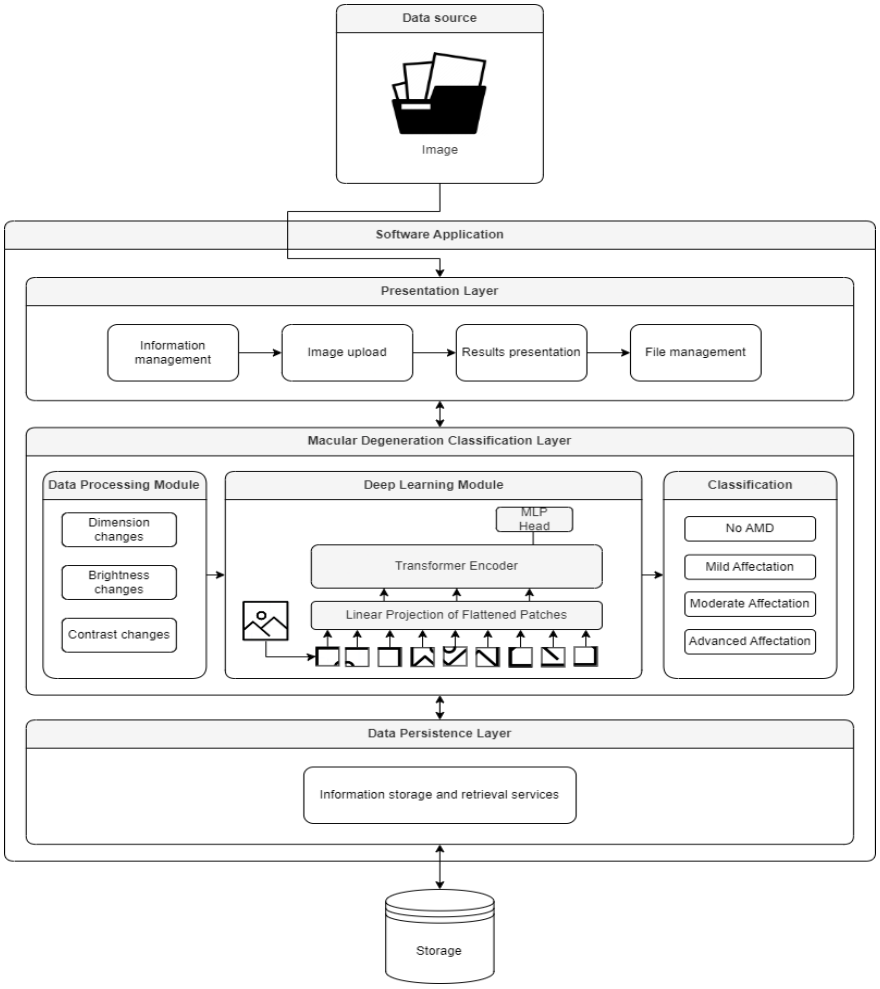

Fig. 1. Architecture design

A. Architecture Design

Fig. 1 depicts the architecture design for the development of the module for multi-class AMD detection using ViT.

Below is a brief description of the architecture layers and their respective components.

Presentation Layer: In this layer, the users (medicals) interact with the system, in this case, a web application. In this layer, registration and user authentication data are entered, as well as the data of the subjects whose fundus images will be classified by the system. Additionally, in this layer, system data management can be controlled.

Macular Degeneration Classification Layer: It is responsible for image classification. It receives, processes, analyzes, and returns the result which is stored in the data persistence layer and is displayed in the presentation layer. This layer, is integrated by two sub-modules:

1. **Data Processing Module:** This module receives the fundus image and processes it, applying conversions such as scaling to 300x300 pixels, brightness adjustments, and contrast changes. This is done to highlight the details that the trained model looks for in classification.

2. **Deep Learning Module:** This is the module that classifies the previously processed image. The model is a fine-tuned version of ViT, based on a previously trained model from Google [14]. This fine-tuned model was trained with a set composed of various databases, and in turn, the new dataset underwent a data augmentation process. This module finally outputs the classification obtained by the model. The architecture of the model is depicted in Fig. 2.

Data Persistence Layer: This layer stores all data that is added to the system. It includes user data, data of the subjects to whom the fundus images belong, the fundus images themselves, and the classification results obtained by the Macular Degeneration Classification Layer.

The following section includes information about the training ViT model including the Data Augmentation.

IV. Training Validation with Data Augmentation and Vision Transformer

The implementation of the Transformers library, using PyTorch, enabled the training of the ViT model. For the creation of the dataset, 266 fundus images were selected, with 160 images allocated for the training segment, and 53 each for the validation and test segments, aiming to achieve a distribution close to 60%/20%/20%. However, 4 images were generated from each original image of the training segment

ISSN 2395-8618

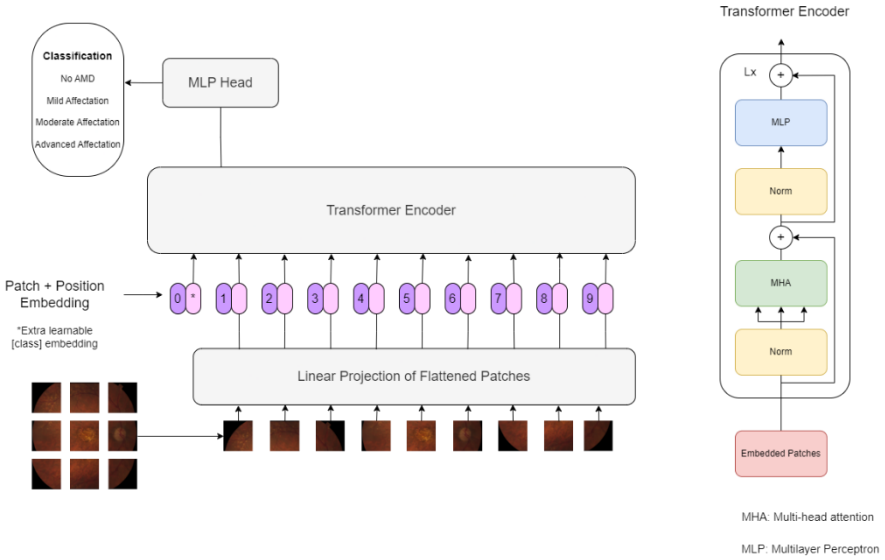

Fig. 2. Architecture of the vision transformer (based on [6])

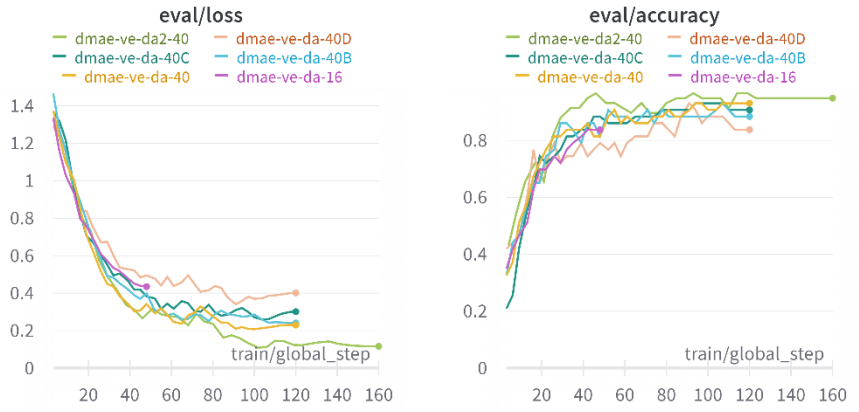

Fig. 3. Results obtained from the ViT model training process

using Data Augmentation, However, a portion of the images were discarded due to chromatic aberrations.

As a result, the final dataset for this study comprised of 684 images, with 578 assigned for the training segment, and 53 for validation and testing. The images were extracted from the iChallenge-AMD dataset [15] and a dataset available on Kaggle published by Mujib [16], which, in turn, incorporates images from various datasets. The multi-class labeling of the images was carried out based on the available literature [17, 18], and then validated by ophthalmology experts.

Through iterative training and validation, the model was successively refined, achieving high accuracy in the identification and classification of different stages of AMD. Emphasis was placed on the optimization and fine-tuning processes, particularly on hyperparameter tuning. To ensure both accuracy and efficiency, we employed the “Weights and Biases” (W&B) tool, a renowned platform for monitoring and tracking experiments in machine learning.

This tool facilitated intuitive visualization and comprehensive tracking of each hyperparameter variation. Consequently, we were able to discern the most optimal configurations that augmented the model's performance in AMD detection. The resulting outcomes are illustrated below.

Fig. 3 illustrates the results obtained from the training routine of the model based on the vision transformer. The accuracy is depicted in the right plot, with distinct curves representing successive iterations of the model's training. The left plot displays the loss value, which attains a level of 0.2660 by the sixth cycle.

Table 2 presents a thorough analysis comparing our proposed method with previously analyzed works in the State-of-the-Art Section, in the domain of age-related macular degeneration (AMD) detection. The comparison aims to demonstrate the differences and similarities in terms of approaches, techniques, accuracy, and other pertinent metrics.

By comparing our method with previous research, our goal is to validate its effectiveness and identify potential improvements and advantages over existing approaches. The table provides a concise and informative summary of current research in this area, as well as the placement of our method within it.

Table 2 shows that the results obtained by our proposal achieve a performance comparable to that of the literature

ISSN 2395-8618

Fig. 4. Multi-class classification results of the model (fundus image extracted from [16])

TABLE II
COMPARISON OF THE DIFFERENT WORKS ANALYZED WITH OUR VIT TRAINED MODEL

Author	Medical condition	Methods	Results
Rivu Chakraborty [7]	Wet and dry AMD	DCNN	0.9000 (iChallenge AMD) 0.9303 (iChallenge AMD augmented) 0.9955 (ARIA)
Jen Hong Tan [8]	Wet and dry AMD	CNN-Blindfold CNN-Cross-validation	0.9117 0.9545
Maximilian Treder [9]	AMD does not specify the type	DCNN	0.9970
Joel C. De Goma [10]	AMD does not specify the type	Neural Network-SLIC Random Forest-SLIC Naive Bayes-SLIC Support Vector Machine-SLIC	0.9561 0.9474 0.9123 0.9035
Yao-Mei Chen [11]	Wet and dry AMD DME Drusen CNV	Alexnet Googlenet VGG16 VGG19 Resnet18 Resnet50 Resnet101	0.9550 0.9531 0.9594 0.9942 0.9847 0.9909 0.9919
Zhencun Jiang [12]	AMD DME	ViT	0.9969
Ordax Galindo [13]	AMD	HOG	0.9150
Our proposal (Multiclass)	No AMD Mild AMD Moderate AMD Advanced AMD	ViT	0.9655

analyzed, with a precision of 0.9655. Four of these works report a precision higher than 0.99 with different datasets and methodologies, while the rest are between 0.90 and 0.95. It is important to note that our study is the only one that focuses on multiclass detection of a single state, as shown in Fig. 4.

V. DISCUSSION

In this research, we have developed a model for the detection of Age-Related Macular Degeneration (AMD) based on Vision by Transformers (ViT). Our findings show a remarkable accuracy of 0.9655 and a loss of 0.2660, suggesting the robustness and reliability of ViT in medical image classification tasks, consistent with recent trends in the field. For instance, Zhencun Jiang [12] has also employed ViT for AMD

identification, further supporting the significance of this method. However, it is worth noting that several studies, including those conducted by Rivu Chakraborty [7] and Maximilian Treder [9], have favored the application of Variable Depth Convolutional Networks (DCNN) for comparable conditions. This suggests a variety of viable methods for this application.

When comparing our approach to other methods, it should be noted that while our accuracy surpasses many comparable

ISSN 2395-8618

studies, the selection of methodology is greatly influenced by the context of the application and dataset characteristics. For instance, Jen Hong Tan [8] examines the application of CNNs with cross-validation and blinding techniques, which may provide benefits in terms of generalization to new data sets. Additionally, Joel C. De Goma [10] utilizes a combination of techniques, such as Neural Networks, Random Forests, and Support Vector Machines, all adapted through the SLIC method. This emphasizes the significance of experimenting with multiple approaches.

Even though classical methods have been effective, our utilization of ViT displays a substantial improvement, especially in handling extended connections among visual data. This ability holds great significance in detecting AMD, where image characteristics can be faint and meager. Comparing with Yao-Mei Chen's [11] methods, which encompass various convolutional neural network structures like AlexNet, GoogleNet, and different iterations of ResNet, confirms the competitiveness of ViT, particularly in capturing intricate patterns and minimizing loss in extensive and diverse datasets.

It is important to acknowledge that selecting an appropriate deep learning model is only one aspect in the battle against AMD. Additional factors, such as interpretability, inference speed, and implementation ease, must also be considered in integrating these tools into clinical systems. In future studies, it would be beneficial to investigate combining ViT with other techniques, perhaps within an ensemble learning framework, to utilize complementary strengths and enhance accuracy and generalization in the automatic detection of AMD.

VI. Conclusions

Age-related macular degeneration (AMD) is emerging as a growing public health challenge, particularly in populations with increasing life expectancy. This situation is particularly critical in communities with a high prevalence of conditions associated with macular degeneration, such as obesity, diabetes and hypertension.

A clear example of this is the Mexican population. In this context, there is an urgent need to emphasize the prevention and early detection of this disease in order to reduce the risk of visual impairment. In this scenario, the implementation of artificial intelligence seems to be a potentially effective solution.

In the present research, a detailed analysis of several studies incorporating cutting-edge technologies for the detection and classification of macular degeneration has been carried out. The results show that computer vision (Data Augmentation) combined with deep learning (ViT) provides encouraging results, with accuracy levels comparable to clinical specialists, establishing itself as a diagnostic tool for healthcare professionals.

VII. Future Work

Based on the comprehensive assessment of the problem and current technologies, we propose the creation of a specialized module for the analysis of specific features for the early detection of age-related macular degeneration. This module will use advanced computer vision and deep learning techniques to analyze fundus images. The selected platforms and tools include Python, PHP, PostgreSQL, PyTorch, Keras, and Node.js. In addition, we decided to implement the XP methodology to guide the development of the project.

As part of our future directions, we intend to further refine and validate the module under discussion. This refinement encompasses the expansion of the dataset and the establishment of a web platform to streamline the integration of the machine learning model. Designed with scalability in mind, this platform seeks to extend its capabilities to detect other ophthalmic conditions, including but not limited to diabetic retinopathy and glaucoma, through multi-class classification techniques.

Our primary objective is to equip the medical community with a support tool for the early detection of macular degeneration, thereby playing a pivotal role in mitigating the onset of visual impairments in the aging population.

Acknowledgments

The authors are deeply grateful to the National Technological Institute of Mexico (TecNM) for supporting this work. In addition, this research was supported by the National Council Humanities Science and Technology (CONAHCYT) and the Secretary of Public Education (SEP) through the PRODEP program.

References

[1]. K.L. Ávila-Heras, Y.K. Carrillo-Mora, S.N. Cely-Jadan, and M. Arcos, "Revisión bibliográfica: Degeneración macular relacionada con la edad. Prevención y tratamiento temprano", *Revista Médica del Hospital José Carrasco Arteaga*, vol. 10, no. 2, pp. 145–149, (2018). DOI: 10.14410/2018.10.2.rb.32.

[2]. World Health Organization "*World report on vision*" (2019) https://apps.who.int/iris/ handle/10665/328717.

[3]. J.Á. Rivera-Dommarco, M.A. Colchero, M. Luis-Fuentes, T. González, C.A. Aguilar-Salinas, G. Hernández-Licona, and S. Barquera, "*La obesidad en México. Estado de la política pública y recomendaciones para su prevención y control*", Instituto Nacional de Salud Pública, (2018).

[4]. C.J. Thomas, R.G. Mirza, and M.K. Gill, "Age-related macular degeneration", *Medical Clinics*, vol. 105, no. 3, pp. 473–491, (2021). DOI: 10.1016/j.mcna.2021.01.003.

[5]. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale", *Computer Science, Computer Vision and Pattern Recognition*, (2020). DOI: 10.48550/ arXiv.2010.11929.

[6]. R. Chakraborty and A. Pramanik, "DCNN-based prediction model for detection of age-related macular degeneration from color fundus images", *Medical & Biological Engineering & Computing*, vol. 60, no. 5, pp. 1431–1448, (2022). DOI: 10.1007/s11517-022-02542-y.

[7]. J.H. Tan, S.V. Bhandary, S. Sivaprasad, Y. Hagiwara, A. Bagchi, U. Raghavendra, and U.R. Acharya, "Age-related macular degeneration detection using deep convolutional neural network", *Future Generation Computer Systems*, vol. 87, pp. 127–135, (2018). DOI: 10.1016/j.future.2018.05.001.

[8]. M. Treder, J.L. Lauermann, and N. Eter, "Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning", *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 256, no. 2, pp. 259–265, (2018). DOI: 10.1007/s00417-017-3850-3.

[9]. J.C. De-Goma, O.J.D. Binsol, A.M.T. Nadado, and J.P.A. Casela, "Age-related macular degeneration detection through fundus

ISSN 2395-8618

image analysis using image processing techniques", In: *Proceedings of the 2019 3rd International Conference on Software and e-Business*, pp. 146–150, (2019). DOI: 10.1145/3374549.3374577.

[10]. Y.M. Chen, W.T. Huang, W.H. Ho, and J.T. Tsai, "Classification of age-related macular degeneration using convolutional-neural-network-based transfer learning", *BMC Bioinformatics*, vol. 22, pp. 1–16, (2021). DOI: 10.1186/s12859-021-04001-1.

[11]. Z. Jiang, L. Wang, Q. Wu, Y. Shao, M. Shen, W. Jiang, and C. Dai, "Computer-aided diagnosis of retinopathy based on vision transformer", *Journal of Innovative Optical Health Sciences*, vol. 15, no. 2, (2022). DOI: 10.1142/S1793545822500092.

[12]. G. Ordax-Galindo, "Análisis de retinografías para la ayuda al diagnóstico de la degeneración macular asociada a la edad," *DMAE*, (2023).

[13]. *Hugging Face*, https://huggingface.co/google/vit-base-patch16-224, (2023).

[14]. National Eye Institute "*Age-related macular degeneration (AMD)*", https://www.nei.nih.gov/ learn- about-eye-health/eye-conditions-and-diseases/age-related-macular-degeneration, (2023).

[15]. W.M. Al-Zamil and S. Yassin, "Recent developments in age-related macular degeneration: a review", *Clinical Interventions in Aging*, vol. 12, pp. 1313–1330, (2017). DOI: 10.2147/ CIA.S143508.