Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19922-1

Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT

Umer Sadiq Khan[1,2,3] · **Saif Ur Rehman Khan**[4]

Received: 9 April 2024 / Revised: 6 July 2024 / Accepted: 19 July 2024

Abstract
Damage to the retinal blood vessels is critical in diabetic retinopathy, a progressively emerging health concern that often advances quietly without explicit symptoms. Optical coherence tomography-OCT has emerged as a favored noninvasive imaging technique for diagnosing diabetic retinopathy promptly and accurately. However, timely and precise diagnoses from OCT images are essential in prevention of blindness. Moreover, accurate interpretation of OCT images is challenging. Single model learning debilitates in managing diverse data types and structures, constraining its adaptability to varied environments. Its limitations become apparent in tasks requiring expertise from multiple domains, delaying overall performance. Moreover, learning may exhibit susceptibility to overfitting with large and heterogeneous datasets, resulting in compromised generalization capabilities. In this study, we propose a hybrid learning model for the classification of four distinct classes of retinal diseases in OCT images with improved generalization capabilities. Our hybrid model is constructed upon the well-established architectural foundations of ResNet50 and EfficientNetB0. By pre-training the hybrid model on extensive datasets like ImageNet and then fine-tuning it on publicly available OCT image datasets, we capitalize on the strengths of both architectures. This empowers the hybrid model to excel in discerning intricate image patterns while efficiently extracting hierarchical prediction from various regions within the images. To enhance classification accuracy and mitigate overfitting, we eliminate the fully connected layer from the base model and introduce a concatenate layer to combine two objective learning prediction. A dataset comprising 84,452 OCT images, each expertly graded for illnesses. we conducted training and evaluation of our proposed model, which demonstrated superior performance compared to existing methods, achieving an impressive overall classification accuracy of 97.50%. Notably, we address the interpretability issue, providing clear and insightful results, making our model suitable for deployment in ophthalmology clinics, where it can significantly contribute to the diagnosis of retinal disorders with enhanced precision and efficiency.

Keywords Deep Learning · Optical coherence tomography · Multi Class

Extended author information available on the last page of the article

1 Introduction

In recent years, a notable surge in the prevalence of retinal diseases has drawn considerable attention [1]. Among these conditions [2], DME and age-related macular degeneration-AMD stand out as significant contributors to irreversible visual impairment. AMD, encompassing both dry and wet forms, is the leading cause of blindness in adults aged 65 and above. Drusen on the retina characterizes the dry AMD variant, while CNV typifies the wet AMD manifestation [3].

The intricate neurovascular networks within the retina undergo structural alterations in the presence of DME, a condition closely associated with diabetes that leads to compromised vision [4]. This ailment stems from disruptions in retinal blood vessel integrity, culminating in the accumulation of fluid and proteins within the retinal layers [5]. Disturbingly, research has indicated that approximately a quarter of individuals afflicted by diabetic retinopathy eventually progress to the more severe DME stage [6]. The significance of early detection and intervention in retinal pathologies cannot be overstated, as timely management can potentially mitigate disease progression and associated vision loss.

OCT is a sophisticated and pivotal ophthalmic imaging modality, offering a unique glimpse into the intricate cross-sectional architecture of retinal layers. The merits of OCT are underscored by its non-contact nature, minimal invasiveness, and rapid imaging capabilities [7]. Notably, OCT has assumed a paramount role within the realm of ophthalmology, serving as a linchpin for quantification, analysis, and the development of therapeutic interventions targeting a spectrum of retinal pathologies. The manual interpretation of retinal OCT images is a multifaceted challenge that warrants meticulous attention. The imperatives of diagnosis and treatment in the face of escalating patient volumes underscore the inadequacy of relying solely on a cadre of certified medical professionals [8]. Furthermore, specific lesion attributes may elude direct observation, engendering potential ambiguities and instances of diagnostic oversight. Alas, the confluence of restricted healthcare accessibility in specific regions has consigned a substantial cohort of patients to the shadows of early-stage ailment, accentuating their conditions' gravity.

Convolutional Neural Networks-CNN has emerged as a promising avenue for automating the diagnostic process of OCT images. In [9], an OCT-Net, a specialized CNN architecture, with the primary goal of categorizing retinal images into healthy and three prevalent retinal disorder classes. This network architecture adeptly gathered and presented data that can be utilized for accurate medical diagnoses.

Ultimately, the collective body of research showcases the compelling potential of DL methodologies in OCT image analysis. These approaches have consistently demonstrated diagnostic accuracy on par with or surpassing that of experienced ophthalmologists [10], opening new horizons for efficient and accurate medical diagnoses in the realm of retinal diseases.

In medical research, acquiring supervised data poses a formidable challenge, often necessitating specialized expertise within the medical domain. Many advanced DL methodologies [11, 12] have been meticulously devised to overcome formidable challenges to address the scarcity of annotated data [13]. One promising avenue for expanding the training dataset is strategically implementing data augmentation techniques. These encompass a spectrum of approaches, including geometric transformations and the replication of image distributions, effectively boosting the diversity and volume of training examples [14, 15].Furthermore, unsupervised learning offers strategic possibilities, including semi-supervised, multi-instance, and TL paradigms [16]. Among these, TL has emerged as a particularly potent strategy due to its ability to benefit pre-existing knowledge from one domain to another effectively. TL minimizes the need for extensive retraining by judiciously transferring model information across tasks, even when those tasks are divergent or seemingly unrelated. This distinctive attribute has amplified the prominence of TL in recent times.

An illustrative case that underscores the potential of TL was demonstrated in [17]. This study harnessed a DL framework to categorize OCT images depicting normal eyes and those afflicted with three distinct macular disorders. Leveraging a dataset of 4,000 OCT images, the researchers showcased TL's efficacy in enhancing classification accuracy. This empirical validation underscores TL methodologies' profound impact on medical image analysis tasks, particularly where limited annotated data hinders conventional supervised learning approaches. The contribution of this study is listed below:

- The paper presents an enhanced hybrid learning method aimed at enhancing the accuracy of retinal disease detection in OCT images. The primary aim is to achieve more precise disease classification, mitigating issues and overfitting that often arise in single-model learning. The benefit of hybrid learning over single-model learning lies in its ability to combine the strengths of multiple models, resulting in improved generalization and robustness in disease detection tasks.
- The hybrid model exploits the robust capabilities of both the ResNet-50 and EfficientNet-B0 models. This fusion of strengths empowers the hybrid model to excel in discerning intricate image patterns while efficiently extracting hierarchical prediction from various regions within the images.
- Our model effectively resolves the interpretability challenge by delivering transparent and enlightening interpretability outcomes, demonstrating the significant advantage of understanding complex processes by utilizing GRADCAM analysis.

2 Related work

Different methods have been used to identify retinal OCT lesions throughout the past few decades [18]. These diagnostic techniques can be roughly divided into two categories. The first category consists of algorithms for spotting retinal OCT lesions using ML methods. These methodologies advantage sophisticated image processing techniques for higher precision and recall.

Numerous established ML methodologies rely on sophisticated feature extraction techniques to generate distinctive and discriminative representations, effectively enhancing image classification tasks. A feature extraction pipeline was meticulously devised in a seminal work by [19], incorporating the local binary mode of OCT and a directional gradient histogram. This intricate pipeline facilitated the creation of a unique feature set, subsequently harnessed by a linear Support Vector Machine-SVM classifier for precise image classification predictions.

In a parallel pursuit, [20] introduced a compelling approach to image classification, methodology hinged on a multiclass linear SVM classification strategy coupled with an innovative universal retinal image alignment and cropping technique. This initial step differentiated between ARMD and DME. A global visual representation was meticulously constructed by strategically incorporating a spatial pyramid framework and advanced sparse coding techniques.

An automated technique for categorizing retinal eye diseases was introduced [21] by implementing a CNN-based model. This approach involved initial image denoising and mask creation using morphological dilation and thresholding techniques to mitigate noise effects. The preprocessed images, alongside their corresponding masks, were utilized for training the CNN model, resulting in the generation of surrogate images. The proposed model achieved impressive Area Under the AUC values of 0.9856 and 0.9783 on the duke and local datasets, respectively.

In parallel research efforts [14, 22], classified DME and AMD as three distinct retinal eye disorders using TL-based technology. Employing ten distinct experiments, the researchers employed the GoogleNet architecture with TL, achieving a commendable accuracy rate of 96%.

Furthermore, [23] developed a robust Deep CNN model for identifying AMD in OCT images. Their model underwent validation through blindfold and ten-fold cross-validation techniques, resulting in accuracy rates of 91.17% and 95.45%, respectively. Notably, [24] proposed a DL methodology to identify DME and DR from fundus images. The authors trained their model from scratch, necessitating a substantial dataset and an extended training phase. Despite these challenges, the model achieved a remarkable mean AUC of 0.991 when evaluated on the EyePACS-1 dataset. These collective efforts underscore the substantial advancements in accurately categorizing retinal eye diseases through advanced DL methodologies.

Several advanced methods using deep learning have been proposed for the automated identification of various diseases through optical OCT images, showcasing the remarkable potential of these techniques in the medical field. Authors [25] introduced a CNN-based model for diagnosing retinal disorders. Their method achieved a noteworthy accuracy index of 97.10% across five distinct retinal eye diseases.

Further advancements were made by [26], who proposed an innovative deep network employing iterative fusion within an international fact-checking network-IFCN framework. This approach achieved an accuracy of 93.25% for the automatic classification of four retinal eye illnesses. Author [27] employed a TL-based approach to discern various eye diseases, employing a comprehensive suite of five distinct DL models. Impressively, their meticulously developed methodology yielded an exceptional maximum accuracy of 93.32% across four distinct disease classes. With a focus on OCT images. Building on this trajectory, [28] introduced a enhance diagnosis enhancement strategy for OCT image classification. By innovatively substituting the traditional residual connections with the EdgeEn block and cross-activation technique, three ResNet [29] topologies underwent significant improvements in contrast enhancement, derivative refinement, and feature extraction, ultimately leading to a marked boost in classification accuracy.

Notably, [30] contributed to the field by introducing the GABNet architecture, a novel and lightweight classification model uniquely tailored for retinal OCT images. The comprehensive experimentation encompassed the analysis of an extensive dataset comprising 108,312 OCT images from a cohort of 4686 patients sourced from the universally recognized UCSD repository. Remarkably, the proposed GABNet model exhibited a substantial 3.7% enhancement in classification accuracy, signaling a promising advancement in the domain of OCT image analysis.

Deep learning-DL, a subset of AI, excels in revealing complex data representations, notably in object recognition and image classification. CNNs, a key DL architecture, employ convolution and pooling layers for feature extraction, often complemented by ReLU activation functions. CNNs streamline computation through techniques like depthwise separable convolution and pooling layers, enhancing efficiency in image processing tasks. Additional components like dropout layers and batch normalization further refine CNNs, making them a robust tool for various image-related applications. TL [31, 32] is a powerful solution in DL to mitigate overfitting, especially with limited data. TL leverages pre-trained models' knowledge from extensive datasets like ImageNet, employing two primary approaches. Feature Extraction involves using pre-trained models as feature extractors, while Fine-Tuning allows for retraining select layers or adjusting weights. Both methods expedite model development and enhance performance by leveraging pre-existing knowledge. Table 1 presents the comparative studies review.

3 Materials and methods

3.1 Dataset detail

The dataset utilized in our study, as referenced in [37], was meticulously curated from a comprehensive study and subsequently made publicly available on Mendeley. Comprising a vast collection of 84,452 high-resolution retinal images, this dataset has been extensively organized into distinct training, testing, and validation sets, ensuring its usability across various research frameworks. The images within this dataset represent four crucial categories of OCT retinal scans: DRUSEN, NORMAL, CNV, and DME, providing a diverse and comprehensive representation of retinal pathology. These OCT images were carefully selected from a retrospective cohort of adult patients, ensuring a robust and representative sample for analysis. Figure 1 visually presents a representative collection of images from this dataset, offering a sight into the affluence and complexity of the retinal scans incorporated within this invaluable research resource [37]. Moreover, it is noteworthy that this dataset has been widely utilized in numerous previous studies [38] to validate various proposed models and algorithms. Its extensive usage across different research endeavors underscores its reliability, relevance, and applicability in advancing the understanding and diagnosis of retinal diseases.

3.2 Data preprocessing and augmentation

The preprocessing phase is pivotal in the data preparation pipeline, serving as the foundational step to ensure data cleanliness and model readiness. CNN demand substantial data to excel in performance while mitigating the risk of overfitting. Data augmentation techniques were employed to strengthen the training dataset size without the need for additional data collection efforts. Data augmentation serves the dual purpose of enriching the dataset with diverse examples and secure the model's resilience against overfitting. In this research, three distinct augmentation strategies were applied. These strategies collectively augment the model capacity to discern features within the data while preventing it from the risk of overfitting, thereby fostering robust and accurate model performance. Figure 2 illustrates various stages of image augmentation techniques employed in this study.

- **Rescale OCT Image:** The ImageDataGenerator class is crucial in preprocessing the original OCT images within the dataset. These original OCT images inherently possess diverse dimensions, which can pose challenges for subsequent analysis and model compatibility. To address this issue, the ImageDataGenerator class is to standardize the image dimensions uniformly. Specifically, it rescales the original OCT image, ensuring

Table 1 Overview of studies related to OCT base retinal disease diagnosis

Reference	Model	Contribution
Karn et al. [33]	ML Classifier	The paper utilizes ML to analyze OCT images for retinal diseases, addressing challenges in understanding OCT biomarkers, particularly outside clinical domains
Liu et al. [34]	TSSK-Net	A novel approach in image-level weakly supervised biomarker localization and segmentation introduces a self-supervised learning technique in this work
Bansal et al. [35]	OCT-CNN	The OCT-CNN technique represents a CNN approach devised by the author. It comprises a total of 17 layers
Naik et al. [36]	Ensemble model	U-Net segmentation, images are inputted into an ensemble model combining Xception and InceptionV3 networks, augmented with a self-attention layer, to improve classification accuracy

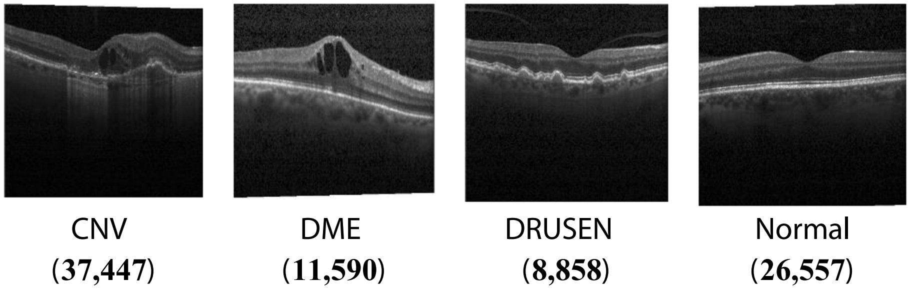


Fig. 1 OCT Retinal Scans Samples

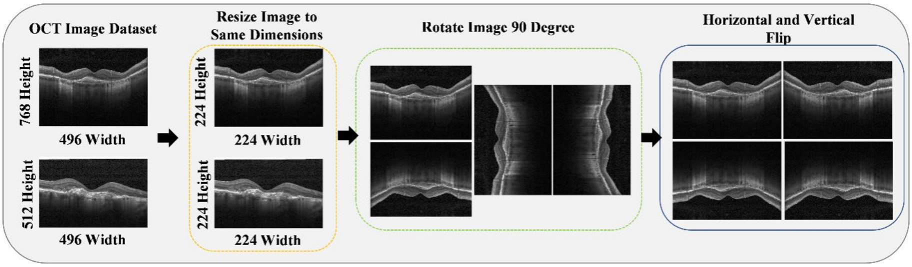


Fig. 2 Image Augmentation OCT Retinal Scans Samples

all images conform to a consistent format (224, 224, 3). This standardization is essential for facilitating effective data handling, compatibility with TL models, and maintaining the integrity of the dataset during the training process.

- **Rotation**: A data augmentation step was undertaken to enhance the dataset's robustness and improve the model's ability to handle variations in object orientations. This augmentation involved the rotation of each scan in the dataset by 90 degrees, creating multiple new versions of the original scans. Introducing these rotated variants makes the model better equipped to identify objects from diverse viewpoints and orientations. This augmentation technique effectively broadens the model's perspective. It equips it with the adaptability required to accurately recognize objects regardless of their orientation, ultimately contributing to improved classification performance and generalization capabilities.
- **Flip**: A mirroring operation was applied to the scans to enhance the training data and augment the dataset, generating mirror-image duplicates of every original image. This augmentation strategy enabled the model to learn from the original images and their horizontally and vertically flipped parallel. By doing so, the model becomes more adept at recognizing objects from various perspectives, encompassing top-to-bottom and bottom-to-top orientations and left-to-right and right-to-left viewpoints. This augmentation technique broadens the model's ability to generalize and detect features irrespective of the image's orientation, ultimately contributing to improved robustness and adaptability in image recognition tasks.

The distribution of images for each class after applying augmentation techniques is as follows: CNV (112,341 images), DME (34,770 images), DRUSEN (26,574 images), and

Normal (79,671 images). In our study, we implemented these techniques to enhance the diversity and robustness of our dataset. These augmentation methods have been widely employed in the literature [10, 36, 39–41] and are recognized for their efficacy in improving model generalization and performance. By incorporating these techniques, we aim to ensure that our model learns invariant features and is better equipped to handle variations in orientation and perspective commonly encountered in real-world scenarios.

Augmentation strategies like rotation and flipping have been extensively validated across various domains, ranging from computer vision to natural language processing. Their utility lies in their ability to artificially expand the dataset without the need for additional data collection, thereby mitigating issues related to overfitting and data scarcity. Furthermore, by mimicking realistic transformations that images may undergo in practical settings, these techniques contribute to the model's adaptability and resilience to unseen variations during inference. Furthermore, while these augmentation methods are well-established, their application requires careful consideration of the specific characteristics of the dataset and the task at hand.

3.3 Methodology

Deep Convolutional Neural Networks-DCNN have emerged as a potent approach for tackling diverse and complex image categorization tasks, significantly reinforcing accuracy in this domain. Despite the widespread adoption of DCNNs, their training can present formidable challenges. Two prominent issues that often afflict these networks are accuracy saturation and overfitting. Accuracy saturation arises when a network reaches a point where it no longer improves its performance despite additional training. Conversely, overfitting occurs when a network becomes too specialized in learning from its training data, causing a decrease in its generalization ability to unseen data. One of the fundamental limitations of deep networks is the vanishing gradient problem. This problem stems from the intensive computational demands and the number of layers in deep networks. It results in the gradients during the training process diminishing to the point where they no longer effectively guide the network's weight updates. As a result, addressing these challenges and mitigating the vanishing gradient problem remain pivotal research issues in DL and neural network design.

While TL offers several significant benefits to DL models, first and foremost, it accelerates the training process by benefiting pre-trained models on vast datasets, allowing the model to inherit valuable feature representations. TL saves computational resources and reduces the need for extensive labeled data, a scarce resource in many domains. TL enhances generalization capabilities, as models pre-trained on diverse tasks tend to capture higher-level patterns and nuances within data. Furthermore, TL fosters model robustness, as it enables the adaptation of existing knowledge to new domains or problem instances, making it a valuable tool for real-world applications where data distribution may change or evolve. TL empowers DL models to learn more efficiently, effectively, and flexibly across various domains and tasks.

Our proposed methodology presents a comprehensive approach for classifying retinal OCT images, benefiting advanced fine-tuning techniques to construct a hybrid learning model. The foundational architecture for extracting discriminative features from these OCT retinal image scans consists of two primary models: ResNet50 $W_{ResNet50}$ and EfficientNetB0 $W_{EfficientNetB0}$. In a subsequent phase. n initial experiments, ResNet-50 and EfficientNetB0 showcased remarkable performance in classification tasks, achieving the highest accuracy scores of 96.70% and 95.19%, respectively. Notably, ResNet-50 exhibited superior precision and recall rates at 95.68% and 95.50%, while EfficientNetB0 demonstrated outstanding precision and recall scores at 96.85% and 96.70%. Consequently, these models yielded an impressive F1 Score of 95.17%, leading us to designate them as the backbone of our proposed architecture. The benefit of ResNet-50 lies in its deep residual learning framework, facilitating more straightforward training of deep neural networks and enabling better convergence. On the other hand, EfficientNetB0's advantage stems from its efficient scaling approach, achieving superior performance with significantly fewer parameters, making it computationally more efficient and suitable for resource-constrained environments. we employ a stacking procedure to benefit the extracted prediction from both models. The overarching objective of this stacking process is to form a powerful ensemble that capitalizes on the strengths and diversity of these individual learners, thereby enhancing the system's classification performance. This method ensures a robust and accurate classification of retinal OCT images, addressing the complex task of diagnosing retinal diseases with high precision and reliability. Figure 3 is presenting an architecture diagram of proposed methodology.

ResNet50 [42], a 50-layer deep neural network-DNN architecture, employs bottleneck residual blocks to reduce model parameters and accelerate training effectively. The model involves the input OCT retinal scan being processed through convolution layers. Simultaneously, it takes a shortcut through an identity connection, and the resulting outputs are passed through an activation function. This innovative approach enhances the model's capacity for generalized classification while mitigating overfitting issues. ResNet50's architecture commences with the input layer, which accommodates images up to a size of 224 × 224 pixels. The first convolutional layer employs 64 filters with a stride of two and incorporates maximum pooling. Subsequently, ResNet50 utilizes four distinct types of convolutional blocks, each characterized by varying numbers of layers, filter sizes, and kernel sizes.

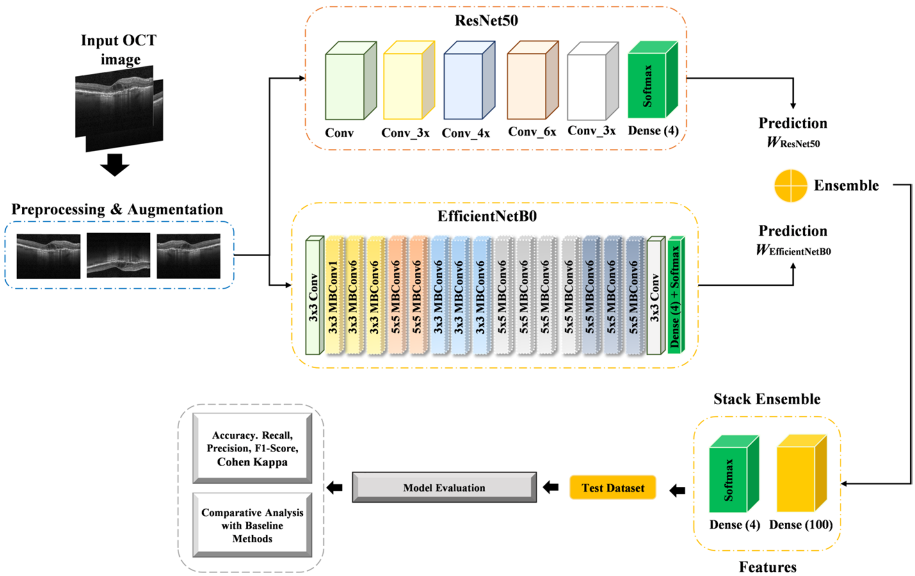


Fig. 3 Architecture diagram of Proposed Methodology

The first convolutional block comprises three separate convolution layers with kernel sizes of 1×1, 3×3, and 1×1 and corresponding filters of 64, 64, and 256, respectively. This block is repeated three times. The second convolutional block consists of three layers repeated four times, employing filters of 128, 128, and 512 in the respective layers. The third convolutional block repeats six times and utilizes three convolutional layers with 256, 256, and 1024 filters, respectively. The fourth and final convolutional block is repeated thrice and features three convolutional layers with 512, 512, and 2048 filters. Ultimately, Customize ResNet50 concludes with a dense layer of four neurons with a Softmax activation function designed for ImageNet classification tasks.

EfficientNetB0 [43] is a CNN model that advance innovative scaling techniques to scale each dimension of the network uniformly. It builds upon the inverted residual blocks originally introduced in MobileNetV2, adapting them to create a highly efficient architecture. With approximately 237 layers, EfficientNetB0 employs advanced features like squeeze-and-excitation mechanisms integrated within its blocks. This model's design prioritizes computational efficiency and superior performance, making it a valuable asset in various computer vision tasks, particularly in scenarios where resource constraints are a concern.

3.3.1 Ensemble architecture

In the ensemble learning [44] strategy known as stacking, a meta classifier or meta-regressor is employed to consolidate predictions from multiple classification or regression models. The process involves training a meta-model using the outputs generated by the base level models, which are initially trained on the complete training dataset. In this particular study, the base level models consist of ResNet50 and EfficientNetB0, which serve as fundamental classifiers for the first level of prediction. These models generate predictions that subsequently function as input prediction for the second level of prediction. This two-level approach allows for a more refined and accurate prediction, exploit the strengths of ResNet50 and EfficientNetB0 to enhance the overall predictive capability of the ensemble model. In this study, our approach centers on a feature ensemble technique as a pivotal part of our hybrid learning strategy. This ensemble method involves binding deep prediction extracted from two distinct pre-trained CNNs into a unified sequence. As illustrated in Fig. 4, our prediction-level ensemble procedure involves concatenating prediction extracted from ResNet-50 $W_{ResNet50}$ and EfficientNet-B0 $W_{EfficientNetB0}$, merging them into a single sequence. To predict the output, this concatenated deep feature sequence is subsequently passed through a dense layer with a dimensionality of 100, complemented by the ReLU activation function. This feature ensemble mechanism captures comprehensive information from multiple sources, enhancing the model's ability to make accurate predictions based on the combined knowledge extracted from ResNet50 and EfficientNetB0 prediction. Stacking, often referred to as Super Learning or Stacked Regression, comprises a set of DL algorithms that introduce a unique approach to ensemble learning. Unlike conventional ensemble techniques like bagging and boosting, which emphasize the combination of weak learners to form a strong learner, stacking takes a more robust route. In stacking, the focus shifts towards assembling a diverse set of strong base learners. These base learners are then subjected to a second-level meta learner, which is responsible for determining the optimal way to combine the predictions from the base learners. This two-tiered architecture allows stacking to the strengths of different base learners effectively and, in turn, enhance the overall predictive performance of the model.

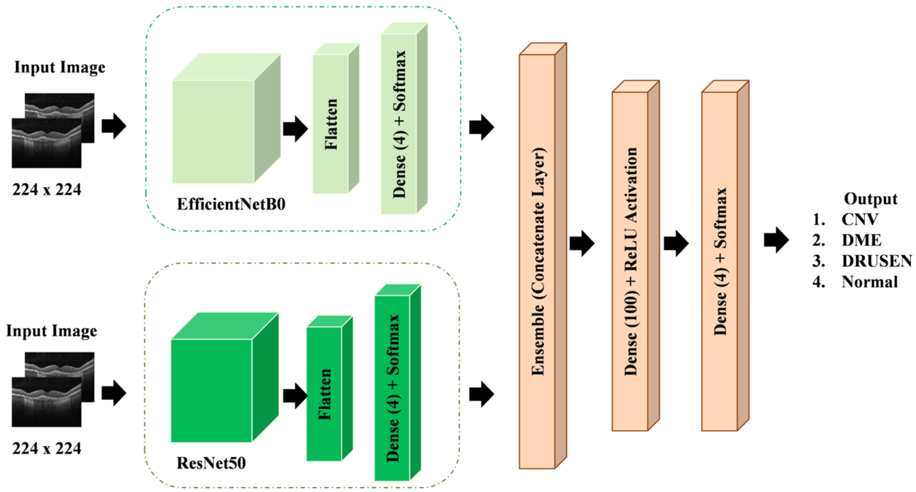


Fig. 4 Stacking of EfficentNetB0 and ResNet50 for final prediction

4 Experiments and results

An exhaustive evaluation was conducted, employing pivotal metrics, including Accuracy, Precision, Recall, and F1-score, to assess the models' proficiency in multi classification comprehensively. These metrics collectively provide a multidimensional view of the model's effectiveness, factoring in false positives and negatives. Additionally, a confusion matrix was employed to enhance the insight into the diagnostic efficacy of the proposed methodologies. This matrix visually illustrates the distribution of true positive, true negative, false positive, and false negative classifications, clearly delineating the model's diagnostic strengths and weaknesses.

A comprehensive evaluation of a hybrid learning model with the strengths of both ResNet50 and EfficientNetB0 architectures was conducted. This evaluation entailed the assessment of the diagnostic capabilities of the model using a dedicated test dataset, ensuring a thorough and meticulous examination. The assessment employed established equations and statistical metrics to quantitatively measure the model's overall performance, facilitating a comprehensive understanding of its effectiveness in classifying diseased cells.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1Score = 2 \times \frac{Precion \times Recall}{Precision + Recall} \tag{4}$$

Table 2 Performance Analysis of Proposed Models

Models	Accuracy	Precision	Recall	F1 Score
EfficientNet-B0-Base	96.70	96.85	96.70	96.69
ResNet-50-Base	95.19	95.68	95.50	95.17
Hybrid Learning	**97.50**	**98.10**	**97.23**	**98.21**

Table 3 Class wise Classification Performance of proposed model

Class	Accuracy	Precision	Recall	F1 Score	Support
CNV		0.93	1.00	0.96	250
DME		1.00	0.97	0.98	250
DRUSEN		1.00	0.93	0.96	250
NORMAL		0.98	1.00	0.99	250
Macro Avg	**97.50%**	0.98	0.97	0.98	1000
Weighted Avg		**0.98**	**0.97**	**0.98**	1000

4.1 Experimental setup

This section presents the test results and details the comprehensive evaluation process of the proposed hybrid learning model. It underwent comparisons with various competing analyses, including TL models and state-of-the-art CNN techniques, to ensure the model's robustness and effectiveness. The evaluation utilized the publicly available OCT retinal scans dataset [37], considered one of this research's most relevant and significant datasets. Model training was conducted using Python and the Keras package, with optimization for GPU runtime environments to enhance computational efficiency. The model was successfully trained on the NVIDIA GeForce RTX 3080, equipped with 13 MiB/1024 MiB of RAM, ensuring optimal performance and reliable results.

4.2 Performance evaluation of proposed architecture

4.2.1 Classification Performance

Table 2 presents an evaluation of different models for a classification task. Three models are compared based on four performance metrics as mention in Sect. 5: Accuracy, Precision, Recall, and F1 Score. The EfficientNet-B0-Base model achieves an accuracy of 96.70%, with corresponding precision, recall, and F1 Score values of 96.85%, 96.70%, and 96.69%, respectively. The ResNet-50-Base model also outperform it significantly, with an accuracy of 95.19% and higher precision, recall, and F1 Score of 95.68%, 95.50%, and 95.17%, respectively. The proposed model demonstrates the best performance, with an accuracy of **97.50%** and the highest precision, recall, and F1 Score values of **98.10%, 97.23%, and 98.21%,** indicating its effectiveness in the classification task.

Table 3 below presents the comprehensive classification performance of our proposed hybrid model applied to the retinal OCT dataset, with results provided for each class label. Precision, recall, f1-score, and macro-average scores have been meticulously calculated for each class type. Notably, the performance metrics exhibit remarkable consistency, with values ranging from 0.93 to 1.00. This remarkable stability underscores the robustness of our model in the task of classifying retinal images. Specifically, our model demonstrates exceptional precision, receiving a perfect score of 1.0 in both the DRUSEN and DME classes. Moreover, it achieves the highest recall of 1.00 for the NORMAL and CNV classes, further affirming its efficacy in correctly identifying these conditions.

When we consider the macro-average scores across all evaluation metrics, our hybrid model exhibits impressive performance metrics, including a precision of 0.98, a recall of 0.97, and an F1 score of 0.98. These results underscore the model's accuracy and its strong performance in the classification of retinal diseases. Notably, our proposed hybrid learning system attains an impressive overall classification accuracy of **97.50%** when tested on the classification of retinal diseases in OCT images.

In order to comprehend class-specific results in terms of the number of correctly and wrongly classified classes, confusion matrices are frequently utilized, a confusion matrix for the ResNet50-Base EfficientNetB0-Base model, which was suggested for the classification of retinal OCT images, is displayed in Fig. 5. This graph demonstrates that 250 + 242 + 233 + 250 = 975 OCT images are appropriately identified. The classification accuracy of the proposed deep residual network is calculated to be 975/1000, or 97.50%, for the remaining 25 images. As observed, eight DME photos, seventeen DRUSEN images, zero CNV, and one NORMAL image were incorrectly classified.

The graph in Fig. 6 illustrates the accuracy, recall, and f1-score metrics for the four distinct classes. Our hybrid learning-based approach has yielded auspicious results across all categories of retinal disorders.

The ROC curve plots the true positive rate (recall) against the false positive rate, serving as a graphical tool to evaluate diagnostic test performance. The AUC quantifies how well a test distinguishes between CNV, DME, DRUSEN, and normal samples, thereby assessing diagnostic accuracy. Figure 7 displays ROC curves illustrating the diagnostic performance of both the suggested model and base models. For the base models, EfficientNetB0 and ResNet50 exhibit AUC values of 0.9978 and 0.9977, respectively. Meanwhile, our suggested model achieves highest AUC value of 0.9983 as compared to base models.

GRAD-CAM analysis was utilized to inspect images and the corresponding GRAD-CAM maps generated by proposed model, which combines the power of CNN models ResNet-50 and EfficientNet-B0. As these models analyze images from our dataset, they each yield unique Grad-CAM visualizations. To offer behavior of each model, we've used a small selection of images and generated Grad-CAM representations for each. Figure 8 displays these sample images alongside their corresponding GRAD-CAM maps, as produced by our models. This observation is the noticeable divergence in the

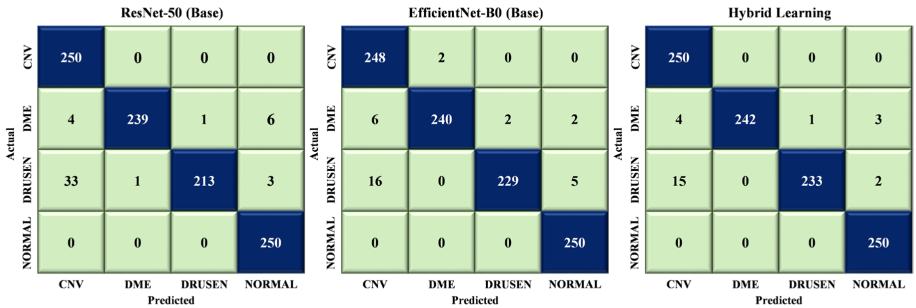


Fig. 5 Confusion Matrix Analysis

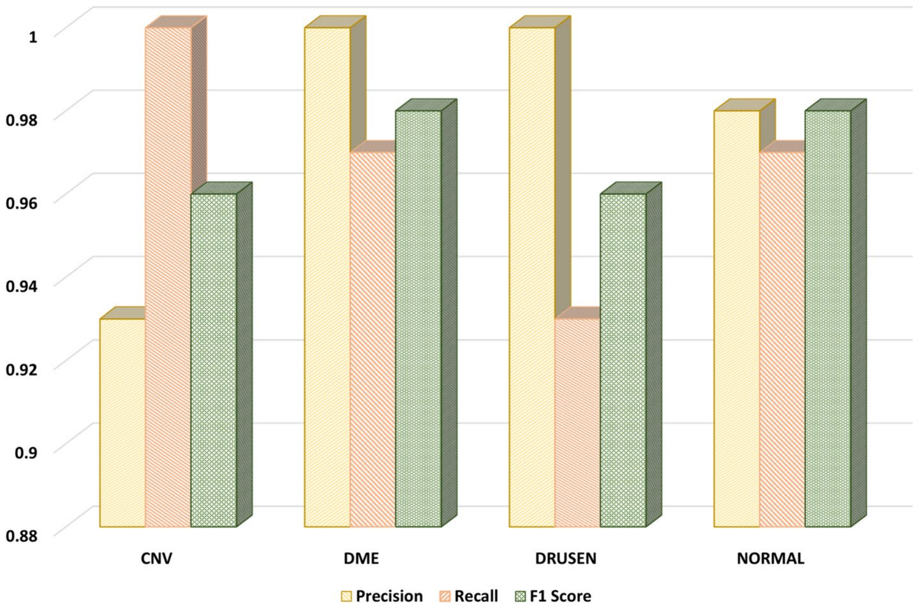
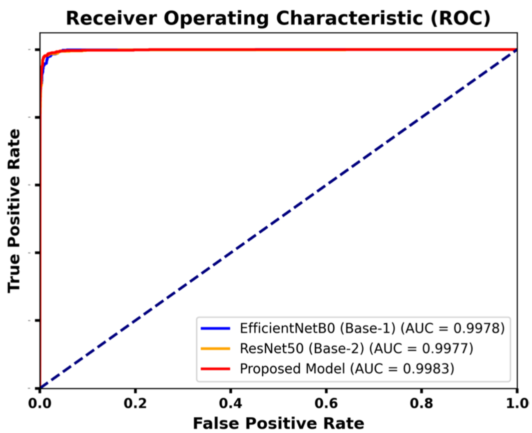


Fig. 6 Class wise precision, recall, and f1 score analysis

Fig. 7 Illustrate the ROC curve between ResNet50 (Base), EfficientNetB0 (Base), and Proposed (Our) Model



specific image regions highlighted by each model. This divergence underscores the fact that each model has ability to discern distinct features crucial for making accurate predictions. Due to this, our proposed models exhibit a capability to accurately identify infected areas in images afflicted with CNV, DME, and DRUSEN.

GRAD-CAM analysis offers valuable insights into DL model decision-making processes, aiding medical professionals in understanding how diagnoses are reached. By highlighting regions of interest within medical images, GRAD-CAM enables clinicians to verify and interpret model predictions, enhancing trust and facilitating collaborative decision-making. Ultimately, this transparency promotes the adoption of AI-driven tools in clinical settings [45, 46], potentially leading to more accurate diagnoses and

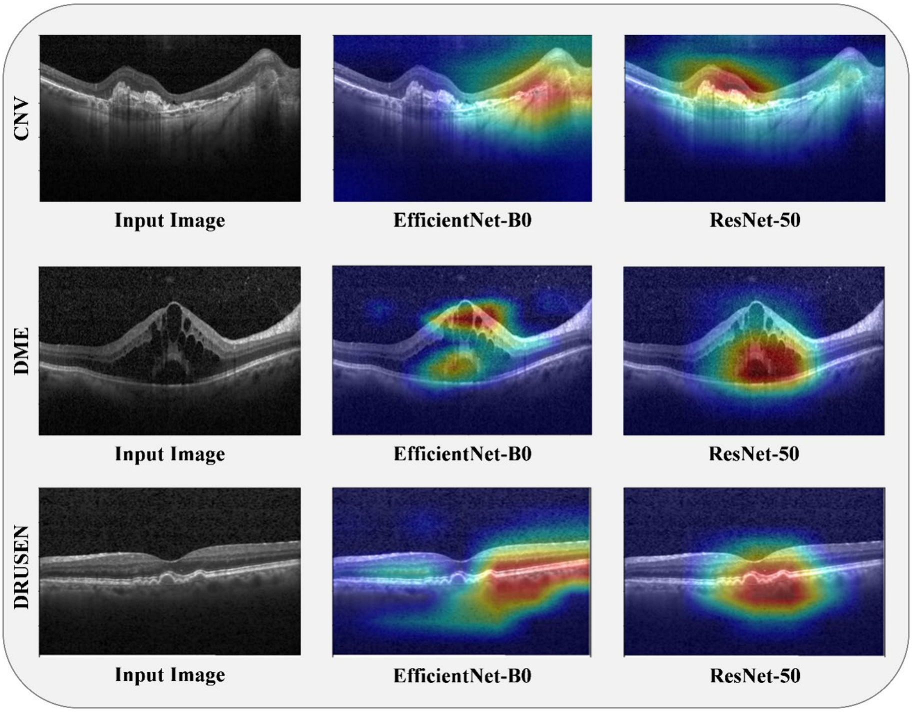


Fig. 8 Few sample images extracted from the dataset generated by the two models employed in our ensemble model, along with their corresponding GRAD-CAM visualizations

improved patient outcomes. While Fig. 9 illustrate the feature map visualization of backbone models.

Table 4 provides a comparative overview of various methods and their corresponding accuracies in a classification task, likely related to a specific domain or dataset. These methods are compared based on their performance in terms of accuracy. The methods listed in Table 3 include IFCNN proposed by [26] with an accuracy of 87.30% and the use of ResNet50, VGG16, and InceptionV3 by [47]. Achieving 90.00%, TL [48] with an accuracy of 87.40%, and DL [49]. Reaching 89.10%, and CNN employed by [50]. They were resulting in an accuracy of 93.30%. The final row represents the Proposed Architecture that utilizes a Hybrid Learning approach and achieves the highest accuracy of 97.50%. Hassan et al. [51] achieved an accuracy of 97.47% using a Random Forest model with TL, indicating the effectiveness of leveraging pre-trained models for improved performance. Özdaş et al. [52] achieved a high accuracy of 95.70% using the Firefly Algorithm, showcasing the efficacy of nature-inspired optimization algorithms in solving complex problems. Meanwhile, Han et al. [53] attained an accuracy of 97.40% by employing a Conditional Generative Adversarial Network-CGAN) to generate OCT images, demonstrating the power of generative models in creating synthetic data for training classifiers. These results underscore the diverse approaches and methodologies that can yield high accuracy in machine learning tasks, each leveraging unique strengths and techniques tailored to the specific problem domain. Table 4 serves as a comparative reference for evaluating the effectiveness of different methods

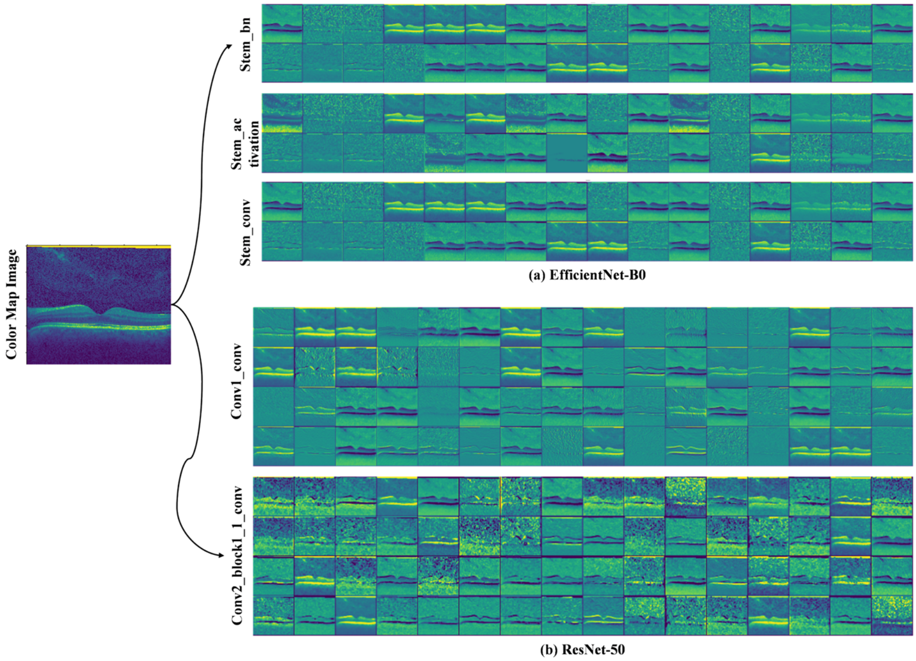


Fig. 9 Feature Map visualization: (a) EfficientNetB0-Base, and (b) ResNet50-Base

Table 4 Performance Comparison with existing studies

Reference	Method	Accuracy
Fang et al. [26]	IFCNN	87.30
Hwang et al. [47]	ResNet50, VGG16, and InceptionV3	90.00
J Han et al. [48]	Transfer Learning	87.40
DDJ Hwang et al. [49]	Deep Learning	89.10
Saraiva et al. [50]	CNN	93.30
Hassan et al. [51]	Random Forest with TL	97.47
Özdaş et al. [52]	Firefly Algorithm	95.70
Han et al. [53]	CGAN-generated OCT	97.40
Proposed Architecture	**Hybrid Learning**	**97.50**

in the specific classification task, highlighting the superior performance of the proposed Hybrid Learning approach with the highest accuracy.

4.3 Performance analysis with baseline methods

Table 5 shows the performance metrics for various models used in a classification task. These models include well-known architectures like DenseNet121, EfficientNetB0, EfficientNetB5, ResNet-50, MobileNet, VGG16, and a hybrid learning model. The evaluation metrics include Accuracy, Precision, Recall, and F1 Score. ResNet-50, and EfficientNetB0

Table 5 Performance Comparison with pre-trained models

Models	Accuracy	Precision	Recall	F1 Score
DenseNet-121-Base	80.20	82.04	80.19	80.11
EfficientNetB0-Base	**96.70**	**96.85**	**96.70**	**96.69**
EfficientNetB5-Base	93.20	94.12	93.20	93.15
ResNet50-Base	**95.19**	**95.68**	**95.50**	**95.17**
MobileNet-Base	80.50	83.13	80.49	80.14
VGG16-Base	92.70	92.69	92.69	92.66
Hybrid Learning	**97.50**	**98.10**	**97.23**	**98.21**

achieved the highest accuracy at 96.70%, 95.19%, demonstrating its effectiveness in classification respectively. It also showed the highest Precision and Recall at 95.68%, 96.85% and 95.50%, 96.70% respectively, resulting in an impressive F1 Score of 95.17%. On the other hand, DenseNet-121 had the lowest accuracy and F1 Score at 80.20% and 80.11%, respectively, indicating lower overall performance. The Hybrid Learning model outperformed all the base models, achieving an exceptional accuracy of **97.50%** and high Precision, Recall, and F1 Score of **98.10%**, **97.23%**, and **98.21%**, respectively. These metrics provide valuable insights into the models' capabilities in accurately classifying data, with the hybrid learning model emerging as the top performer.

5 Discussion

In the realm of disease diagnosis, precisely categorizing medical images poses a significant challenge. DL models, known for their exceptional performance in this field, typically require extensive datasets with accurate annotations. However, the availability of well-labeled medical image data is often scarce. To tackle this issue, this study utilizes TL to develop a robust model for classifying retinal diseases. The study leverages TL to develop a hybrid learning model, merging ResNet-50 and EfficientNet-B0 architectures for retinal disease classification from OCT images. Evaluation on a public dataset demonstrates superior performance in precision, recall, and F1-score. This model enhances classification accuracy, even with limited annotated data, reducing the dependency on extensive specialized datasets. ResNet50-Base and EfficientNetB0-Base offer superior performance due to their deeper architectures and efficient parameter usage, respectively, resulting in more accurate feature extraction and classification from images compared to other pre-trained models. Prediction Fusion of ResNet50-Base and EfficientNetB0-Base combines the strengths of these architectures, enhancing prediction accuracy by leveraging diverse feature representations, leading to more robust and reliable detection of retinal diseases in OCT images. Moreover, this combination creates a robust, adaptable model for diverse retinal images. Notably, while individual models may struggle with generalization, our ensemble approach mitigates this issue. The ensemble model offers enhanced classification accuracy and reliability, aiding in tailored handling strategies and broadening accessibility to retinal disease diagnosis in medical applications. ResNet-50 and EfficientNetB0 have demonstrated highest effectiveness in classification tasks, achieving the highest accuracies at 96.70% and 95.19%, respectively. These models also exhibit superior Precision and Recall rates, with ResNet-50 scoring 95.68% and 96.85%, and EfficientNetB0 achieving 95.50%

and 96.70% respectively, and impressive F1 Score of 95.17%. However, our Hybrid Learning model outperform all base models, boasting an highest accuracy of 97.50%, alongside high Precision, Recall, and F1 Score metrics, standing at 98.10%, 97.23%, and 98.21%, respectively. Notably, both our proposed model and its backbone models outperform not only other pretrained models but also existing state-of-the-art models in terms of classification performance.

6 Conclusion

Tackling the vanishing gradient problem in deep learning is a critical challenge, especially when it comes to classifying medical images, particularly in the diagnosis of retinal diseases using OCT images. To overcome this challenge, we have devised a robust hybrid learning approach that synergizes the capabilities of two well-established models, ResNet-50 and EfficientNet-B0, aimed at achieving precise classification of four distinct retinal illnesses. Our methodology revolves around merging the ResNet-50 and EfficientNet-B0 models, leading to a significant enhancement in classification accuracy. This fusion utilizes the strengths of both models and creates a framework for identifying retinal diseases. In our experiments, we employed a substantial dataset comprising 84,452 OCT images, categorized into four distinct retinal illness groups: DRUSEN, NORMAL, CNV, and DME. Thorough evaluations of our proposed approach resulted in impressive outcomes, boasting an average precision of 98.10%, a recall rate of 97.23%, an F1-score of 98.21%, and an overall classification accuracy of 97.50%. These outstanding results carry significant implications for the field of ophthalmology. Not only does our approach facilitate accurate identification of a range of eye disorders, but it also introduces an interpretable dimension to the decision-making process. This interpretability factor sets our method apart from non-interpretable approaches, instilling trust and confidence in healthcare professionals who use it. As part of our ongoing research, we intend to expand the scope of our approach by incorporating additional retinal eye diseases, further enhancing its effectiveness in classifying such disorders. Ultimately, our research contributes significantly to the advancement of medical image classification techniques, benefiting both healthcare practitioners and patients alike. Beside of this, our model proposed some limitation such as Using multiple pre-trained models in an ensemble can demand substantial processing power, which poses a significant challenge for researchers with limited computing resources. This high computational requirement may hinder their ability to effectively implement and experiment with such ensembles.

Abbreviations OCT: Optical coherence tomography; DL: Deep Learning; DME: Diabetic macular edema; CNV: Choroidal neovascularization; AMD: Age-related macular degeneration; SVM: Support Vector Machine; CNN: Convolutional Neural Network; IFCN: International fact-checking network; TL: Transfer Learning; TN: Transformer network; ML: Machine Learning; DCNN: Deep Convolutional Neural Networks; DNN: Deep Neural Network; CGAN: Conditional Generative Adversarial Network

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 61972136, in part by the Natural Science Foundation of Hubei Province under Grant 2020CFB497, in part by the Hubei Provincial Department of Education Outstanding Youth Scientific Innovation Team Support Foundation under Grant T201410 and Grant T2020017, in part by the Natural Science Foundation of Education Department of Hubei Province under Grant B2020149, in part by the Science and Technology Research Project of the Education Department of Hubei Province under Grant Q20222704, and in part by the Natural Science Foundation of Xiaogan City under Grant XGKJ2022010095 and Grant XGKJ202201 0094.

Authors' contributions Saif Ur Rehman Khan: Conceptualization, Methodology, Data curation, Writing-Original draft preparation, Validation, Investigation, Reviewing and Editing: Umer Sadiq Khan: Reviewing and Editing.

Funding Not Applicable.

Data availability Data are available from the authors upon reasonable request and with permission of Author.

Declarations

Ethics approval Approved.

Conflict of interest There is no conflict of interest.

References

1. Li T et al (2021) Applications of deep learning in fundus images: A review. Med Image Anal 69:101971
2. Zhao J et al (2022) Emerging trends and research foci in artificial intelligence for retinal diseases: Bibliometric and visualization study. J Med Internet Res 24(6):e37532
3. Kim, J. and L. Tran (2021) Retinal disease classification from oct images using deep learning algorithms. in 2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE
4. Anoop B (2022) Binary classification of DR-diabetic retinopathy using CNN with fundus colour images. Mater Today: Proc 58:212–216
5. Tan T-E, Wong TY (2023) Diabetic retinopathy: Looking forward to 2030. Front Endocrinol 13:1077669
6. Tang L, Xu G-T, Zhang J-F (2023) Inflammation in diabetic retinopathy: possible roles in pathogenesis and potential implications for therapy. Neural Regen Res 18(5):976–982
7. Yasuno Y (2023) Multi-Contrast Jones-Matrix Optical Coherence Tomography—The Concept, Principle, Implementation, and Applications. IEEE J Sel Top Quantum Electron 29:1–18
8. Tsuji T et al (2020) Classification of optical coherence tomography images using a capsule network. BMC Ophthalmol 20(1):1–9
9. Perdomo O et al (2019) Classification of diabetes-related retinal diseases using a deep learning approach in optical coherence tomography. Comput Methods Programs Biomed 178:181–189
10. Li F et al (2019) Deep learning-based automated detection of retinal diseases using optical coherence tomography images. Biomed Opt Express 10(12):6204–6226
11. Khan SUR, Asif S (2024) Oral cancer detection using feature-level fusion and novel self-attention mechanisms. Biomed Signal Process Control 95:106437
12. Khan SUR et al (2024) Deep hybrid model for Mpox disease diagnosis from skin lesion images. Int J Imaging Syst Technol 34(2):e23044
13. An G et al (2021) Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images. Sci Rep 11(1):4250
14. Goodfellow I et al (2020) Generative adversarial networks. Commun ACM 63(11):139–144
15. Raza A, Meeran MT, Bilhaj U (2023) Enhancing Breast Cancer Detection through Thermal Imaging and Customized 2D CNN Classifiers. VFAST Trans Softw Eng 11(4):80–92
16. Cheplygina V, de Bruijne M, Pluim JP (2019) Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Med Image Anal 54:280–296
17. Raut P et al (2024) Machine Learning-Based Detection and Classification of Eye Diseases: A Comprehensive Review and Novel Algorithm. Int J Intell Syst Appl Eng 12(13s):622–629
18. Das V, Dandapat S, Bora PK (2020) A data-efficient approach for automated classification of OCT images using generative adversarial network. IEEE Sens Lett 4(1):1–4
19. Yoo TK et al (2019) The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment. Med Biol Eng Compu 57:677–687

20. Sun Y, Li S, Sun Z (2017) Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning. J Biomed Opt 22(1):016012–016012
21. Rong Y et al (2018) Surrogate-assisted retinal OCT image classification based on convolutional neural networks. IEEE J Biomed Health Inform 23(1):253–263
22. Ker J et al (2017) Deep learning applications in medical image analysis. Ieee Access 6:9375–9389
23. Tan JH et al (2018) Age-related macular degeneration detection using deep convolutional neural network. Futur Gener Comput Syst 87:127–135
24. Hua R et al (2022) Development and validation of a deep learning algorithm based on fundus photographs for estimating the CAIDE dementia risk score. Age Ageing 51(12):afac282
25. Alqudah AM (2020) AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. Med Biol Eng Compu 58:41–53
26. Fang L et al (2019) Iterative fusion convolutional neural networks for classification of optical coherence tomography images. J Vis Commun Image Represent 59:327–333
27. Roy, K., et al (2020) Transfer learning coupled convolution neural networks in detecting retinal diseases using OCT images. Intelligent computing: Image processing based applications, 153–173
28. Karthik K, Mahadevappa M (2023) Convolution neural networks for optical coherence tomography (OCT) image classification. Biomed Signal Process Control 79:104176
29. Khan SUR, Zhao M, Asif S, Chen X, Zhu Y (2023) GLNET: global–local CNN's-based informed model for detection of breast cancer categories from histopathological slides. J Supercomput 80:7316–7348
30. Huang X et al (2023) GABNet: global attention block for retinal OCT disease classification. Front Neurosci 17:1143422
31. Li F et al (2019) Fully automated detection of retinal disorders by image-based deep learning. Graefes Arch Clin Exp Ophthalmol 257:495–505
32. Fki Z, Ammar B, Ayed MB (2023) Towards automated optimization of residual convolutional neural networks for electrocardiogram classification. Cogn Comput 16(3):1334–1344
33. Karn PK, Abdulla WH (2023) On machine learning in clinical interpretation of retinal diseases using oct images. Bioengineering 10(4):407
34. Liu X et al (2023) TSSK-Net: Weakly supervised biomarker localization and segmentation with image-level annotation in retinal OCT images. Comput Biol Med 153:106467
35. Bansal P et al (2024) Utilization of big data classification models in digitally enhanced optical coherence tomography for medical diagnostics. Neural Comput Appl 36(1):225–239
36. Naik, G., et al (2024) Eye Disease Prediction Using Ensemble Learning and Attention on OCT Scans. in Future of Information and Communication Conference. Springer
37. Kermany DS et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122–1131
38. Choudhary, A., et al (2023) A deep learning-based framework for retinal disease classification. in Healthcare. MDPI
39. Shahzad I, Khan SUR, Waseem A, Abideen ZU, Liu J (2024) Enhancing ASD classification through hybrid attention-based learning of facial features. Signal, Image Video Process, pp 1–14
40. Lu W et al (2018) Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. Transl Vis Sci Technol 7(6):41–41
41. Song D et al (2021) Deep relation transformer for diagnosing glaucoma with optical coherence tomography and visual field function. IEEE Trans Med Imaging 40(9):2392–2402
42. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
43. Tan, M. and Q. Le (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. in International conference on machine learning. PMLR
44. Huang M-L, Liao Y-C (2023) Stacking Ensemble and ECA-EfficientNetV2 Convolutional Neural Networks on Classification of Multiple Chest Diseases Including COVID-19. Acad Radiol 30(9):1915–1935
45. Wani NA, Kumar R, Bedi J, Rida I (2024) Explainable AI-driven IoMT fusion: unravelling techniques, opportunities, and challenges with Explainable AI in healthcare. Inf Fusion 102472
46. Loh HW et al (2022) Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Comput Methods Programs Biomed 226:107161
47. Hwang D-K et al (2019) Artificial intelligence-based decision-making for age-related macular degeneration. Theranostics 9(1):232

48. Han J et al (2022) Classifying neovascular age-related macular degeneration with a deep convolutional neural network based on optical coherence tomography images. Sci Rep 12(1):2232
49. Hwang DD-J et al (2021) Distinguishing retinal angiomatous proliferation from polypoidal choroidal vasculopathy with a deep neural network based on optical coherence tomography. Sci Rep 11(1):9275
50. Saraiva AA, Santos DBS, Pimentel PMC, Sousa JVM, Ferreira NM, Batista Neto JDES, Soares S, Valente A (2020) Classification of optical coherence tomography using convolutional neural networks. In Proceedings
51. Hassan E et al (2023) Enhanced deep learning model for classification of retinal optical coherence tomography images. Sensors 23(12):5393
52. Özdaş MB, Uysal F, Hardalaç F (2023) Classification of retinal diseases in optical coherence tomography images using artificial intelligence and firefly algorithm. Diagnostics 13(3):433
53. Han K, Yu Y, Lu T (2024) Transfer Learning and Interpretable Analysis-Based Quality Assessment of Synthetic Optical Coherence Tomography Images by CGAN Model for Retinal Diseases. Processes 12(1):182

Authors and Affiliations

Umer Sadiq Khan[1,2,3] · Saif Ur Rehman Khan[4]

✉ Umer Sadiq Khan
umersadiq@hbeu.edu.cn

✉ Saif Ur Rehman Khan
saifurrehman.khan@csu.edu.cn

[1] School of Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China

[2] School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China

[3] Institute for AI Industrial Technology Research, Hubei Engineering University, Xiaogan 432000, China

[4] School of Computer Science and Engineering, Central South University, Yuelu District, 932 Lushan S Rd, Changsha Hunan, China