

Enhanced Retinal-Choroidal Disorders Classification Model via Temporal Sequence Analysis of OCT Images across Multiple Lines of Fovea

Kwankhao Tangprasert

*Department of Computer Engineering
Khon Kaen University
Khon Kaen, Thailand
kwankhao.t@kkumail.com*

Pitchaya Wiratchotisatian

*Department of Statistic
Khon Kaen University
Khon Kaen, Thailand
pitcwi@kku.ac.th*

Panita Khakhai, M.D.

*Department of Ophthalmology
Khon Kaen University
Khon Kaen, Thailand
panikha@kku.ac.th*

Wipada Laovirojjanakul, M.D.

*Department of Ophthalmology
Khon Kaen University
Khon Kaen, Thailand
wipala@kku.ac.th*

Sarun Paisarnsrisomsuk*

*Department of Computer Engineering
Khon Kaen University
Khon Kaen, Thailand
sarunpa@kku.ac.th*

Abstract—This study focuses on the classification of subretinal fluid (SRF) accumulation in conditions such as Central Serous Chorioretinopathy (CSC), Polypoidal Choroidal Vasculopathy (PCV), and Vogt-Koyanagi-Harada (VKH) disease. Optical coherence tomography (OCT) is a widely used imaging modality by ophthalmologists to diagnose these conditions; however, overlapping clinical features often make accurate differentiation challenging. This work aims to develop a robust classification model capable of making predictions from sequences of OCT scans, leveraging a dataset of 642 records that include over 23,000 retinal OCT images obtained from Srinagarind Hospital in Khon Kaen, Thailand. The model was trained and validated using stratified 10-fold cross-validation, achieving an overall classification accuracy of 96.57%, with per-class precision, recall, and F1 scores averaging 96.60%, 96.63%, and 96.60%, respectively. Additionally, gradient-weighted class activation mapping (Grad-CAM) was employed to highlight the anatomical regions relevant to the model's classification, which closely matched the diagnostic features identified by ophthalmologists. These findings indicate that the proposed approach holds substantial promise for advancing automated retinal-choroidal disorder analysis, thereby supporting clinical decision-making in ophthalmology.

Index Terms—Deep Learning, LSTM, CNN, Medical Image Analysis, Retinal OCT Image, Subretinal Fluid Classification, Central Serous Chorioretinopathy, Polypoidal Choroidal Vasculopathy, Vogt-Koyanagi-Harada.

I. INTRODUCTION

Vision is essential for performing a wide range of tasks in daily life, from navigating complex environments to engaging in social interactions. Even fundamental activities, such as reading or recognizing familiar faces, are dependent on clear visual clarity. Unfortunately, many individuals experience visual impairment due to retinal conditions. Naqeeb and Naser [1] reported a significant increase in the admission rates

for retinal and choroidal disorders in Australia, as determined by their ecological analysis. One prominent type of retinal-choroidal disorder involves the accumulation of subretinal fluid (SRF), which impairs vision across various age groups and populations. SRF is frequently associated with conditions such as Central Serous Chorioretinopathy (CSC), Polypoidal Choroidal Vasculopathy (PCV), and Vogt-Koyanagi-Harada (VKH) diseases.

Traditionally, ophthalmologists utilize optical coherence tomography (OCT) images to diagnose SRF conditions by evaluating the retinal and choroidal layers through multiple scans across the fovea. OCT employs both raster and radial scanning techniques, each serving a specific role in capturing the retinal architecture. Raster scans (Fig. 1a) consist of parallel, evenly spaced lines, providing high-resolution cross-sectional views that detail localized structures within the retina. Conversely, radial scans (Fig. 1b) are organized in a spoke-like configuration centered around the fovea, offering a comprehensive geometric overview from multiple orientations.

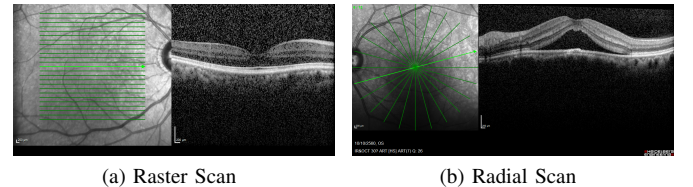


Fig. 1: Comparison of OCT scanning techniques. (a) Raster Scan: Parallel, high-resolution cross-sectional lines providing detailed views of localized retinal structures. (b) Radial Scan: Spoke-like arrangement around the fovea, offering comprehensive orientation and geometric insights essential for detecting subtle retinal abnormalities.

* Corresponding author.

These complementary scanning methods enhance diagnostic accuracy by enabling the detection of abnormalities that may be missed using a single scanning approach. The analysis of these images typically requires ophthalmologists to interpret numerous scans, a process that may take approximately 15 minutes per patient, thereby increasing the potential for human error and diagnostic delays. Despite advancements in OCT technology, SRF disorders often present subtle and complex features that remain challenging to differentiate, even for experienced practitioners.

A. Related Work

Deep learning methodologies, particularly Convolutional Neural Networks (CNNs), have gained prominence for automating image classification tasks, including OCT imaging in ophthalmology. For instance, [2]–[5] proposed CNN models for diagnosing choroidal neovascularization (CNV), diabetic macular edema (DME), and drusen, achieving high levels of accuracy, sensitivity, and specificity. Other studies have utilized CNNs to detect and classify age-related macular degeneration (AMD), successfully differentiating between diseased and healthy retinas [6], [7]. Elkholy and Marzouk [8] developed a CNN-based model trained to simultaneously classify DME, CNV, and AMD, achieving an impressive accuracy of over 93%. These Artificial Intelligence (AI) techniques have significantly reduced the time required for manual diagnosis and mitigated issues related to inter-observer variability among ophthalmologists.

Recurrent Neural Networks (RNNs), and their variants such as Long Short-Term Memory (LSTM) networks, have also been employed in OCT image analysis to capture temporal dependencies within sequential scans. This approach enhances the analysis by leveraging consecutive scans, revealing contextual details that are critical for evaluating disease progression across tissue layers. For example, Wang et al. [9] introduced a model that combined CNN-based feature extraction with LSTM to classify OCT images, effectively demonstrating the utility of RNNs in temporal analysis.

In the context of SRF accumulation, Hassan et al. [10] proposed a CNN-based approach for the automatic detection of CSC from OCT images, achieving a precision ranging from 92.39% to 98.91%. Khakhai et al. [11] developed CNN models to classify CSC, PCV, VKH, and normal patients, attaining an accuracy of 87.10%. However, a detailed analysis of the confusion matrix indicated frequent misclassification between CSC and PCV, underscoring the inherent challenges in differentiating these two conditions due to overlapping features. This limitation reflects a broader issue in existing methodologies that primarily rely on single-line scans across the fovea per patient, thereby restricting predictive accuracy in complex cases. As illustrated in Fig. 2a, an image of a PCV patient was misclassified as CSC due to the absence of distinct disease characteristics in that particular image, whereas Fig. 2b depicted another scan of the same patient showing the presence of these characteristics. This underscores the challenges associated with relying solely

on single-line scans across the fovea, which may constrain prediction accuracy in complex cases. Leveraging multiple OCT scans across multiple foveal lines presents a promising opportunity for more comprehensive analysis and improved diagnostic performance.

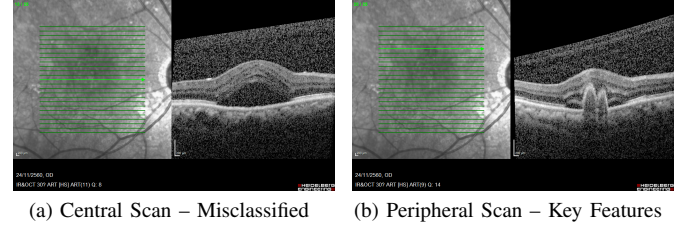


Fig. 2: Illustration of classification limitations in a single-image model. (a) Central OCT scan passing through the foveal region of a PCV patient, misclassified as CSC due to the absence of distinct PCV morphological markers. (b) Peripheral OCT scan from the same patient, exhibiting defining characteristics of PCV that were missed in the central scan alone.

B. Contributions of this paper

We present an approach integrating CNN with LSTM networks to analyze OCT raster and radial scan sequences across multiple foveal lines. Our model achieved an overall accuracy of 96.57%, outperforming existing CNN-based methods for SRF accumulation classification. The integration of temporal analysis via LSTM networks enabled the model to capture structural changes in retinal layers, thereby enhancing differentiation between CSC, PCV, VKH, and normal cases. Gradient-weighted class activation mapping (Grad-CAM) was employed to identify relevant anatomical regions for classification, aligning closely with diagnostic features identified by ophthalmologists.

II. METHODOLOGY

A. Data Description

The dataset was obtained from the Department of Ophthalmology, Faculty of Medicine, Srinagarind Hospital, Khon Kaen, Thailand, encompassing patient visits from January 2012 to 2022. The collection and use of this dataset were reviewed and approved by the Khon Kaen University Ethics Committee (approval number HE664021). The dataset was annotated by experienced ophthalmologists, who classified each case into one of four categories: CSC, PCV, VKH, or no SRF (normal), thereby indicating whether the patient exhibited one of the three pathological conditions or was classified as normal.

The dataset comprises 371 unique patients, with some patients having multiple visits, resulting in a total of 642 records, each representing a distinct set of images from a single visit. Each patient visit involved diagnostic imaging using both raster and radial scans. Specifically, each record may contain a sequence of 19, 25, or 31 raster images, along with a sequence of 6 or 12 radial images. In total, the study utilized 15,977 raster images and 7,044 radial images.

B. Data Preprocessing

Each record initially contained a varying number of raster and radial sequential images, leading to input with inconsistent dimensions, which posed significant challenges for computational efficiency and model performance. To address this, the number of sequential scans was standardized to 25 raster images and 12 radial images per record. Raster scans were systematically ordered from bottom to top, while radial scans were arranged beginning from the north direction and proceeding in a clockwise manner. Where necessary, excess scans were removed or additional scans were duplicated to achieve uniformity across the dataset, thereby enhancing the efficiency of the learning process and mitigating performance inconsistencies.

To be specific, let A_i ($i = 1, \dots, n$) represent the original sequence of images of length n and let B_j ($j = 1, \dots, m$) denote the new sequence of images of desired length m . To generate the new sequence, we select the j^{th} image in the new sequence to correspond to the k^{th} index of the original sequence, where $k = \text{round}(\frac{j}{m} \cdot n)$. This process is equivalent to performing a linear interpolation on the indices of the original sequential images.

After standardizing the number of images, we conducted image preprocessing by cropping a 600×600 pixel square centered on the retinal region of the OCT scan, as illustrated by the orange box in Fig. 3a. Subsequently, data augmentation techniques were employed to expand the dataset and enhance the model's generalization capabilities. To simulate a diverse range of imaging conditions, we applied adjustments in brightness, horizontal flipping, and random rotations to the dataset (Fig. 3b). Specifically, brightness adjustments involved setting the alpha value to 0.75 (dimming) and 1.25 (brightening), as shown in Fig. 3c. Furthermore, the images were horizontally flipped (Fig. 3d) and randomly rotated within a range of -45 to 45 degrees (Fig. 3e).

C. Model Architecture

The model architecture comprises two primary components: a feature extraction model and a classification model. The feature extraction model utilizes a CNN to extract informative features from retinal OCT images, while the classification model processes these features to predict retinal-choroidal disorders. Fig. 4 illustrates the overall model architecture.

1) *Feature Extraction Model*: The feature extraction model employs a pre-trained EfficientNetB7, originally trained on a diverse set of general images. In this study, the weights of the layers from the 7th block onward were unfrozen to allow for fine-tuning, enabling the model to adapt specifically to the OCT dataset. The extracted feature maps were then subjected to a global average pooling (GAP) layer to reduce their spatial dimensions. Subsequently, fully-connected layers were appended to refine these features, making them more contextually relevant to OCT images. ReLU activation functions were applied following each fully-connected layer, with a dropout layer incorporated to mitigate overfitting and reduce

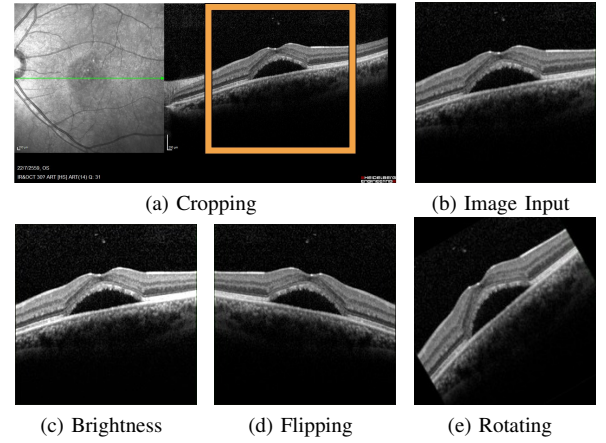


Fig. 3: Data preprocessing and augmentations for input images. (a) Cropping a 600×600 pixel square centered on the retinal region. (b-e) Data augmentation techniques applied to enhance model generalizability: (c) Brightness adjustments; (d) Horizontal flipping; and (e) Rotation within a range of -45 to 45 degrees

model complexity. A softmax layer was appended at the end to produce class probabilities.

2) *Classification Model*: For each record, individual OCT images from sequences of 25 raster and 12 radial scans were processed through the feature extraction model, and the extracted features were then input into the classification model. Features from raster and radial scans were fed into two separate two-layer LSTM networks to capture temporal dependencies. The outputs from these LSTM networks were concatenated, followed by fully-connected layers to produce the final prediction. ReLU activation functions were applied after each LSTM layer, and softmax activation was used after the final fully-connected layer to generate probabilistic predictions. The output of the complete model was a probabilistic classification into one of four categories: CSC, PCV, VKH, or Normal, representing three abnormal subretinal fluid accumulation conditions and a healthy condition.

D. Experimental Setup

In this study, we use a subject-wise stratified 10-fold cross-validation during the model training for both the feature extraction model and the classification model. First, the list of 371 patients is randomly divided into ten folds, each of which consists of approximately 12 patients with CSC condition, 9 patients with PCV condition, 6 patients with VKH, and 11 perfectly normal patients. Be aware that certain patients may have made multiple visits to the hospital, which lead to different amounts of records in each fold. During each iteration of the cross-validation process, one of the folds is fixed as the test set, and the remaining folds are separated into eight folds for training and one fold for validating. After repeating the process ten times – one iteration per each fold, the confusion matrices are aggregated and reported as the performance of the model.

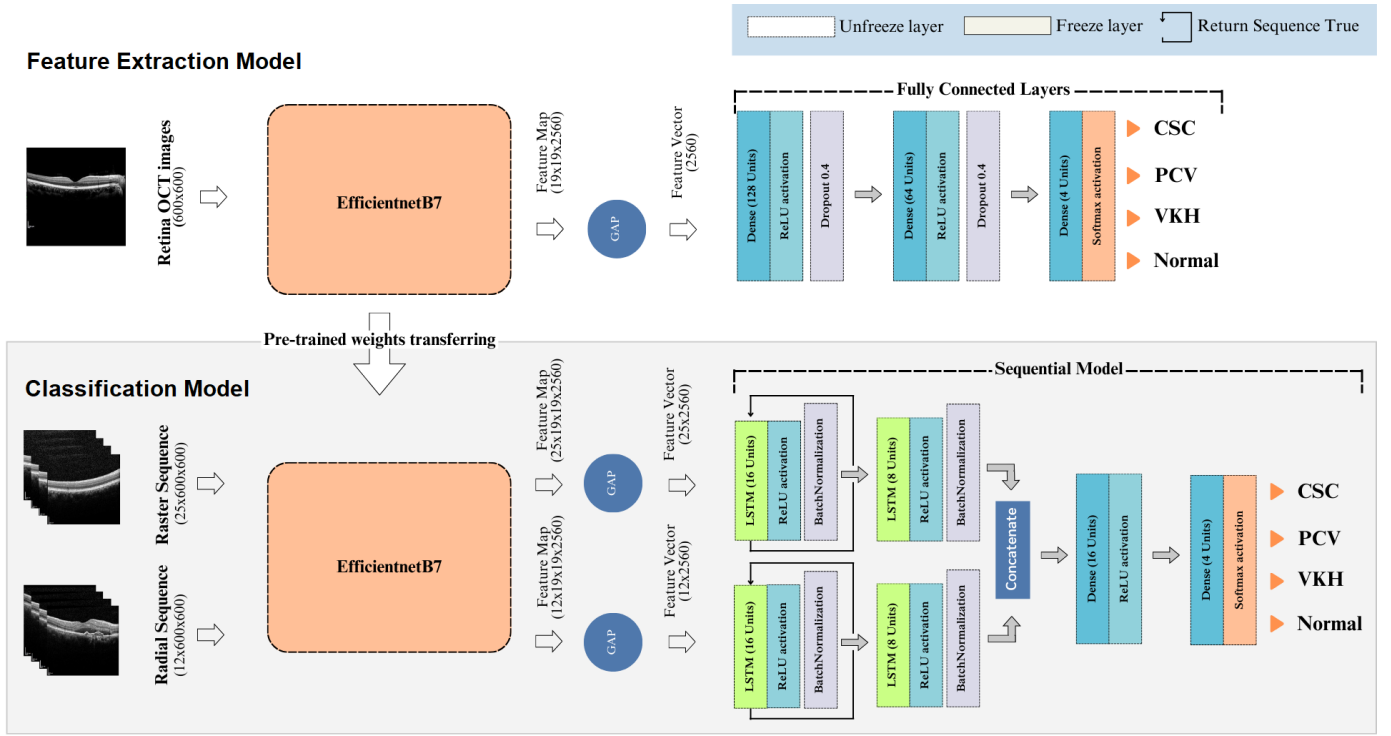


Fig. 4: Model architecture, consisting of a feature extraction model (top) that captures spatial features relevant to retinal structures, and a classification model (bottom) that utilizes these extracted features from sequential OCT images to classify them into one of four categories: CSC, PCV, VKH, or Normal.

The implementation of both models leveraged the CUDA parallel computing platform to expedite computations, using an NVIDIA A100 GPU with 40GB of memory, Python version 3.10.14, and Keras version 2.9.0. The Adam optimizer was employed to train both models. The feature extraction model was initialized with a learning rate of 10^{-5} , which was reduced by a factor of 0.2 after four epochs without improvement. Similarly, the classification model was trained with an initial learning rate of 10^{-3} , reduced by a factor of 0.4 after four epochs without improvement. The sparse categorical cross-entropy loss function was utilized to optimize model performance. A batch size of 64 was selected to enhance computational efficiency, with training conducted for up to 100 epochs, incorporating early stopping if validation performance failed to improve after ten epochs.

III. RESULTS AND DISCUSSION

A summary of the model's predictive performance is presented in Table I. The proposed model achieved an overall accuracy of 96.57%, significantly outperforming previous work [], which utilized a CNN trained on the same dataset and reported an overall accuracy of 87.10%. The average per-class precision, recall, and F1-score were 96.30%, 96.63%, and 96.60%, respectively, demonstrating robust predictive performance. These metrics highlight the model's high capability in effectively differentiating among the four classes, including

the three pathological SRF conditions (CSC, PCV, VKH) and normal cases. The incorporation of LSTM layers for temporal analysis provided a notable advantage over previous CNN-only approaches by enabling the model to capture complex temporal dependencies inherent in sequential OCT images, thus improving overall diagnostic accuracy.

To further interpret the results, we compared our work with that of Kakai et al. [11], as both studies utilized the same dataset, enabling a direct comparison. Their model encountered significant challenges in differentiating between CSC and PCV, misclassifying 59 out of 528 PCV records as CSC and, conversely, 34 out of 522 CSC records as PCV. This difficulty likely stems from the shared morphological features between CSC and PCV, which make distinguishing between these two conditions particularly challenging. This observation is consistent with previous studies by Lee et al. [12], which highlighted the inherent challenges in accurately differentiating CSC from PCV due to their overlapping morphological characteristics of the choroid. In contrast, our model demonstrated a marked improvement in differentiating between these two classes, with only a few instances of misclassification. This underscores the effectiveness of incorporating temporal analysis across multiple sequential scans, which significantly enhanced the model's predictive performance.

To validate the models' focus on relevant anatomical regions and ensure the reliability of their outputs, Gradient-weighted

	CSC	PCV	VKH	Normal	Recall	F1 score
CSC	152	4	3	0	95.60%	94.41%
PCV	6	156	1	2	97.04%	97.33%
VKH	4	0	164	1	97.04%	96.00%
Normal	1	0	0	148	99.33%	98.67%
Precision	93.25%	97.62%	97.50%	98.01%		

TABLE I: Performance metrics for the classification model, including cumulative confusion matrix, precision, recall, and F1-score for each class: CSC, PCV, VKH, and Normal. With an overall accuracy of 96.57%, these results highlight the effectiveness of incorporating temporal analysis across OCT scan sequences in this work, significantly enhancing the model’s ability to distinguish between the four classes compared to prior single-image models used in previous work.

Class Activation Mapping (Grad-CAM) [13] was employed. Grad-CAM visualized the input regions most influential in the model’s classification decisions by providing an overlaid heatmap on the OCT scans. This enabled the identification of critical anatomical features that the model considered significant during decision-making. The results in Fig. 5 shows that the model primarily focused on key anatomical regions, such as the retinal pigment epithelium and the choroid, rather than on imaging artifacts or irrelevant regions, enhancing confidence in the predictions. These findings align well with the regions ophthalmologists emphasize during diagnosis, indicating that the model’s learning is consistent with expert clinical practice. Furthermore, the use of Grad-CAM provided much-needed transparency in the model’s decision-making process, which is essential for building clinician trust in automated systems.

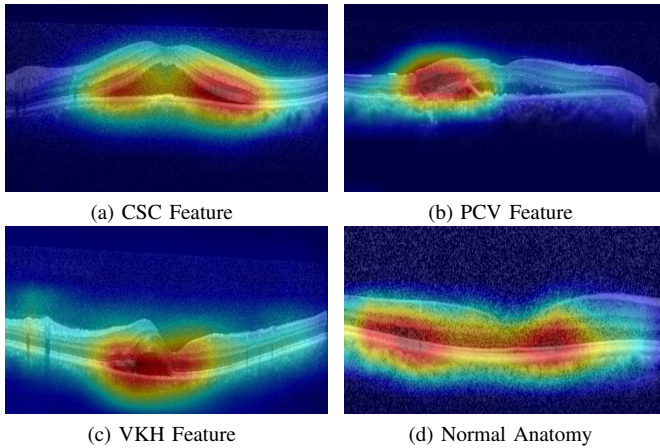


Fig. 5: Grad-CAM visualization of the feature extraction model, highlighting regions of OCT scans most influential in decision-making. Heatmaps overlay key anatomical features, such as the retinal pigment epithelium and choroid, which align closely with areas examined by ophthalmologists for diagnosing (a) CSC, (b) PCV, (c) VKH, and (d) a normal patient.

IV. CONCLUSION

The proposed hybrid CNN-LSTM model demonstrates a substantial advancement in the classification of subretinal fluid (SRF) accumulation by effectively leveraging temporal information present in sequential OCT images. This approach enables the model to capture nuanced structural changes in the retinal layers that are often missed by traditional single-image models. The integration of convolutional and recurrent layers allowed for the utilization of both spatial and temporal features, resulting in a robust predictive capability with an accuracy of 96.57%, thereby significantly outperforming existing approaches.

Furthermore, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to validate the model’s attention to relevant anatomical regions, thereby enhancing the interpretability of the model’s predictions. The regions highlighted by Grad-CAM closely aligned with those typically evaluated by ophthalmologists, which is crucial for building trust in the model’s decision-making process.

The proposed model demonstrates a significant advancement in automated retinal disorder classification by effectively integrating spatial and temporal features from sequential OCT images. This capability not only enhances diagnostic accuracy but also streamlines the process, offering the potential to support ophthalmologists with real-time decision-making and reducing the reliance on time-intensive manual evaluations.

One consideration in this study is the reliance on data primarily from a Thai population, which could affect the model’s generalizability to other demographics. Expanding validation efforts to include diverse datasets and integrating multimodal inputs, such as patient demographics and clinical histories, could enhance its predictive accuracy. Leveraging longitudinal data to analyze disease progression and adopting semi-supervised approaches to utilize unlabeled data are also promising directions for future research.

REFERENCES

- [1] M. R. Naqeeb and A. Y. Naser, "Hospitalisation trends for choroid and retina diseases in the past 20 years: An ecological study in australia," *Clinical Optometry*, pp. 247–259, 2023.
- [2] F. Li, H. Chen, Z. Liu, X. d. Zhang, M. s. Jiang, Z. z. Wu, and K. q. Zhou, "Deep learning-based automated detection of retinal diseases using optical coherence tomography images," *Biomedical optics express*, vol. 10, no. 12, pp. 6204–6226, 2019.
- [3] M. Toğaçar, B. Ergen, and V. Tümen, "Use of dominant activations obtained by processing OCT images with the CNNs and slime mold method in retinal disease detection," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 646–666, 2022.
- [4] S. A. Kamran, S. Saha, A. S. Sabbir, and A. Tavakkoli, "Optic-Net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 2019, pp. 964–971.
- [5] T. S. Apon, M. M. Hasan, A. Islam, and M. G. R. Alam, "Demystifying deep learning models for retinal oct disease classification using explainable AI," in *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. IEEE, 2021, pp. 1–6.
- [6] N. Motozawa, G. An, S. Takagi, S. Kitahata, M. Mandai, Y. Hirami, H. Yokota, M. Akiba, A. Tsujikawa, M. Takahashi *et al.*, "Optical coherence tomography-based deep-learning models for classifying normal and age-related macular degeneration and exudative and non-exudative age-related macular degeneration changes," *Ophthalmology and therapy*, vol. 8, pp. 527–539, 2019.
- [7] A. Thomas, P. Harikrishnan, R. Ramachandran, S. Ramachandran, R. Manoj, P. Palanisamy, and V. P. Gopi, "A novel multiscale and multipath convolutional neural network based age-related macular degeneration detection using OCT images," *Computer methods and programs in biomedicine*, vol. 209, p. 106294, 2021.
- [8] M. Elkholy and M. A. Marzouk, "Deep learning-based classification of eye diseases using convolutional neural network for OCT images," *Frontiers in Computer Science*, vol. 5, p. 1252295, 2024.
- [9] C. Wang, Y. Jin, X. Chen, and Z. Liu, "Automatic classification of volumetric optical coherence tomography images via recurrent neural network," *Sensing and Imaging*, vol. 21, no. 1, p. 32, 2020.
- [10] S. A. E. Hassan, S. Akbar, S. Gull, A. Rehman, and H. Alaska, "Deep learning-based automatic detection of central serous retinopathy using optical coherence tomographic images," in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*. IEEE, 2021, pp. 206–211.
- [11] P. Khakhai, S. Paisarnsrisomsuk, P. Wiratchotisatian, K. Tangprasert, and W. Laovirojjanakul, "Diagnostic performance of machine learning technology using optical coherence tomographic image in retinal diseases presented with subretinal fluid," *[Manuscript submitted for publication]*.
- [12] H. Lee, K. Bae, S. W. Kang, S. J. Woo, N.-K. Ryoo, S. J. Kim, and G. Han, "Morphologic characteristics of choroid in the major choroidal thickening diseases, studied by optical coherence tomography," *PLoS One*, vol. 11, no. 1, p. e0147139, 2016.
- [13] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [14] D. Lu, M. Heisler, S. Lee, G. Ding, M. V. Sarunic, and M. F. Beg, "Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network," *arXiv preprint arXiv:1710.04778*, 2017.
- [15] C. Otte, S. Otte, L. Wittig, G. Hüttmann, C. Kugler, D. Drömann, A. Zell, and A. Schlaefer, "Investigating recurrent neural networks for OCT A-scan based tissue analysis," *Methods of Information in Medicine*, vol. 53, no. 04, pp. 245–249, 2014.
- [16] Z. Xu, W. Wang, J. Yang, J. Zhao, D. Ding, F. He, D. Chen, Z. Yang, X. Li, W. Yu *et al.*, "Automated diagnoses of age-related macular degeneration and polypoidal choroidal vasculopathy using bi-modal deep convolutional neural networks," *British Journal of Ophthalmology*, vol. 105, no. 4, pp. 561–566, 2021.