RESEARCH ARTICLE

Revista Mexicana de Ingeniería Biomédica

VOL. 47 | NO. 1 | SPECIAL ISSUE ON TECHNOLOGICAL ADVANCES AND INNOVATION IN HEALTH | PP 1-23

https://dx.doi.org/10.17488/RMIB.47.SI-TAIH.1520 E-LOCATION ID: e1520

WAMDS2: Early detection of wet AMD using Swin Transformer V2

WAMDS2: Detección temprana de la DMAE húmeda utilizando Swin Transformer V2

Roberto Márquez Castro[1], *José Luis Sánchez Cervantes*[1], *Giner Alor Hernández*[1], *Augusto Javier Reyes Delgado*[1], *Alfonso Flores Leal*[1], *Martín Mancilla Gomez*[2], *Jorge Ernesto Gonzalez Diaz*[1]

[1]I.T. Orizaba, Tecnológico Nacional de México, Orizaba, Veracruz - México

[2]Facultad de Negocios y Tecnologías, Campus Ixtaczoquitlan, Universidad Veracruzana, Veracruz - México

ABSTRACT

Age-related macular degeneration (AMD) is a progressive eye disease that primarily affects individuals over 50 years old. Among the AMD variants, wet is the most severe, as it represents the advanced stage of dry AMD and can cause severe vision loss if not detected in time. This study focuses on the development of WAMDS2, a web module designed to identify characteristics associated with Wet AMD, facilitating early and accurate detection. To achieve this, a literature review was conducted on AMD and advanced techniques in computer vision and deep learning. The proposed model integrates Swin Transformer V2, a vision transformer implemented in PyTorch, to analyze fundus images and classify the different stages of the disease. The system's performance was evaluated using metrics such as accuracy, recall, and F1-Score. An accuracy of 84.76 % was achieved on the test set, suggesting its feasibility in clinical settings. The obtained results highlight the potential of WAMDS2 in ophthalmology and computer vision, demonstrating its capability to enhance automated diagnosis and patient care.

KEYWORDS: age-related macular degeneration, swin transformer, vision transformer

RESUMEN

La degeneración macular asociada con la edad (DMAE) es una enfermedad ocular progresiva que afecta principalmente a personas mayores de 50 años. Entre sus variantes, la DMAE húmeda es la más grave, pues representa la evolución avanzada de la DMAE seca y puede causar una pérdida visual severa si no se detecta a tiempo. Este estudio se centra en el desarrollo de WAMDS2, un módulo web diseñado para identificar características asociadas con la DMAE húmeda, lo que facilita una detección temprana y precisa. Para ello, se llevó a cabo una revisión de literatura sobre la DMAE y técnicas avanzadas de visión por computadora y aprendizaje profundo. El modelo propuesto integra el Swin Transformer V2, un transformador de visión implementado en PyTorch, para analizar imágenes de fondo de ojo y clasificar los diferentes estadios de la enfermedad. El rendimiento del sistema se evaluó mediante métricas como precisión, sensibilidad y F1-Score, logrando una precisión del 84.76 % en el conjunto de prueba, lo que sugiere su viabilidad en entornos clínicos. Los resultados obtenidos resaltan el potencial de WAMDS2 en el ámbito de la oftalmología y la visión por computadora, evidenciando su capacidad para mejorar el diagnóstico automatizado y la atención al paciente.

PALABRAS CLAVE: degeneración macular asociada a la edad, transformador swin, transformador de visión

Corresponding author

TO: **JORGE ERNESTO GONZALEZ DIAZ**
INSTITUTION: **TECNOLÓGICO NACIONAL DE MÉXICO/ I. T. ORIZABA**
ADDRESS: **AV. ORIENTE 9 NO. 852. COL EMILIANO ZAPATA, CP. 94320, ORIZABA, VERACRUZ - MÉXICO**
EMAIL: **d04010291@orizaba.tecnm.mx**

Received:

21 February 2025

Accepted:

5 September 2025

Published:

13 January 2026

INTRODUCTION

Age-related macular degeneration (AMD) is a progressive and degenerative retinal disorder that primarily affects individuals over the age of 50, representing one of the leading causes of vision loss globally[1]. As life expectancy continues to rise, the prevalence of AMD is expected to grow, posing significant challenges to healthcare systems worldwide. Current estimates indicate that approximately 30 % of individuals over the age of 70 are affected by AMD, with notable variations across populations and epidemiological studies[2]. This condition significantly impacts the quality of life of those affected, leading to difficulties in daily activities such as reading, driving, and recognizing faces[3]. Furthermore, the economic burden of AMD is substantial, encompassing direct medical costs, loss of productivity, and the need for long-term care in advanced cases[4].

AMD is categorized into two types: (1) dry AMD and (2) Wet AMD. Dry AMD is characterized by drusen deposits, melanin redistribution in the retinal pigment epithelium (RPE), and geographic atrophy, leading to gradual vision loss[5]. Wet AMD is characterized by choroidal neovascularization (CNV), where new blood vessels from the choroid penetrate the RPE and Bruch's membrane, extending beneath the retina[6]. Wet AMD is more aggressive and causes most severe vision loss cases. Despite advancements in imaging technologies, diagnosing AMD at early stages remains challenging due to subtle biomarkers and overlapping characteristics between disease subtypes[7]. Recent advances in artificial intelligence (AI) and computer vision offer opportunities to tackle these challenges. Image-based object detection algorithms have shown transformative potential in medical imaging[8]. Transformers, originally developed for natural language processing, have emerged as a leading approach for image analysis, outperforming traditional Convolutional Neural Networks (CNNs) in several tasks[9]. These architectures are especially effective in handling large, high-dimensional datasets, making them ideal for analyzing retinal imaging data[10]. The Swin Transformer V2 (Swin 2) is one of the most advanced Transformer models, demonstrating strong capabilities in object detection, feature extraction, and classification, making it an appropriate choice for AMD-related research[11].

Current diagnostic methods like Optical Coherence Tomography (OCT) and fluorescein angiography are critical but have limitations. OCT, although effective for imaging, struggles to detect early-stage Wet AMD, and is expensive and not widely available. Fluorescein angiography is invasive, and while it helps visualize advanced neovascularization, it's less effective for early detection and non-vascular abnormalities. These challenges highlight the need for non-invasive, cost-effective, and early detection solutions. Transformers, particularly Swin 2, was selected due to its unique ability to handle visual data through its hierarchical shifted windowing mechanism. This feature allows the model to capture spatial relationships at different scales, which is particularly useful in ophthalmic image segmentation and classification tasks, where retinal structures have features at various resolutions. Swin Transformer V2 outperforms traditional convolutional neural networks (CNNs) by enabling better segmentation of pathological patterns in fundus images, resulting in higher diagnostic accuracy compared to previous models. This ability to adapt to different image scales is crucial for the early detection of Wet AMD, where pathological patterns are more subtle and require a model capable of analyzing fine details at the pixel level.

The primary contribution of this study is the development of a web-based module for Wet AMD detection using the Swin Transformer V2 architecture. This innovative approach combines state-of-the-art deep learning techniques with practical, real-world applicability. Unlike previous studies, WAMDS2 (Web-based AMD Detection System) not only improves the diagnostic accuracy by leveraging the Swin 2 model but also introduces a scalable, non-invasive, and cost-effective solution for clinical settings. In this study, we focus on the early detection of Wet AMD. Although the model classifies five stages of the disease, the primary objective is to detect early-stage AMD, particularly the transition from No AMD to Mild AMD, which is crucial for timely medical intervention. Early detection is vital as it allows for treatment to slow the progression of the disease before irreversible damage occurs.

The remainder of this paper is organized as follows: Section 2 reviews the state of the art in AMD detection, highlighting the limitations of current methodologies. Section 3 describes the WAMDS2 architecture and its layers, including Presentation, Classification, and Data Persistence. Section 4 outlines the WAMDS2 workflow, from uploading fundus images to obtaining diagnostic results. Section 5 details the materials and methods, including fundus image acquisition, Swin Transformer V2 model training, and the development of the WAMDS2 web application. Section 6 presents a case study where a family doctor uses the DiFO ophthalmoscopy adapter to capture images, upload them to WAMDS2, and perform AMD detection. Section 7 discusses the results, including the model's classification accuracy (84.76 %), comparisons with prior studies, and its clinical applicability. Finally, Section 8 concludes, highlighting the importance of WAMDS2 in early AMD detection and its potential impact on ophthalmology.

State of art

In the literature, there are reported several researches related to our initiative. Most of them use approaches such as CNNs while others, although to a lesser extent, make use of vision transformers for the detection of visual disease by analyzing different types of eye images such fundus, Optical Coherence Tomography (OCTs), Color Fundus Photography (CFP), to mention but a few. In this section, we summarize in descending order several initiatives published in the last five years, to identify the main differences between them and our work.

An innovative deep learning method for the automated detection and quantification of atrophic features associated with macular atrophy (MA) in wet age-related macular degeneration (AMD) using Optical Coherence Tomography (OCT) was proposed by Wei W. *et al.*[12]. Although the model effectively identified atrophic morphological changes, it faced challenges due to annotation complexity and feature overlap, such as subretinal fluid, pigment epithelial detachment, and subretinal hyperreflective material. Despite these limitations, the model achieved an average Dice similarity coefficient (DSC) of 0.706, precision of 0.834, and sensitivity of 0.615, demonstrating its potential for early detection and monitoring of MA progression in Wet AMD. This method offers significant implications for early diagnosis and clinical monitoring, though future research should aim to incorporate additional features of Wet AMD for broader applicability. A hybrid approach combining a fully dense convolutional neural network (FD-CNN) and a deep support vector machine (D-SVM) was presented in [13], for retinal disease classification in OCT images. This hybrid model outperformed other methods, such as D-KNN, in precision, sensitivity, specificity, and F1 scores on datasets like UCSD and Duke. The integration of explainable AI techniques, such as LIME, enhanced transparency, making the model a promising tool for early diagnosis and effective treatment of retinal disorders. On the other hand, a novel framework combining a scale-adaptive autoencoder with a ResNet50-based classifier for early detection of AMD was proposed by El-Den N. *et al.*[14]. This model effectively distinguished normal retinas from various AMD stages, including intermediate AMD, geographic atrophy (GA), and Wet AMD, achieving superior accuracy compared to earlier methods. This innovative approach demonstrated significant potential for integration into existing retinal imaging systems for non-invasive and efficient AMD detection. Similarly, to [13][14] in [15], the Conv-ViT model introduced a hybrid feature extraction approach by integrating Convolutional Neural Networks (CNNs) and transformers for improved detection of retinal diseases. Using pre-trained models like Inception-V3 and ResNet-50, Conv-ViT achieved high F1 scores for diabetic macular edema (DME) and demonstrated enhanced precision in drusen classification. The study emphasized the benefits of combining texture and shape features, making Conv-ViT a promising tool for accurate retinal disease detection. The initiative of Lopez-Varela E. *et al.*[16] consisted of a computer-aided diagnostic approach incorporating 3D visualization for rapid identification of Wet AMD. Using a CNN trained on specific OCT datasets, the methodology achieved precise segmentation of fluid accumulations associated with AMD. Data augmentation and transfer learning further enhanced model robustness, making it a valuable tool for clinical diagnostics and workload optimization.

The automated application of DARC (Detection of Apoptosing Retinal Cells) technology with CNNs to predict subretinal fluid (SRF) formation in Wet AMD patients presented by Corazza P. *et al.*[17] is an interesting initiative that outperformed clinical specialists, achieving 94 % sensitivity and demonstrating DARC's potential as a biomarker for early retinal angiogenic activity. SAE-wAMD, an advanced CNN model with self-attention mechanisms for detailed classification of Wet AMD subtypes using OCT was presented by Haigong E. *et al.*[18]. The SAE-wAMD model outperformed standard CNNs and clinical expertise in detecting neovascular AMD and PCV, achieving higher F1 scores and demonstrating robust lesion detection capabilities. SAE-wAMD offers significant promise for improving clinical diagnostics and enhancing the accuracy of ophthalmic disease classification. A multi-modal CNN (MM-CNN) was introduced in [19], combining OCT and color fundus photography (CFP) data for enhanced AMD classification. Incorporating GAN-based image synthesis and loose pairing techniques, the model demonstrated superior accuracy and sensitivity for detecting Wet AMD and PCV compared to single-modality methods. This approach emphasizes the advantages of multi-modal data integration in retinal disease classification. Similarly, to [14], a ResNet50 model pre-trained on ImageNet and the Kermany dataset achieved 96.56 % accuracy in classifying dry and Wet AMD was presented in [20], which involved transfer learning and batch normalization techniques significantly improved model performance, demonstrating the potential of automated retinal image analysis for clinical applications. In [21], the research was focused on the early identification of eye diseases such as macular degeneration, cataracts, diabetes, glaucoma, hypertension, and myopia, which can lead to vision loss if not diagnosed in time. Such research proposed a deep learning-based approach utilizing Vision Transformer (ViT) architectures to classify fundus images, achieving greater classification reliability compared to previous methods. The results showed that the Vision Transformer-14 model achieved an F1-score of 83.49 %, with a sensitivity of 84 % and a precision of 83 %, indicating a significant improvement. The ODIR (Ocular Disease Intelligent Recognition) dataset was used, along with image augmentation techniques to expand the dataset. The methodology included custom transformer encoder blocks with multi-head attention. The study concluded that Vision Transformers enhance the classification of eye diseases and suggested future work to optimize these architectures, apply them to other types of medical images, and integrate more clinical data to improve diagnostic accuracy. A U-Net-based model with ResNeSt blocks and spatial clustering pyramids for automated segmentation of choroidal neovascularization in OCTA images was developed by Feng W. *et al.*[22]. Their model significantly outperformed traditional methods, emphasizing the role of AI and deep learning in enhancing diagnostic precision and treatment planning for AMD. Similarly to [15][16][18][20], in [23], cutting-edge CNN architectures, including ResNet18 and InceptionV3, were employed for automated retinal disease detection. InceptionV3 achieved the highest accuracy (99.79 %), underscoring the effectiveness of deep learning models in improving diagnostic precision. Wu M. *et al.*[24] examined various machine learning models for AMD classification, with ConvNeXT achieving the best performance (96.89 % accuracy). Data augmentation had a crucial role in enhancing model accuracy, emphasizing ConvNeXT's potential for clinical deployment. In [25] explored AI-driven risk stratification for AMD progression. Deep learning algorithms outperformed clinical experts in predicting disease progression and identifying patients at high risk for advanced AMD, highlighting the transformative impact of AI in ophthalmology. A deep learning-based method using AlexNet and ResNet for automated classification of dry and Wet AMD was introduced in [26]. ResNet outperformed AlexNet, achieving an AUC of 94 % for dry AMD and 63 % for Wet AMD, highlighting its effectiveness for precise and timely diagnosis. Such study emphasized the critical role of deep CNNs in advancing ophthalmological care and disease management. In [27], the effectiveness of CNN and ViT-based Systems for detecting glaucoma in fundus images was evaluated. The authors tested various CNN architectures, such as VGG19, ResNet50, InceptionV3, and Xception, along with ViT variants like Swin Transformer and Twins-PCPVT, as well as hybrid systems like CaiT, DeiT (Data-efficient image transformer), CeiT (Convolution-enhanced image transformer), and ConViT (Convolutional Vision Transformer), and the ResMLP architecture. In [28], a method for classifying retinal diseases using optical coherence tomography (OCT) images was introduced, utilizing a Swin-Poly Transformer network. The findings indicated that this method facilitated accurate and efficient retinal classification, highlighting the value of artificial intelligence in ophthalmic diagnostics and the potential of ViT networks in this field. Finally, in [29], a study focused on the classification of glaucomatous eye conditions using ViT models in full and cropped fundus images of the optic disc. They evaluated ViT architectures such as Swin, CaiT, crossViT, XciT, ResMlp, and DeiT, both individually and in ensembles. In addition to glaucoma, they addressed other ophthalmic diseases such as diabetes, cataracts, hypertension, pathological myopia, and other anomalies. Table 1 it presents a comparative analysis of the articles related to the proposed study, highlighting aspects such as medical condition, architecture, image type, and accuracy.

TABLE 1. **Comparative analysis of the articles related to the proposed study.**

Author	Eye disease	Architecture	Image type	Accuracy
Wei W *et al.*[12]	Wet AMD	U-net	OCT	83.40 %
Kayadibi, İ *et al.*[13]	Wet AMD Diabetic Macular Edema (DME)	FD-CNN	OCT	99.60 %
El-Den N. *et al.*[14]	Wet AMD Dry AMD	ResNet50	Color Retinal Fundus Images	96.2 %
Dutta *et al.*[15]	Wet AMD Dry AMD	Conv-ViT: Inception-V3 ResNet-50	OCT	94.47 %
López-Varela E. *et al.*[16]	Wet AMD	U-net	OCT	83.40 %
Corazza P *et al.*[17]	Wet AMD	U-net	OCTA	98.91 %
Haigong E. *et al.*[18]	Wet AMD Dry AMD	SAE-VGG16	OCT	92.56 %
Wang, W. *et al.*[19]	Wet AMD	MM-CNN (multi-modal CNN)	OCT + CFP	Not specified
Abdullahi. *et al.*[20]	Wet AMD	U-net	OCT	94.92 % 96.03 %
S. D. Gummadi. *et al.*[21]	AMD, cataracts, diabetes, glaucoma, hypertension, and myopia	Vision Transformer (ViT)	Fundus Images	83.18 %
Feng W. *et al.*[22]	Wet AMD Dry AMD	ResNet50	OCT	96.56 %
Haq, A. *et al.*[23]	Retinal diseases	ResNet18, InceptionV3, ResNext50	OCT	99.80 %
Wu M. *et al.*[24]	Wet AMD	DARC-CNN	OCT DARC	97.00 %
Yim, J. *et al.*[25]	Wet AMD	Models based on three-dimensional (3D)	OCT	Not specified
A. Serener. *et al.*[26]	Wet AMD Dry AMD	AlexNet ResNet	OCT	94.00 % 63.00 % 96.50 %
Alayon S. *et al.*[27]	Glaucoma	ViT, Swin Transformer, TwinsPCPVT, CaiT	Fundus images	81.67 % 81.67 % 84.17 %
He J. *et al.*[28]	Diabetic retinopathy, Diabetic macular edema, Glaucoma, Ocular abnormalities	ViT Swin Transformer	OCT	99.80 %
Wassel M.*et al.*[29]	Glaucoma Cataracts, Pathological Myopia	Cait, crossViT, XciT, ResMlp, DeiT, ViT	Fundus images	95.40 %
Our Study	Wet AMD	Swin Transformer V2	Fundus images	84.76 %

The comparison presented highlights that CNN architectures are the leading choice for analyzing Wet AMD, with U-net and ResNet50 being the most employed in conjunction with OCT imaging. These models have consistently delivered high accuracy, affirming their reliability in this area. While ViT-based architectures are less studied, they show promising potential. For example, Dutta *et al.*[15] achieved 94.47 % accuracy with a ViT-based model, making it a suitable alternative to CNN models. This study used OCT images, which provide high-resolution retinal layer visualization, helping to detect subtle changes more easily than fundus images. Additionally, Dutta *et al.*[15] worked with a large dataset of 109,309 images and applied data augmentation. In contrast, WAMDS2 used a smaller dataset of 522 images, also augmented, and achieved 84.76 % accuracy. Despite the smaller dataset, WAMDS2 demonstrated the potential of transformer-based architectures in data-limited settings, aligning with Gonzalez Diaz *et al.*[30], who showed that transformers outperform CNNs. This positions WAMDS2 as a valuable contribution to advancing Vision Transformers in Wet AMD detection, maintaining competitiveness with CNN models even with limited data. This data-efficiency is one of WAMDS2's key innovations, as it shows that Vision Transformers, while typically datahungry, can perform effectively in resource-constrained environments.

Furthermore, WAMDS2 introduces architectural improvements over previous ViT implementations. While Dutta *et al.*[15] used a ViT-based model, WAMDS2 incorporates Swin Transformer V2, a more efficient and scalable version of ViT, tailored specifically for retinal image analysis. The shifted window mechanism in Swin 2 allows it to capture both local and global features, improving its ability to segment and classify subtle patterns in fundus images. This improvement addresses some of the challenges faced by previous ViT-based models, making WAMDS2 a stronger model for Wet AMD detection.

Additionally, WAMDS2 highlights the adaptability of Vision Transformers, especially when large datasets are unavailable. Unlike previous studies, like Dutta *et al.*[15], which used extensive OCT datasets, WAMDS2 focused on fundus imaging, offering a broader spatial view for detecting AMD-related changes. This focus maximizes transformer performance in diverse imaging modalities and addresses gaps left by OCT-dominant approaches. The use of fundus images with Vision Transformers also positions WAMDS2 as a pioneering approach in leveraging this technology for retinal disease detection.

The limited size of the dataset can increase the risk of overfitting, meaning the model may learn patterns specific to the training set that do not generalize well to unseen data. To mitigate this risk, advanced data augmentation techniques, such as rotations, random cropping, and contrast adjustments, were applied to diversify the dataset and improve the model's robustness. Additionally, transfer learning was employed, using pre-trained weights. This strategy, combined with data augmentation, allowed WAMDS2 to generalize better and make more accurate predictions on unseen data, overcoming common challenges when working with small datasets. The findings suggest that Vision Transformers can play a key role in accurate Wet AMD diagnosis, particularly when paired with fundus imaging, and emphasize the importance of adapting architectures to specific imaging types. By focusing on fundus imaging and transformer architectures, WAMDS2 offers a complementary perspective to existing CNN-based approaches, paving the way for more versatile and data-efficient solutions in Wet AMD analysis. To summarize, WAMDS2 not only complements CNN-based approaches but also offers significant improvements, especially in data efficiency and architectural design, by introducing Swin Transformer V2 and focusing on fundus images. This study highlights the potential of Vision Transformers in ophthalmology and their clinical relevance, providing a more adaptable, data-efficient, and scalable solution for wet AMD diagnosis in clinical practice.

Architecture of WAMDS2

This section describes the WAMDS2 architecture with the integration of Swin Transformer V2, selected for its computational efficiency, ability to handle image resolution variations, capacity to identify complex patterns, and adaptability to data augmentation techniques[31]. These characteristics make it suitable for early detection of Wet AMD. Compared to ViT and Swin Transformer V1, its hierarchical structure, flexibility in image processing, and strong generalization make it a robust choice for early Wet AMD detection. Figure 1 shows the WAMDS2 architecture.

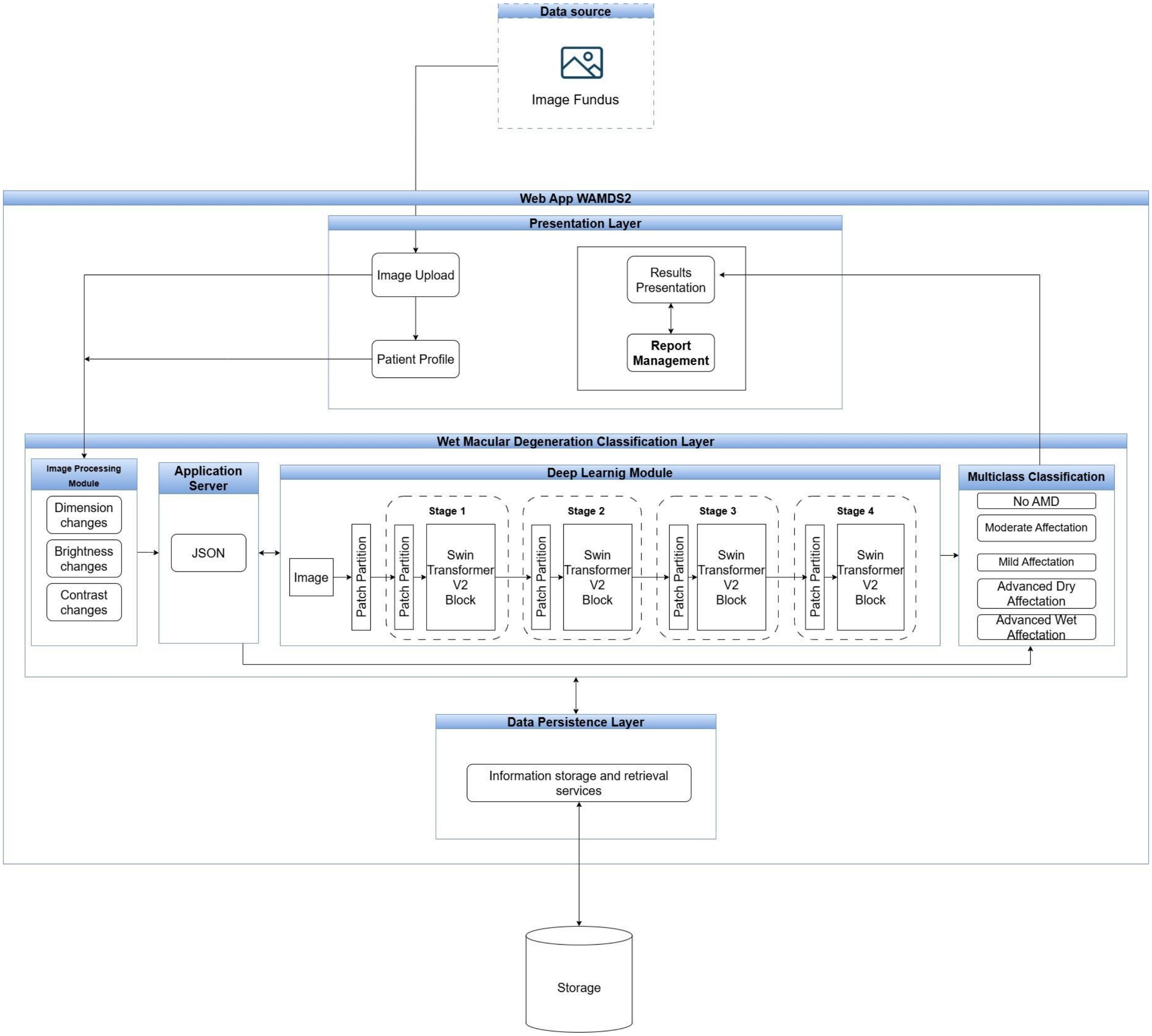

FIGURE 1. **WAMDS2 Architecture**

The architecture depicted in Figure 1 is structured in four layers to facilitate the WAMDS2 maintenance. Each layer is briefly described below.

- **Presentation Layer:** This layer represents the user interface for WAMDS2. Users, primarily family doctors, can upload fundus images along with their registration details to receive a classification indicating the level of Wet AMD.

Wet AMD Classification Layer: This layer classifies the images and manages the reception, processing, analysis, and return of results. The outcomes are stored in the data persistence layer for later retrieval. It consists of two key submodules:

- **Data Processing Module:** This module prepares the fundus image by performing essential transformations, including resizing to 224x224 pixels and making brightness and contrast adjustments. These modifications help highlight the critical details required for accurate classification by the trained model.

- **Deep Learning Module:** This module uses the Swin 2 model (Figure 2), an advanced version of Swin Transformer[32], trained using transfer learning with pre-trained weights to improve classification accuracy, especially on small datasets. The model classifies Wet AMD stages into five levels, from no degeneration (No AMD) to Advanced Wet AMD, using thresholds determined during training. For classification, key features such as drusen, geographic atrophy, and neovascularization are evaluated. Accuracy, sensitivity, and specificity were optimized during training, with classification thresholds adjusted using the F1 score to maximize precision and recall. Results are displayed to the user, and a report is generated and stored in the data persistence layer. Additionally, data augmentation techniques are used to prevent overfitting. Figure 2 shows the Swin 2 architecture.

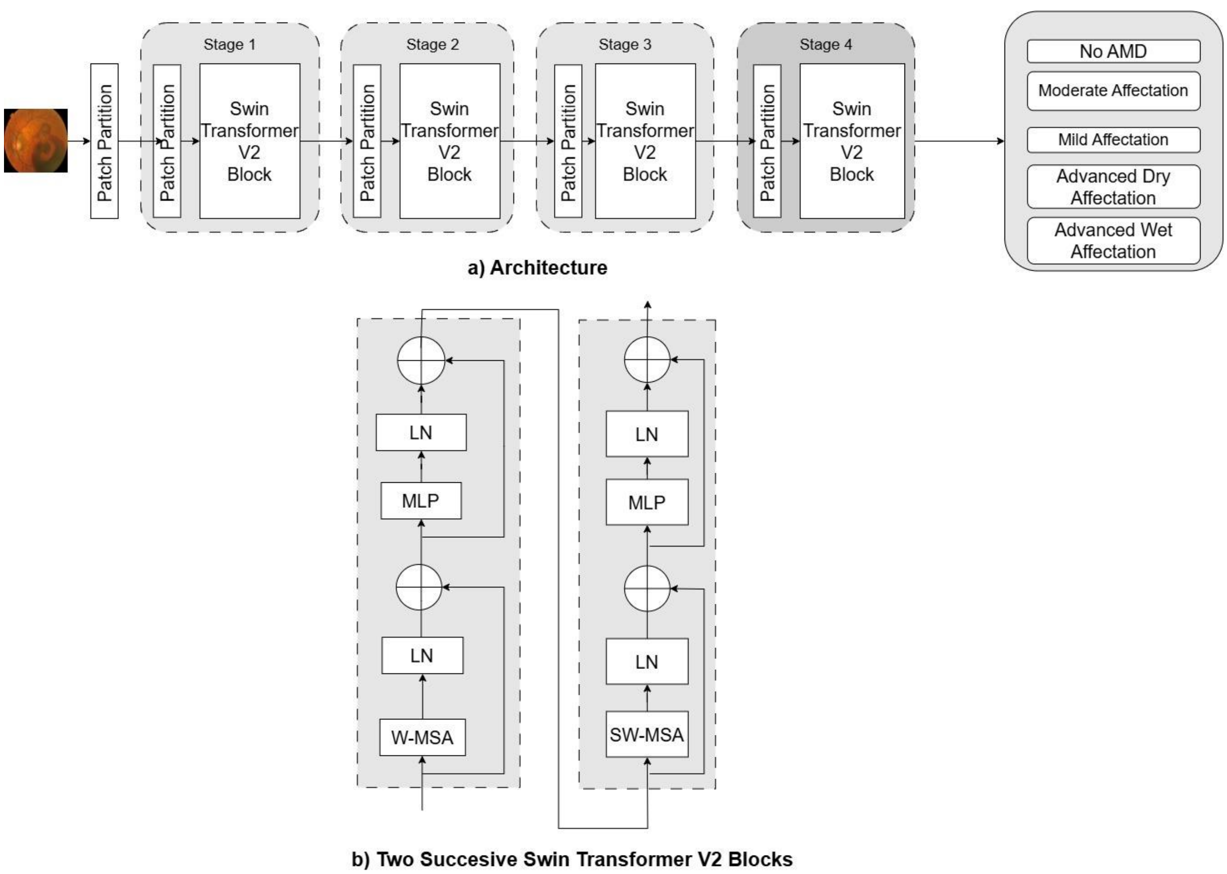

FIGURE 2. **Swin 2 architecture (based on [33])**

Data Persistence Layer: This layer stores all information received by WAMDS2, including user details, patient records, fundus images, and results from the Wet AMD classification layer. To ensure the security of patient data, data encryption protocols are applied both in transit and at rest, in compliance with the LFPDPPP (Ley Federal de Protección de Datos Personales en Posesión de los Particulares). Only authorized personnel can access the stored data, providing a secure and privacy-compliant environment for sensitive medical information.

Overall, the WAMDS2 architecture offers key advantages over traditional CNN-based approaches, especially in the early detection of AMD. The use of Swin Transformer V2 allows the model to capture local and global features from fundus images, improving the detection of subtle changes at early stages. Furthermore, by employing transfer learning, WAMDS2 achieves competitive performance with smaller datasets, making it more efficient than other systems. The architecture is also more robust and scalable, better adapting to diverse capture conditions and image quality. Compared to other approaches, WAMDS2 offers higher accuracy and sensitivity, optimizing the early detection of Wet AMD and improving clinical decision-making.

Workflow

The relationships between the components of the WAMDS2 architecture shown in Figure 1 define the workflow for the process of characterizing and classifying Wet AMD that the user may suffer from. Starting with the capture of general and authentication data, as well as the fundus image, up to the representation of the obtained results. The following is a brief description of the architecture workflow:

1. The user accesses the system through a web interface by entering their credentials. If the credentials are valid, access is granted.

2. The user provides patient information, including a fundus image, which will be processed for classification. This data is securely stored and encrypted in a PostgreSQL database, where both the patient's medical history and images are stored, ensuring compliance with the LFPDPPP regulations.

3. The uploaded image is processed to enhance its visual features by adjusting parameters such as contrast, color, and brightness, highlighting the characteristics required by the trained model. This step uses OpenCV for image preprocessing and enhancement.

4. Using the Swin Transformer V2 model implemented in PyTorch, the system classifies the image into one of the AMD progressions levels: No AMD, Mild AMD, Moderate AMD, Advanced Dry AMD or Advanced Wet AMD. The results were validated by comparing the model’s predictions with medical diagnoses from ophthalmologists to ensure the model's accuracy and reliability.

5. After classification, the system generates a report detailing the results, which is stored in the database.

6. The user is then presented with the classification results through a detailed report, which includes the probability percentages for each of the identified levels of severity. Figure 3 shows the deployment diagram that describes the previously mentioned workflow.

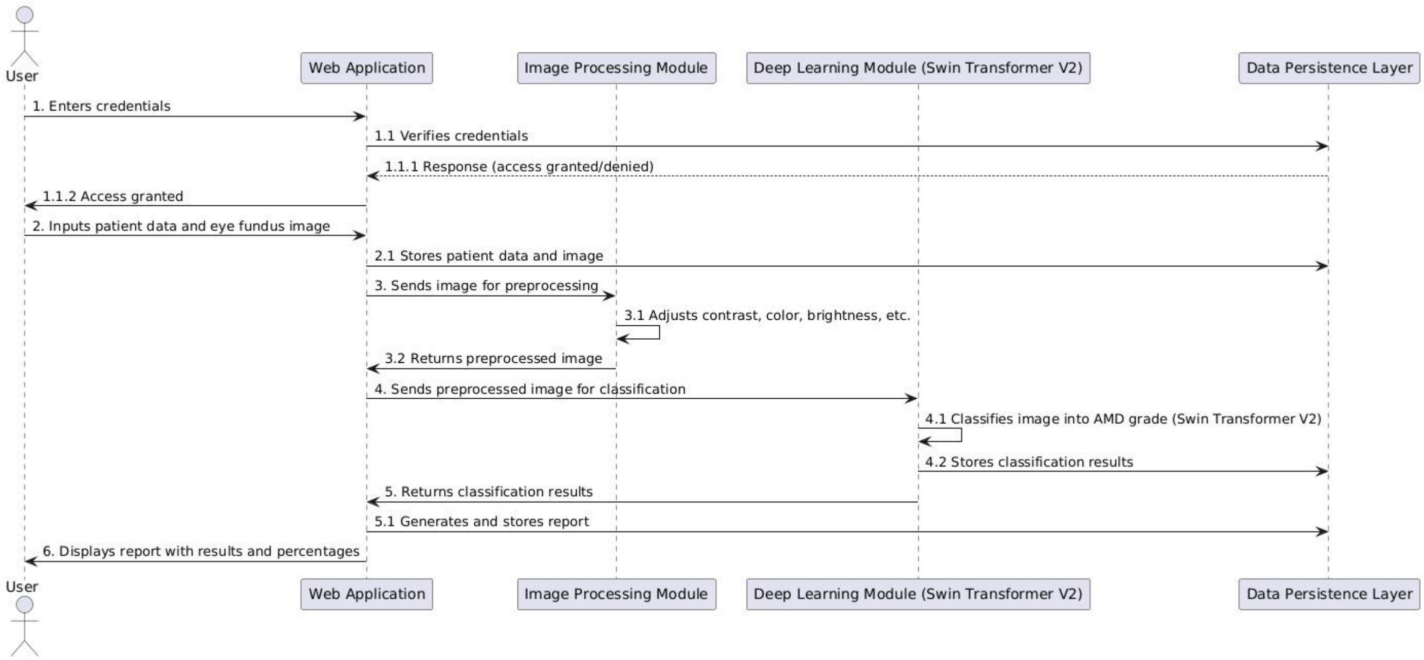

FIGURE 3. **Deployment diagram of the WAMDS2 workflow**

The WAMDS2 workflow provides an efficient process for characterizing and classifying Wet AMD in patients. From user authentication to generating a detailed report, each stage ensures the optimal capture, processing, and classification of fundus images. Integrating the Swin Transformer V2 model enables automated diagnosis through deep image analysis, offering reliable results on disease progression. To ensure effectiveness and accuracy, it is crucial to have proper infrastructure and tools to implement the workflow. The next section, Materials and Methods, details the resources used to develop the WAMDS2 platform, the database, image preprocessing algorithms, and the Swin Transformer V2 model's configuration parameters, offering insights into the methodology for classifying Wet AMD and validating the system.

MATERIALS AND METHODS

Figure 4 depicts the process of Materials and Methods used to develop WAMDS2. Each phase is briefly described in these sections.

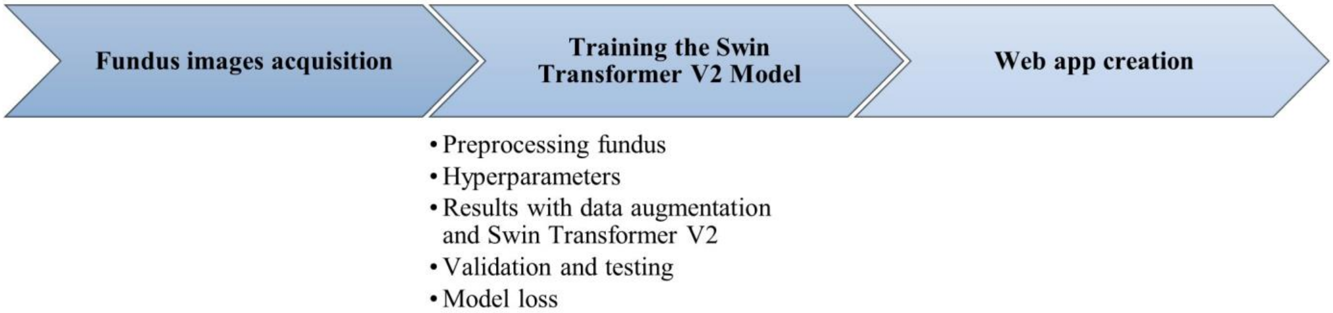

FIGURE 4. **Process of Materials and Methods used to WAMDS2**

1. Fundus images acquisition

Fundus images were obtained from three main sources: **1) Images from medical institutions:** These images were provided through collaborations with the Ophthalmology Institute Conde de Valenciana Foundation[34] and the Doctor Hernandez Zurita Foundation, I.B.P.[35]. They were carefully selected and preprocessed for model training; **2) Fundus images captured by the family doctor:** The family doctor used a smartphone and pupil dilation drops to capture the images under controlled conditions to ensure clarity and uniformity; **3) Public fundus image datasets:** Two publicly available datasets were used: the iChallenge-AMD dataset[36] and another from Kaggle[37], both containing annotated fundus images relevant to the classification task.

2. Training the Swin Transformer V2 Model

The pre-trained Swin Transformer V2 model was chosen for the classification task due to its ability to handle high-resolution images efficiently and its proven performance in computer vision tasks. The dataset for training consisted of 522 fundus images, with 64 % (333 images) used for training, 16 % (84 images) for validation, and 20 % (105 images) for testing. Data augmentation techniques were applied to enhance image diversity and improve model generalization. The training process followed these steps:

1) **Preprocessing fundus:** In the preprocessing step, all fundus images were resized to a uniform dimension of 224×224 Pixels. Additionally, a process of Dynamic Data Augmentation was applied to improve the model's generalization. Techniques including transformations such as contrasts, rotations, cropping, color adjustments, and horizontal flipping were used to simulate real-world variations in retinal images, enhancing the model's ability to handle unseen data effectively.

2) **Hyperparameters tuning:** The selection of hyperparameters for the Swin Transformer V2 model was carried out through a combination of manual tuning and grid search. During this process, different configurations were tested to optimize the model's performance on the validation set. The main hyperparameters adjusted were the learning rate, batch size, number of epochs, and warm-up rate. Below are the ranges of values tested and the final selected values:

TABLE 2. **Ranges of tested values**

Hyperparameter	Range of Tested Values	Final Selected Value
learning rate	1e-5 to 1e-3	4e-5
per_device_train_batch_size	8, 16, 32	16
num_train_epochs	20, 40, 60	40
warmup_ratio	0.05, 0.1, 0.2	0.1

3) **Tuning process:** The learning rate was tested from 1e-5 to 1e-3. It was observed that a learning rate of 4e-5 provided the best balance between convergence speed and training stability. The batch size was set to 16, as it provided good performance without memory issues. The number of epochs was set to 40 to ensure sufficient training without overfitting. The warm-up rate of 0.1 was selected to stabilize the early training cycles.

The final hyperparameter values were selected after several iterations of trial and error. These values were chosen for their ability to optimize the model's performance without compromising generalization. The configuration of the final hyperparameters used in the training process is summarized in the Table 3 below:

TABLE 3. **Hyperparameters for Training the Model based on Swin Transformer V2 for AMD-Wet classification**

Hyperparameters	Configuration / value
remove_unused_columns	False
evaluation_strategy	Epoch
save_strategy	Epoch
learning_rate	4e-05
per_device_train_batch_size	16
gradient_accumulation_steps	4
per_device_eval_batch_size	16
num_train_epochs	40
warmup_ratio	0.1
logging_steps	10
load_best_model_at_end	True
metric_for_best_model	Accuracy

4) **Validation and testing:** During training, an independent validation set was used to monitor performance metrics such as accuracy and F1-score. The final model was evaluated on the test set to ensure that it did not influence hyperparameter tuning.

5) **Results with data augmentation and Swin Transformer V2**: The Swin Transformer V2 model was trained and evaluated using the dataset splits (64 % training, 16 % validation, and 20 % testing). The results show robust performance in fundus image classification. Key metrics include: 1) Test accuracy: The model achieved 84.76 %, indicating good generalization to unseen data; 2) Test metrics: The model identified classes without bias, as shown in Table 4.

TABLE 4. **Test set metrics applied to model based on Swin Transformer V2 for AMD-wet classification**

Accuracy	Class	Precision	Recall	F1-Score
0.8476	No amd	0.8333	0.8333	0.8333
	Mild	0.8275	0.8275	0.8275
	Moderate	0.8095	0.8095	0.8095
	Advanced Dry	0.6666	0.3333	0.4444
	Advanced Wet	0.8913	0.9534	0.9213

Figure 5 shows the normalized confusion matrix for the validation dataset, where each column represents the percentages of actual instances of each class, while the rows show the predictions made by the model.

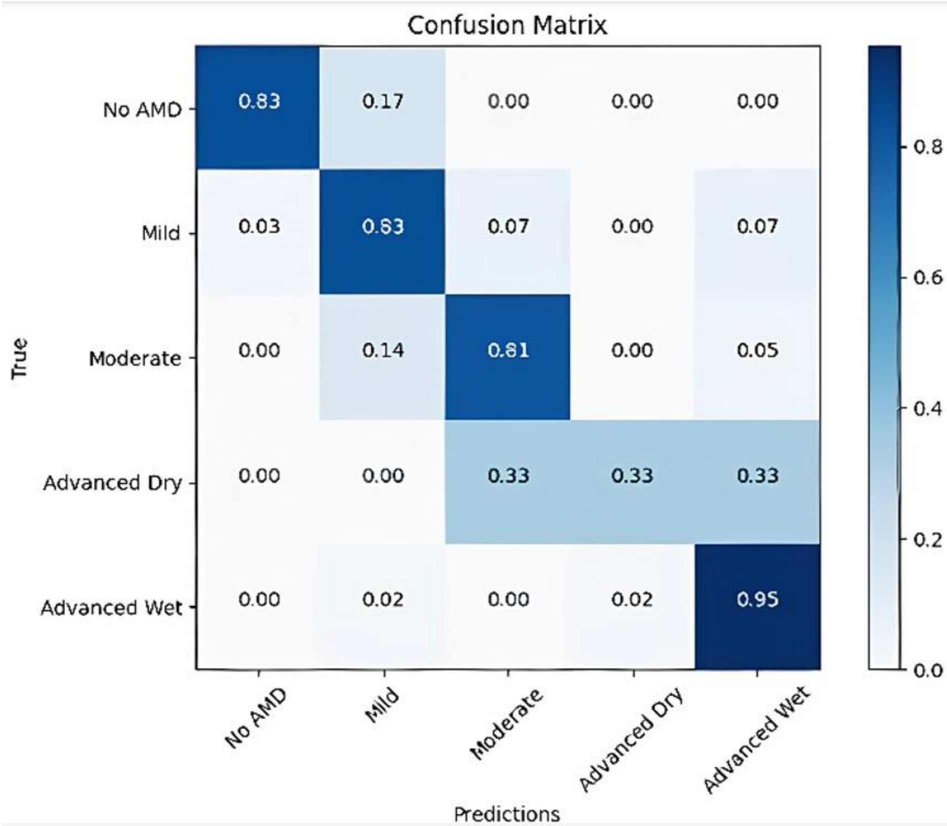

FIGURE 5. **Normalized confusion matrix on the test set**

6) **Model loss:** Training Loss consistently decreased, reaching values below 0.5 in the final epochs. Furthermore, the validation Loss was minimized (~0.62) at epoch 29, where the model achieved its optimal validation performance.

3. Web app creation

To carry out the detection process, the WAMDS2 web application was developed using the following Technological Stack:

- **Frontend:** The user interface was built using React.js, offering an interactive and intuitive experience for image uploads and results visualization.

- **Backend:** The backend was developed using Node.js, which handled requests, processed the uploaded images, and integrated the trained Swin Transformer V2 model for predictions.

- **Database:** The metadata and results were stored in a PostgreSQL database for efficient retrieval and management.

The Swin Transformer V2 model showed strong performance in classifying Wet AMD. Dynamic data augmentation improved the model's ability to generalize, achieving 84.76 % accuracy on the test set. These results confirm the effectiveness of the Swin Transformer V2 in classifying Wet AMD in fundus images, with balanced identification across severity levels. To assess its real-world applicability, a case study is provided in the next section, evaluating the model's reliability and usefulness in a clinical setting.

Case study

For this case, it is assumed that a person over the age of 50 with diabetes noticed the appearance of a dark spot affecting their central vision range. Before this, they perceived those objects appeared distorted. The affected person did not have access to nearby ophthalmology specialists to conduct an examination that would allow them to determine the level of impairment in their eyes. However, they were able to visit their family doctor, who had a portable device for capturing fundus images.

Based on this, the following questions arise:

- How can this individual recognize the potential presence of a condition, taking their age as a significant factor into account?

- What methods can be employed to assess the extent of ocular impairment?

- How can the process of ocular analysis be expedited prior to a medical consultation, especially considering the limited access to nearby specialists?

An alternative to address each of the questions raised is the use of WAMDS2, which provides the user with a web application that integrates the Swin Transformer V2 model for fundus image analysis in the multiclass classification of AMD. As shown in Figure 6, the family doctor captures a fundus image using the DiFO ophthalmoscopic adapter[38], an innovative device designed to efficiently capture fundus images using a smartphone attached to an ophthalmoscope.

1) Fundus captured by the user (family doctor)

The process begins when the family doctor places the adapter on the ophthalmoscope and adjusts the smartphone's position to obtain a sharp fundus image. Once the image is captured, it is uploaded to the WAMDS2 web application interface through a form where important patient data is entered, enabling the detection process. As depicted in Figure 6.

FIGURE 6. **Fundus image captured by the family doctor using the DiFO ophthalmoscopic adapter based on [38]**

2) Results of WAMDS2

As shown in Figure 7, WAMDS2 generates a real-time report that can be exported in PDF format, clearly and neatly displaying the patient's data and the results of the AMD detection analysis. This allows the user to understand the level of impact of the disease on the patient and enables them to bring the document to ophthalmology specialists for further evaluation and to streamline the monitoring of their visual health status.

Epidemiological Relevance

The individual in this case is over 50 years old and has diabetes, both of which significantly increase the likelihood of developing Wet AMD. According to research, individuals with diabetes are two to three times more likely to develop AMD compared to the general population. In fact, studies show that over 25 % of individuals aged 60 and above with diabetes are at risk of developing Wet AMD. Furthermore, the prevalence of AMD in individuals over 50 is substantial. In Mexico 3.4% of people over 65 years old are affected by AMD (IMSS) and approximately 19.3 % of people over 50 years old have diabetes, and the prevalence is expected to rise to 34.0 % by 2050[39].

Model Precision in Clinical Settings

The WAMDS2 model classifies images into five levels of AMD progression: No AMD, Mild AMD, Moderate AMD, Advanced Dry AMD, and Advanced Wet AMD. The system also provides confidence levels for each classification, offering probabilities that range from 80 % to 99 %, depending on the certainty of the classification. However, the model's performance can be affected by factors such as image quality, lighting conditions, and the presence of other retinal conditions.

Additionally, WAMDS2 has been compared to ophthalmologist diagnoses to validate its predictions. Although the model performs well in detecting Wet AMD, its accuracy may vary depending on the quality of the images and the experience of the clinician.

Impact of the Report on Decision-Making

The WAMDS2-generated report provides a detailed breakdown of the classification results, including the probabilities for each level of severity. This report serves as a valuable tool for both general practitioners (GPs) and ophthalmologists, enhancing the communication between the two. It allows the GP to refer the patient to an ophthalmologist with a well-documented diagnosis, facilitating quicker intervention and better monitoring of the patient's visual health. The report streamlines the decision-making process, helping ensure that Wet AMD is detected early and treated promptly.

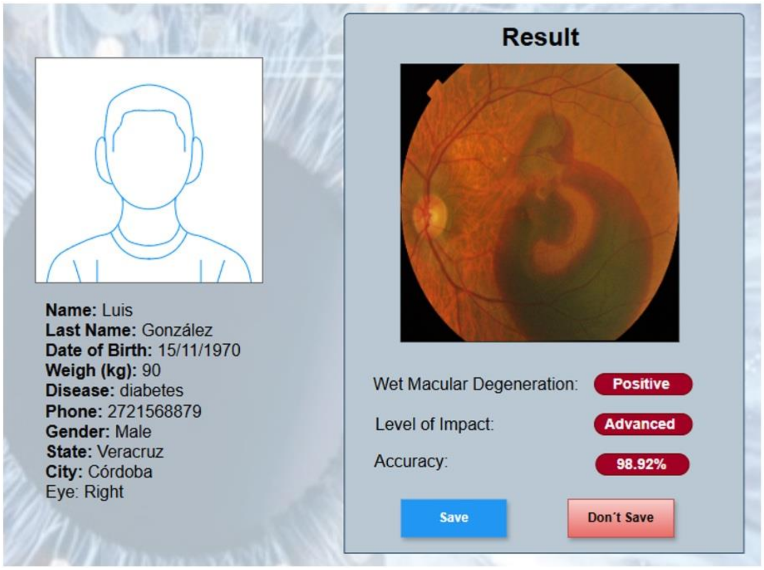

FIGURE 7. **Results Report**

The integration of WAMDS2 and the DiFO ophthalmoscopic adapter represents an effective alternative for individuals with limited access to ophthalmology specialists, particularly those at risk of developing Wet AMD. With this technology, family doctors can capture fundus images using an ophthalmoscope attached to a smartphone, allowing them to be uploaded to WAMDS2, where the evaluation and classification process of potential visual impairments is optimized efficiently and accurately.

Limitations and Future Improvements

Despite the promising results of the WAMDS2 model, several limitations must be considered. The model was trained using a relatively small dataset of 522 fundus images, which may affect its ability to generalize to unseen data. Additionally, the dataset did not fully represent the diversity of clinical conditions and patient demographics, potentially limiting its performance in real-world scenarios with more varied data. Another limitation is the quality of some of the fundus images, as they were captured under controlled conditions that might not always be replicable in clinical settings.

While the model performs acceptably, one notable challenge observed during testing was its difficulty in distinguishing between the moderate and advanced dry stages of AMD. This issue could be attributed to the subtle visual differences between these stages, which can be difficult to identify, even for experienced ophthalmologists. Moreover, the quality of fundus images plays a significant role in the model's ability to detect early-stage features, such as drusen or geographic atrophy, which are crucial for accurately distinguishing between these stages. Future improvements could include expanding the dataset by incorporating more diverse and larger collections of fundus images, including different stages of AMD and other retinal diseases. Additionally, exploring the use of advanced augmentation techniques or more robust preprocessing steps could help improve the model's generalization ability.

Several approaches could further improve the performance of WAMDS2. One possible future improvement is to increase the size and diversity of the data set by incorporating images from multiple clinical centers and different retinal diseases. This would expose the model to a wider variety of cases, improving its generalizability. Another strategy is to explore transfer learning using pretrained models on larger image datasets to improve feature extraction and classification accuracy.

Moreover, incorporating semi-supervised or unsupervised learning approaches could help leverage unannotated data, increasing the model's training set and improving its robustness. A key area for future work would be the real-time deployment and evaluation of the model in clinical environments, assessing its impact on clinical workflows and patient outcomes. Furthermore, an important future improvement is the integration of a feedback mechanism to allow clinicians to provide input on the accuracy of the diagnosis. This feedback could be used to adjust future predictions, continuously enhancing the model's performance and improving its accuracy over time. Implementing this system would ensure that the model learns from diagnostic errors, fine-tuning its predictions and adapting to new data, ultimately increasing its reliability and effectiveness in clinical settings. Finally, integration of multimodal images (combining OCT and fundus images) could provide more robust information for the model and improve its ability to detect more complex patterns associated with Wet AMD.

RESULTS AND DISCUSSION

The results obtained from the implementation of the WAMDS2 module, based on the Swin Transformer V2 architecture, demonstrated effectiveness in early detection of Wet AMD. The system achieved 84.76 % accuracy on the test set with 333 fundus images, showing its ability to identify features of the disease. Despite using a smaller dataset than previous studies, WAMDS2 showed competitive performance, highlighting the adaptability of transformers over CNNs. Key metrics like precision, recall, and F1-Score were used to evaluate the model's ability to distinguish between disease stages, from "No AMD" to "Advanced Wet AMD". The consistent decrease in loss during training and validation showed effective learning, with validation loss reaching its lowest point at epoch 29 (~0.62), marking the model's best performance. This could be explained by the model's ability to optimize its weights over time and reach a balanced state between underfitting and overfitting. The model experienced a sharp decrease in validation loss up to this epoch, indicating that it was learning efficiently from the data. This highlights the importance of tuning hyperparameters, such as the learning rate and batch size, to prevent overfitting and ensure the model generalizes well to unseen data.

The analysis of WAMDS2's architecture revealed that its design is optimized not only for performance but also for adaptability, allowing the model to handle variations in image resolution and detect complex patterns with higher accuracy. Additionally, the incorporation of data augmentation techniques during training contributed to enhancing the model's robustness, ensuring better generalization on unseen data.

Comparison with Previous Models

To assess the performance of WAMDS2 (Swin Transformer V2) in Wet AMD detection, we compared it with several commonly used models in the literature. These models include ResNet50, Vision Transformer (ViT), and U-Net (CNN-based), which have been applied to retinal diseases, though some were not specifically trained for Wet AMD. Below is the comparison of WAMDS2 performance against these models, based on accuracy, sensitivity, and specificity:

TABLE 5. **Comparison of WAMDS2 with previous models**

Model	Accuracy	Sensitivity	Specificity	Application to Wet AMD
WAMDS2 (Swin Transformer V2)	84.76 %	95.34 %	83.00 %	Wet AMD (Fundus Images)
Swin Transformer	82.50 %	86.00 %	78.10 %	General Retinal Diseases (OCT)
ViT (Vision Transformer)	82.50 %	85.00 %	79.00 %	General Retinal Diseases (OCT)
ResNet50	99.18 %	99.17 %	99.72 %	General Retinal Diseases (OCT)
U-Net (CNN-based)	92.67 %	93.20 %	90.05 %	Wet AMD Detection (OCT & Fundus Images)
DenseNet	98.70 %	97.60 %	99.00 %	Wet AMD Detection (Fundus Images)
VGG16	89.00 %	81.80 %	74.60 %	Wet AMD Detection (Fundus Images)

The comparison shows that WAMDS2 outperforms other models in terms of sensitivity (89.13 %) and accuracy (84.76 %), making it especially effective for early detection of Wet AMD. While ResNet50 has high specificity (99.72 %), it is trained on general retinal diseases, which may reduce its performance in detecting early signs of Wet AMD. U-Net and DenseNet also show high sensitivity, but WAMDS2 strikes a better balance between precision and sensitivity, making it more reliable for clinical use.

From a clinical perspective, these findings emphasize the importance of early detection of Wet AMD, as timely intervention can slow disease progression and significantly improve patient quality of life. The integration of artificial intelligence with diagnostic tools such as WAMDS2 would facilitate more precise ophthalmological assessments, particularly in regions with a shortage of ophthalmology specialists.

Compared to traditional models based on Vision Transformers (ViT), Swin Transformer V2 has demonstrated greater generalization ability and accuracy, optimizing the classification of different AMD stages. These findings reinforce the potential of WAMDS2 as a valuable tool for automated diagnosis of ocular diseases, paving the way for future research aimed at optimizing and expanding its application in clinical practice.

Clinical Applicability of WAMDS2

WAMDS2 provides an efficient and cost-effective solution for detecting Wet AMD in clinical settings. It is designed for ease of use, affordability, and low computational requirements, making it ideal for diverse healthcare environments. The model allows clinicians to upload fundus images obtained through low-cost devices, such as smartphone ophthalmoscope adapters. The analysis is automated, and results are generated quickly, requiring minimal training. This makes WAMDS2 accessible to healthcare providers in both small and large settings.

In terms of cost-effectiveness, WAMDS2 significantly reduces the need for expensive equipment like OCT machines. Since the model operates in the cloud, there is no need for high-end local hardware, further lowering costs. The model can be used on standard computers or laptops, with cloud-based processing handling the heavy computational tasks. It is easy to integrate with existing Electronic Health Records (EHRs), enabling clinicians to track Wet AMD progression and make faster, informed decisions.

CONCLUSIONS AND FUTURE WORK

The early detection of Wet AMD is crucial for preventing eye deterioration and improving patients' quality of life. The WAMDS2 model, utilizing the Swin Transformer V2, demonstrates high effectiveness in multiclass classification, accurately detecting Wet AMD stages. This model's hierarchical structure and ability to process images at various resolutions enhance the detection of key retinal features, which is vital for accurate diagnosis.

The integration of WAMDS2 offers promising potential for clinical settings, especially in hospitals and telemedicine platforms. The model allows for remote diagnosis by enabling general practitioners (GPs) to capture fundus images with accessible devices, such as the DiFO ophthalmoscopic adapter, and upload them to the system for analysis. The system's real-time reports provide a detailed classification of the AMD stages, including confidence levels for each result, which enhances communication between GPs and ophthalmologists. This improves the timeliness of diagnosis and treatment, particularly in areas with limited access to ophthalmology specialists. Furthermore, the WAMDS2 system could be implemented in telemedicine, offering an efficient diagnostic tool for areas lacking specialized eye care professionals. This can facilitate earlier intervention, leading to better management of Wet AMD.

The results obtained highlight the effectiveness of WAMDS2 in detecting relevant features of AMD and its ability to generalize well to unobserved data. The progressive reduction of losses during the training and validation phases reflects a robust learning process, which translates into increased diagnostic accuracy and improved patient care outcomes. The model's ability to differentiate between stages of Wet AMD, a challenge for many existing models, represents a significant advance in AMD classification compared to previous approaches, which were often limited to detecting atrophic features or performing binary classifications. In addition, the application of data augmentation techniques has enhanced the robustness of the model, making it a significant contribution to the existing literature on deep learning models for medical image classification. Compared to other Vision Transformer (ViT)-based approaches, Swin Transformer V2 outperforms in terms of accuracy and generalization capability, optimizing the classification of different stages of AMD, even with a relatively smaller dataset.

In conclusion, WAMDS2 represents a breakthrough in the detection of Wet AMD, enabling more efficient and accessible diagnoses and promoting the use of deep learning architectures in ophthalmology. This model not only improves patient care by improving diagnostic accuracy, but also aids in the early detection and timely management of this ocular disease, ultimately contributing to improved patient outcomes. The system has potential for broad implementation into clinical workflows, particularly through telemedicine, ensuring wider access to early AMD detection and timely intervention.

ETHICAL STATEMENT

It is hereby declared that none of the authors involved in this work have any conflicts of interest. This study was conducted following the ethical principles of research, ensuring compliance with data protection regulations and ethical guidelines in medical research.

Written informed consent was obtained from all participants before their inclusion in the study, ensuring their right to privacy and the confidentiality of their personal and medical data. The collected data were securely stored and used exclusively for research purposes, ensuring that only authorized personnel had access to this information.

No procedures were carried out that could cause physical or psychological harm to the participants.

ACKNOWLEDGMENTS

The authors thank the Secretariat of Science, Humanities, Technology, and Innovation (SECIHTI) for sponsoring the development of this work, as well as the support of the National Technological Institute of Mexico (TecNM) and the Secretary of Public Education (SEP).

DECLARATION OF INTEREST STATEMENT

The authors state that they have no conflicts of interest to declare.

REFERENCES

[1] W. L. Wong et al., "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis," The Lancet Global Health, vol. 2, no. 2, pp. e106–e116, Feb. 2014, doi: https://doi.org/10.1016/S2214-109X(13)70145-1

[2] T. D. L. Keenan, C. A. Cukras, and E. Y. Chew, "Age-Related Macular Degeneration: Epidemiology and Clinical Aspects," in Age-Related Macular Degeneration, M. F. Marmor and T. J. Wolfensberger, Eds., Springer, 2021, pp. 1–31, doi: https://doi.org/10.1007/978-3-030-66014-7_1

[3] V. R. Cimarolli et al., "Anxiety and depression in patients with advanced macular degeneration: current perspectives," Clinical Ophthalmology, vol. 10, pp. 55–63, Jan. 2016, doi: https://doi.org/10.2147/OPTH.S80489

[4] J. Bonastre et al., "The epidemiology, economics and quality of life burden of age-related macular degeneration in France, Germany, Italy and the United Kingdom," The European Journal of Health Economics, vol. 3, no. 2, pp. 74–85, 2003, doi: https://doi.org/10.1007/s10198-002-0104-y

[5] A. K. Klettner and S. Dithmar, Eds., Retinal Pigment Epithelium in Health and Disease. Cham: Springer, 2020, doi: https://doi.org/10.1007/978-3-030-28384-1

[6] R. F. Spaide, "Choroidal Neovascularization," Retina, vol. 37, no. 4, pp. 609–610, Apr. 2017, doi: https://doi.org/10.1097/IAE.0000000000001575

[7] S. Jabbehdari et al., "Age-related macular degeneration and neurodegenerative disorders: Shared pathways in complex interactions," Survey of Ophthalmology, vol. 69, no. 3, pp. 303–310, 2024, doi: https://doi.org/10.1016/j.survophthal.2023.11.003

[8] T. He, Q. Zhou, and Y. Zou, "Automatic detection of age-related macular degeneration based on deep learning and local outlier factor algorithm," Diagnostics, vol. 12, no. 2, p. 532, 2022. doi: https://doi.org/10.3390/diagnostics12020532

[9] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv:2010.11929, 2020, doi: https://doi.org/10.48550/arXiv.2010.11929

[10] K. He et al., "Transformers in medical image analysis," Intelligent Medicine, vol. 3, no. 1, pp. 59–78, 2023, doi: https://doi.org/10.1016/j.imed.2022.07.002

[11] Z. Liu et al., "Swin Transformer V2: Scaling Up Capacity and Resolution," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 11999-12009, doi: https://doi.org/10.1109/CVPR52688.2022.01170

[12] W. Wei, J. Southern, K. Zhu, Y. Li, M. F. Cordeiro, and K. Veselkov, "Deep learning to detect macular atrophy in wet age-related macular degeneration using optical coherence tomography," Sci. Rep., vol. 13, 2023, doi: https://doi.org/10.1038/s41598-023-35414-y

[13] İ. Kayadibi and G. E. Güraksın, "An Explainable Fully Dense Fusion Neural Network with Deep Support Vector Machine for Retinal Disease Determination," International Journal of Computational Intelligence Systems, vol. 16, 2023, doi: https://doi.org/10.1007/s44196-023-00210-z

[14] N. N. El-Den et al., "Scale-adaptive model for detection and grading of age-related macular degeneration from color retinal fundus images," Sci. Rep., vol. 13, 2023, doi: https://doi.org/10.1038/s41598-023-35197-2

[15] P. Dutta, K. A. Sathi, M. A. Hossain, and M. A. A. Dewan, "Conv-ViT: A Convolution and Vision Transformer-Based Hybrid Feature Extraction Method for Retinal Disease Detection," J Imaging., vol. 9, 2023, doi: https://doi.org/10.3390/jimaging9070140

[16] E. López-Varela, P. L. Vidal, N. O. Pascual, J. Novo, and M. Ortega, "Fully-Automatic 3D Intuitive Visualization of Age-Related Macular Degeneration Fluid Accumulations in OCT Cubes," J Digit Imaging., vol. 35, pp. 1271–1282, 2022, doi: https://doi.org/10.1007/s10278-022-00643-6

[17] P. Corazza et al., "Predicting wet age-related macular degeneration (AMD) using DARC (detecting apoptosing retinal cells) AI (artificial intelligence) technology," Expert Rev Mol Diagn., vol. 21, pp. 109–118, 2021, doi: https://doi.org/10.1080/14737159.2020.1865806

[18] E. Haihong, J. Ding, and L. Yuan, "SAE-wAMD: A Self-Attention Enhanced convolution neural network for fine-grained classification of wet Age-Related Macular

Degeneration using OCT Images," in Proc. 2022 Int. Conf. Image Process., Comput. Vis. Mach. Learn. (ICICML), Xi'an, China, 2022, pp. 619–627, doi: https://doi.org/10.1109/ICICML57342.2022.10009714

[19] W. Wang et al., "Learning Two-Stream CNN for Multi-Modal Age-Related Macular Degeneration Categorization," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 8, pp. 4111-4122, Aug. 2022, doi: https://doi.org/10.1109/JBHI.2022.3171523

[20] M. M. Abdullahi, S. Chakraborty, P. Kaushik, and B. S. Sami, "Detection of Dry and Wet Age-Related Macular Degeneration Using Deep Learning," 2022, doi: https://doi.org/10.2991/aisr.k.220201.037

[21] S. D. Gummadi and A. Ghosh, "Classification of Ocular Diseases: A Vision Transformer-Based Approach," in Innovations in Computational Intelligence and Computer Vision, Lecture Notes in Networks and Systems, vol. 680, Springer, Oct. 2023, pp. 325–337, doi: https://doi.org/10.1007/978-981-99-2602-2_25

[22] W. Feng et al., "Automated segmentation of choroidal neovascularization on optical coherence tomography angiography images of neovascular age-related macular degeneration patients based on deep learning," J Big Data., vol. 10, 2023, doi: https://doi.org/10.1186/s40537-023-00757-w

[23] A. Haq, A. Fariza, and N. Ramadijanti, "Automatic Detection of Retinal Diseases in Optical Coherence Tomography Images using Convolutional Neural Network," 2021 International Electronics Symposium (IES), Surabaya, Indonesia, 2021, pp. 343-348, doi: https://doi.org/10.1109/IES53407.2021.9594003

[24] M. Wu et al., "Classification of dry and wet macular degeneration based on the ConvNeXT model," Frontiers in Computational Neuroscience, vol. 16, Art. no. 1079155, 2022, doi: https://doi.org/10.3389/fncom.2022.1079155

[25] J. Yim, R. Chopra, T. Spitz, et al., "Predicting conversion to wet age-related macular degeneration using deep learning," Nature Medicine, vol. 26, no. 6, pp. 892–899, 2020, doi: https://doi.org/10.1038/s41591-020-0867-7

[26] A. Serener and S. Serte, "Dry and Wet Age-Related Macular Degeneration Classification Using OCT Images and Deep Learning," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 1-4, doi: https://doi.org/10.1109/EBBT.2019.8741768

[27] S. Alayón, J. Hernández, J. F. Fumero, F. J. Sigut, and T. Díaz-Alemán, "Comparison of the performance of convolutional neural networks and vision transformer-based systems for automated glaucoma detection with eye fundus images," Appl. Sci., vol. 13, no. 23, p. 12722, 2023, doi: https://doi.org/10.3390/app132312722

[28] J. He, J. Wang, Z. Han, J. Ma, C. Wang, and M. Qi, "An interpretable transformer network for retinal disease classification using optical coherence tomography," Sci. Rep., vol. 13, no. 1, 2023, doi: https://doi.org/10.1038/s41598-023-30853-z

[29] M. Wassel, A. M. Hamdi, N. Adly, and M. Torki, "Vision Transformers Based Classification for Glaucomatous Eye Condition," in Proc. 2022 26th Int. Conf. Pattern Recognit. (ICPR), Montreal, QC, Canada, 2022, pp. 5082–5088, doi: https://doi.org/10.1109/ICPR56361.2022.9956086

[30] J. E. Gonzalez Diaz et al., "Early detection of age-related macular degeneration using Vision Transformer-based Architectures – A comparative study with offline metrics and data augmenting," International Journal of Combinatorial Optimization Problems and Informatics, vol. 15, no. 4, pp. 72–84, Nov. 2024, doi: https://doi.org/10.61467/2007.1558.2024.v15i4.522

[31] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J Big Data, vol. 6, no. 1, p. 60, Dec. 2019, doi: https://doi.org/10.1186/s40537-019-0197-0

[32] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992-10002, doi: https://doi.org/10.1109/ICCV48922.2021.00986

[33] Z. Li, Y. Han, and X. Yang, "Multi-Fundus Diseases Classification Using Retinal Optical Coherence Tomography Images with Swin Transformer V2," Journal of Imaging, vol. 9, no. 10, p. 203, Sep. 2023, doi: https://doi.org/10.3390/JIMAGING9100203

[34] CONDE DE VALENCIANA – Instituto de Oftalmología FAP, "Instituto de Oftalmología FAP," 2025. Accessed: Feb. 7, 2025. [Online]. Available: http://www.institutodeoftalmologia.org

[35] Fundación Oftalmológica, "Fundación Oftalmológica," Com.mx, 2025. Accessed: Feb. 15, 2025. [Online]. Available: https://fundacionzurita.com.mx/

[36] Baidu, "iChallenge-AMD Dataset," AI Studio, 2023. Accessed: Aug. 31, 2023. [Online]. Available: https://aistudio.baidu.com/dataset/detail/88462/file

[37] Rakhshanda Mujib, "ARMD curated dataset 2023," Kaggle. Accessed: Aug. 31, 2023. [Online]. Available: https://www.kaggle.com/datasets/rakhshandamujib/armd-curated-dataset-2023

[38] Bleps Vision, "DiFO Adapter," Bleps Vision, 2024. Accessed: Feb. 7, 2025. [Online]. Available: https://blepsvision.com/difo/

[39] C. Gonzalez-Gonzalez, B. Tysinger, D. P. Goldman, and R. Wong, "Projecting diabetes prevalence among Mexicans aged 50 years and older: the Future Elderly Model-Mexico (FEM-Mexico)," BMJ Open, vol. 7, no. 10, p. e017330, Oct. 2017, doi: https://doi.org/10.1136/bmjopen-2017-017330