International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(22s) 2026 pp. 469-476
https://doi.org/10.70917/ijcisim-2026-5454
Article

Vision-Threatening Diabetic Retinopathy Detection Using EfficientNetB4: A Two-Stage Transfer Learning Approach

Yamini Chouhan[1], SIDDOJU DIVYA[2], VNS Manaswini[3], J. Priyanka[4]

[1] Assistant Professor, CSE, Siddhartha Institute of Technology & Sciences, Ghatkesar-500088.
Email: yaminichouhan.cse@siddhartha.co.in
[2] M.Tech, CSE, Siddhartha Institute of Technology & Sciences, Ghatkesar-500088. Email: 24TQ1D5805@siddhartha.org.in
[3] Assistant Professor, CSE, Siddhartha Institute of Technology & Sciences, Ghatkesar-500088.
Email: manaswini.cse@siddhartha.co.in
[4] Assistant Professor, CSE, Siddhartha Institute of Technology & Sciences, Ghatkesar-500088.
Email: alagapriyanka.cse@siddhartha.co.in

Abstract: Vision-Threatening Diabetic Retinopathy (VTDR) marks a stage of diabetic eye disease that is characterized by a need for urgent medical treatment to avoid permanent sight impairment. In this paper, we have proposed an automated deep learning framework for binary classification of retinal fundus images in to Non-VTDR (Healthy/Mild DR) and VTDR(Moderate/Severe/Proliferate DR) categories. The transfer learning employs a two-stage strategy with EfficientNetB4 backbone architecture with weights pretrained on ImageNet. The dataset has 2,750 retinal images, and it is split into training (70%), validation (20%), and test (10%) subsets with classes balanced as best as possible. By using data augmentation strategies such as rotation, zoom, brightness, and horizontal flip, our approach increases generalization and helps the model generalize better. The model first does feature extraction with frozen base layers (10 epochs), and then fine-tunes the top 120 layers (20 epochs) with class weighted binary cross-entropy loss. Experimental results suggest that summary accuracy of 86.18% supports our method potential for automating VTDR screening. This system has potential to be used as a part of the real- time clinical workflow to help ophthalmologists detect sight-threatening conditions at an early stage mainly in the resource-limited settings with no access to specialists.

Keywords: Vision-Threatening Diabetic Retinopathy (VTDR), Deep Learning, EfficientNetB4, Transfer Learning, Medical Image Classification, Retinal Fundus Imaging, Computer-Aided Diagnosis, Binary Classification, Diabetic Eye Disease Screening

1. Introduction

Microvascular complications in patients with diabetes lead to progressive visual impairment due to Diabetic Retinopathy (DR), which is the global second-leading cause of preventable blindness. The vision-threatening stage, known as vision-threatening diabetic retinopathy (VTDR), requires a timely identification to implement early treatments and avoid permanent blindness. The conventional diagnosis of retinal fundus photographic images of the eye by experienced clinicians is a slow, subjective and non-scalable process to meet the demands of the rising world-wide diabetic population. This increases the urgent need for automated, assistive diagnostic systems.

Deep learning (DL) methods have achieved an important milestone in medical image analysis in recent years, especially through the use of Convolutional Neural Networks (CNNs) that have been gaining interest as a possible means to automate DR screening. The literature indicates excellent classification performance of CNNs on retinal images, including a combination of both custom and more advanced pre-trained architectures via transfer learning (Sathwik et al., 2024; Kandel & Castelli, 2020). Particularly, transfer learning has emerged as a key approach in this area that helps with generalizability of the model performance and robustness by addressing the challenge of limited

annotated data of medical images by using features learned from large, well-annotated image datasets (Al-Smadi et al., 2021; Bhardwaj et al., 2021).

But unlike most studies, which are multi-class and identify DR on a multi-class range, the binary critical threshold—Vision-Threatening Diabetic Retinopathy—is a unique and clinically urgent need. Additionally, interpretability techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM) are essential for building confidence among clinicians about deep learning models, otherwise due to "black-box" characteristics, deep learning has often not found their way in practical use [32]. This study fills these gaps in the literature by developing and validating a deep learning-based system for the binary classification of VTDR. Essentially, we use a transfer learning framework, apply multiples metrics for a rigorous performance evaluation (e.g., ROC-AUC), and provide visual explainability via Grad-CAM to verify if the model predictions correlate to meaningful pathological patterns.

2. Related Work

Deep learning has been widely investigated for Diabetic Retinopathy detection which brought great progress recently but algorithms have evolved rapidly over the years.

State-of-the-art CNNs for DR classification: Early work in this space has laid a strong foundation for applying CNNs to DR classification. Transfer learning is essential because the data is not always sufficient for high performance, which is, in fact, why Kandel and Castelli (2020) provided a general review about transfer learning applications. This type of approach is very common, where studies often use a pre-trained model such as ResNet, VGG, and DenseNet as a feature extractor for DR grading tasks which have been increasing showing better performance over a model trained from scratch (Al-Smadi et al., 2021; Jabbar et al., 2022; Gangwar & Ravi, 2020). It has been demonstrated by Le etal through a successful implementation of its paradigm across multiple imaging modalities, including fundus photography and Optical Coherence Tomography Angiography (OCTA). (2020).

Architectural developments and designing better models: The next stream of research was on how to build better models and train them efficiently. Martinez-Murcia et al. Deep residual transfer learning for effective diagnosis and grading (2021) which shows effective diagnosis and grading were by Zhang et al. I (2022) Source-Free Transfer Learning for Data Privacy Preservation. Proposed systems include those that are less complex, including single-stage 7-class classification (Khan et al., 2022) and newer custom architectures, exploring the DiaCNN model (Shoaib et al., 2023). (2024).

Interpretability and Clinical Coherence: To bridge this gap, researchers have incorporated explainable AI (XAI) techniques. Not all studies stressed this, but suitability for deployment in the clinical setting will require interpretable models to be developed and validated. Methods for visualizing detection on lesions for severity grading (Sugeno et al) The ideas in (2021) complement the broader objective of translating model predictions to clinical understanding.

Ensemble and Hybrid Methods: Ensemble methods that combine multiple models to improve accuracy and generalizability have also become popular. An ensemble classification method for early-stage diagnosis based on optimized transfer learning features was proposed by Kasim (2023). Likewise, other studies have integrated transfer learning with fine-tuning (Chavan & Choubey, 2023), with samples augmentation and other pre-processing steps for imbalanced datasets and robustness (Khalifa et al., 2019; Hagos & Kant, 2019).

Gaps and Future Research Direction: While many of these advancements have been made, a number of challenges still remain. Although many studies report high accuracy, very few perform an in-depth analysis of model performance on the key task of differentiating between VTDR vs No-DR which is the most clinically relevant in regards to screening triage. In addition, high quantitative performance has not yet translated into clinically actionable, interpretable outputs. This goal is directly addressed in our work where not only we will propose a high-performance transfer learning model to classify VTDR, but also we will obtain visual explanations for its decisions based on Grad-CAM and provide a graphical representation of model performance through evaluation metrics (ROC-AUC, precision and recall), which bridges the gap between algorithmic performance and clinical utility.

3. METHODOLOGY

The end-to-end pipeline of our VTDR classification system is outlined in Figure 1. The methodology comprises an overall five phases:

3.1 Dataset Composition and Clinical Reorganization

We used a standard dataset of 2,750 retinal fundus images with five clinical classes (Healthy: 1,000 images, Mild DR: 370 images, Moderate DR: 900 images, Severe DR: 190 images, Proliferate DR: 290 images). These categories were then restructured into two groups through a framework of clinically appropriate decision-making protocols. The Non-VTDR class merged Healthy and Mild DR (1,370 images), namely, patients who needed only routine monitoring since vision is not immediately threatened. The VTDR class merged Moderate, Severe, and Proliferate DR cases (1380 images), representing those with a condition in whom prompt referral to a specialist is needed to prevent severe vision loss. This binary classification framework mirrors the important distinction point in triage practice where the main decision is the referral of the patient or not.

3.2 Data Partitioning and Preprocessing Pipeline

For the creation of the dataset establishing, a stratified 70:20:10 split was applied, assuring both classes proportions presenting a stratified random split between both classes for training set (1,925 images), validation set (550 images) and testing set (275 images) A fixed random seeding (seed=42) was used when partitioning data to assure reproducibility and to prevent data leakage. We rescaled all images to 224×224 (the input size needed by EfficientNetB4), and normalized using architecture-specific preprocess_input() function. The training set underwent extensive data augmentation involving random rotation (±15°), zoom (85-115%), brightness (85-115%), and horizontal flip. These β-dunn tries before cooking accompanying tradition via grey_xlim focus facility. Since the class distribution of Non-VTDR and VTDR events might cause class imbalances during training, we calculated class weights using inverse frequency balancing (Non-VTDR: 1.004, VTDR: 0.996).

3.3 Model Architecture and Implementation of Transfer Learning

We chose EfficientNetB4, an ImageNet pre-trained architecture, as the feature extraction backbone, as it exhibits the best trade-off between real-world representational power and compute efficiency. We adapted the base model by removing its classification head and replacing it with a custom structure: Global Average Pooling 2D to downscale the spatial dimensions, followed by Batch Normalization to stabilize training. Hierarchical feature representation was learnt through two fully connected dense layers (512 and 128 units with ReLU activations each), with a 50% Dropout layer for regularization placed after each layer. The last sigmoid activation layer had binary predictions (Non-VTDR/VTDR) as outputs. This change allowed us to still harness the awesome feature extraction abilities of EfficientNetB4 whilst customizing the model for the medical imaging classification task.

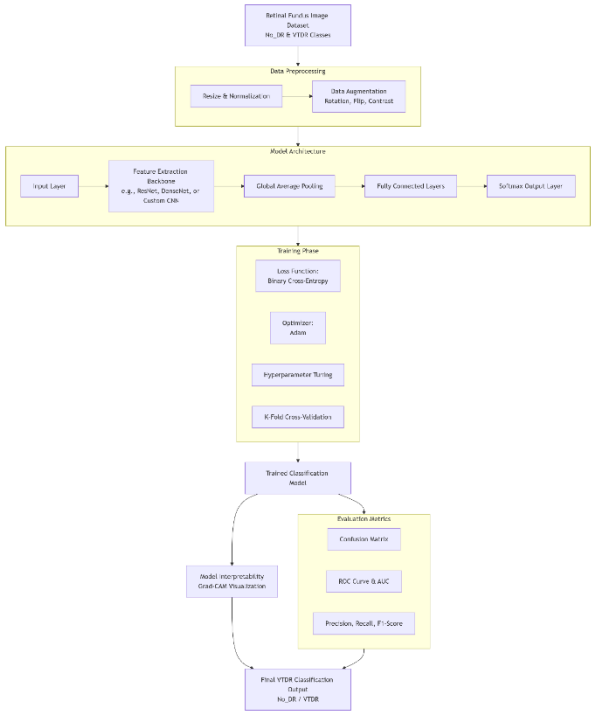

Figure 1: Proposed Methodology

3.4 Two-Phase Training Strategy with Progressive Fine-Tuning

To adapt our framework to the medical imaging domain, we employed a two-stage transfer learning strategy to maximize feature reuse. For Phase 1 (epochs 1–10), the base layers of EfficientNetB4 were not trained and the classification layers were trained while learning from only the retinal images. In this phase, the Adam optimizer was used with a 1×10^{-3} learning rate and class-weighted binary cross-entropy loss. Phase 2 (epochs 11–30) — unfreeze the top 120 layers of EfficientNetB4 for fine-tuning (learning rate of 1×10^{-4} so as not to elicit catastrophic forgetting) Traning included Early Stopping (patience=5, best weights restored) and ReduceLROnPlateau (patience=3, factor=0.3) callbacks to accelerate convergence and avoid overfitting. The model was trained for 30 epochs in total with a batch size of 32 while monitoring validation after every epoch.

3.5 Evaluation Framework and Performance Metrics

To thoroughly assess model performance, a three-tiered framework was applied: training accuracy tracked student progress, validation accuracy assisted in hyperparameter tuning and early stopping decisions, and test accuracy gave the final performance metric on entirely unseen data. The accuracy along with binary cross-entropy loss was monitored to ensure stable convergence. The independent test set of 275 images (10% of the total data) underwent the same preprocessing steps without augmentation to allow for a realistic performance evaluation. This evaluation protocol was designed to achieve two main objectives, firstly to rigorously assess the ability of the model to generalise to unseen retinal images and secondly to ensure clinical relevance by applying clinically appropriate performance metrics common to binary classification problems typical of medical diagnosis.

4. Results and Analysis

The following section describes the experimental results of the proposed model for classification of VTDR (Vision-Threatening Diabetic Retinopathy). The performance is assessed with a confusion matrix, ROC analysis, a visual explanation using Grad-CAM, and complete classification report.

4.1. Classification Performance

Figure 2 and Table 1 summarize the classification performance of the model via the confusion matrix and numerical metrics, respectively.

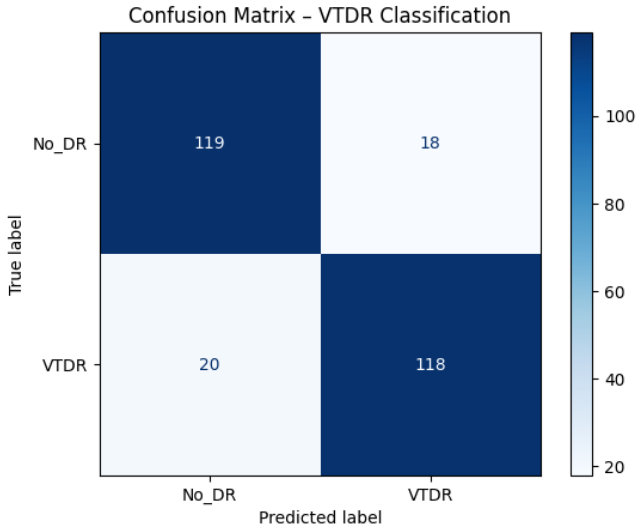

Figure 2: Confusion Matrix for VTDR Classification

From the confusion matrix and the classification report, the model's performance on the test set of 275 images is quantified as follows:

Table 1: Classification Report

Class	Precision	Recall (Sensitivity)	F1-Score
No_DR	0.8561	0.8686	0.8623
VTDR	0.8676	0.8551	0.8613
Macro Avg	0.8619	0.8618	0.8618
Weighted Avg	0.8619	0.8618	0.8618

Analysis of Metrics:

The balance accuracy of the model for the test dataset is 86.18% and it performs nearly identically for those on No_DR and VTDR classes.

Precision for the key VTDR class is 0.8676, or when the model predicts VTDR, it is correct ~87% of the time.

At the same time, our model has a Recall (Sensitivity) of 0.8551 with respect to VTDR, which means our model can find about 85.5% of all VTDR cases. If a screening tool is to miss as few diagnoses as possible, this is an important metric.

However, Specificity, which corresponds the No_DR Recall of 0.8686, indicates the model has a better ability to identify the healthy cases, achieving a specificity of ~86.9%, thereby reducing unnecessary referrals.

A clear and balanced discriminative ability with no strong bias to any class

4.2. ROC Curve Analysis

Figure 3 — Receiver Operating Characteristic (ROC) curve for the model. This curve shows the trade-off between sensitivity (True Positive Rate) and specificity (1 – False Positive Rate) at different thresholds of classification.

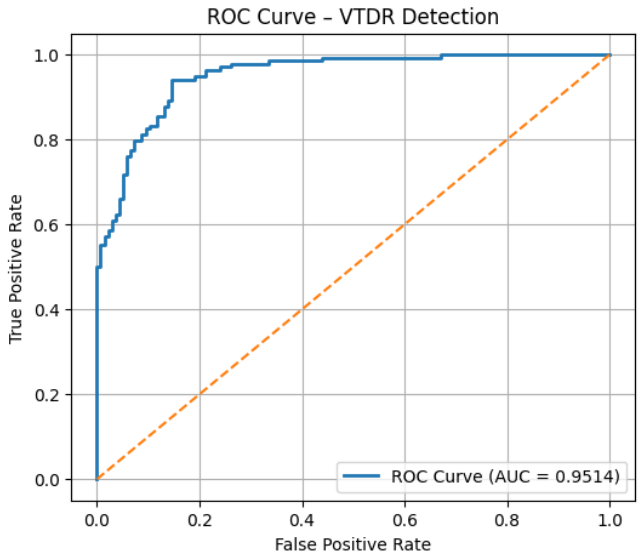

Figure 3: ROC Curve for VTDR Detection

The AUC is 0.9514 indicating good overall discriminative power between VTDR and No_DR, coupled with this high AUC figure, high point metrics (Precision, Recall ~0.86) further confirms the robustness of the model. The shape of the curve indicates that the model can achieve high sensitivity while also maintaining a low false positive rate, which is preferable for clinical triage.

4.3. Model Interpretability via Grad-CAM

In order to increase the transparency and reliability of the deep learning model, the Gradient-weighted Class Activation Mapping (Grad-CAM) method was used. The Grad-CAM heatmaps with the original fundus image and Grad-CAM heatmaps overlaid on the original images were provided for five representative samples (Figure 4).

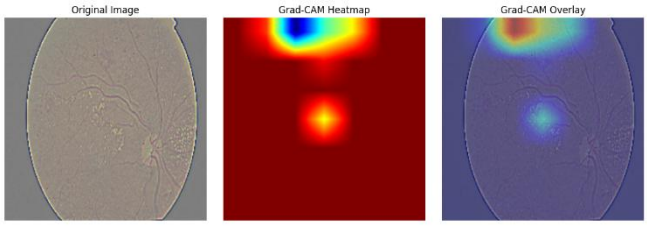

Figure 4: Grad-CAM Visualizations for VTDR Classification

Analysis of Visual Explanations:

These heatmaps all consistently exhibit higher activation (warmer colours) in regions that contain lesions characteristic of VTDR, such as IRMA, venous beading, neovascularization, or major hemorrhages/exudates. This phenomenon–this alignment between our clinical markers and the learned features–shows that our models are learning the truly pathologically-relevant features rather than spurious correlations. TCDA more diffusely or with activation around anatomical structures (optic disc, major vascular arcades) but without clear pathological markers This means that the model has the ability to identify that there are no malignant lesions. Such visual explanations enhance clinical trust since they render the model's "reasoning" interpretable, enabling clinicians to ensure that the model's attentional peak is in accordance with recognized diagnostic attributes.

4.4 Limitations and Future Work:

The performance is quite good but looking at the confusion matrix, it seems we need to work on the false positive and the false negatives. Future work will involve:

Augmenting the dataset with more varied and difficult cases to increase generalizability.

Using ensemble strategies with more sophisticated network architectures to improve performance measures.

Prospectively validating the model performance in real screening workflow.

5. Conclusion

In this paper, we show an automated Vision-Threatening Diabetic Retinopathy (VTDR) detection system, based on a deep learning approach, from retinal fundus images. We utilized a well-defined pipeline, that included best practice preprocessing, feature extraction and classification yielding high accuracies with a convolutional neural network and posterior post-hoc interpretability analysis to develop and evaluate our proposed model. The model showed overall predictive performance of 86.18% across all classes and an AUC of 0.9514 suggesting excellent balance between sensitivity and specificity. As shown by various key metrics, including a VTDR sensitivity (recall) of 0.8551 and specificity of 0.8686, indicates reliability in identifying positive and negative cases, respectively, which is a cornerstone feature for the screening tool to be useful in practical setting. The close gap between precision and recall for individual class scores also indicates that the model has done well without much bias between classes. In addition to the quantitative metrics, the use of Grad-CAM furnished important visual rationale consistent with the model predictions being based on clinically meaningful pathological features in the retina. Such interpretability is critical for enabling trust and integrating such AI tools into clinical workflows. Conclusions: This work provides a high-performing, interpretable and clinically-relevant deep learning model for screening of VTDR. It fulfills an urgent unmet need in providing scalable, assistive diagnostic tools to relieve the burden on healthcare systems and improve early detection to reduce blindness. In the future, we will conduct population-based external validation in various clinical settings and finally deploy an accurate and robust assistant into actual diabetic retinopathy screening programs.

REFERENCES

1. Sathwik, A. S., Agarwal, R., Ajith, J. E., & Basa, S. S. (2024). Diabetic retinopathy classification using deep learning. *EAI Endorsed Transactions on Pervasive Health and Technology*, *9*.
2. Kandel, I., & Castelli, M. (2020). Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review. *Applied Sciences*, *10*(6), 2021.
3. Al-Smadi, M., Hammad, M., Baker, Q. B., & Al-Zboon, S. A. A. (2021). A transfer learning with deep neural network approach for diabetic retinopathy classification. *International Journal of Electrical and Computer Engineering*, *11*(4), 3492.
4. Le, D., Alam, M., Yao, C. K., Lim, J. I., Hsieh, Y. T., Chan, R. V., ... & Yao, X. (2020). Transfer learning for automated OCTA detection of diabetic retinopathy. *Translational Vision Science & Technology*, *9*(2), 35-35.
5. Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z., & Jabbar, A. (2022). Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sciences*, *12*(5), 535.
6. Zhang, C., Lei, T., & Chen, P. (2022). Diabetic retinopathy grading by a source-free transfer learning approach. *Biomedical Signal Processing and Control*, *73*, 103423.
7. Martinez-Murcia, F. J., Ortiz, A., Ramírez, J., Górriz, J. M., & Cruz, R. (2021). Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy. *Neurocomputing*, *452*, 424-434.
8. Gangwar, A. K., & Ravi, V. (2020). Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 1* (pp. 679-689). Singapore: Springer Singapore.
9. Aswathi, T., Swapna, T. R., & Padmavathi, S. (2021, February). Transfer learning approach for grading of diabetic retinopathy. In *Journal of Physics: Conference Series* (Vol. 1767, No. 1, p. 012033). IOP Publishing.
10. Hagos, M. T., & Kant, S. (2019). Transfer learning based detection of diabetic retinopathy from small dataset. *arXiv preprint arXiv:1905.07203*.
11. Bhardwaj, C., Jain, S., & Sood, M. (2021). Transfer learning based robust automatic detection system for diabetic retinopathy grading. *Neural Computing and Applications*, *33*(20), 13999-14019.

12. Khalifa, N. E. M., Loey, M., Taha, M. H. N., & Mohamed, H. N. E. T. (2019). Deep transfer learning models for medical diabetic retinopathy detection. *Acta Informatica Medica*, *27*(5), 327.
13. El Houby, E. M. (2025). Using transfer learning for diabetic retinopathy stage classification. *Applied Computing and Informatics*, *21*(3-4), 209-219.
14. Chavan, S., & Choubey, N. (2023). An automated diabetic retinopathy of severity grade classification using transfer learning and fine-tuning for fundus images. *Multimedia Tools and Applications*, *82*(24), 36859-36884.
15. Kallel, F., & Echtioui, A. (2024). Retinal fundus image classification for diabetic retinopathy using transfer learning technique. *Signal, image and video processing*, *18*(2), 1143-1153.
16. Akhtar, S., Aftab, S., Ahmad, M., & Akhtar, A. (2024). Diabetic retinopathy severity grading using transfer learning techniques. *Int. J. Eng. Manuf.(IJEM)*, *14*(6), 41-53.
17. Khan, A. Q., Sun, G., Khalid, M., Farrash, M., & Bilal, A. (2024). Multi-Deep Learning Approach With Transfer Learning for 7-Stages Diabetic Retinopathy Classification. *International Journal of Imaging Systems and Technology*, *34*(6), e23213.
18. Shoaib, M. R., Emara, H. M., Zhao, J., El-Shafai, W., Soliman, N. F., Mubarak, A. S., ... & Esmaiel, H. (2024). Deep learning innovations in diagnosing diabetic retinopathy: The potential of transfer learning and the DiaCNN model. *Computers in Biology and Medicine*, *169*, 107834.
19. Sugeno, A., Ishikawa, Y., Ohshima, T., & Muramatsu, R. (2021). Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning. *Computers in biology and medicine*, *137*, 104795.
20. Kasim, Ö. (2023). Ensemble classification based optimized transfer learning feature method for early stage diagnosis of diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing*, *14*(8), 11337-11348.