Digital Object Identifier

Graph Attention Mechanisms For Macular Degeneration Disease Diagnosis Using Self-Routing Capsule Knowledge Distillation

ABONYO MITCHELL NINA, ID-0009-0001-4428-9593[1], NAKIGULI ELIZABETH GLORIA, ID-0009-0005-4537-0878[2]
[1]College of Computing and Information Sciences, Makerere University, Uganda, Kampala (mtchllnina@gmail.com)
[2]College of Computing and Information Sciences, Makerere University, Uganda, Kampala (nakiguligloria5@gmail.com)

ABSTRACT Macular degeneration, commonly referred to as Age-related Macular Degeneration (AMD), is one of the leading causes of vision-related complications particularly in older people. The technique that is used for diagnosing macular degeneration is known as Optical Coherence Tomography (OCT) which works by capturing detailed images of the retina to aid doctors in identifying any AMD-related changes. However, the diagnosis and treatment of macular degeneration is costly and interpretation of the OCT scans requires trained personnel, who are scarce in many rural regions of Uganda. To address these challenges, a Graph Attention Self-Routing Capsule Knowledge Distillation framework which integrates graph attention networks, self-routing capsule networks, knowledge distillation and vision transformers was developed to aid the early detection of macular degeneration through the automatic interpretation of OCT scans. The graph attention networks enhanced feature extraction by capturing complex relationships between retinal structures while the vision transformers improved segmentation accuracy by leveraging global and local dependencies in the OCT scan images. Additionally, knowledge distillation compressed the high-performance larger models into a smaller models that detected complex patterns in images, such as retinal layer deformations while the capsule networks employed their self-attention mechanisms to highlight important details of the OCT scan image in order to identify any lesions such as the sub retinal fluid. The study evaluated multiple segmentation models such as the U-Net, U-Net with Vision Transformers, U-Net with Graph Attention Mechanisms, ResU-Net model, RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation, the ATT_CNN_KD model integrated with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation, U-Net++_KD model with knowledge distillation and the U-Net++_CNN_KD model integrated with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation. Among these, the the RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation emerged as the best-performing model. It was concluded that it was the most suitable choice for real-world clinical applications since it outperformed the other models by achieving the lowest distillation loss of 0.0002, highest dice score of 0.9815 and had a longer training duration of 17 epochs which suggested greater robustness over the other models. This macular degeneration disease lesion detection framework is therefore expected to enhance diagnostic cost-effectiveness and improve early detection of the disease not only among the older patients but among all the various age groups of individuals. Future work will focus on improving model generalization and efficiency in order to ensure accessibility in regions with limited computational resources such as the rural parts of Uganda.

INDEX TERMS Age-Related Macular Degeneration, Optical Coherence Tomography, Graph Attention Networks, Capsule Networks, Knowledge Distillation, Vision Transformers.

I. BACKGROUND AND INTRODUCTION

VISION loss due to Age-Related Macular Degeneration (AMD) is a critical global challenge [1], [2], [3]. The number of AMD patients worldwide was estimated at 196 million in 2020 and is expected to reach 288 million by 2040 [2]. This paper explores how machine learning can improve early detection, enhance diagnosis accuracy and the overall management of macular degeneration.

Age-Related Macular Degeneration (AMD) is the leading cause of central vision loss mainly among individuals over the age of fifty primarily affecting the macula, the central part of the retina which is responsible for detailed vision [4], [5]. AMD is of two types, that is, dry AMD which progresses slowly due to thinning of the macula, and wet AMD which is characterized by the growth of abnormal blood vessels under the retina that leak fluid. [6], [7], [8]. The most common symptoms of macular degeneration include blurred central vision, difficulty reading and even difficulty in recognizing faces with the key risk factors including aging, family history, smoking and poor diet [9], [10]. One of the ways of managing AMD and preserving vision is early detection which can be done through regular eye exams. The key tool for diagnosing and monitoring AMD is Optical Coherence Tomography(OCT) [11].

Optical Coherence Tomography (OCT) is the non invasive imaging technique that provides high-resolution cross-sectional images of the retina [11], [12]. It enables doctors to detect subtle changes in the macula, assess the thickness of the retinal layers and monitor the progression of both the dry and wet forms of macular degeneration. By accurately mapping and measuring the thickness of the retina's layers, OCT provides detailed visualizations of retinal changes which aids in distinguishing between the various stages of macular degeneration and therefore assists ophthalmologists in diagnosing macular degeneration and providing timely and appropriate treatment to patients [13], [14]. In clinical practice, this technology has surpassed traditional imaging methods for diagnosing macular degeneration and has become essential for the early diagnosis and treatment planning of macular degeneration.

Graph Attention Networks (GATs) also play a significant role in the diagnosis and treatment of macular degeneration. GATs learn the relationship between various regions of the retina by assigning different attention weights to connected nodes which gives them the capability of handling complex graph-structured data, such as retinal images. By considering the importance of different data points, GATs assist ophthalmologists in the early detection of macular degeneration. [15] [16].

Capsule knowledge distillation is a technique that integrates the use of capsule networks and knowledge distillation that can be used to enhance diagnostic systems for macular degeneration [17]. In macular degeneration, where precise identification of subtle changes in the retina is critical, the capsule networks' ability to detect complex patterns in images, such as retinal layer deformations and fluid accumulation, is especially valuable [14], [18]. These capsule networks also use their self-routing mechanisms to analyze the OCT scan images and decide which parts of the image are most important to focus on. Instead of following fixed paths like regular neural networks, they highlight the key features and organize these features into capsules, which capture both what the object is and where it is which in turn would help make accurate predictions regarding the stage of the macular degeneration disease. Additionally, the implementation of knowledge distillation where a larger well-trained "teacher" model transfers learned knowledge to a smaller "student" model [19] enables healthcare providers to screen and diagnose AMD with high accuracy and reduce computational costs even in resource constrained environments such as the rural parts of Uganda.

Vision Transformers is a deep learning approach that applies the principles of transformers in image recognition tasks [20]. They treat images as sequences of patches by applying transformer based mechanisms to capture the global relationships across these image patches [20], [21]. In the context of macular detection, vision transformers enhance the analysis of retinal images in the OCT scans by capturing subtle changes in the retina that could be an indicator of early-stage AMD.

II. LITERATURE REVIEW

The recent advancements in Age-Related Macular Degeneration (AMD) research have integrated various deep learning techniques to enhance the accuracy in prediction of the disease [22], [23]. A notable development is the application of vision transformers to analyze OCT scans. For instance, the DeepDrAMD model which employs a hierarchical vision transformer-based deep learning approach that employs data augmentation techniques in the detection of AMD [24], [25].This model has demonstrated high performance in detecting macular degeneration and distinguishing between its various forms. We also have the ViT-AMD model which divides images into small patches and processes them with self-attention mechanisms to capture long-range image dependencies. [26]

Additionally, the Graph Attention U-Net has been introduced for the precise segmentation of retinal layers in the OCT scan images in order to address the challenges posed by retinal layer deformation [15], [27]. It employs a graph de-correlation module to remove information unrelated to retinal layers, thereby improving retinal surface detection. By analyzing the OCT scans and effectively modeling complex relationships within diverse data types, the graph attention networks contribute to more accurate diagnostics, a deeper understanding of the disease mechanisms, and the development of personalized treatment strategies for AMD [28].

There has also been the integration of adversarial learning techniques in which models are trained by introducing "adversarial examples" to mislead the model which gives it the ability to withstand a wide range of conditions and still function properly [29]. This has been done to ensure

that automated diagnostic systems for macular degeneration can maintain high accuracy even in the presence of such distortions [30]. Adversarial learning has also been employed to improve the quality of OCT images through the use of unsupervised super-resolution techniques which has facilitated better visualization of retinal layers [31].

Knowledge distillation has also proven beneficial for the detection of AMD. AMD detection which typically uses Optical Coherence Tomography (OCT) scans can be computationally expensive. This makes knowledge distillation a useful approach because of the reduced computational costs associated with it. A study highlights the application of knowledge distillation in medical image classification tasks, demonstrating its ability to transfer knowledge from larger models to smaller models while maintaining high accuracy, improving model performance and enhancing model generalization even when data sets are limited. [32].

In addition to the above, capsule networks have also been effectively utilized in detecting macular degeneration through the analysis of the retinal OCT scan images. A notable study introduced the CLAHE-CapsNet model, which integrates capsule networks with Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance image clarity [18]. This model demonstrated superior performance in retinal OCT image classification which suggested that capsule networks are highly effective in identifying retinal diseases such as macular degeneration. Furthermore, the application of capsule knowledge distillation which combines the use of capsule networks and knowledge distillation has enhanced the accuracy in the diagnosis of macular degeneration. For instance, the Synergic Adversarial Label Learning (SALL) method leverages knowledge distillation and multitask learning to improve the grading of retinal diseases, including AMD. By utilizing additional monitoring signals from various sources, SALL trains a robust model with less data which effectively enhances diagnostic accuracy [33], [34].

Overall, these studies highlight how the various machine learning techniques are improving the diagnosis of macular degeneration. By leveraging advanced architectures such as vision transformers and graph-based neural networks, researchers are addressing complex challenges in retinal imaging, leading to more accurate diagnostics and personalized treatment strategies. This ensures that quality healthcare remains accessible thus bridging the gap between advanced diagnostic capabilities and the resource-constrained environments such as rural Uganda.

III. PROBLEM STATEMENT

Macular Degeneration is among the leading causes of irreversible blindness in older adults worldwide [35]. Despite the various advancements in medical technology, the early detection and accurate diagnosis of macular degeneration remains a significant challenge [35], [36]. The complexity of analyzing retinal images from Optical Coherence Tomography (OCT) scans and variation in the disease progression among patients makes it difficult for health-care providers to provide timely and personalized treatments. This work aims to address these issues through the development of a machine learning-based system that can automate the early detection of macular degeneration and accurately predict the progression of the disease by leveraging deep learning techniques such as graph attention networks, capsule networks, knowledge distillation, and vision transformers.

The major challenge is developing a model that can efficiently handle variations in image quality and patient demographics such as age, ethnicity and gender. Additionally, it must efficiently handle large volumes of imaging data and generalize across diverse patient populations while ensuring real-time performance. To achieve all this, the system will incorporate advanced techniques such as adversarial learning while leveraging self-supervised learning methodologies that will further enhance the system's ability to recognize early changes in the retina which will ultimately contribute to the early detection of AMD.

This system will improve the accuracy of diagnoses, enhance treatment planning and reduce the burden of AMD-related vision loss particularly in regions with limited access to health-care providers and diagnostic facilities.

IV. PROPOSED SOLUTION

Macular degeneration has presented significant challenges in its early diagnosis due to the subtle nature of its early stage symptoms. To address these challenges, a machine-learning based solution that integrates vision transformers, graph attention networks, adversarial learning, capsule networks and knowledge distillation has been developed.

The vision transformers that demonstrate exceptional ability in capturing global contextual information from images will enhance the analysis of Optical Coherence Tomography (OCT) images to identify early stage AMD with high accuracy as demonstrated by models like ViT-AMD [26]. Meanwhile, the Graph Attention Networks will improve retinal layer segmentation by treating the OCT images as graphs by dynamically weighing regions of interest for more precise segmentation, as seen in more recent studies [15] [37].The capsule networks which are designed to capture spatial relationships between features and preserve the important structural information through dynamic routing between capsules will use their self-attention mechanism to highlight important details of the OCT scan image in order to identify the lesion areas of the retinal image [17].

Collectively, these techniques will create an approach that will improve the detection accuracy of macular degeneration thus leading to better patient outcomes.

A. METHODOLOGY

After establishing the problem to be solved, the steps involved in the implementation of our framework begun with the selection of the dataset, preprocessing of the data, model development and training and finally evaluation and deployment of our best performing model as shown in figure 1. Reti-

nal images were sourced from the publicly available Age-Related Macular Degeneration Segmentation Dataset (AMD-SD) [38]. Since these retinal images vary in resolution and quality, preprocessing was an essential step that involved resizing images and normalizing pixel values. Additionally, image augmentation techniques, including rotation and flipping were carried out in order to improve model generalization. The processed images were then further split into training and validation sets to ensure proper model evaluation.

Following preprocessing, the images were passed through a Vision Transformer (ViT) model for feature extraction where images were transferred into fixed-size patches as shown in figure 3. Here, each patch was embedded into a high-dimensional space using positional encodings and a self-attention mechanism which allowed the model to capture complex spatial relationships across different retinal regions. A Graph Attention Network (GAT) was then employed to enhance the extracted features. Here, the retinal image was transformed into a graph representation where each node corresponded to a region of interest, and edges defined relationships between these regions. The attention mechanism in the graph attention networks assigned different importance levels to these connections, which enabled the model to prioritize critical areas such as lesions and other early indicators of macular degeneration. This approach allowed the model to understand how different retinal regions interact which improved its ability to detect the early signs of AMD. Capsule Networks were also integrated in the development of our model as shown in figure 4. They utilized their self-routing mechanisms to enhance the analysis of Optical Coherence Tomography (OCT) image scans by preserving spatial hierarchies and capturing the complex structural relationships in the retinal image in order to identify subtle deformations in the retinal layers that indicate the presence of macular degeneration.

Since vision transformers and graph attention networks are computationally intensive, knowledge distillation was used to create a lightweight yet effective model. In this approach, a large, complex model was trained first, and a smaller model learned from it as shown in figure 2. To further enhance the distillation process, capsule knowledge distillation was employed. Capsule knowledge distillation which combines both capsule networks and knowledge distillation, ensured that the student model learnt richer representations of the OCT scan by preserving the structural information from the teacher model. This allowed the student model to replicate the teacher's performance while maintaining a much smaller architecture and ensuring that there is proper information flow such that there is proper identification and classification of specific features in the retinal scan images. This ensured better feature retention and made the model more effective for detecting subtle retinal abnormalities associated with macular degeneration.

Following training, the models underwent a thorough evaluation process to assess their performance based on various metrics such as accuracy, dice score, intersection over union(IOU), distillation loss and Haurdsoff distance. The best-performing model was then carefully selected and optimized for deployment. Once fully prepared, the model will then be integrated into a real-time system which will assist healthcare professionals in identifying early signs of macular degeneration.

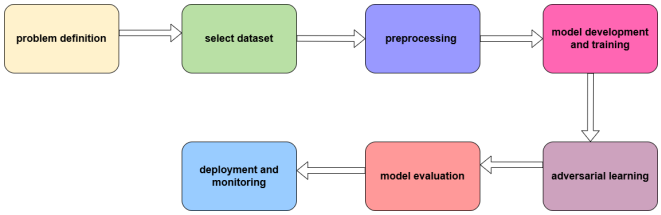


FIGURE 1. Methodology flowchart showing the steps involved in the development of our model.

Figure 1 illustrates the process that was followed during the implementation of our framework from the problem definition, selection of the dataset to be used, preprocessing of the data, model training, model evaluation and finally model deployment.

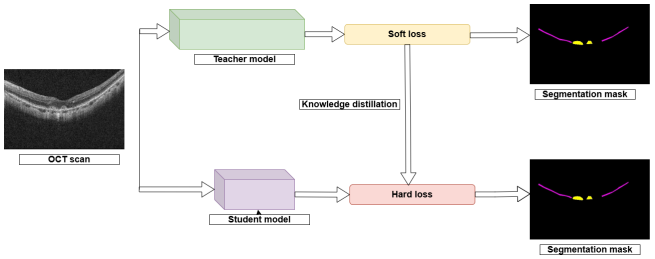


FIGURE 2. A diagram showing the knowledge distillation process.

Figure 2 represents the knowledge distillation process in which an Optical Coherence Tomography scan image was fed into the teacher and student models. The teacher model then generated a segmentation mask and provided soft labels as additional supervision for the student model. The student model then learnt from both the teacher's outputs (soft loss) and the ground truth labels (hard loss).

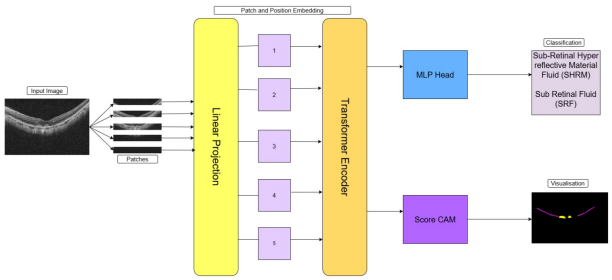


FIGURE 3. A diagram showing the working of vision transformers.

Figure 3 illustrates how vision transformers processed the OCT scan images for the classification of macular degeneration. They divided the scan image into patches, applied

linear projection and positional embeddings, and then used a transformer encoder to extract features. The MLP head then classified the image into segmentation mask categories while the Score-CAM highlighted the important regions for interpretation.

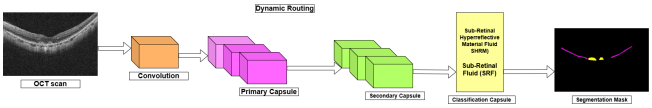


FIGURE 4. A diagram showing the working of capsule networks.

Figure 4 illustrates the architecture of the capsule network for analyzing OCT scans. The process began with a raw OCT scan, which was processed through a series of convolutional layers and capsule layers (primary and secondary). Dynamic routing was then employed to facilitate information flow between these layers leading to the accurate identification and classification of specific features in the OCT scan image.

V. DATASET DESCRIPTION

Our model was trained on the Age-Related Macular Degeneration Segmentation Dataset (AMD-SD) [38], which is a compressed file. This data set which was obtained from 138 patients, that is to say, 67 female patients and 77 male patients, comprised a total of 3,049 OCT B-scan images that had a resolution of 570x380 pixels along with their corresponding lesion area segmentation masks. Upon extraction, the dataset was organized into four components, a folder which contained a variety of images ("Images"), a Microsoft excel file which contained demographic data ("Demographics of participants"), a training text file ("training") and a validation text file ("validation").

The "Images" folder contained 156 sub-folders, each corresponding to a unique eye ID and within each sub-folder, were Optical Coherence Tomography (OCT) image scans with their corresponding segmentation masks. Images were named in the form "n_x.png", where "n" represented eye ID and "x" represented the image number. The Microsoft Excel file contained patient demographic information that was organized into six columns such as patient ID, eye ID, eye category, age, sex, and time acquired (indicating when the eye scan was taken). The "training" and "validation" text files contained specific image names for model training and testing respectively, following an 80:20 split of the dataset. Each OCT B- scan segmentation mask was a six-color image where different colors represented distinct features. These colors included red pixels that represented sub retinal fluid (SRF), pink pixels that represented the inner segment / outer segment (IS/OS), blue pixels that represented pigment epithelium detachment (PED), green pixels that represented intraretinal fluid (IRF), yellow pixels that represented sub retinal hyperreflective material (SHRM) and finally the black color that represented the background of the OCT scan.

TABLE 1. A table summarizing the AMD-SD dataset characteristics.

Characteristic	Details
Dataset Name	AMD-SD Dataset
Number of patients	138 patients
Number of images	3049 OCT B-scan images
Dataset Components	Images folder, Demographics file, Training text file, Validation text file
Demographics considered	Patient ID, Eye ID, Eye category, Age, Sex, Time acquired
Color representations	Red-SRF, Pink-IS/OS, Blue-PED, Green-IRF, Yellow-SHRM, Black-Background

VI. EXPLANATORY DATA ANALYSIS

The exploratory data analysis (EDA) provided crucial insights that directly related to our goal of detecting macular degeneration. Since AMD predominantly affects older adults, the dataset presented an age distribution that is approximately normal, with a mean age of 67 years and a standard deviation of 9 years. The population was primarily composed of older adults, with the highest concentration of patients being in the age range of 70 years as shown in figure 5. In regards to the sex distribution, an initial analysis revealed a significant imbalance, with a higher count of male patients compared to the female patients as shown in figure 7. However, this was expected to introduce bias in model predictions. To address this, we applied oversampling techniques to achieve a balanced representation ensuring that our findings and model generalizations remained valid for both sexes as shown in figure 8. Additionally, the dataset included information on the timing of eye scans 9, where a majority of the patients fall under the "naive" group, meaning that they had no prior treatment before the scan. This distribution suggested that our model would primarily learn from the early-stage AMD cases, but it may have limited ability to differentiate between untreated and previously treated conditions due to the under representation of the latter group.

When considering the distribution of eye categories, the dataset was well-balanced between Oculus Sinister (OS) and Oculus Dexter (OD) ensuring that analyses were not skewed towards one eye as shown in figure 10. Furthermore, in figure 11 we see the stratification of age distribution by sex which revealed a slightly higher median age for males patients than the female patients. The sample OCT B-scan images and their corresponding segmentation masks further illustrated the presence of different retinal lesions with intraretinal fluid (IRF) being the least frequent lesion type and pigment epithelial detachment (PED) being the most common as indicated by its high pixel count in segmentation masks as shown in figure 12.

A detailed analysis of lesion co-occurrence, visualized through heat maps and stacked bar charts, highlighted strong associations between certain lesion types such as the sub retinal hyperreflective material (SHRM) and pigment epithelial detachment (PED). Similarly, sub retinal fluid (SRF) was frequently observed alongside PED and SHRM, while intraretinal fluid exhibited lower co-occurrence rates with the other lesions. The percentage distribution of lesion types across all images confirmed that PED was the most dominant lesion, whereas IRF was the least represented. A normalized lesion co-occurrence heat map further emphasized the strong associations between PED and SHRM (0.49), as well as SRF with PED (0.34). This provided valuable cues for automated detection and thus our model was optimized to recognize not just individual lesion types but also their co-occurrence patterns to improve diagnostic accuracy.

Finally in figure 17, an analysis of demographic correlations revealed strong positive associations between patient ID and eye ID, while weak correlations were observed between age and both patient and eye IDs. The weak correlation between age and other variables suggested that age alone may not be a strong predictor of lesion presence. A weak negative correlation (-0.17) was noted between eye category and the time when the scan was acquired which also suggested that treatment timing played a minor role in lesion distribution. Additionally, sex showed minimal association with other demographic variables in the dataset. By understanding the findings and relationships from our EDA, we were able to gain insight on the composition and potential biases of the dataset and were therefore able to refine our model training strategies to ensure robust and unbiased detection of macular degeneration.

Overall, the age distribution in the dataset aligned with the target population and the prevalence of pigment epithelial detachment (PED) as the most common lesion highlighted its importance in the diagnosis of AMD. Finally the co-occurrence of lesions all suggested that our model should be able to detect multiple lesions at once. These findings helped shape our work by providing insights into the dataset's key characteristics thereby guiding us in developing a more accurate and effective model for the early detection of macular degeneration.

A. EXPLANATORY DATA ANALYSIS RESULTS

Figure 5 represents the age distribution of patients in the dataset. The data revealed a mean age of 67 years, standard deviation age of 9 years and a roughly normal distribution, with the highest concentration of patients being in the 70-year age range.

TABLE 2. A table summarizing the EDA key findings.

Aspect	Key Findings
Age	Mean age is 67 years
Sex Distribution	Males > Females
Eye Scan Timing	Majority are in naive group (no prior treatment)
Eye Category	OS-right eye, OD-left eye
Lesion Distribution	PED is the most common, IRF is the least common
Lesion Co-occurrence	PED and SHRM (0.49), PED and SRF(0.34)
Demographic Correlation	Weak correlation between age and other variables

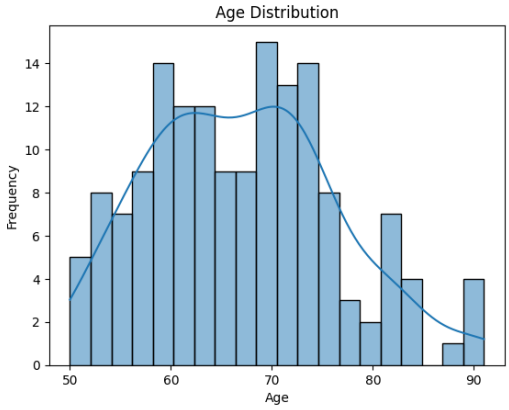


FIGURE 5. A histogram showing age distribution of patients in the dataset.

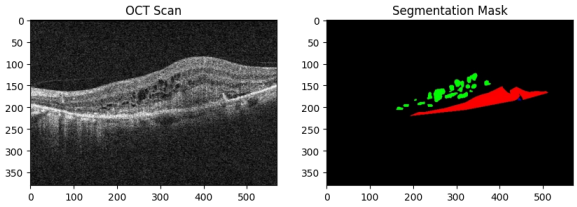


FIGURE 6. A sample OCT B scan image of a patient on the left with its corresponding segmentation mask on the right.

Figure 6 illustrates the gray scale OCT scan image on the left and the corresponding segmentation mask on the right. It highlights intra retinal fluid(IRF) and sub-retinal fluid(SRF) which are represented by the colors green and red respectively.

Figure 7 illustrates the gender distribution in the dataset. There was a substantially higher count of male individuals compared to the female individuals.

Figure 8 demonstrates a balanced representation of both genders which was acquired by oversampling. This strengthened

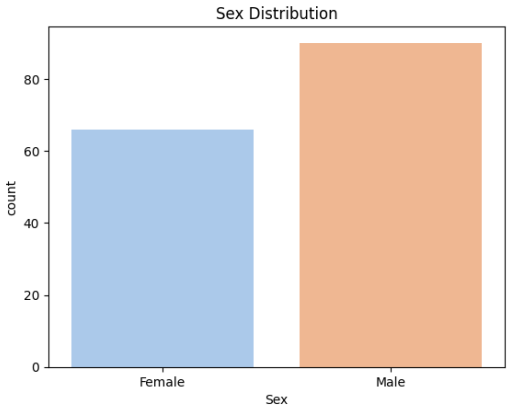


FIGURE 7. A column graph showing sex distribution of patients in the dataset.

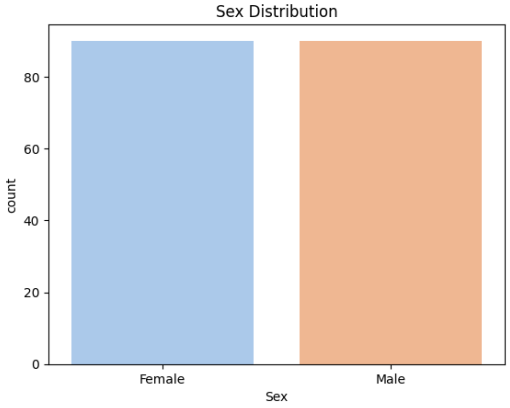


FIGURE 8. A column graph showing balanced sex distribution of patients in the dataset

the generalization of the findings and reduced potential bias in the diagnosis of AMD.

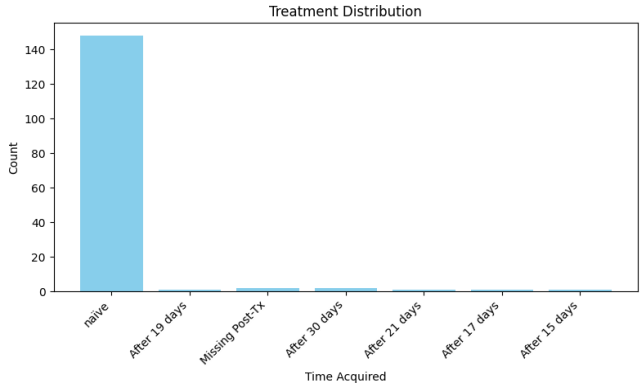


FIGURE 9. A column graph showing the time distribution of when patients acquired treatment in the dataset

Figure 9 illustrates the distribution of the time when patients acquired eye scans. A striking imbalance was observed, with significantly higher number of individuals in the "naive" group (people who acquired eye scans without record of any prior treatment) compared to the other groups.

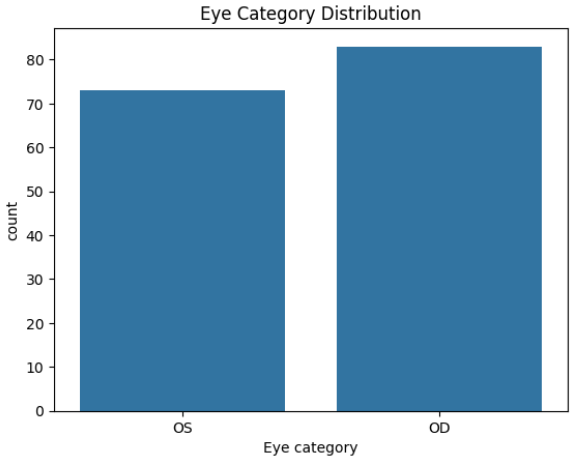


FIGURE 10. A column graph showing eye category distribution of patients in the dataset.

Figure 10 illustrates the distribution of the eye categories Oculus Sinister (OS) which is the right eye and Oculus Dexter (OD) which is the left eye in the dataset. The approximately equal number of observations in each category ensured that our analysis was not skewed towards one eye or the other.

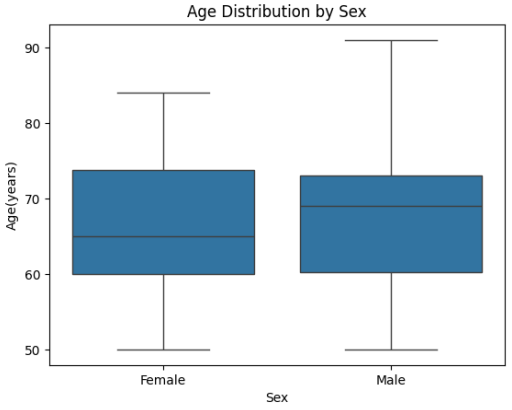


FIGURE 11. A box plot showing age distribution by sex of patients.

Figure 11 illustrates the age distribution of patients, stratified by sex. The box plot revealed a slightly higher median age for males compared to that of females which was indicated by the horizontal line within each box. Additionally, the box for females was slightly taller than that for males which suggested that the inter quartile range(the middle 50 percent of ages) was slightly larger for females. The upper whisker for males seemed to extend a bit higher than the upper whisker for females, hinting that there were potentially older ages in the male group.

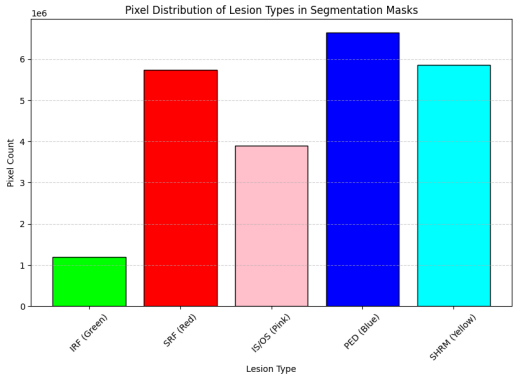


FIGURE 12. A column graph showing pixel distributions of lesions in the segmentation masks of eye scans from patients.

Figure 12 represents the distribution of lesion types by pixel counts within the segmentation masks with the pigment epithelial detachment (PED) lesion exhibiting the highest pixel count and the intraretinal fluid (IRF) lesion type showing the lowest pixel count. This showed that PED was the most frequently encountered lesion type in the dataset.

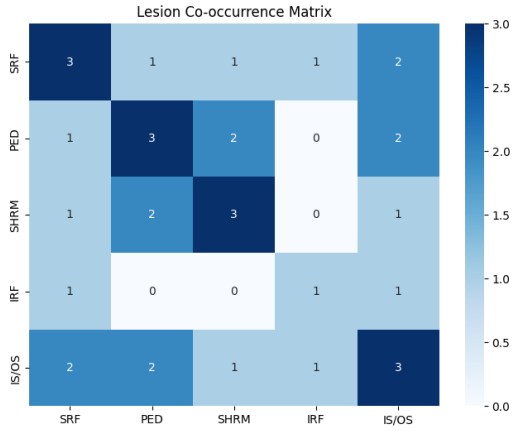


FIGURE 13. A heat map representation that shows correlation between lesion types in segmentation masks.

Figure 13 displays the co-occurrence matrix. It illustrates the relationships between the five retinal lesion types, that is to say, Sub Retinal Fluid (SRF), Inner Segment/ Outer Segment (IS/OS), Pigment Epithelium Detachment (PED), Intra Retinal Fluid (IRF), and Sub Retinal Hyper Reflective Material (SHRM). The matrix demonstrates the frequency with which lesions were observed at the same time within the dataset. The diagonal elements reflect the individual prevalence of each lesion, while the off-diagonal elements show their co-occurrence. From the heat map, there is a high co-occurrence of SHRM and PED which is represented by the dark blue cell at their intersection. In contrast, the co-occurrence of IRF and PED was notably low.

Figure 14 demonstrates the lesion co-occurrence in the segmentation masks. Sub retinal fluid (SRF) had the highest overall frequency of co-occurrence and was frequently observed with pigment epithelial detachment (PED), sub retinal

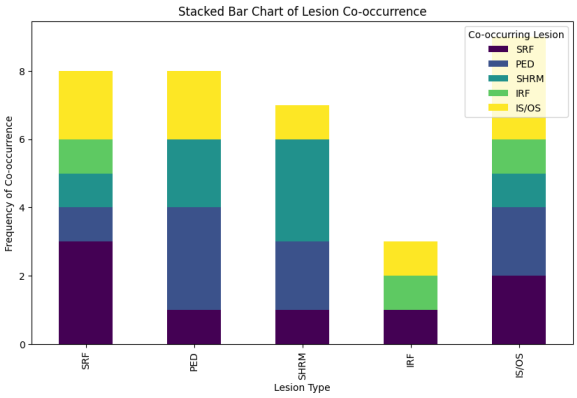


FIGURE 14. A stacked bar chart showing lesion co-occurrence in the segmentation masks in the dataset

hyperreflective material (SHRM), and inner segment / outer segment (IS / OS) lesions. PED also showed a high frequency of co-occurrence with SRF and IS/OS. On the other hand, intraretinal fluid (IRF) showed a relatively lower frequency of co-occurrence compared to SRF and PED.

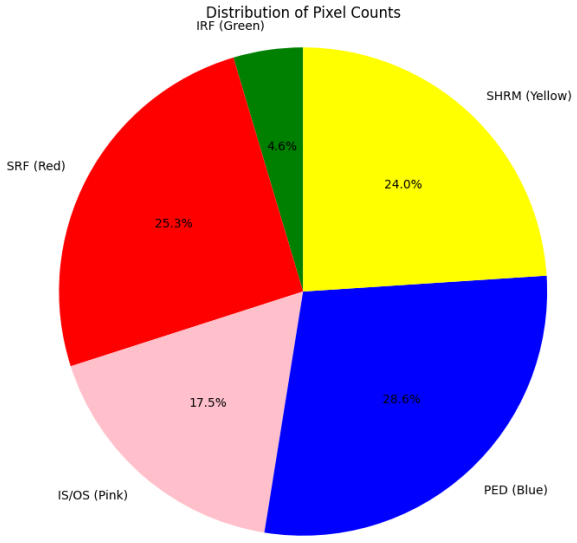


FIGURE 15. A pie chart showing the percentage of lesion types across all images in the dataset.

Figure 15 displays the percentage of lesion types across all images in the dataset with PED lesion having the highest percentage and the IRF lesion having the smallest percentage. Figure 16 represents the normalized co-occurrence of different lesion types. The color intensity represents the strength of co-occurrence, with darker shades indicating higher association. The most prominent co-occurrence (0.49) was found between sub retinal hyperreflective material(SHRM) and pigment epithelium detachment(PED). Other notable co-occurrences include sub retinal fluid (SRF) with PED (0.34) and SHRM (0.25), pigment epithelium detachment with intraretinal fluid (0.28) and SHRM with inner segment/ outer

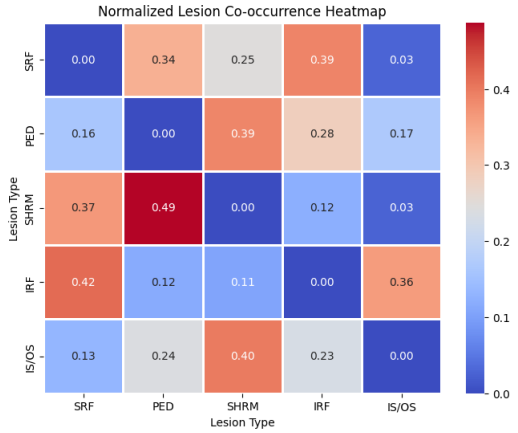


FIGURE 16. A heat map representation that showing "normalized" lesion co-occurrence.

segment (0.40). Some lesion pairs such as sub retinal fluid and inner segment/ outer segment (0.03) exhibited a minimal co-occurrence.

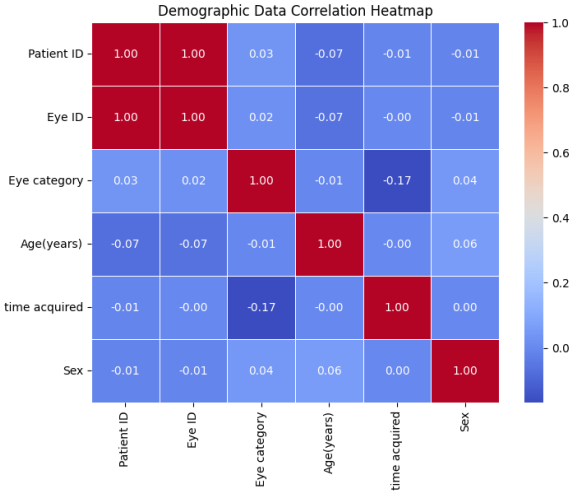


FIGURE 17. A correlation heat map visualizing relationships between different demographic variables.

Figure 17 displays a correlation heat map of demographic variables. Strong positive correlations were observed between patient ID and eye ID while weak correlations were noted between age and both patient/eye ID and also between eye category and time acquired. It was also noted that the sex variable showed minimal association with other variables.

VII. MODEL TRAINING

In this work, various models such as the U-Net model, the U-Net with Vision Transformers, U-Net with Graph Attention Mechanisms, the ResU-Net model, RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation, the ATT_CNN_KD model integrated with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation, U-Net++_KD model with knowledge distillation and the U-Net++_CNN_KD model integrated with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation were trained and evaluated in order to identify the most effective for the detection of AMD.

During training of the models, the segmentation masks of the OCT scans were analyzed using six distinct colors, each representing a different retinal feature. They used a class distribution where Class 0 (red) was representing sub-retinal fluid (SRF), Class 1 (pink) was indicating the inner segment/outer segment (IS/OS), Class 2 (blue) was representing pigment epithelium detachment (PED), Class 3 (green) was representing the intraretinal fluid (IRF), Class 4 (yellow) was an indicator of the sub-retinal hyperreflective material (SHRM) and Class 5 (black) was denoting the background of the OCT scan. Model evaluation was then carried out in order to understand the performance of each model and the evaluation results got from different metrices highlighted the areas of strength and potential weaknesses of these models.

A. U-NET MODEL

The training process of our U-Net model began with data preprocessing, where each image was loaded and checked to ensure it met the expected width of 1140 pixels. Each image was then split into two halves, the left half containing the raw OCT scan and the right half containing the corresponding segmentation mask. To ensure uniform input sizes for the deep learning model, both the OCT scan and the mask were resized to 192×288 pixels. The segmentation masks were then converted into class label masks using a predefined color-to-class mapping. This conversion ensured that each pixel was assigned a unique class index representing a specific retinal abnormality. The images were further normalized to a [0,1] range for efficient training while the masks were one-hot encoded to facilitate the use of categorical cross-entropy loss.

The U-Net model which is built around an encoder-decoder framework employed the encoder which consists of several convolutional layers and max-pooling layers that helped the model learn the hierarchical features of the OCT scan while extracting increasingly complex features from the input scan and retaining essential information as shown in equations (1) and (2) respectively. Following equations (3) and (4) respectively, the image size was progressively increased through the up-sampling layers and a pixel-wise probability map was generated by a final 1×1 convolution layer with softmax activation. Here each pixel was assigned a probability of belonging to one of the six segmentation classes corresponding to different retinal abnormalities. The model was then compiled using the Adam optimizer which dynamically adjusted the learning rate for efficient convergence and categorical cross-entropy loss in order to ensure accurate pixel-wise classification. It was then trained for 10 epochs with a batch size of 8 and evaluated by computing its validation accuracy as shown in equation(5).

Convolutional layer

$$C_i = \sigma(W_i * X + b_i) \tag{1}$$

where:

- C_i = Feature map after convolution
- W_i = Convolution filter weights
- X = Input image
- b_i = Bias term
- σ = ReLU activation function

Max Pooling layer

$$P_i = \max(C_i) \tag{2}$$

Upsampling Layer:

$$U_i = \text{UpSampling}(P_i) \tag{3}$$

Softmax Output:

$$\hat{Y} = \frac{e^{C_i}}{\sum_{j=1}^{6} e^{C_j}} \tag{4}$$

Validation Accuracy:

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} \mathbb{1}(\hat{Y}_i = Y_i)}{N} \times 100 \tag{5}$$

By effectively automating the segmentation of key retinal structures and abnormalities such as sub retinal fluid (SRF), intraretinal fluid (IRF) and pigment epithelial detachment (PED), this U-Net model will serve as a valuable tool in the clinical diagnosis of macular degeneration.

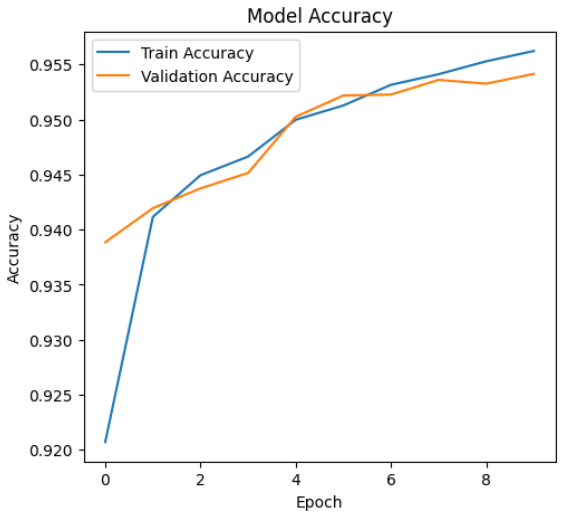


FIGURE 18. A line graph showing the training and validation accuracy trends during training of the U-NET model.

Figure 18 displays the trends in training and validation accuracy over 10 epochs. The training accuracy (blue line) started lower but increased steadily reaching more than 95.5 percent while the validation accuracy (orange line) followed a similar trend, slightly lagging but maintaining a strong alignment with the training accuracy.

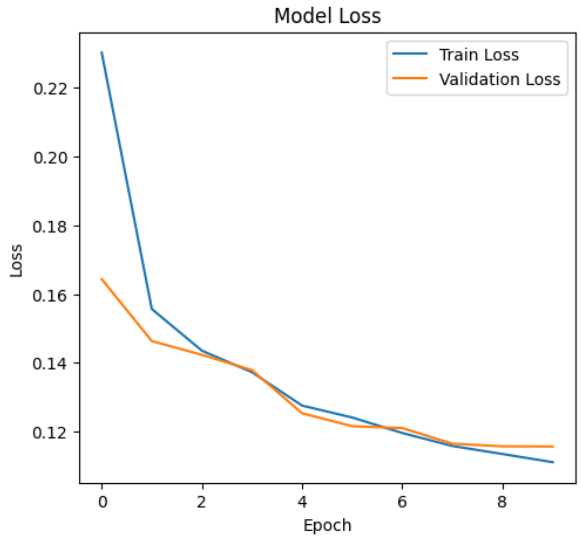


FIGURE 19. A line graph showing the loss reduction during the U-NET model training.

Figure 19 represents the loss reduction over 10 epochs. The training loss (blue line) started high but rapidly decreased and stabilized around 0.11 while the validation loss (orange line) followed a similar decreasing pattern, confirming that the model was effectively minimizing errors during training and validation. Since the validation loss did not increase or diverge from the training loss, the model was efficiently learning robust features for retinal lesion segmentation.

TABLE 3. A table showing the evaluation of the U-NET model over different metrics.

Class	Dice	IoU	Hausdorf	Accuracy
0	0.739512	0.597072	36.94272	0.952541
1	1	1	inf	1
2	1	1	inf	1
3	0.88478	0.88478	inf	0.999975
4	0.46799	0.463599	inf	0.996643
5	0.974154	0.949786	13.37094	0.953752

B. RES U-NET MODEL

The ResNet U-Net model is a hybrid architecture that combines the strengths of ResNet and U-Net for image segmentation. This ResNet-UNet model worked by processing an OCT scan image beginning with data loading and preprocessing. The raw OCT scan image, typically sized 380×1140×3 was first read and then split into the left half (OCT B-scan) and the right half (segmentation mask). In order to ensure uniform input dimensions, the OCT scan was resized using bilinear interpolation while the segmentation mask was resized with nearest-neighbor interpolation to preserve segmentation labels. Hierarchical features from the OCT scan image were then extracted using the convolutional layers of the encoder in the U-Net while the bottleneck layer at the core of the U-Net captured deep semantic information before the decoder

progressively reconstructed the segmentation mask. A final sigmoid-activated output layer then generated a binary segmentation mask, distinguishing retinal structures from the background of the scan. The model was then compiled using Adam optimization and a binary cross-entropy loss function and trained for 10 epochs . Once trained, the model was then able to predict segmentation masks on new OCT scans, where predictions were visualized by comparing the original scan and the predicted mask.

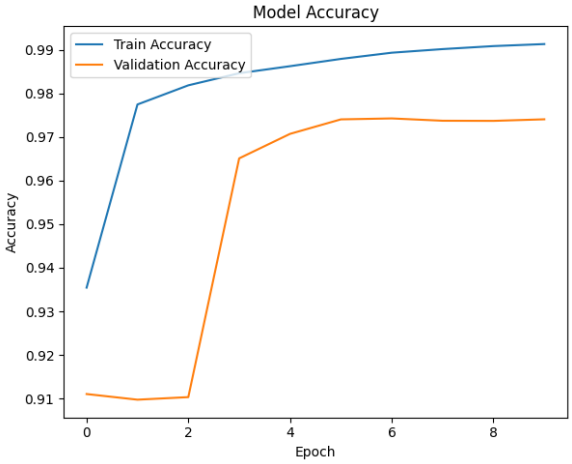


FIGURE 20. A line graph showing the training and validation accuracy during training of the ResU-Net model.

Figure 20 illustrates the accuracy of the model over a variety of training epochs, with the blue line representing training accuracy and the orange line representing validation accuracy. Initially, the training accuracy increased rapidly, nearly reaching 99 percent which was an indicator that the model was learning effectively from the training data while the validation accuracy started lower, around 91 percent, but improved significantly stabilizing at approximately 97 percent. However, a noticeable gap between the two suggested potential overfitting.

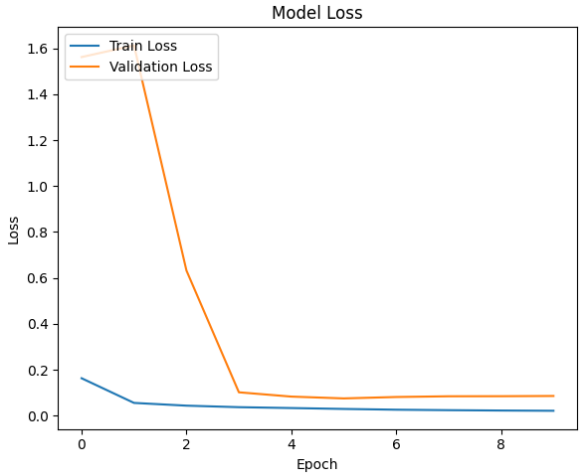


FIGURE 21. A line graph showing the loss reduction during ResU-Net model training.

Figure 21 illustrates the training and validation loss curves. Initially, both losses started relatively high, with validation loss being significantly higher than training loss. However, within the first few epochs, validation loss dropped steeply indicating rapid learning and adaptation of the model to the dataset. After approximately the fourth epoch, both losses stabilized at a low value suggesting that the model had effectively minimized the error. However, a slight upward trend in validation loss towards the later epochs was an indicator of the beginning of overfitting.

TABLE 4. A table showing the evaluation of the ResU-NeT model over different metrics.

Class	Dice	IoU	Hausdorff	Accuracy
0	0.849149	0.750161	25.97448	0.974086
1	1	1	inf	1
2	1	1	inf	1
3	0.88478	0.88478	inf	0.999975
4	0.693264	0.667503	inf	0.997449
5	0.986968	0.974351	9.716761	0.976566

C. U-NET MODEL WITH VISION TRANSFORMERS

This model integrated both the U-Net and Vision Transformer architectures for enhanced feature extraction in the OCT scan images. The training process started with preprocessing where the OCT scan images were first loaded and split. They were then resized to a fixed dimension of 192×288 pixels following equation(6). The segmentation mask then underwent further processing where the specific RGB (Red, Green, Blue) digital imaging colors were converted into class labels based on a predefined color mapping so that the model to learnt meaningful pixel-wise classifications using equation(7).

$$I_B, I_M = \text{Resize}(I_B, h, w), \quad \text{Resize}(I_M, h, w). \qquad (6)$$

$$L(x, y) = f(I_M(x, y)), \tag{7}$$

where f is the function mapping the digital imaging color values to discrete class indices.

For feature extraction, the model leveraged EfficientNetB0 as its encoder, a lightweight and high-performance convolutional neural network (CNN) known for efficient feature representation using equation(8). To further enhance the extracted features, a Vision Transformer (ViT) module was incorporated using equation (9) to capture long-range dependencies in the image.

$$F_{\text{enc}} = \text{EfficientNet}(I_B). \tag{8}$$

where, F_{enc} represents the encoded feature map.

$$Z = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{9}$$

where:

- $Q = W_Q X$ = Query Matrix
- $K = W_K X$ = Key Matrix
- $V = W_V X$ = Value Matrix

and d_k is the embedding dimension for scaling.

The decoder then followed a traditional U-Net architecture that gradually recovered spatial information, followed by convolutional layers that refined the feature maps. Following equation (10), a final softmax activation function produced a multi-class segmentation output which ensured that the model learnt to distinguish the various retinal layers and abnormalities in the OCT scans.

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i), \tag{10}$$

where:

- C is the number of classes.
- y_i is the ground truth label for class i.
- $\hat{y}_i$ is the predicted probability for class i.

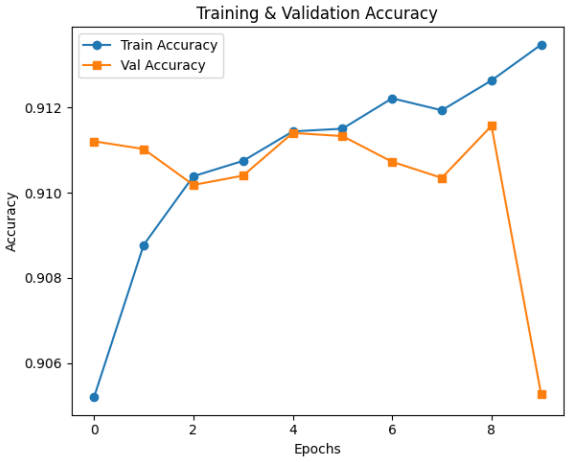


FIGURE 22. A line graph showing the training and validation accuracy during training of a U-NET Model with Vision Transformers.

Figure 22 represents the training accuracy (blue line) which showed a steady increase over epochs, indicating that the model was learning and improving on the training data while the validation accuracy (orange line) initially followed a similar trend but began to fluctuate, especially towards the final epoch where it suddenly dropped.

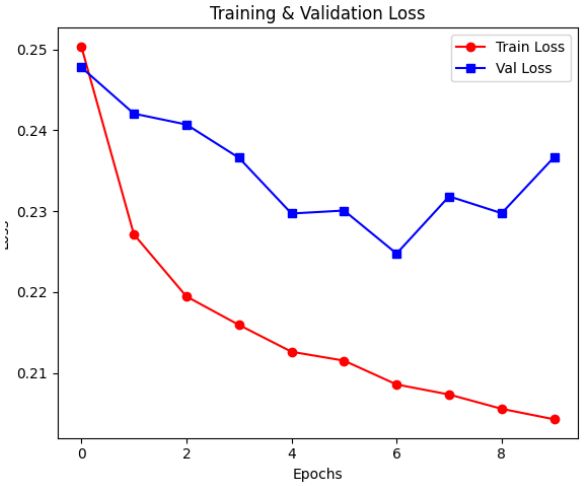


FIGURE 23. A line graph showing the loss reduction during training of a U-NET Model with Vision Transformers.

Figure 23 represents the training and validation loss curves. The training loss (red line) decreased consistently, showing that the model was minimizing error on the training set. However, the validation loss (blue line) decreased initially but then stabilized and even started increasing towards the end. This divergence between training and validation loss supported the idea of overfitting.

TABLE 5. A table showing the evaluation of the U-NeT model with Vision Transformers over different metrics.

Class	Dice	IoU	Hausdorf	Accuracy
0	0.395262	0.256116	70.10607	0.906958
1	1	1	inf	1
2	1	1	inf	1
3	0.88478	0.88478	inf	0.999975
4	0.651494	0.651494	inf	0.996778
5	0.949168	0.9036	15.329468	0.906804

D. U-NET MODEL WITH GRAPH ATTENTION MECHANISMS

This model that combines the U-Net mechanisms and the Graph Attention Mechanisms ensured that both localized lesions and global structural changes were effectively captured. At the core of the model is the Graph Attention Layer (GAT) that transformed pixel features into a graph-based representation and connectivity determined by a spatial adjacency matrix. The adjacency matrix was dynamically constructed based on pixel neighborhood relationships ensuring that each

pixel node was connected to its adjacent pixels in the image grid. Within the GAT layer, node features underwent a learnable linear transformation followed by an attention mechanism that computes importance scores for neighboring nodes. The attention scores were then normalized using a softmax function which allowed the model to focus more on relevant spatial regions while suppressing less important connections.

After applying the Graph Attention Module, the transformed feature representation was reshaped back into a spatial format and fed into the U-Net-inspired decoder. This decoder was responsible for up-sampling the learned features to a higher spatial resolution which made it suitable for segmentation tasks. The up-sampling process was then followed by the convolutional layers that refined the output feature maps. The final segmentation prediction was generated using a 1×1 convolution layer with a softmax activation function which produced a pixel-wise classification map with multiple classes that represent the different retinal layers in the OCT scan.

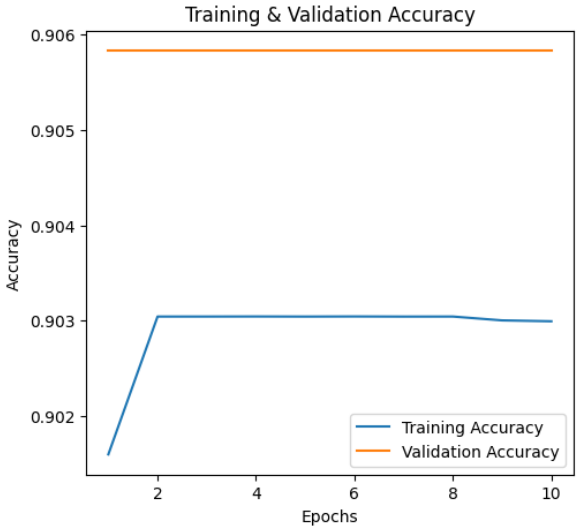


FIGURE 24. A line graph showing the training and validation accuracy during training of a U-NET Model with GATs.

Figure 24 shows that the training accuracy improved quickly in the early epochs and then stabilized at approximately 0.903. In contrast, the validation accuracy remained almost constant at around 0.906 from the beginning, indicating that the model's generalization performance did not improve further after the initial training phase. This suggested that the model reached its peak accuracy early and additional training did not contribute to better performance.

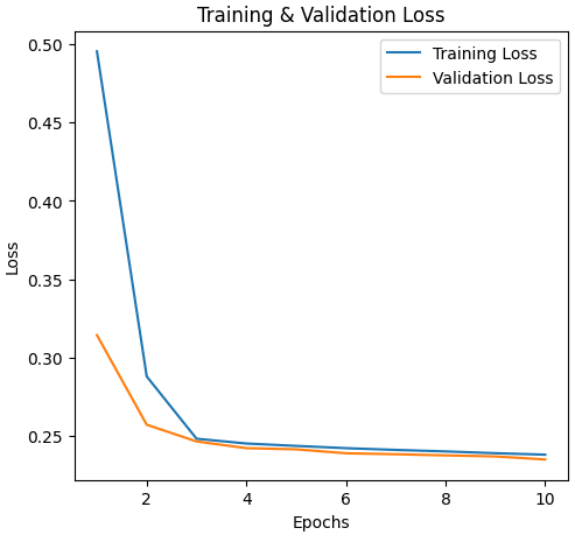


FIGURE 25. A line graph showing the loss reduction during training of a U-NET Model with GATs.

Figure 25 illustrates the training and validation performance of the model over ten epochs. The first plot, which represents training and validation loss, shows that the training loss started high at around 0.5 but rapidly decreased within the first few epochs, indicating effective learning. Similarly, the validation loss began at approximately 0.3 and steadily decreased. By the final epochs, both losses converged around 0.23, suggesting that the model had reached stability without significant overfitting.

TABLE 6. A table showing the evaluation of the U-NeT model with graph attention mechanisms over different metrics.

Class	Dice	IoU	Hausdorf	Accuracy
0	0.004482	0.002246	68.47627	0.902424
1	9.78E-12	9.78E-12	0	0.991121
2	1.67E-10	1.67E-10	0	0.99948
3	4.86E-05	2.43E-05	154.1104	0.032238
4	2.68E-11	2.68E-11	0	0.996762
5	0.031963	0.016241	20.61553	0.107916

E. RNN_CNN_KD MODEL WITH SELF-ROUTING CAPSULE NETWORKS, GRAPH ATTENTION MECHANISMS, AND KNOWLEDGE DISTILLATION

The RNN_CNN_KD model is a knowledge distillation framework where a ResUNet serves as the teacher model, and a Convolutional Neural Network (CNN) acts as the student model. The Knowledge Distillation (KD) process enables the CNN to learn from the RNN, enhancing its performance while maintaining computational efficiency.

It is enhanced with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation which are powerful techniques for accurate and efficient image segmentation. In order to improve feature selection, the model employed Graph Attention Networks (GATs) which operated

by treating the image as a graph where each pixel or region is a node. The adjacency matrix connected neighboring pixels and allowed the Graph Attention Layer to dynamically assign importance to different regions in the scan. This enhanced the model's ability to focus on the retinal abnormalities in the OCT scan. The backbone of the model followed a ResU-Net structure where the encoder extracted deep features and the decoder reconstructed the segmented output. The capsule networks then employed their self-routing mechanisms in order to improve spatial hierarchies and reduce misclassifications in fine retinal structures. Finally, knowledge distillation enabled a smaller, computationally efficient student model to learn from a larger, more complex teacher model while maintaining high segmentation accuracy. By combining these advanced techniques, the model is expected to deliver precise, robust and efficient OCT scan analysis, making it ideal for diagnosing macular degeneration.

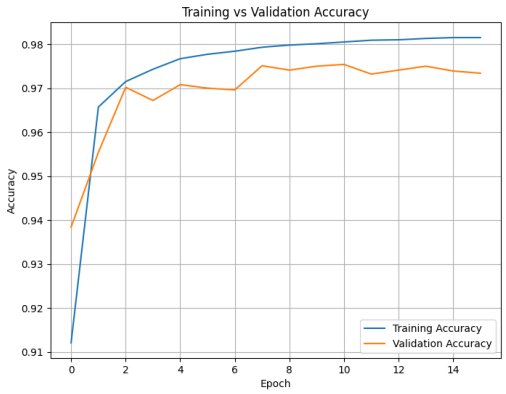


FIGURE 26. A line graph showing the training and validation accuracy during training of the RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation.

Figure 26 shows the training and validation accuracy of the model over 15 epochs. The training accuracy (blue line) increased rapidly at first and then stabilized while the validation accuracy (orange line) fluctuated and plateaued slightly lower.

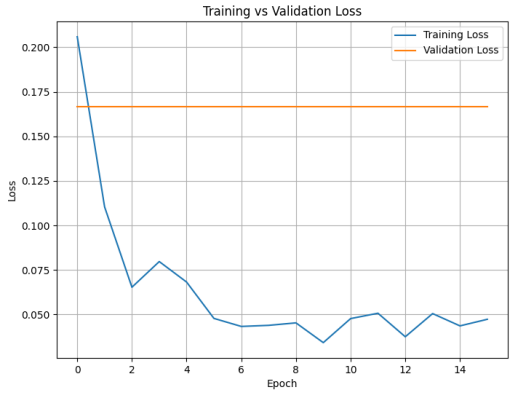


FIGURE 27. A line graph showing the loss reduction during training of the RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation.

Figure 27 illustrates the training and validation loss over 15 epochs. The training loss (blue line) decreased rapidly, showing effective learning while the validation loss (orange line) plateaued early, indicating limited generalization. This divergence suggested overfitting where the model performed well on training data but struggled with unseen data.

TABLE 7. A table showing the evaluation of the the RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation over different metrics.

Dice	IoU	Accuracy	Distillation Loss
0,9122	0.8522	0.9121	0.0033
0.9657	0.9338	0.9657	0.0016
0.9715	0.9446	0.9715	0.0018
0.9743	0.9499	0.9743	0.002
0.9767	0.9545	0.9767	0.0017
0.9777	0.9565	0.9777	0.0018

F. AN ATT_CNN_KD MODEL THAT INTEGRATES SELF-ROUTING CAPSULE NETWORKS, GRAPH ATTENTION MECHANISMS, AND KNOWLEDGE DISTILLATION

The ATT_CNN_KD model is a knowledge distillation-based framework designed for efficient and precise lesion detection and segmentation. It integrates an Attention U-Net (ATT) as the teacher model, a Convolutional Neural Network (CNN) as the student model, and incorporates Self-Routing Capsule Networks and Graph Attention Mechanisms to enhance feature representation. The encoder in the U-Net extracted multi-scale features through convolutional layers, batch normalization, and max pooling, while the decoder reconstructed the segmented image using Conv2DTranspose layers. To enhance feature learning, the model incorporated Graph Attention Mechanisms, treating retinal layers as nodes in a graph to capture spatial dependencies and improve segmentation accuracy. Additionally, Self-Routing Capsule Networks ensured that the positional relationships and hier-

archical structure of retinal layers were preserved. Attention gates were also integrated to selectively focus on important retinal regions while suppressing irrelevant background features thereby improving segmentation quality. To optimize performance, knowledge distillation was applied where a large complex teacher model transferred learned knowledge to a smaller, efficient student model. Ultimately, the model produced a precisely segmented OCT scan highlighting the different retinal layers and abnormalities.

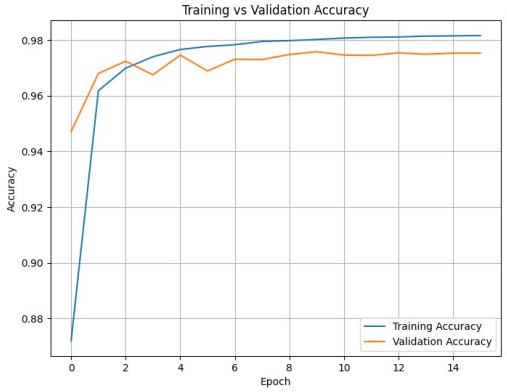


FIGURE 28. A line graph showing the training and validation accuracy during training of an ATT_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation.

Figure 28 shows that the training accuracy increased from 0.88 to 0.96 in the first two epochs and gradually reached 0.98. The validation accuracy also improved but fluctuated slightly before stabilizing around 0.975. It consistently remained just below training accuracy which indicated slight overfitting but good generalization.

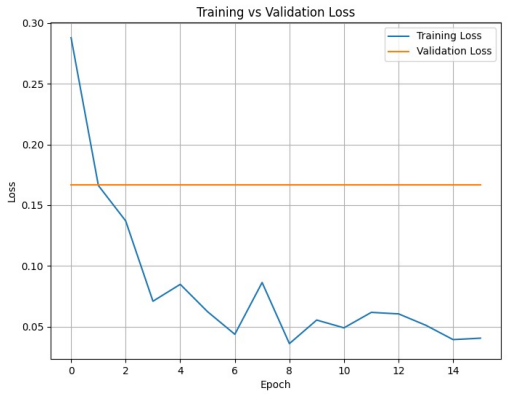


FIGURE 29. A line graph showing the loss reduction during training of an ATT_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation

Figure 29 illustrates the rapid decrease in training loss from 0.29 to below 0.1 in the early epochs which eventually stabilized at 0.04. However, validation loss remained nearly constant at 0.17 indicating overfitting. This suggested the need for regularization in order to enhance performance on unseen data.

TABLE 8. A table showing the evaluation of an ATT_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation over different metrics.

Dice	IoU	Accuracy	Distillation loss
0.8722	0.8033	0.8720	0.0032
0.9618	0.9265	0.9618	0.0025
0.9699	0.9416	0.9699	0.0019
0.9740	0.9493	0.9740	0.0018
0.9766	0.9543	0.9766	0.0023
0.9777	0.9564	0.9777	0.0021

G. U-NET++_CNN_KD MODEL WITH SELF-ROUTING CAPSULE NETWORKS, GRAPH ATTENTION MECHANISMS, AND KNOWLEDGE DISTILLATION

The U-Net++_CNN_KD model is a knowledge distillation-based framework designed to enhance lesion detection and segmentation in OCT scans while maintaining computational efficiency. It features U-Net++ as the teacher model, a Convolutional Neural Network (CNN) as the student model, and integrates Self-Routing Capsule Networks (SRCNs) and Graph Attention Mechanisms (GATs) to improve feature representation and spatial relationships. U-Net++ enhances the traditional U-Net architecture with nested skip pathways enabling better feature propagation and precise segmentation of structures like retinal layers. Self-Routing Capsule Networks improve spatial feature recognition by dynamically routing information between capsules while the graph attention mechanisms refine feature interactions by focusing on the most relevant parts of the image. Finally, Knowledge Distillation transfers knowledge from a larger, more powerful teacher model to a smaller student model improving performance and generalization on OCT scans.

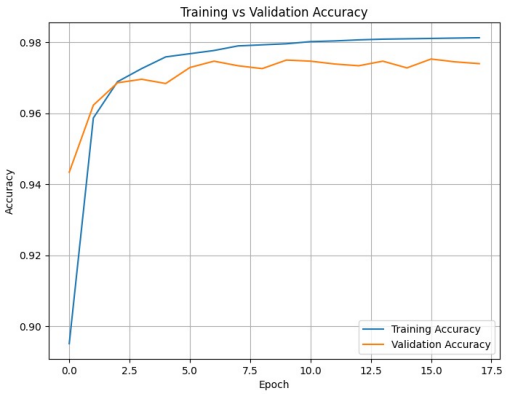


FIGURE 30. A line graph showing the training and validation accuracy during training of a U-Net++_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation

Figure 30 shows the performance of the U-Net++ model during training. The training accuracy (blue line) increased rapidly in the first few epochs surpassing 96 percent by epoch 5 and stabilized near 98 percent while the validation accuracy (orange line) followed a similar trend but stabilized slightly lower around 97 percent indicating strong generalization with minor variations. The close alignment between the two curves suggests minimal overfitting meaning the model effectively learns features relevant to OCT image segmentation.

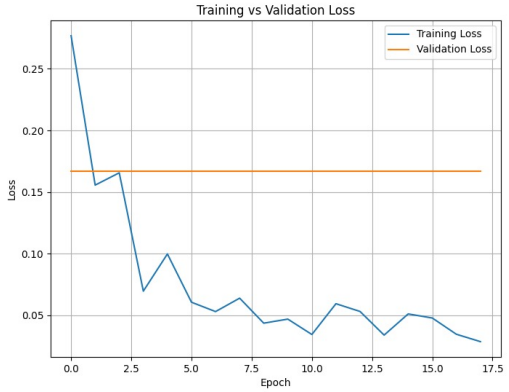


FIGURE 31. A line graph showing the loss reduction during training of a U-Net++_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation

Figure 31 illustrates the reduction in model error over time. The training loss (blue line) started high but dropped significantly within the first few epochs indicating rapid learning. As training progressed, the loss continued to decline stabilizing near zero suggesting that the model had learned to accurately segment OCT scans. However, the validation loss (orange line) remained relatively constant implying that while the model performed well on training data, it may not improve much further on unseen data. This slight discrepancy could suggest minor overfitting.

TABLE 9. A table showing the evaluation of the U-Net++_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation over different metrics.

Dice	IoU	Accuracy	Distillation loss
0.8951	0.8240	0.8951	0.0071
0.9587	0.9208	0.9587	0.0024
0.9688	0.9396	0.9689	0.0018
0.9726	0.9467	0.9726	0.0011
0.9758	0.9528	0.9759	0.0009
0.9768	0.9546	0.9768	0.0008

H. U-NET++_KD MODEL WITH KNOWLEDGE DISTILLATION

The U-Net model with knowledge distillation processes OCT scans by first training a large teacher U-Net on a segmentation dataset where it learns to predict soft probability distributions for each pixel. In this case, the teacher model minimized segmentation loss using a Dice loss function and a smaller, efficient Student U-Net was then trained to mimic the teacher's predictions using a combination of segmentation loss and distillation loss, typically measured using KL-Divergence. The total loss function balanced both terms ensuring the student model retained high accuracy while being computationally efficient. During inference, the student model segmented new OCT scans and identified retinal layers and fluid accumulations. The output was refined through post-processing and overlaid on the original OCT scan for visualization.

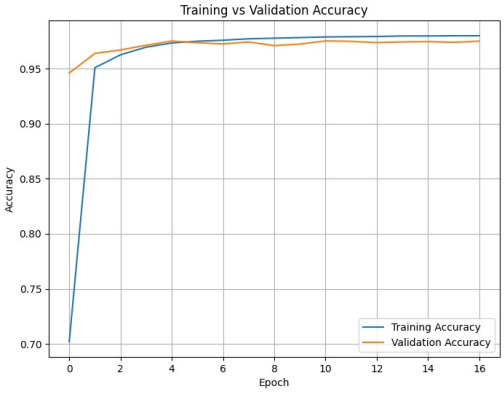


FIGURE 32. A line graph showing the training and validation accuracy during training of a U-Net++_KD model with knowledge distillation

Figure 32 shows the relationship between training accuracy and validation accuracy depicted across 17 epochs. The training accuracy, represented in blue, demonstrated a steady increase, reaching about 98 percent while the validation accuracy, shown in orange experienced a notable increase, closely following the training accuracy but leveling off around 95 percent. This slight discrepancy between training and validation accuracy suggested that while the model performed well on training data, there might have been some overfitting occurring, as the validation accuracy did not reach the same high level as training accuracy.

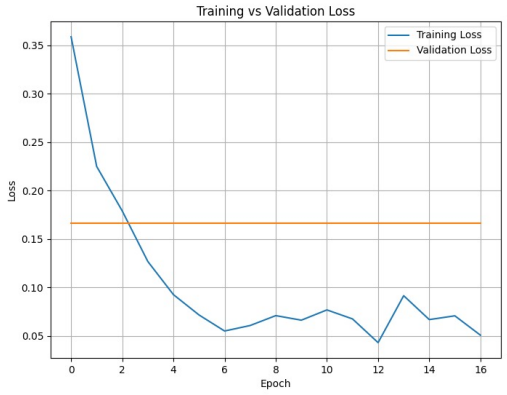


FIGURE 33. A line graph showing the loss reduction during training of a U-Net++_KD model with knowledge distillation.

Figure 33 illustrates the training loss and validation loss over 17 epochs. The training loss, depicted in blue, started at a higher value and decreased dramatically to around 0.05 indicating that the model effectively minimized the error on the training dataset. In contrast, the validation loss, represented in orange, stabilized around 0.15 after an initial drop. The significant gap between training and validation loss reinforced the idea that while the model was improving on the training set, its performance on unseen data could be less robust.

TABLE 10. A table showing the evaluation of theU-Net++_KD model with knowledge distillation over different metrics.

Dice	IoU	Accuracy	Distillation loss
0.7027	0.6105	0.7020	0.01
0.9510	0.9067	0.951	0.006
0.9627	0.9282	0.9628	0.004
0.9797	0.9413	0.9697	0.0021
0.9735	0.9485	0.9735	0.0014
0.9751	0.9515	0.9751	0.001

VIII. MODEL SELECTION AND DEPLOYMENT

The model selection process was driven by a balance between accuracy, efficiency, and real-world applicability. The models were trained and evaluated using key metrics such as Dice Score, Intersection over Union (IoU), Hausdorff distance, precision, F1 score, distillation loss and accuracy. Among these, the RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation emerged as the best-performing model due to its superior segmentation accuracy and superior training stability over the other models. It achieved the lowest distillation loss of 0.0002, highest dice score of 0.9815 and had a longer training duration of 17 epochs which suggested greater robustness over the other models.

Given these findings, the RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation was ultimately chosen not only for its strong performance but also for its computational efficiency and interpretability. Its balance of high segmentation precision and lower computational demands made it the most suitable choice for lesion detection.

We evaluated, selected, and implemented the ResU-Net model for detecting and segmenting lesions in OCT scan images, specifically focusing on the diagnosis of macular degeneration. The MLOps pipeline involved training, fine-tuning, and evaluating the model's architecture on a dataset of OCT scans, followed by model serialization into a deployable format using TensorFlow. The trained model was deployed on Render to enable scalable, on-demand access to lesion detection via a Flask-based web service. This service accepts the gray-scale OCT images, processes them through the model, and returns segmented lesion masks as ".png" images.

IX. EXPLAINABLE AI

This section explores the various techniques that were used to make the model operations transparent and understandable. As seen above, multiple models were trained. However, for comparison the best two and worst two models were analyzed. In this analysis, the best two models considered were the RNN_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation and the U-Net++_CNN_KD model with Self-Routing Capsule Networks, Graph Attention Mechanisms, and Knowledge Distillation. While for the worst two models the U-NET Model with Graph Attention Mechanisms and U-NET Model with Vision Transformers were selected. The techniques used to determine the transparency of our best two and worst two models were the Grad-CAM and Saliency maps techniques.

Saliency Maps and Grad-CAM, also known as, Gradient-weighted Class Activation Mapping are explainable AI (XAI) techniques that help interpret deep learning models applied to Optical Coherence Tomography (OCT) scans, which are used for diagnosing retinal diseases like macular degeneration. Grad-CAM works by generating a heatmap over the OCT scan, highlighting the most critical regions that influenced the model's decision. The warm colors such as red and yellow in the heatmap indicate the areas of high relevance, typically corresponding to structural abnormalities like fluid accumulation or retinal layer disruptions. On the other hand, Saliency Maps compute the gradient of the model's output with respect to each pixel in the input image, identifying the most important regions at a finer level. Bright regions in the Saliency Map indicate pixels that had the highest impact on classification, allowing for a more detailed understanding of the model's focus. By visualizing these techniques on OCT scans, clinicians can gain insights into how AI models make decisions, improving trust and aiding in the validation of automated retinal disease diagnosis.

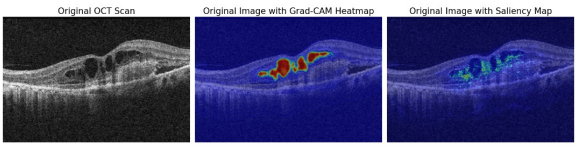


FIGURE 34. An image demonstrating the application of Grad-CAM and Saliency maps techniques on an OCT scan using the RNN_CNN_KD model with Self-Routing Capsule Networks,Graph Attention Mechanisms, and Knowledge Distillation.

Figure 34 demonstrates the application of Grad-CAM and Saliency maps techniques on an OCT scan. The leftmost image presents the original grayscale OCT scan, displaying retinal layers with visible fluid-filled spaces, which are indicative of retinal pathologies. The middle image overlays the same OCT scan with a Grad-CAM heatmap, which highlights the most influential regions used by the deep learning model in its classification. The red and yellow areas represent the most critical features contributing to the diagnosis, likely corresponding to disease-affected regions. On the other hand, the rightmost image applies a Saliency Map technique, which identifies pixel-level importance by computing gradients with respect to the model's output. The bright green and blue regions indicate the areas that had the highest impact on the classification, allowing for fine-grained analysis of structural abnormalities.

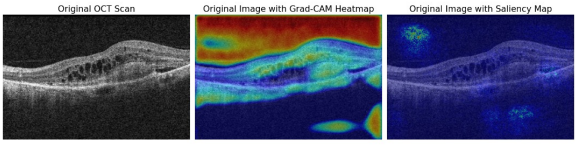


FIGURE 35. An image illustrating the use of Grad-CAM and Saliency maps techniques on an OCT scan using the U-Net++_CNN_KD model with Self-Routing Capsule Net- works, Graph Attention Mechanisms, and Knowledge Distillation

Figure 35 illustrates the use of Grad-CAM and Saliency maps techniques on an OCT scan. The leftmost image presents the original grayscale OCT scan. The middle image applies Grad-CAM which generates a heatmap to indicate the most important regions contributing to the model's decision. The red and yellow areas highlight where the model focuses most, corresponding to potential disease markers such as fluid accumulation or retinal layer disruptions. The rightmost image employs a Saliency Map technique, which computes pixel-wise gradients to determine the most significant features influencing the model's classification. The bright green and blue areas represent regions of high relevance, providing fine-grained insight into subtle retinal changes.

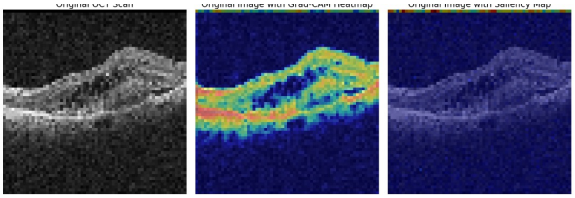


FIGURE 36. An image showing the application of Grad-CAM and Saliency maps techniques on an OCT scan using the U-NET_GAT model

Figure 36 demonstrates the application of Grad-CAM and Saliency maps techniques on an OCT scan. The leftmost image represents the original OCT scan in grayscale, showing a cross-section of the retina with potential abnormalities. The middle image applies Grad-CAM which generates a heatmap highlighting the most relevant areas the model used to make its classification. The yellow and red regions indicate high activation, corresponding to structural abnormalities within the retinal layers. The rightmost image employs a Saliency Map, which calculates pixel-level gradients to determine the most significant regions influencing the model's decision. The areas with higher intensity indicate regions that the model found most relevant, providing a fine-grained understanding of how it interprets retinal abnormalities.

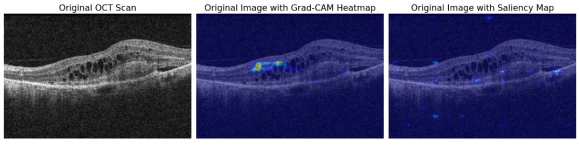


FIGURE 37. A An image showing the Grad-CAM and Saliency maps techniques on an OCT scan using the U-NET_VIT model

Figure 37 shows Grad-CAM and Saliency maps techniques on an OCT scan. The first is the original OCT scan that provides a detailed black-and-white representation of the internal structures, the second is the original image augmented with a Grad-CAM heatmap, which highlights areas of interest using color gradations to showcase their significance in the analysis and the third visualization features a saliency map applied to the original image, emphasizing specific regions with bright spots that indicate features deemed important for understanding the underlying characteristics of the scan.

X. LIMITATIONS

Although the developed model for the early detection of macular degeneration showed promising results, it is crucial to acknowledge the limitations and challenges associated with it. Key concerns include dataset limitations, model complexities and ethical considerations, all of which impact the reliability, fairness, and real-world applicability of this AI-driven diagnostic tool.

Macular degeneration primarily affects older adults, so it is expected that the dataset that was used predominantly consists of individuals aged fifty years and above. However,

this demographic analysis limited the model's ability to generalize to younger individuals who may also be at risk due to genetic predisposition or environmental factors. If younger individuals with early signs of AMD are underrepresented, the model may struggle to accurately detect and classify their cases. This limitation could reduce the clinical applicability of the model for broader screening purposes, especially in cases where early intervention is crucial.

Additionally, certain lesion types such as the intraretinal fluid (IRF), had a significantly lower representation in the dataset. Since deep learning models rely heavily on exposure to diverse examples during training, underrepresented lesion types may not have been well learned, resulting in lower sensitivity for detecting these specific abnormalities. This imbalance could potentially lead to a higher false-negative rate for cases with rare lesions thus reducing the model's effectiveness in identifying less common but clinically significant manifestations of macular degeneration in real world settings.

Despite employing techniques like knowledge distillation during training to reduce model size and improve efficiency, the combination of vision transformers and graph attention networks still made the full model computationally expensive. These deep learning techniques require significant processing power which will make real time inference challenging, especially in low resource settings with limited access to high performance Graphics Processing Units (GPUs). This may hinder deployment in clinical environments where rapid diagnosis is essential.

Although adversarial learning enhances the model's robustness against minor perturbations and adversarial attacks, extreme distortions and low-resolution scans can still negatively affect the model's performance. In real-world scenarios, variations in image quality due to differences in scanning devices, patient movement or even improper acquisition settings could lead to poor classifications of the lesion types. The model may struggle to generalize well to such noisy inputs thus affecting its reliability in practical clinical settings.

Furthermore, even if the model demonstrated strong performance, its predictions must be validated by ophthalmologists before being used in real world clinical settings. Machine learning models are designed to assist, not replace, medical professionals. This means that clinicians will need to verify the system's outputs before making diagnostic or treatment decisions related to macular degeneration. The variations in clinical expertise, institutional protocols, and regional healthcare practices may influence how the model is adopted. Without widespread clinical acceptance and validation, integration into standard ophthalmic practice may be slow.

The integration with existing workflows could also pose as a challenge. Deploying an AI-based solution in real clinical settings requires more than just model accuracy. It must be seamlessly integrated into existing hospital infrastructures and diagnostic workflows. Many healthcare facilities operate on legacy systems, making it challenging to incorporate AI-driven tools without significant modifications to software and hardware. In addition to the above, regulatory approval processes can be time-consuming and resource intensive. This can cause delays and thus hinder the quick deployment of the model in environments where rapid diagnosis is essential.

The fact that the AMD-SD dataset that was used is one that does not adequately represent diverse populations may result in biases that affect the model's performance across different demographic groups. The model might struggle to generalize to patients from underrepresented populations such as the rural parts of Uganda. Such biases can lead to disparities in diagnostic accuracy, potentially resulting in misdiagnoses or inconsistent performance.

Another possible limitation is the potential difficulty in handling coexisting eye conditions. Many patients with macular degeneration also have other retinal diseases, such as diabetic retinopathy or glaucoma, which could introduce overlapping features in OCT scans. The model may misclassify such cases due to the presence of confounding features that were not well represented during training.

Finally, patient compliance and accessibility to screening remain critical challenges. Although machine learning models can provide rapid and accurate assessments, their utility depends on the participation of the patient in regular screenings. In many parts of Uganda, the limited awareness of the Age Related Macular Degeneration(AMD) and the reluctance to undergo routine eye examinations could lead to delayed detection.

XI. CONCLUSION AND FUTURE WORKS

This study presented a deep learning-based framework for the early detection of macular degeneration by leveraging a combination of artificial intelligence techniques such as Graph Attention Networks(GATs), capsule networks, knowledge distillation, and vision transformers. Our model demonstrated promising results in diagnosing the presence of AMD from OCT scan images. The findings above highlight the potential of machine learning in assisting ophthalmologists with early diagnosis. To achieve robust feature extraction and improve model interpretability, we incorporated the Graph Attention Networks (GATs) to capture spatial dependencies within retinal images, allowing for a more refined understanding of the subtle structural changes in the macula. Additionally, the self routing mechanisms in capsule networks were employed to enhance the model's ability to recognize hierarchical patterns within the OCT scans thus reducing susceptibility to variations in image quality and noise. The integration of knowledge distillation enabled us to transfer knowledge from a complex teacher model to a lightweight student model, ensuring that the model remains computationally efficient while preserving high accuracy. Despite these advancements, several areas for improvement remain.

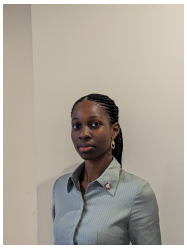
Future work will focus on enhancing model generalization by increasing dataset diversity and optimizing the model's efficiency for deployment on low-resource devices in order to

ensure accessibility in regions with limited computational resources. By addressing this, we hope to refine our model and contribute to the widespread adoption of artificial intelligence in ophthalmology, ultimately aiding in the early detection and management of macular degeneration thus reducing the risk of severe vision loss.

REFERENCES

[1] M. Eisenstein, "Age-related macular degeneration," Nature, 2021.
[2] R. M. Hussain, B. A. Shaukat, L. M. Ciulla, A. M. Berrocal, and J. V. Sridhar, "Vascular endothelial growth factor antagonists: promising players in the treatment of neovascular age-related macular degeneration," Drug Design, Development and Therapy, pp. 2653–2665, 2021.
[3] R. H. Guymer and T. G. Campbell, "Age-related macular degeneration," The Lancet, 2023.
[4] M. Fleckenstein, S. Schmitz-Valckenberg, and U. Chakravarthy, "Age-related macular degeneration: a review," JAMA, vol. 331, no. 2, pp. 147–157, 2024.
[5] N. Boopathiraj, I. V. Wagner, S. K. Dorairaj, D. D. Miller, and M. W. Stewart, "Recent updates on the diagnosis and management of age-related macular degeneration," Mayo Clinic Proceedings: Innovations, Quality & Outcomes, vol. 8, no. 4, pp. 364–374, 2024.
[6] T. A. C. De Guimaraes, M. D. Varela, M. Georgiou, and M. Michaelides, "Treatments for dry age-related macular degeneration: therapeutic avenues, clinical trials and future directions," British Journal of Ophthalmology, vol. 106, pp. 297–304, 2022.
[7] H. Xie, W. Xu, Y. X. Wang, and X. Wu, "Deep learning network with differentiable dynamic programming for retina oct surface segmentation," Biomedical Optics Express, vol. 14, pp. 3190–3202, 2023.
[8] M. W. MacCumber and et al., "Antivascular endothelial growth factor agents for wet age-related macular degeneration: an iris registry analysis," Canadian Journal of Ophthalmology, vol. 58, pp. 252–261, 2023.
[9] C. J. Flaxel, R. A. Adelman, S. T. Bailey, A. Fawzi, J. I. Lim, G. A. Vemulakonda, and G.-s. Ying, "Age-related macular degeneration preferred practice pattern®," Ophthalmology, vol. 127, no. 1, pp. P1–P65, 2020.
[10] F. Monika, K. T. DL, R. H. Guymer, C. Usha, S.-V. Steffen, C. C. Klaver, W. T. Wong, and E. Y. Chew, "Age-related macular degeneration (primer)," Nature Reviews: Disease Primers, vol. 7, no. 1, 2021.
[11] S. Girgis and L. R. Lee, "Treatment of dry age-related macular degeneration: a review," Clinical & Experimental Ophthalmology, vol. 51, no. 8, pp. 835–852, 2023.
[12] M. Zeppieri, S. Marsili, E. S. Enaholo, A. O. Shuaibu, N. Uwagboe, C. Salati, L. Spadea, and M. Musa, "Optical coherence tomography (oct): a brief look at the uses and technological evolution of ophthalmology," Medicina, vol. 59, no. 12, p. 2114, 2023.
[13] P. J. Rosenfeld, M. Shen, O. Trivizki, J. Liu, G. Herrera, F. E. Hiya, J. Li, A. Berni, L. Wang, O. S. El-Mulki et al., "Rediscovering age-related macular degeneration with swept-source oct imaging: The 2022 charles l. schepens, md, lecture," Ophthalmology Retina, 2024.
[14] A. R. C. Celebi, E. Bulut, and A. Sezer, "Artificial intelligence based detection of age-related macular degeneration using optical coherence tomography with unique image preprocessing," European Journal of Ophthalmology, vol. 33, no. 1, pp. 65–73, 2023.
[15] Y. Shen, J. Li, W. Zhu, K. Yu, M. Wang, Y. Peng, Y. Zhou, L. Guan, and X. Chen, "Graph attention u-net for retinal layer surface detection and choroid neovascularization segmentation in oct images," IEEE Transactions on Medical Imaging, vol. 42, no. 11, pp. 3140–3154, 2023.
[16] C. Yin, S. E. Moroi, and P. Zhang, "Predicting age-related macular degeneration progression with contrastive attention and time-aware lstm," in Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 4402–4412.
[17] Z. Xiao, H. Tong, R. Qu, H. Xing, S. Luo, Z. Zhu, F. Song, and L. Feng, "Capmatch: Semi-supervised contrastive transformer capsule with feature-based knowledge distillation for human activity recognition," IEEE Transactions on Neural Networks and Learning Systems, 2023.
[18] M. Opoku, B. A. Weyori, A. F. Adekoya, and K. Adu, "Clahe-capsnet: Efficient retina optical coherence tomography classification using capsule networks with contrast limited adaptive histogram equalization," Plos one, vol. 18, no. 11, p. e0288663, 2023.
[19] X. Xu, M. Li, C. Tao, T. Shen, R. Cheng, J. Li, C. Xu, D. Tao, and T. Zhou, "A survey on knowledge distillation of large language models," arXiv preprint arXiv:2402.13116, 2024.
[20] A. Khan, Z. Rauf, A. Sohail, A. R. Khan, H. Asif, A. Asif, and U. Farooq, "A survey of the vision transformers and their cnn-transformer based variants," Artificial Intelligence Review, vol. 56, no. Suppl 3, pp. 2917–2970, 2023.
[21] Y. Shen, P. Guo, J. Wu, Q. Huang, N. Le, J. Zhou, S. Jiang, and M. Unberath, "Movit: Memorizing vision transformers for medical image analysis," in International Workshop on Machine Learning in Medical Imaging. Springer, 2023, pp. 205–213.
[22] Y. Wang, Y. Zhang, Z. Yao, R. Zhao, and F. Zhou, "Machine learning based detection of age-related macular degeneration (amd) and diabetic macular edema (dme) from optical coherence tomography (oct) images," Biomedical optics express, vol. 7, no. 12, pp. 4928–4940, 2016.
[23] 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, pp. 259–265, 2018.
[24] X. Guo, Y. Ding, W. Xu, D. Wang, H. Yu, Y. Lin, S. Chang, Q. Zhang, and Y. Zhang, "Predicting brain age gap with radiomics and automl: A promising approach for age-related brain degeneration biomarkers," Journal of Neuroradiology, vol. 51, no. 3, pp. 265–273, 2024.
[25] M. H. Wang, K. K.-l. Chong, Z. Lin, X. Yu, and Y. Pan, "An explainable artificial intelligence-based robustness optimization approach for age-related macular degeneration detection based on medical iot systems," Electronics, vol. 12, no. 12, p. 2697, 2023.
[26] N. T. Le, T. Le Truong, S. Deelertpaiboon, W. Srisiri, P. F. Pongsachareonnont, D. Suwajanakorn, A. Mavichak, R. Itthipanichpong, W. Asdornwised, W. Benjapolakul et al., "Vit-amd: A new deep learning model for age-related macular degeneration diagnosis from fundus images," International Journal of Intelligent Systems, vol. 2024, no. 1, p. 3026500, 2024.
[27] M. Li, Y. Shen, R. Wu, S. Huang, F. Zheng, S. Chen, R. Wang, W. Dong, J. Zhong, G. Ni et al., "High-accuracy 3d segmentation of wet age-related macular degeneration via multi-scale and cross-channel feature extraction and channel attention," Biomedical Optics Express, vol. 15, no. 2, pp. 1115–1131, 2024.
[28] Y. Zhao and A. Kumar, "Graph attention networks for predicting amd progression from oct scans," Artificial Intelligence in Medicine, vol. 36, no. 4, pp. 212–224, 2023.
[29] Y. Wang, T. Sun, S. Li, X. Yuan, W. Ni, E. Hossain, and H. V. Poor, "Adversarial attacks and defenses in machine learning-empowered communication systems and networks: A contemporary survey," IEEE Communications Surveys & Tutorials, 2023.
[30] Y. Cao, E. Rizk, S. Vlaski, and A. H. Sayed, "Multi-agent adversarial training using diffusion learning," in ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023, pp. 1–5.
[31] V. Das, S. Dandapat, and P. K. Bora, "Unsupervised super-resolution of oct images using generative adversarial network for improved age-related macular degeneration diagnosis," IEEE Sensors Journal, vol. 20, no. 15, pp. 8746–8756, 2020.
[32] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.
[33] H. Jiang and L. Xu, "Capsule knowledge distillation for improved retinal disease diagnosis: Synergic adversarial label learning (sall) for amd grading," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 9, pp. 4500–4512, 2021.
[34] H. Tan and Z. Chen, "Multi-task learning and knowledge distillation for enhanced diagnosis of retinal diseases in low-resource settings," Journal of Artificial Intelligence in Medicine, vol. 58, no. 2, pp. 127–140, 2024.
[35] S. Zhang, J. Ren, R. Chai, and et al., "Global burden of low vision and blindness due to age-related macular degeneration from 1990 to 2021 and projections for 2050," BMC Public Health, vol. 24, p. 3510, 2024.
[36] L. Kodjikian et al., "Early predictive factors of visual loss at 1 year in neovascular age-related macular degeneration under anti–vascular endothelial growth factor," Ophthalmology Retina, vol. 6, pp. 109–115, 2022.
[37] Y. Gao, F. Xiong, J. Xiong, Z. Chen, Y. Lin, X. Xia, Y. Yang, G. Li, and Y. Hu, "Recent advances in the application of artificial intelligence in age-related macular degeneration," BMJ Open Ophthalmology, vol. 9, no. 1, 2024.
[38] Y. Hu, Y. Gao, W. Gao, W. Luo, Z. Yang, F. Xiong, Z. Chen, Y. Lin, X. Xia, X. Yin et al., "Amd-sd: An optical coherence tomography image dataset

for wet amd lesions segmentation," Scientific Data, vol. 11, no. 1, p. 1014, 2024.



ABONYO MITCHELL NINA is currently pursuing a Bachelor of science in Computer Science with Makerere University. Her research interests include artificial intelligence and machine learning.

She has experience in software development and data science, having worked on various projects involving computer vision and natural language processing. She completed a summer internship in machine learning research and has participated in various fellowship programs such as the Ubuntu-Net Alliance 2024. She has also contributed to research publications, including Advancements in Artificial Intelligence for Medical Imaging. Her current work focuses on deep learning applications in health care, while her previous research explored graph neural networks and knowledge distillation.

NAKIGULI ELIZABETH GLORIA is currently pursuing a Bachelor of science in Computer Science with Makerere University. Her research interests include artificial intelligence and machine learning.

She has experience in software development and data science, having worked on various projects involving computer vision and natural language processing. She completed a summer internship in machine learning research and has participated in various fellowship programs such as the Ubuntu-Net Alliance 2024. She has also contributed to research publications, including Advancements in Artificial Intelligence for Medical Imaging. Her current work focuses on deep learning applications in health care, while her previous research explored graph neural networks and knowledge distillation.