An Explainable Deep-Learning Model to Aid in the Diagnosis of Age Related Macular Degeneration

María Herrero-Tudela[1(✉)], Roberto Romero-Oraá[1,2], Roberto Hornero[1,2], Gonzalo C. Gutiérrez-Tobal[1,2], María I. Lopez[1], and María García[1,2]

[1] Biomedical Engineering Group, University of Valladolid, Valladolid, Spain
maria.herrero.tudela@uva.es
[2] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Valladolid, Spain
https://gib.tel.uva.es/members.php#herrerotudela_m

Abstract. Age-related macular degeneration (AMD) is the most frequent cause of blindness in people of advanced age. As AMD is asymptomatic in its early stages, this condition is normally identified in advanced stages of the disease, when treatments are less effective. To address this challenge, automated AMD image assessment systems offer the potential to significantly reduce the time, costs, and effort involved in screening. While previous works have demonstrated success for AMD detection using convolutional neural networks, their lack of explainability mechanisms limits their use in clinical settings. To address this limitation, we propose an explainable deep-learning approach using Local Interpretable Model-agnostic Explanations (LIME). Our model, based on RegNetY-320, achieved 86.5% accuracy, 85.21% sensitivity, and 91.01% specificity on the Automatic Detection challenge on Age-related Macular degeneration dataset. Through the LIME technique, we identified the specific areas in retinal images that influence the prediction of the model, providing a tool for clinical interpretation and enhancing diagnostic confidence.

Keywords: Age-related macular degeneration · deep learning · eXplainable Artificial Intelligence · Local Interpretable Model-agnostic Explanations · retinal imaging

1 Introduction

Age-related macular degeneration (AMD) is a progressive and chronic disease that is nowadays responsible for most cases of blindness in people of advanced age [1]. There are no symptoms present at an early stage of AMD. However, as the disease progresses, it can culminate in blindness due to the development of abnormal blood vessels that damage the central region of the retina [1]. Hence, an early detection of AMD is of outmost importance to prevent progressive visual impairment in the elderly. In the literature, different AMD detection procedures have been proposed. Tan et al. [2] proposed a fourteen-layer

T. Jarm et al. (Eds.): EMBEC 2024, IFMBE Proceedings 112, pp. 85–94, 2024.
https://doi.org/10.1007/978-3-031-61625-9_10

deep Convolutional Neural Network (CNN) to automatically classify the fundus images into normal or AMD classes. In the ADAM challenge, the Team Tiger team utilized ResNet101 to classify images into non-AMD and AMD. In the same challenge, the WWW and the Airamatrix teams pre-trained an EfficientNet-B7 and an EfficientNet-B4, respectively, on ImageNet. Then, they applied fine-tuned using the clinical dataset [3]. Although previously mentioned deep-learning systems have achieved high accuracy in the detection of AMD, they have yet to reach deployment into clinical practice. The main difficulties are ethical, medico-legal and logistical barriers, and lack of interpretability [4]. Consequently, there has been an increasing emphasis on the need for eXplainable Artificial Intelligence (XAI) to enhance the transparency and accountability of artificial intelligence systems [4]. The importance of XAI has also been consistently emphasized across international frameworks. A relevant example are the Ethics Guidelines for Trustworthy Artificial Intelligence by the European Union High-Level Expert Group on Artificial Intelligence, included transparency and accountability as part of the 7 key requirements necessary for trustworthy artificial intelligence [4].

As deep learning is gaining prominence for medical imaging [5] and ophthalmology [6], this work assesses the efficacy of CNNs as a potential method for AMD detection and the use of Local Interpretable Model-agnostic Explanations (LIME) to explain the decisions made by the model. Such an algorithm could be used not only in clinical environments, but also in public settings such as under-resourced areas where there is limited access to health care.

2 Dataset

The Automatic Detection challenge on Age-related Macular degeneration (ADAM) dataset [3], also known as iChallenge-AMD, comprises 1200 fundus images with 267 of them belonging to patients diagnosed with AMD. These images have a resolution of 2124×2056 pixels, and 1444×1444 pixels [3]. All images have labels indicating the presence or absence of AMD in the patient [3]. The diagnosis of AMD is determined using the fundus images and supplementary information such as medical history and optical coherence tomography. However, this material is not included in the dataset and it is not publicly available [3]. Additionally, coarse segmentation maps for several lesions were also provided. Currently, to the best of our knowledge, this is the only public dataset that provides pixel-level annotations of different AMD-associated lesions, including drusen (154 images), exudates (130 images), hemorrhages (72 images), scars (68 images), and other lesions (29 images) [3]. The dataset was divided in three subsets: a training set (400 images), a validation set (400 images), and a test set (400 images) [3]. We used the same subset division in this study.

3 Methods

The proposed method starts with a preprocessing stage to normalize the input images. Then, a CNN model was developed to detect AMD using techniques

such as data augmentation, transfer learning and fine-tuning. Finally, the LIME technique was used to visualize the image areas that influence the model predictions.

3.1 Preprocessing

The model was trained, validated, and tested using preprocessed versions of the original images. To standardize input data and reduce processing time, all images were resized to a resolution of 512 × 512 pixels [3].

3.2 Data Augmentation

Deep neural networks require a large amount of training data. To increase the number of images available for training the model, we applied online data augmentation, creating synthetic images in each epoch through simple transformations like rotations and flips [7].

3.3 Model Architecture

CNNs are neural networks capable of extracting representative features of images [8]. In this work, a CNN architecture was developed using a pretrained RegNetY-320 architecture as a base model [9]. RegNet is a network design approach that seeks to create fast, simple and efficient networks by means of a linear parameterization [9]. These networks outperform popular models such as EfficientNet in performance and GPU speedup [9]. To adapt the architecture to the binary classification problem of AMD, we added an average pooling layer, a dropout layer with an average pooling layer, a dropout layer with a factor of 0.5, a dense layer of 2048 neurons, another dropout layer with a factor of 0.5 and a dense layer of 2 neurons [10]. The number of neurons of the last layer corresponds to the number of classes we want to discriminate: non-AMD patient and AMD patient. A ReLU-type activation function was used in the first dense layer, and a softmax activation function was used in the last dense layer [10]. The softmax returns a probability distribution over the classes, indicating the probability that the input image belongs to the non-AMD class or the class with AMD [10]. In the fine-tuning phase, the early-stopping technique was used to minimize overfitting of the network [8]. In this way, the training process was automatically stopped when the validation loss did not improve after 5 consecutive epochs. We used the categorical cross-entropy as a loss function and Adam as the optimization algorithm [10]. To minimize overfitting in advanced epochs, the learning rate was reduced by a factor of 10 every time the validation error reached a minimum and kept constant [8]. A batch size of 4 images was used, since it was the maximum size that our GPU NVIDIA GeForce RTX 4080 allowed.

3.4 Local Interpretable Model-Agnostic Explanations

Currently, in medical domain, XAI functionality is a necessary requirement for automatic detection systems. XAI provides explanations for the decisions of deep-learning methods, allowing interpretability. One of the algorithms that enables visual explainability is LIME [11]. In this work, we studied the image version of LIME. In order to explain the prediction of a model f for an example ξ, LIME [11]:

1. decomposes ξ in d superpixels, that is, small homogeneous image patches;
2. creates a number of new images $x_1, \ldots, x_n$ by randomly turning on and off these superpixels;
3. queries the model, getting predictions $y_i = f\ (x_i)$;
4. builds a local weighted surrogate model $\hat{\beta}_n$ fitting the y_is to the presence or absence of superpixels.

Each coefficient of $\hat{\beta}_n$ is associated to a superpixel of the original image ξ and, the more positive the more important the superpixel is for the prediction at ξ according to LIME.

4 Results

4.1 AMD Detection

To evaluate the performance of the automated AMD detection model, we used the test set of the ADAM dataset, consisting of 400 images. Results are summarized in Table 1. It shows sensitivity, specificity, accuracy and area under the receiver operating characteristic curve (AUC) values.

4.2 LIME Interpretation

We further tried to provide a visual explanation for interpreting the developed model using LIME. We obtained the image regions that were more relevant for AMD prediction. Some examples of the visualizations obtained with this method can be seen in Figs. 1, 2, 3 and 4. Figure 1 and Fig. 2 show examples that were correctly predicted using the proposed method. Figure 3 and Fig. 4 show examples of LIME explanations that illustrate the most common errors observed in our study.

Table 1. Performance of the proposed method.

Sensitivity	Specificity	Accuracy	AUC
91.01%	85.21%	86.50%	0.95

5 Discussion

This work presents a novel and robust method for the automatic detection of AMD. We present a model with high diagnostic performance, while requiring a relatively small number of images for training. Moreover, we emphasize the incorporation of LIME as a key element, enabling the model to provide visual explanations.

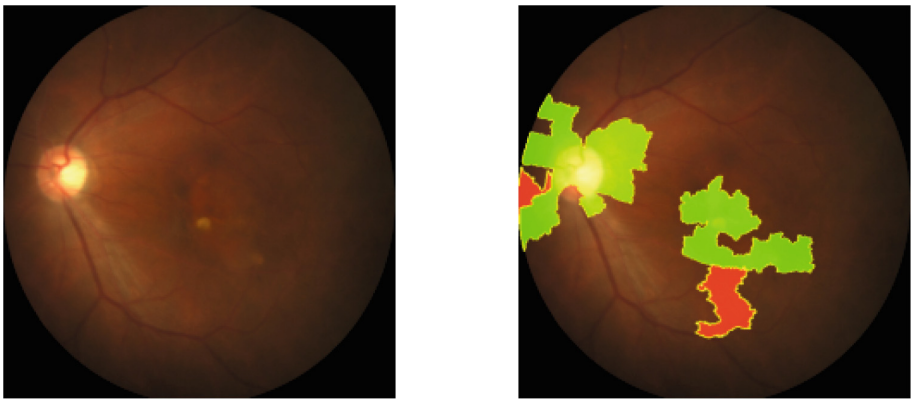

Fig. 1. LIME of correct AMD predicted. (a) Original image. (b) LIME visualization.

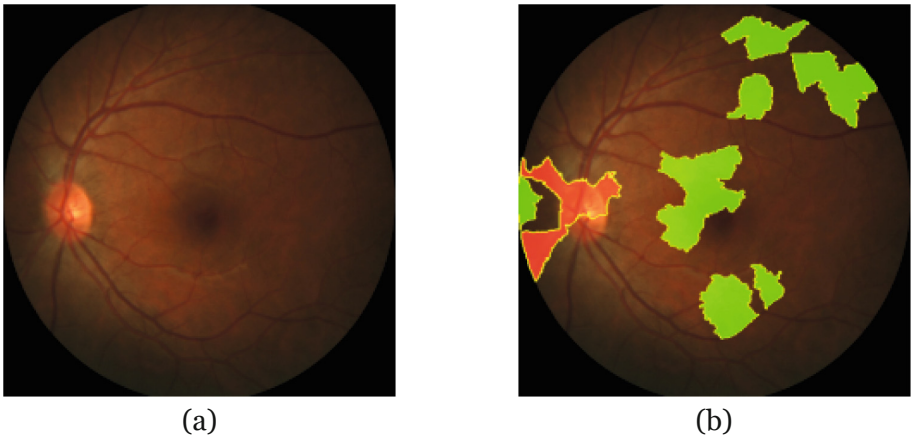

Fig. 2. LIME of correct non-AMD predicted. (a) Original image. (b) LIME visualization.

5.1 AMD Detection Performance

The method has been developed exclusively using fundus images from the ADAM database. Results in Table 1 demonstrate that the proposed method achieves high performance in detecting AMD from fundus images. In fact, the model only misclassifies 8 images as not having AMD when they have the pathology. Table 2 illustrates that our results are well in line with previous studies of those obtained in previous studies. However, it is important to note that the methods previously proposed in the literature are based on more complex networks than our chosen model. The five top-ranked teams used ensemble methods, with the top three teams using at least 5 models. This approach involves training multiple different architectures or the same neural network architecture under different settings

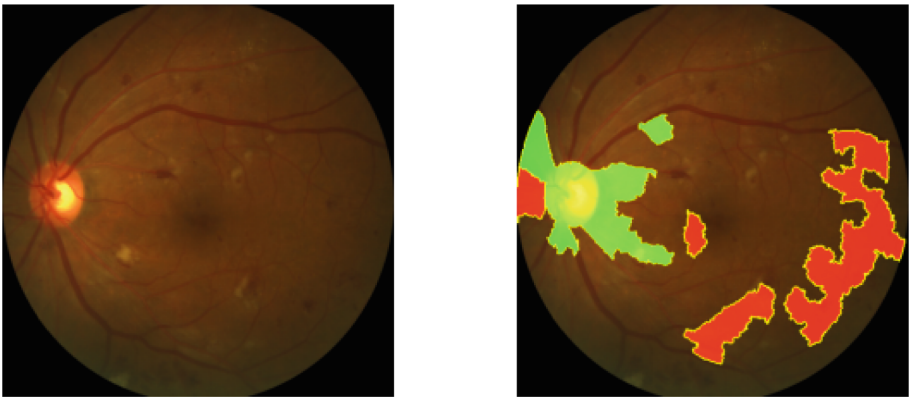

Fig. 3. LIME of misclassified AMD. (a) Original image. (b) LIME visualization.

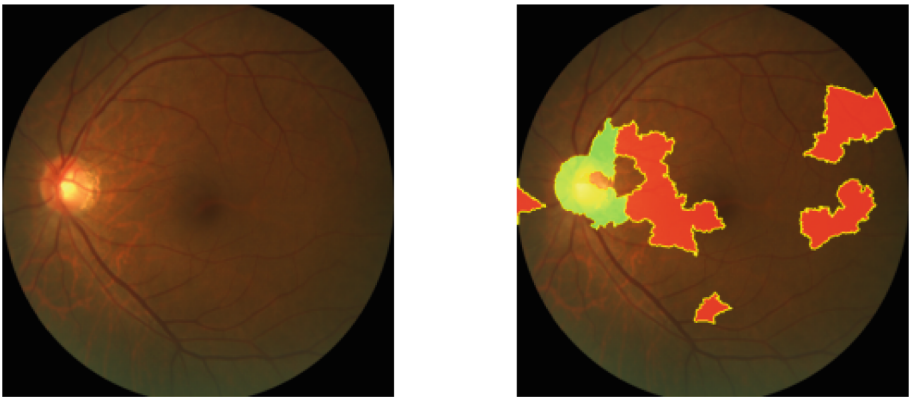

Fig. 4. LIME of misclassified non-AMD. (a) Original image. (b) LIME visualization.

and combining their outputs to produce a single prediction per image. Moreover, the VUNO EYE TEAM used clinical annotations for 15 types of lesions to train independent models, which enabled better AMD classification results [3]. In contrast, our method only requires image-level annotations indicating the presence or absence of the disease.

5.2 LIME Explanations of the Model

The provided approach to XAI for AMD image analysis can help clinicians to trust the decisions made by the deep-learning model and also help them interpreting the decisions in the form of visualizations. To the best of our knowledge, LIME has not been applied in the field of automatic AMD detection yet. Some representative examples of the LIME explanations obtained in our study are shown in Figs. 1, 2, 3 and 4. Figure 1 shows an example of a correctly classified AMD image. Green colour indicates the regions that were crucial for the positive AMD diagnosis made by the network. Specifically, the macula and the optic disc emerge as important areas for classification. In this case, there are some lesions similar to drusen near the macula. Figure 2 depicts an image without AMD correctly classified by the network. Green regions near the macula and the periphery of the retina were the most significant for classification. In this scenario, the macula exhibits no lesions. In Fig. 3, an image with AMD that was misclassified by our method is depicted. Apparently, this image lacks lesions near the macula. It is worth noting that there are no lesion annotations within the provided database for this image. In this case, the diagnosis might have relied on supplementary information, such as medical history and optical coherence tomography [3]. Finally, Fig. 4 illustrates the case of a non-AMD image that was considered with AMD by the proposed method. This discrepancy could be due to the presence of other retinal lesions, associated with a different disease. In this image, the periphery of the retina displays a red region, indicating it does not align with the AMD category. Nonetheless, the shape of the optic disc (in green) seems to have some characteristics associated with AMD.

Table 2. Comparison with previous studies.

Study	Number of test images	AUC
VUNO EYE TEAM	ADAM (n = 400)	0.97
ForbiddenFruit	ADAM (n = 400)	0.96
Zasti_AI	ADAM (n = 400)	0.96
Proposed method	**ADAM (n = 400)**	**0.95**
Muenai_Tim	ADAM (n = 400)	0.94
ADAM-TEAM	ADAM (n = 400)	0.93
WWW	ADAM (n = 400)	0.92
XxlzT	ADAM (n = 400)	0.91
TeamTiger	ADAM (n = 400)	0.91
Airamatrix	ADAM (n = 400)	0.88

In light of these results, this study suggests a potential link between the center of the optic disc and the presence of AMD, as the analyzed images display a pattern in this region. Drusen, small deposits of cellular debris, may accumulate in the macula of AMD patients and can also be detected near the optic disc. Thus, the emergence of a pattern in the center of the optic disc in certain images may be associated with the presence of AMD. There could be changes in the structure or shape of the optic disc as the disease progresses. Previous studies have assessed the changes in the optic disc as AMD progresses [12]. Nevertheless, further research is required to validate this hypothesis.

5.3 Clinical Usefulness of Our Proposal

The development of an automated system for AMD detection holds significant clinical promise, particularly considering the need for efficient screening methods in ophthalmology. Our proposed method demonstrates high efficacy in detecting AMD, as shown in Table 1. Notably, the model exhibits a high level of accuracy, misclassifying only 8 AMD images (2% of the test set). This high accuracy rate indicates that the proposed method could be useful in clinical settings, where accurate diagnosis is of utmost importance.

Moreover, the incorporation of XAI techniques, particularly LIME explanations, enhances the clinical utility of our model [11]. Through visualizations provided by LIME, clinicians gain valuable insights into the decision-making process of the deep-learning model. Clinicians will be able to identify the specific areas within the images that are relevant for the model, facilitating its integration into clinical practice.

5.4 Limitations and Future Work

Our study also presents some limitations that should be mentioned. First, our model was developed and tested using a publicly available database. It would be convenient to validate our proposal on a larger and more heterogeneus database in order to assess the robustness of the proposed method. Another limitation arises from the exclusive use of LIME to explain our model. Future goals may include testing other XAI algorithms. Finally, to the best of our knowledge, the quantitative analysis of XAI visual explanations has not been addressed. In this sense, it would be desirable to further investigate on the quantitative assessment of XAI visual maps.

6 Conclusions

In this work, we have proposed a XAI approach for the detection of AMD and the retinal regions that influence the prediction. This is particularly crucial,

as the absence of explainability poses a significant challenge to the practical implementation of automated methods in real-world contexts. Our XAI approach based on LIME allowed us to identify those retinal regions associated with AMD. Our results suggest that the proposed method could represent an important step toward providing an explainable and reliable AMD diagnostic tool. Early AMD detection can facilitate timely access to treatment and, consequently, prevent vision loss in at-risk population.

Acknowledgment. This research has been developed under the grants TED2021-131913B-I00 and PID2020-115468RB-I00 funded by 'Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/10.13039/501100011033/', European Regional Development Fund (ERDF) A way of making Europe and European Union NextGenerationEU/PRTR.; and by 'CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBERBBN)' through 'Instituto de Salud Carlos III' co-funded with ERDF funds. M. Herrero-Tudela was in receipt of a PIF-UVa grant of the University of Valladolid.

Compliance with Ethical Standards. All data comes from public databases [3].

Conflicts of Interest. The authors declare no conflict of interest.

References

1. Fleckenstein, M., et al.: Age-related macular degeneration. Nat. Rev. Dis. Primers. **7**, 31 (2021)
2. Tan, J.H., et al.: Age-related macular degeneration detection using deep convolutional neural network. Future Gener. Comput. Syst. **87**, 127–135 (2018). https://www.sciencedirect.com/science/article/pii/S0167739X17319167
3. Fang, H., et al.: ADAM challenge: detecting age-related macular degeneration from fundus images. IEEE Trans. Med. Imaging **41**, 2828–2847 (2022)
4. Chaddad, A., Peng, J., Xu, J., Bouridane, A.: Survey of explainable AI techniques in healthcare. Sensors **23**, 634 (2023)
5. Celard, P., et al.: A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. Neural Comput. Appl. **35**, 1–33 (2022)
6. De Fauw, J., Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat. Med. **24**, 1342–1350 (2018) by: 1357. All Open Access, Green Open Access
7. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. ArXiv **abs/1712.04621**, https://api.semanticscholar.org/CorpusID:12219403 (2017)
8. Shao, L., Zhu, F., Li, X.: Transfer learning for visual categorization: a survey. IEEE Trans. Neural Netw. Learn. Syst. **26**, 1019–1034 (2015)
9. Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P. Designing network design spaces. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10425–10433 (2020)
10. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning, MIT Press, Cambridge (2016)

11. Ribeiro, M., Singh, S., Guestrin, C.: Why should i trust you?: Explaining the predictions of any classifier, pp. 97–101 (2016)
12. Law, S.K., et al.: Optic disk appearance in advanced age-related macular degeneration. Am. J. Ophthalmol. **138**, 38–45 (2004)