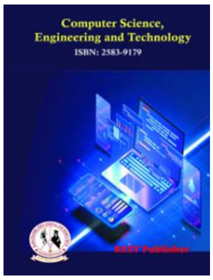
Computer Science, Engineering and Technology

Vol: 3(3), September 2025

REST Publisher; ISSN: 2583-9179

Website: https://restpublisher.com/journals/cset/

DOI: https://doi.org/10.46632/cset/3/3/7



Explainable Deep Learning for Age-related Macular Degeneration Detection in OCT Images: A Review

*** Shipra Singh, Shashwati Banerjea**

Harcourt Butler Technical University, Kanpur, India.

*Corresponding Author Email: shipragkp99@gmail.com

Abstract: *Age-related Macular Degeneration (AMD) is a leading cause of vision loss among older adults. Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have significantly improved the accuracy of AMD detection using Optical Coherence Tomography (OCT) images. Despite their performance, these models often function as black boxes, limiting their trustworthiness in clinical environments. This review highlights key explainable artificial intelligence (XAI) techniques—including Grad-CAM, SHAP, and LIME—that have been utilized to interpret deep learning models in AMD classification tasks. We discuss their methodologies, applications, comparative strengths, and clinical relevance. The paper also identifies current challenges and suggests future directions for integrating interpretable AI into ophthalmic diagnostics.*

Keywords: *Age-related Macular Degeneration, Optical Coherence Tomography, Deep Learning, Explainable AI, Grad-CAM, SHAP, Model Interpretability.*

1. INTRODUCTION

Age-related Macular Degeneration (AMD) is a progressive retinal disorder that impairs central vision, primarily affecting individuals over the age of 60. Early and accurate detection is essential for timely intervention. OCT imaging provides detailed cross-sectional views of the retina, making it a valuable tool for diagnosing AMD. Deep learning models, especially CNNs, have shown impressive results in analyzing OCT scans. However, the opacity of these models has raised concerns among clinicians. Explainable AI (XAI) methods offer a way to understand model decisions, which is critical for fostering trust and supporting clinical integration. This paper provides an overview of key XAI methods used in AMD detection and evaluates their impact on clinical interpretability.

2. BACKGROUND

Age-related Macular

Degeneration AMD manifests in two forms: dry (non-exudative) and wet (exudative). Dry AMD is characterized by the accumulation of drusen and retinal pigment epithelium degeneration, while wet AMD involves abnormal blood vessel growth leading to fluid leakage. OCT imaging is the gold standard for visualizing these pathological changes.

Deep Learning in Retinal Imaging

CNNs are a class of deep learning models tailored for image data. They are capable of automatically learning hierarchical patterns from OCT images, enabling effective classification of retinal diseases, including AMD. Popular CNN architectures like ResNet, VGG, and DenseNet have been widely adopted in this domain.

Importance of Explainability

Although CNNs provide high diagnostic accuracy, their lack of transparency poses challenges for clinical adoption. Explainability techniques allow stakeholders to interpret how models arrive at specific predictions, thereby enabling error analysis, increasing trust, and ensuring ethical deployment.

3. EXPLAINABLE AI TECHNIQUES

Grad-CAM (Gradient-weighted Class Activation Mapping)

Grad-CAM generates visual heatmaps highlighting the regions of an input image that contribute most to the model's prediction. It leverages gradient information flowing into the last convolutional layer to identify important features.

SHAP (SHapley Additive ex Planations)

SHAP uses cooperative game theory to assign importance values to each feature in a model's prediction. It offers both local and global interpretability, making it suitable for understanding model behavior across different inputs.

LIME (Local Interpretable Model-agnostic Explanations)

LIME approximates the behavior of complex models by training interpretable surrogate models on perturbed input samples. This technique reveals how small changes in input features influence predictions.

4. APPLICATIONS IN AMD DETECTION

XAI methods have been increasingly adopted in OCT-based AMD studies: Grad-CAM was integrated with ResNet-50 by Kermany et al. to validate that the model focused on AMD-related retinal regions. Ahnaf et al. utilized SHAP to interpret a custom CNN, identifying which pixels contributed to dry AMD predictions. Other researchers have combined Grad-CAM and LIME to cross-verify model attention mechanisms

5. COMPARATIVE EVALUATION OF XAI METHODS

TABLE 1.

Method	Strengths	Limitations	Typical Use Case
Grad-CAM	Offers intuitive visual explanations	Requires specific CNN structures	Visual validation on OCT
SHAP	Model-agnostic with strong theory	Computationally expensive	Feature contribution analysis
LIME	Flexible across model types	Unstable with high-dimensional data	Local prediction interpretation

Clinical Significance

- XAI techniques bridge the gap between AI predictions and clinical understanding.
- Visual explanations aid ophthalmologists in validating model outcomes.
- Explainable systems can enhance trust and accountability in AI-assisted diagnosis, crucial for regulatory compliance.

Current Challenges

- Lack of standardized metrics to evaluate explanation quality.
- Difficulty in interpreting explanations without domain expertise.
- Computational costs associated with high-resolution OCT image analysis.

Future Prospects

- Development of hybrid interpretability frameworks
- combining visual and statistical insights.
- Incorporation of XAI tools into clinical decision support systems.
- Educational initiatives to train medical professionals in AI interpretability.

6. CONCLUSION

Explainable AI offers essential tools for interpreting deep learning models in the context of AMD detection. Techniques such as Grad-CAM, SHAP, and LIME enhance model transparency and support clinical decision-making. Continued research and standardization are needed to fully integrate these methods into ophthalmic practice.

REFERENCES

[1]. Kermany, D. S., et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning." Cell, 2018.
[2]. Ahnaf, S. M. A., et al. "Exploring CNN Architectures for AMD Classification Using OCT Images." IEEE Access, 2024.
[3]. Selvaraju, R. R., et al. "Grad-CAM: Visual explanations from deep networks via gradient-based localization." ICCV, 2017.
[4]. Lundberg, S. M., and Lee, S.-I. "A unified approach to interpreting model predictions." NeurIPS, 2017.
[5]. Ribeiro, M. T., et al. "Why should I trust you? Explaining the predictions of any classifier." ACM SIGKDD, 2016.