An Explainable Deep Learning Framework for Multi-Disease Ocular Classification and Myopia Severity Estimation

Sharat Acharja, Azhar Hossen Novel, Amikur Rahman
Department of Computer Science and Engineering
Green University of Bangladesh
Dhaka, Bangladesh
sharatacharjee6@gmail.com, rampardnovel83@gmail.com, thamik67@gmail.com

***Abstract*—Visual disorders arising from ocular diseases and refractive abnormalities continue to affect a large portion of the global population, highlighting the importance of automated and interpretable screening solutions. . Although deep learning techniques have shown strong performance in fundus image analysis, many existing approaches lack interpretability and do not address disease severity, limiting their clinical adoption. This work introduces an explainable deep learning-based framework for precision eye care designed to jointly analyze multiple ocular conditions using retinal fundus images and further estimates myopia severity based on diagnostic keyword-derived annotations. A transfer learning strategy is employed to extract discriminative retinal features, while explainable artificial intelligence techniques are integrated to visualize regions influencing model decisions. In addition, a guideline-based clinical decision support mechanism is designed to provide preliminary treatment suggestions for myopia-related cases. Experimental evaluation on a publicly available ocular dataset demonstrates consistent classification performance, indicating that the proposed system can serve as a supportive tool for early eye disease screening and severity-aware clinical analysis.**

***Index Terms*—Fundus Image Analysis, Ocular Disease Classification, Myopia Severity Analysis , Explainable Artificial Intelligence, Deep Learning, Clinical Decision Support**

I. Introduction

Vision impairment caused by ocular diseases and refractive errors remains a major global health concern, affecting millions of individuals across different age groups. Disorders such as diabetic retinopathy, glaucoma, cataract, age-related macular degeneration, hypertensive retinopathy, and myopia can lead to progressive vision loss if not detected and managed at an early stage. Retinal fundus imaging has become a widely adopted, non-invasive diagnostic modality for visualizing structural and pathological changes in the eye, making it particularly suitable for large-scale screening and automated analysis.

Recent progress in machine learning, particularly deep learning, has enabled more effective automated analysis of retinal fundus images. . In particular, convolutional neural networks have demonstrated strong capability in learning discriminative retinal features directly from image data, enabling accurate ocular disease classification [1], [2]. However, a large number of existing studies remain confined to single-disease prediction or prioritize accuracy improvement without broader clinical context without addressing model transparency. Such black-box systems often face resistance in clinical practice, where interpretability is essential for validating automated decisions.

Myopia is one of the most prevalent refractive errors worldwide, and its global incidence continues to rise at an alarming rate [3]. While numerous studies have investigated myopia detection using imaging and clinical data, severity-level inference of myopia from retinal fundus images, particularly in the absence of direct refractive measurements, remains relatively underexplored. Severity-aware analysis is crucial for clinical decision-making, personalized treatment planning, and long-term monitoring of disease progression. Integrating myopia power estimation with automated ocular disease classification can therefore enhance the clinical relevance of intelligent eye care systems.

Another significant limitation of many deep learning-based medical imaging models is the lack of explainability. Explainable artificial intelligence techniques aim to address this issue by providing visual insights into model behavior, typically by highlighting image regions that contribute most to predictions [4]. In ophthalmic applications, such visual explanations can assist clinicians in verifying whether automated decisions are based on anatomically meaningful retinal structures rather than irrelevant patterns.

Motivated by these challenges, this paper proposes an explainable machine learning framework for precision eye care that jointly performs multi-disease ocular classification and severity-aware myopia analysis using retinal fundus images. The proposed approach integrates explainable visualization techniques to improve transparency and incorporates a guideline-based clinical decision support strategy to provide preliminary management suggestions for myopia-related cases. The overall objective of this work is to bridge the gap between high-performance automated analysis and clinically interpretable, severity-aware decision support systems for ocu-

lar health. The main contributions of this work are summarized as follows:

- An explainable deep learning framework for multi-disease ocular classification using retinal fundus images.
- A severity-aware myopia power estimation strategy based on diagnostic keyword-derived annotations without relying on direct refractive measurements.
- Integration of explainable artificial intelligence techniques to visually interpret model predictions and support clinical validation.
- A guideline-driven clinical decision support module for preliminary myopia management recommendations.

II. Related Work

Automated analysis of ocular diseases using retinal fundus images has gained substantial attention in recent years due to advancements in machine learning and deep learning techniques. Early studies primarily relied on handcrafted features combined with traditional classifiers; however, such approaches were often constrained by limited feature expressiveness and poor generalization capability across diverse disease patterns.

With the emergence of deep convolutional neural networks, data-driven feature learning has significantly improved the accuracy and robustness of fundus image-based ocular disease classification. Several influential works have demonstrated the effectiveness of deep learning for detecting specific ocular conditions, including diabetic retinopathy, glaucoma, cataract, and age-related macular degeneration [5], [6]. Despite their strong performance, most of these approaches are designed for single-disease prediction and do not readily extend to multi-disease clinical scenarios.

Recent studies have explored multi-disease ocular classification using fundus images; however, many of these methods primarily focus on improving overall classification accuracy without explicitly modeling disease severity or progression [7]. As a result, their applicability in clinical decision-making remains limited, particularly in cases requiring severity-aware assessment.

Myopia-related research has largely concentrated on binary detection or basic screening tasks. Although a few studies attempt to categorize myopia severity, they often depend on external clinical measurements or rule-based criteria, which restrict their integration into fully automated image-based systems. Estimating refractive power severity directly from retinal fundus images thus remains an open and underexplored research problem.

Another major limitation of existing deep learning-based eye care systems is their lack of interpretability. Most models operate as black-box predictors, providing limited insight into the visual cues driving their predictions. Explainable artificial intelligence techniques, such as gradient-based visualization and saliency mapping, have been introduced to address this issue [4], [8]. While these methods improve model transparency, they are rarely integrated with multi-disease ocular classification and severity-aware myopia analysis within a unified framework.

In contrast to prior work, this study proposes an explainable machine learning framework that jointly performs multi-disease ocular classification, myopia severity estimation, and guideline-based clinical decision support. By combining severity-aware analysis with explainable visualization, the proposed approach aims to enhance both the clinical interpretability and practical usability of automated eye care systems.

III. Methodology

A. Dataset Description

This study utilizes the Ocular Disease Intelligent Recognition (ODIR-5K) dataset [9], a publicly available benchmark designed for automated ocular disease analysis using retinal fundus images. The dataset contains color fundus photographs for both left and right eyes of each subject, accompanied by clinically annotated diagnostic keywords. The availability of bilateral eye images enables comprehensive ocular assessment and increases the diversity of visual patterns available for learning.

The dataset includes annotations for eight distinct ocular disease categories, allowing the formulation of a multi-disease classification task within a unified learning framework. In addition to general disease labels, diagnostic keywords associated with myopia are exploited to facilitate severity-aware refractive power estimation. This design enables the proposed system to extend beyond binary disease detection and incorporate clinically meaningful myopia severity categorization.

ODIR-5K was selected due to its heterogeneous disease manifestations, variations in image quality, and realistic clinical annotation structure, which closely reflect real-world screening scenarios. To ensure data reliability and experimental consistency, only valid retinal image files were included in the study. Incomplete records, corrupted images, and non-image entries were excluded during data preparation, thereby reducing noise and supporting stable model training and evaluation.

B. Data Preprocessing and Label Generation

All retinal images were resized to a uniform resolution of 224×224 pixels to ensure compatibility with the deep learning model input requirements. Image intensity values were normalized to enhance numerical stability and facilitate effective convergence during training. Diagnostic keyword annotations provided in the dataset were processed to generate categorical labels corresponding to different ocular disease classes.

For myopia-related samples, severity levels were grouped into three categories, namely mild, strong, and pathological myopia, using diagnostic keyword-derived annotations aligned with clinically accepted refractive power ranges. Specifically, mild myopia corresponds approximately to refractive power values between -0.5 and -3.0 diopters (D), strong myopia

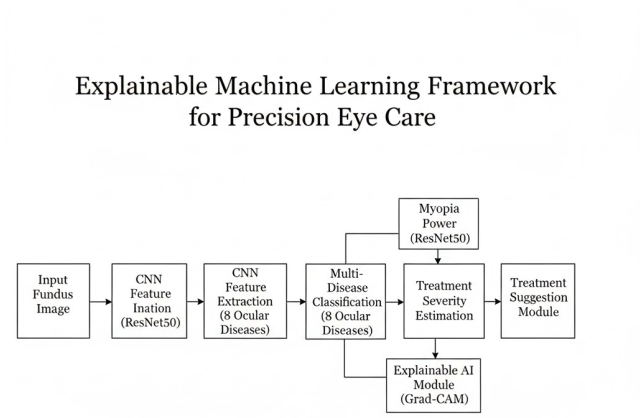

Fig. 1. Overall architecture of the proposed explainable machine learning framework illustrating fundus image input, multi-disease classification, myopia severity estimation, explainable AI integration, and clinical decision support.

corresponds to values between -3.0 and -8.0 D, and pathological myopia corresponds to values ranging from approximately -8.0 to -20.0 D. These ranges are included solely to provide clinical context for severity categorization and are not directly estimated by the proposed model. The dataset was divided into training, validation, and test sets using stratified sampling to preserve class distribution and minimize sampling bias.

C. Deep Learning Model Architecture

A transfer learning approach was adopted to efficiently extract discriminative retinal features. A ResNet50 convolutional neural network pre-trained on the ImageNet dataset was employed [2]. The convolutional layers of the base network were frozen to retain previously learned visual representations and to reduce the risk of overfitting on the ocular dataset.

On top of the pre-trained backbone, custom classification layers were added, including a global average pooling layer to reduce spatial dimensionality, followed by fully connected layers for feature refinement. A softmax activation function was applied at the output layer to generate class probability distributions for multi-disease ocular classification. The model was optimized using categorical cross-entropy loss in conjunction with the Adam optimizer.

D. Explainable Artificial Intelligence Integration

To enhance transparency and clinical interpretability, explainable artificial intelligence techniques were incorporated into the proposed framework. Gradient-based visualization methods were employed to generate heatmaps that highlight retinal regions contributing most significantly to the model's predictions. These visual explanations assist in verifying whether the model focuses on anatomically meaningful structures, such as the optic disc or macular region, rather than irrelevant artifacts.

The integration of explainability allows clinicians to better understand and trust the automated predictions, thereby bridging the gap between deep learning models and real-world clinical decision-making.

E. Clinical Decision Support Strategy

A rule-based clinical decision support mechanism was designed to provide preliminary management suggestions for myopia-related cases. The decision logic maps predicted myopia severity levels to corresponding non-invasive clinical recommendations derived from established ophthalmological guidelines. This module is intended to support early screening and awareness rather than replace professional medical diagnosis.

By integrating multi-disease classification, myopia severity estimation, and explainable visualization, the proposed methodology forms a unified and clinically relevant framework for precision eye care.

IV. Experimental Results

The proposed framework was evaluated to assess its effectiveness in multi-disease ocular classification, myopia severity estimation, and model interpretability. Experimental analysis was conducted using a held-out test set to ensure unbiased performance evaluation. Standard classification metrics were employed to measure predictive reliability and consistency.

A. Classification Performance

The performance of the proposed multi-disease ocular classification framework was evaluated using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). These complementary metrics provide a holistic assessment of overall predictive reliability as well as class-wise discrimination capability under class imbalance conditions.

As summarized in Table I, the proposed framework demonstrates consistent and reliable classification performance across multiple ocular disease categories. The close alignment between precision and recall indicates an effective balance between sensitivity and specificity, while the F1-score reflects stable predictive behavior in the presence of class imbalance.

TABLE I
Overall Performance of the Proposed Framework on the Test Dataset

Metric	Value
Accuracy	84.95%
Precision	0.83
Recall	0.84
F1-score	0.83
AUC	0.8066

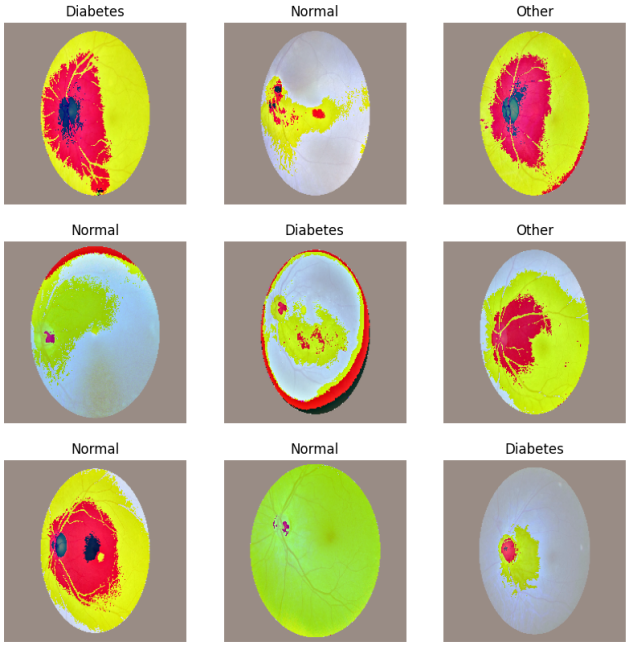

Fig. 2. Grad-CAM visualizations illustrating discriminative retinal regions that influence the proposed model's predictions for representative ocular disease and myopia cases.

Moreover, an AUC value of 0.8066 confirms strong class separability and suggests that the learned representations generalize well to previously unseen retinal fundus images rather than exhibiting chance-level performance.

B. Myopia Severity Estimation

For myopia-related cases, the proposed framework categorizes refractive power into mild, strong, and pathological levels based on diagnostic keyword-derived annotations. The severity-aware analysis enables differentiation between varying degrees of myopia, which is essential for clinical assessment and long-term monitoring. Experimental observations suggest that the model maintains consistent prediction behavior across different myopia severity groups.

C. Explainability Analysis Using Grad-CAM

To enhance interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize regions of retinal fundus images [4] that contribute most significantly to model predictions. The generated heatmaps highlight clinically relevant anatomical structures, such as the optic disc and macular regions, indicating that the model bases its decisions on meaningful retinal features rather than background artifacts. update that

The visual explanations provided by Grad-CAM improve transparency and support clinical validation of automated predictions. By aligning highlighted regions with known ophthalmological structures, the explainability component enhances trust in the proposed system.

Overall, the experimental findings demonstrate that the proposed explainable machine learning framework achieves a balance between predictive performance and interpretability. The integration of multi-disease classification, myopia severity estimation, and explainable visualization supports its applicability as a decision support tool for early ocular disease screening and precision eye care.

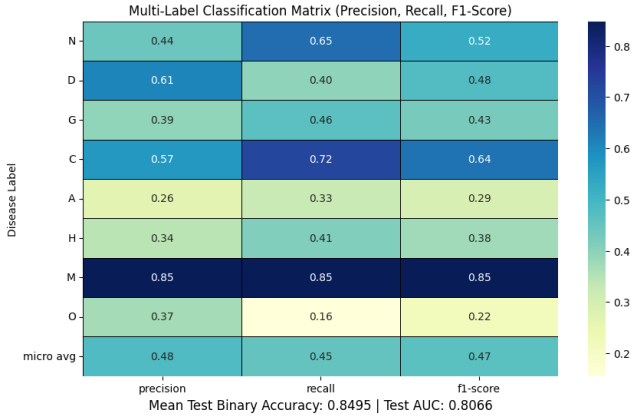

Fig. 3. Disease-wise precision, recall, and F1-score heatmap for the proposed multi-disease ocular classification framework. Higher performance is observed for myopia-related cases, while comparatively lower scores for certain disease categories indicate increased visual ambiguity and class imbalance effects.

D. Disease-wise Performance Analysis

In addition to overall evaluation metrics, disease-wise performance analysis was conducted to examine class-specific behavior of the proposed framework. Precision, recall, and F1-score were computed separately for each ocular disease category to identify performance variability across different clinical conditions.

Figure 3 illustrates that myopia-related cases achieve consistently high performance across all evaluation metrics, whereas comparatively lower scores are observed for certain disease categories with subtle retinal manifestations. This observation emphasizes the importance of class-wise analysis beyond overall accuracy when evaluating multi-disease ocular classification systems.

V. Discussion

The experimental results demonstrate that the proposed explainable machine learning framework effectively analyzes retinal fundus images for multi-disease ocular classification while maintaining clinically meaningful interpretability. The achieved performance metrics indicate that the model successfully captures discriminative retinal patterns associated with diverse ocular conditions, even under variations in image quality and disease presentation.

The use of transfer learning with a pre-trained ResNet50 backbone plays a significant role in stabilizing model performance. Leveraging representations learned from large-scale natural image datasets reduces the reliance on handcrafted features and improves generalization, which is particularly important in medical imaging scenarios where annotated data are limited and heterogeneous.

Incorporating severity-aware myopia analysis further enhances the clinical relevance of the proposed system. Rather than restricting the task to binary myopia detection, the framework differentiates between mild, strong, and pathological myopia categories, aligning more closely with real-world clinical decision-making and long-term patient monitoring. However, it should be noted that refractive power estimation is indirectly inferred from diagnostic keyword-based annotations, which may introduce labeling uncertainty when compared to direct optometric measurements.

Explainable artificial intelligence serves a critical role in bridging the gap between automated prediction and clinical trust. Grad-CAM visualizations demonstrate that the model consistently attends to anatomically meaningful regions such as the optic disc and macular area, suggesting that predictions are guided by relevant retinal structures rather than spurious image artifacts. Nonetheless, these explanations are qualitative in nature and should be viewed as supportive visual evidence rather than definitive clinical justification.

Despite the encouraging results, several limitations remain. The evaluation is conducted on a single public dataset, which may not fully represent the diversity of real-world clinical populations. In addition, class imbalance and annotation variability can negatively affect performance for less prevalent ocular disease categories. Future research may address these limitations through cross-dataset validation, integration of numerical clinical measurements, and collaboration with ophthalmologists for expert-verified annotations.

Overall, the proposed framework achieves a balanced trade-off between predictive performance, interpretability, and clinical applicability. While not intended to replace professional ophthalmological diagnosis, the system demonstrates strong potential as a supportive tool for early ocular disease screening and precision eye care decision support.

VI. Conclusion

This paper presented an explainable machine learning framework for precision eye care that integrates multi-disease ocular classification, myopia severity estimation, and interpretable model visualization using retinal fundus images. By leveraging transfer learning and clinically motivated label generation, the proposed system effectively analyzes diverse ocular conditions while maintaining robust generalization on unseen data.

The experimental evaluation demonstrates that the framework achieves consistent classification performance and provides meaningful insights into myopia severity levels. The incorporation of explainable artificial intelligence techniques, such as Grad-CAM, enhances transparency by visually highlighting anatomically relevant retinal regions influencing model predictions. This interpretability supports clinical validation and improves trust in automated decision-making systems.

Although the proposed approach shows promising results, it is intended as a supportive screening and decision assistance tool rather than a replacement for professional medical diagnosis. Future work may focus on expanding the dataset, incorporating additional clinical metadata, and performing cross-population validation to further improve robustness and real-world applicability.

Overall, the study demonstrates the potential of explainable deep learning methods to support early ocular disease screening and severity-aware analysis, contributing toward more accessible and reliable precision eye care solutions.

Acknowledgment

The authors sincerely acknowledge the Department of Computer Science and Engineering at Green University of Bangladesh for providing the academic infrastructure, laboratory facilities, and a supportive research environment necessary for the successful completion of this work. The authors are grateful for the continuous academic guidance, constructive feedback, and methodological insights received throughout the research process, which significantly contributed to refining the proposed framework.

The authors also appreciate the institutional support that enabled access to computational resources and scholarly materials required for experimental evaluation and analysis. Finally, sincere thanks are extended to all faculty members and peers whose discussions and suggestions indirectly contributed to improving the clarity, structure, and overall quality of this research.

References

[1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[3] B. A. Holden, T. R. Fricke, D. A. Wilson, M. Jong, K. S. Naidoo, P. Sankaridurg, T. Y. Wong, T. J. Naduvilath, and S. Resnikoff, "Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050," *Ophthalmology*, vol. 123, no. 5, pp. 1036–1042, 2016.

[4] 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 (ICCV)*, 2017, pp. 618–626.

[5] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, and J. Cuadros, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[6] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. Tan, N. D. Quang, A. Gan, H. H. Hamzah, R. Garcia-Franco, I. S. Yeo, and S. Y. Lee, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.

[7] W. et al., "Multi-disease detection in retinal fundus images using deep learning," *Computer Methods and Programs in Biomedicine*, 2021.

[8] A. et al., "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[9] L. e. a. Li, "Odir-2019: An ophthalmic disease intelligent recognition dataset," *IEEE Access*, vol. 7, pp. 182 752–182 765, 2019.