International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

A Deep Learning-Based Approach for Retinal Image Classification Using EfficientNet-B0 and Channel Attention

K. Silpaja Chandrasekar
Department of Information Technology, Alpha College of Engineering,
Chennai, Tamil Nadu, India

S. Sudha
Department of Information Technology, Alpha College of Engineering,
Chennai, Tamil Nadu, India

Abstract

Retinal fundus images contain visual information that can assist in the identification of several ocular conditions. This paper presents an image-classification pipeline based on EfficientNet-B0 with Squeeze-and-Excitation (SE) channel attention for the Ocular Disease Intelligent Recognition (ODIR) dataset. The pipeline applies Contrast Limited Adaptive Histogram Equalization (CLAHE) to the luminance channel in LAB colour space, extracts image features using an ImageNet-pretrained EfficientNet-B0 backbone, and applies channel-wise feature recalibration before multi-label prediction. Gradient-weighted Class Activation Mapping (Grad-CAM) is used to visualize image regions associated with model outputs. The original manuscript reports an overall ODIR accuracy of 98.5%. It also reports class-wise precision values from 0.91 to 0.96, recall values from 0.93 to 0.98, and F1-scores from 0.93 to 0.97 across the eight ODIR label categories. These values are reproduced from the supplied manuscript and require confirmation against the original experiment logs and the exact test-set protocol. The study is presented as an applied deep-learning implementation rather than a clinically validated diagnostic system. Further work is required to document the complete experimental configuration and assess performance on independent data.

Keywords: Retinal image classification; ODIR; EfficientNet-B0; Squeeze-and-Excitation; CLAHE; Grad-CAM; Deep learning; Multi-label classification

1 Introduction

Retinal fundus photography provides a non-invasive view of structures such as the optic disc, macula, and retinal blood vessels. Changes in these structures may be associated with ocular conditions including diabetic retinopathy, glaucoma, cataract, age-related macular degeneration, hypertensive retinopathy, and myopia. The interpretation of fundus images requires expertise, and automated image analysis has therefore been investigated as a means of supporting image review and screening workflows.

Convolutional neural networks (CNNs) learn visual features directly from images. Transfer learning allows a model pretrained on a large image collection to be adapted to a more specific task. EfficientNet is one such CNN family, designed to balance network depth, width, and input resolution. Squeeze-and-Excitation (SE) blocks provide channel-wise feature recalibration, while

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Contrast Limited Adaptive Histogram Equalization (CLAHE) can improve local image contrast. Grad-CAM provides a visual representation of image regions associated with a model output.

This study focuses on a practical pipeline for the ODIR dataset. Unlike diabetic-retinopathy grading datasets that assign one severity grade to an image, ODIR includes diagnostic labels that may co-occur. The task is therefore treated as multi-label classification, with eight output categories. The method combines CLAHE preprocessing, EfficientNet-B0 feature extraction, SE channel attention, and sigmoid-based prediction. Grad-CAM is used to inspect selected predictions.

The work makes the following practical contributions:

- It describes an ODIR-focused retinal image classification pipeline using CLAHE and an ImageNet-pretrained EfficientNet-B0 model.
- It incorporates SE channel recalibration into the feature representation used for multi-label prediction.
- It uses Grad-CAM to provide visual summaries of image regions associated with individual model outputs.
- It reports the ODIR class-wise metrics supplied in the original manuscript and identifies the experimental details that must be documented for reproducibility.

2 Related Work

Deep learning has been widely investigated for retinal image analysis. Gulshan et al. studied deep convolutional networks for diabetic-retinopathy detection, while Abramoff et al. evaluated an autonomous system for diabetic-retinopathy screening. Ting et al. examined deep learning for diabetic retinopathy and related retinal conditions across a multi-ethnic population [1– Other studies have explored automated retinal disease detection and grading [4, 5].

CNN architectures have developed through approaches such as Inception, residual learning, MobileNet, and EfficientNet [21– EfficientNet uses compound scaling to balance model depth, width, and resolution. The SE mechanism introduced channel-wise feature recalibration and has been incorporated into CNN architectures to adaptively weight feature channels [12].

Visual explanation methods are also used to inspect model behaviour. Grad-CAM uses gradients flowing into convolutional feature maps to produce class-specific localization maps [27]. LIME and SHAP provide other approaches to model explanation [28, 29]. These methods can assist model inspection, but a heatmap alone does not establish that a model uses clinically valid evidence.

The present study uses established components in a single ODIR-focused pipeline. Its results should be interpreted in relation to the dataset, label definitions, split protocol, and evaluation metrics used. Direct comparison with published results is appropriate only when the underlying experimental settings are sufficiently similar.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

3 Materials and Methods

3.1 Dataset

The study uses the Ocular Disease Intelligent Recognition (ODIR) dataset. The supplied manuscript describes ODIR as a retinal fundus image dataset containing diagnostic keywords for eight categories: Normal fundus, Diabetic Retinopathy, Glaucoma, Cataract, Age-related Macular Degeneration (AMD), Hypertensive Retinopathy, Myopia, and Other abnormalities. An image may have more than one diagnostic label.

Each image is represented by an eight-element binary target vector:

$$\mathbf{y} = [y_1, y_2, \ldots, y_8], \qquad y_j \in \{0, 1\},$$

where a value of 1 indicates that the corresponding category is present. The diagnostic keywords are mapped to the eight output categories. The exact number of images retained after filtering, the class frequencies, and the train, validation, and test counts were not stated in the supplied manuscript and must be added from the experiment records before submission.

3.2 Image Preprocessing

The original method resizes fundus images to 224×224 pixels. Each RGB image is converted to LAB colour space. CLAHE is applied to the luminance (L) channel, while the chromatic channels are retained. The enhanced luminance channel is then recombined with the original chromatic channels to obtain the processed image.

CLAHE performs local histogram equalization while limiting histogram amplification. For an intensity value r, the cumulative distribution function can be written as

$$CDF(r) = \sum_{i=0}^{r} p(i),$$

where $p(i)$ is the normalized frequency of intensity level i. Histogram clipping limits excessive enhancement, and the clipped excess is redistributed across intensity levels. The supplied manuscript does not state the numerical CLAHE clip limit or tile-grid size; these values must be reported from the implementation rather than estimated.

The original manuscript also describes intensity normalization and training-time augmentation, including rotation, horizontal flipping, brightness variation, and geometric transformations. The exact normalization convention and augmentation ranges should be specified in the final submission.

3.3 EfficientNet-B0 Feature Extraction

An ImageNet-pretrained EfficientNet-B0 network is used as the feature extraction backbone. Its original classification head is removed, and the convolutional feature extractor is retained. For an input image X, the extracted feature representation is

$$F = E(X; \theta),$$

where E denotes the EfficientNet-B0 feature extractor and θ denotes its parameters.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

EfficientNet uses mobile inverted bottleneck convolution blocks to extract hierarchical visual features. The use of pretrained weights provides an initialization for adapting the network to retinal images. The supplied manuscript states that the backbone is fine-tuned, but does not provide the number of frozen layers or the exact fine-tuning schedule.

3.4 Squeeze-and-Excitation Channel Attention

An SE operation is used to recalibrate the channels of the extracted feature maps. Let $F \in \mathbb{R}^{H\times W\times C}$ denote a feature tensor. Global average pooling produces a descriptor for each channel:

$$z_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F_c(i,j).$$

The channel weights are generated using two learned transformations:

$$\mathbf{s} = \sigma\left(W_2\delta(W_1\mathbf{z})\right),$$

where δ is ReLU and σ is the sigmoid function. The recalibrated feature maps are

$$\widehat{F}_c = s_c F_c.$$

This operation allows the network to assign different weights to feature channels. It does not, by itself, demonstrate that the emphasized channels correspond to clinically meaningful structures.

3.5 Classification Layer

The refined feature maps are passed through global average pooling and a classification head. Since ODIR permits multiple labels for an image, the output layer uses sigmoid activation to generate an independent score for each of the eight categories:

$$\hat{y}_k = \frac{1}{1+\exp(-z_k)}, \qquad k = 1,\ldots,8.$$

The original manuscript describes binary cross-entropy with class weighting as the training objective:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{8} w_k\left[y_{ik}\log(\hat{y}_{ik}) + (1-y_{ik})\log(1-\hat{y}_{ik})\right],$$

where N is the number of training images and w_k is the weight assigned to class k. The manuscript states that Adam is used for optimization. The learning rate, batch size, number of epochs, weight-decay setting, stopping criterion, and class-weight calculation are not supplied and must be completed from the actual training configuration.

3.6 Grad-CAM Visualization

Grad-CAM is used to visualize regions associated with a selected output category. Let A^k be the k-th feature map in the final convolutional layer and y^c be the score for category c. The feature-map weight is

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k},$$

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

where Z is the number of spatial locations. The class-specific map is

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha^c_k A^k\right).$$

The resulting map can be resized and overlaid on the input image. These visualizations are used for qualitative inspection only. No expert assessment or quantitative validation of the heatmaps is reported in the supplied manuscript.

3.7 Evaluation Metrics

The original manuscript reports precision, recall, F1-score, and accuracy for each ODIR category. Precision, recall, and F1-score are defined as

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN},$$

$$F1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

where TP, FP, and FN denote true positives, false positives, and false negatives for a given label. For a multi-label task, the threshold used to convert sigmoid scores into binary predictions and the averaging convention should be stated. The supplied manuscript does not specify these details. The reported accuracy column is retained below as provided, but its calculation should be checked because class-wise accuracy can be defined in more than one way in multi-label evaluation.

4 Results

4.1 ODIR Classification Performance

The original manuscript reports the class-wise ODIR results shown in Table 1. The reported precision values range from 0.91 to 0.96, while recall ranges from 0.93 to 0.98. The reported F1-scores range from 0.93 to 0.97. The Normal category has the highest reported recall (0.98), while the AMD and Hypertensive Retinopathy categories have reported F1-scores of 0.93.

Table 1: Reported class-wise performance on the ODIR dataset

Class	Precision	Recall	F1-score	Accuracy
Normal	0.96	0.98	0.97	0.98
Diabetic Retinopathy	0.94	0.96	0.95	0.98
Glaucoma	0.95	0.96	0.96	0.98
Cataract	0.93	0.95	0.95	0.97
Age-related Macular Degeneration	0.93	0.93	0.93	0.98
Hypertensive Retinopathy	0.91	0.93	0.93	0.98
Myopia	0.92	0.96	0.96	0.96
Other Abnormalities	0.94	0.95	0.95	0.97

The supplied manuscript also reports an overall ODIR accuracy of 98.5%. This value is retained as originally reported; however, the manuscript does not define the exact overall-

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

accuracy calculation. It should therefore be verified against the saved predictions and test labels before publication. In particular, the authors should report whether the value represents exact-match accuracy, label-wise accuracy, or another aggregation.

4.2 Qualitative Inspection

Grad-CAM was included in the original pipeline to display regions associated with individual category scores. Such maps can help inspect whether the model responds to image regions that appear relevant. They should not be described as proof of clinical interpretability or as evidence that the model has localized lesions correctly unless an expert assessment or a suitable quantitative evaluation has been performed.

5 Discussion

The reported ODIR results indicate promising class-wise precision, recall, and F1-score values for the proposed EfficientNet-B0 and SE-based pipeline. The method combines image contrast enhancement, transfer learning, channel recalibration, and multi-label prediction in a relatively straightforward workflow.

The results should nevertheless be interpreted cautiously. The supplied manuscript does not provide the exact dataset partition, patient-level split information, decision threshold, random seed, or full training configuration. Without these details, it is not possible to assess the risk of data leakage or reproduce the reported metrics. The high reported overall accuracy also requires a clear definition, particularly because ODIR is a multi-label dataset.

The contribution of each component has not been isolated through ablation experiments in the supplied material. Consequently, the current results do not establish that CLAHE or SE attention independently improves performance. A controlled comparison of EfficientNet-B0 alone, EfficientNet-B0 with CLAHE, EfficientNet-B0 with SE, and the complete pipeline would be needed to support such a conclusion. No ablation values are introduced in this revised manuscript because they were not present in the source.

Comparisons with published studies are not included as evidence of superiority. Published results may use different ODIR partitions, label mappings, metrics, and thresholds. A fair comparison would require evaluation under a common protocol or a clearly qualified narrative comparison.

6 Limitations and Future Work

This study has several limitations. First, the exact number of images used, class distribution, and data-splitting procedure are not specified in the supplied manuscript. Second, the training configuration and prediction threshold are incomplete, which limits reproducibility. Third, the class-wise and overall accuracy definitions require clarification for the multi-label setting. Fourth, Grad-CAM provides a post-hoc visualization and does not establish that the model bases its predictions on clinically valid features. Finally, no independent external validation or ophthalmologist assessment is reported.

Future work should document a patient-independent data split where identifiers permit, report the complete training configuration, evaluate class-wise and macro/micro metrics, and

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

perform ablation experiments using identical data partitions. Independent testing and expert review of explanation maps would provide additional evidence about generalization and the relevance of the visual explanations.

7 Conclusion

This paper presented an ODIR-focused retinal image classification pipeline based on CLAHE preprocessing, EfficientNet-B0 feature extraction, SE channel attention, and sigmoid-based multi-label prediction. Grad-CAM was used to visualize image regions associated with model outputs. The original manuscript reports an overall accuracy of 98.5% and class-wise F1-scores between 0.93 and 0.97. These results are reproduced without alteration, but require verification against the original experiment logs and a clearly defined evaluation protocol.

The method provides an applied framework for studying multi-label retinal image classification. The available evidence is not sufficient to establish clinical readiness, external generalization, or the independent contribution of each pipeline component. Complete reporting of the data split, training settings, metric definitions, and independent evaluation is necessary before the results can be reproduced or considered for practical screening use.

Declarations

Funding: The authors report that no external funding was received.

Competing interests: The authors declare no competing interests.

Author contributions: The original manuscript states that all authors contributed to conceptualization, literature review, analysis, and manuscript preparation. The authors should confirm that this statement accurately reflects each person's contribution.

Ethics approval: The source manuscript describes the work as using a publicly available dataset and states that ethics approval was not required. This statement should be checked against the ODIR dataset terms and the target journal's requirements.

Consent to participate: Not applicable, subject to the dataset's terms of use.

Consent to publish: All authors are stated to have approved the manuscript for publication.

Data availability: The study uses the ODIR dataset. The authors should provide the official dataset access information and clarify any restrictions that apply.

Code availability: Not stated in the supplied manuscript. Add a repository link if the code is publicly available; otherwise state the actual availability position.

References

[1] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA (2016)

[2] Abramoff, M.D., et al.: Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy. npj Digital Medicine (2018)

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

[3] Ting, D.S.W., et al.: Development and validation of a deep learning system for diabetic retinopathy and related eye diseases. JAMA (2017)

[4] Gargeya, R., Leng, T.: Automated identification of diabetic retinopathy using deep learning. Ophthalmology (2017)

[5] Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P., Zheng, Y.: Convolutional neural networks for diabetic retinopathy classification. Procedia Computer Science (2016)

[6] al., Q.: Deep learning approaches for retinal image analysis using transfer learning. IEEE Transactions on Medical Imaging (2017)

[7] al., W.: Multi-scale convolutional neural network for retinal lesion detection. IEEE Transactions on Medical Imaging (2017)

[8] al., B.: Deep learning-based detection of age-related macular degeneration and diabetic retinopathy. Ophthalmology (2017)

[9] Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell (2018)

[10] De Fauw, J., et al.: Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine (2018)

[11] Leibig, C., et al.: Supporting ophthalmologists with deep learning for diabetic retinopathy classification. Ophthalmology (2017)

[12] al., Y.: Machine learning-based retinal image analysis for diabetic retinopathy screening. Journal of Diabetes Science and Technology (2018)

[13] al., L.: Deep learning-based diabetic retinopathy detection using retinal fundus images. IEEE Access (2018)

[14] al., R.: Artificial intelligence-based diabetic retinopathy screening in real-world clinical environment. Eye (2018)

[15] al., M.: Automated retinal image grading systems for diabetic retinopathy screening. Eye (2019)

[16] al., S.: Deep learning-based detection of diabetic retinopathy, age-related macular degeneration and glaucoma. Scientific Reports (2019)

[17] al., C.: Deep learning-based framework for automated screening of diabetic retinopathy and diabetic macular edema. IEEE Access (2020)

[18] al., L.: Artificial intelligence-based automated perimetry system for visual field assessment. Ophthalmology (2020)

[19] al., H.: Deep learning algorithm for detecting referable diabetic retinopathy, macular diseases and glaucoma suspects. Ophthalmology (2020)

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

[20] al., L.: Deep learning framework for automated diabetic retinopathy detection. IEEE Access (2021)

[21] Szegedy, C., *et al.*: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)

[22] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE CVPR (2016)

[23] Howard, A.G., *et al.*: Mobilenets: Efficient convolutional neural networks for mobile vision applications. (2017)

[24] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: IEEE CVPR (2018)

[25] Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning (2019)

[26] Dosovitskiy, A., *et al.*: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)

[27] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: IEEE International Conference on Computer Vision (2017)

[28] Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you? explaining the predictions of any classifier (2016)

[29] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions (2017)

[30] Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you? explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144. ACM, ??? (2016). `https://doi.org/10.1145/2939672.2939778`

[31] Tjoa, E., Guan, C.: A survey on explainable artificial intelligence (xai): Toward medical xai. IEEE Transactions on Neural Networks and Learning Systems **32**(11), 4793–4813 (2021) `https://doi.org/10.1109/TNNLS.2020.3027314`

[32] Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature (2017)

(This work is licensed under a Creative Commons Attribution 4.0 International License.)