International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I3P134

Volume 72 Issue 3, 391-401, March 2024

Original Article

Automated Macular Disease Classification from OCT Images using Feature Attention Fusion Network

V. Latha[1], Sreeni K G[2]

[1]*College of Engineering Trivandrum, Kerala, India.*
[2]*Rajiv Gandhi Institute of Technology, Kottayam, Kerala, India*.

[1]*Corresponding Author : vlatha@cet.ac.in*

Received: 25 October 2023 Revised: 21 February 2024 Accepted: 29 February 2024 Published: 17 March 2024

Abstract *- Optical Coherence Tomography (OCT) is a promising and essential tool for retinopathy diagnosis. Ophthalmologists use OCT images to identify, treat, and track macular diseases. Manually analysing and interpreting these illnesses from the enormous volume of OCT images takes time and effort. The Convolutional Neural Network (CNN), a potential deep learning method, has proven exceptionally accurate in classifying images, making them suitable for computer-assisted diagnosis. This paper introduces a CNN-based feature extraction and attention fusion network, the FAF-Net, to classify common macular diseases. This network enhances the flexibility and accuracy of conventional CNN classification systems. The attention module strengthens pre-trained CNN models by emphasising significant features related to anatomical defects in the retina while reducing the importance of irrelevant regions. Combining deep pre-trained models with attention processes further improves classification accuracy. This study used OCT data to diagnose macular disorders using pre-trained models VGG16 and ResNet50. The proposed approach was assessed on the UCSD dataset, a publicly available OCT imaging dataset, achieving a classification accuracy of 98.40%. Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed as a visualisation technique to assess the efficacy of the FAF-Net. These results demonstrate that the proposed FAF-Net approach substantially increases classifier performance. This method has a promising future in the medical field and provides a robust conceptual framework for diagnosing macular diseases.*

Keywords *- CNN, Deep learning, Feature extraction, Image classification.*

1. Introduction

The central region of the retina, known as the macula, plays a vital role in providing clear and sharp vision. The two most prevalent retinal diseases are Age-related Macular Degeneration (AMD), which is the primary cause of blindness in older people, and Diabetic Macular Edema (DME), which is the leading cause of vision loss in younger and middle-aged individuals [1]. Diabetes can damage retinal blood vessels, causing them to leak or constrict, leading to DME, a severe eye disorder that can impair vision and even cause blindness. Dry AMD, also known as drusen, is characterised by the formation of minute yellow deposits of fatty proteins beneath the retina. Over time, these deposits might accumulate, leading to wet AMD impairing central vision. Wet AMD, mainly caused by Choroidal Neovascularisation (CNV), the development of new blood vessels in the choroid layer, is the leading cause of central vision loss. Accurate and prompt retinal screening and early treatment can prevent more than 80% of vision loss [2,3]. OCT imaging [4] produces high-resolution cross-sectional pictures non-invasively and provides a detailed view of structural and textural variations within the layers of the retina [5]. Figure 1 shows some examples of OCT scans from various classes. It aids in early disease detection and enables treatment before the appearance of visual complaints, decreasing the chances of irreversible vision loss [6]. However, manual analysis and diagnosis of retinal diseases from vast volumes of OCT scans is challenging, time-consuming, and susceptible to subjective findings. Computer-assisted detection of retinal illness greatly assists in automatic and accurate OCT image categorisation.

Developing classifier models for automatically discriminating retinal OCT images using deep learning-based methods significantly benefits clinical healthcare [7]. Deep Learning (DL) techniques, especially those based on Convolutional Neural Networks (CNN), have shown remarkable improvements in classifying biological images. The improvement can be credited to their ability to extract features from images using convolution methods. Tayal et al. [8] offered three CNN models to classify four eye disorders based on OCT data. They used 6,678 OCT images and enhanced them before presenting them to the models. With a 96.50% accuracy, the nine-layer CNN model surpassed the five- and seven-layer CNN models.

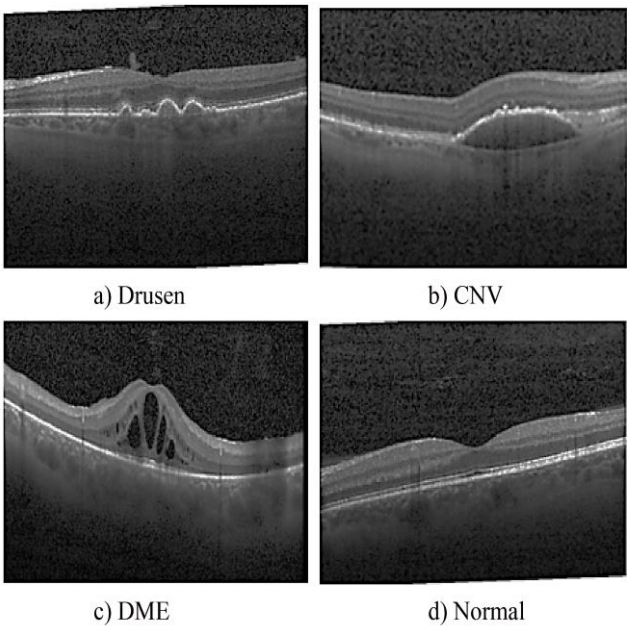


Fig. 1 Some examples of OCT images of Macular disease

For the study of medical images, large, labelled datasets may not always be accessible. Then, Transfer Learning (TL) is advantageous for various biological image categorisation applications. Han et al. [9] used three CNN TL models to differentiate between three subtypes of neovascular AMD (nAMD) and normal. They trained and validated their approach on a dataset of 4,749 SD-OCT images, tested it on 920 images, and achieved an accuracy of 87.4% using the VGG-16 model. Minagi et al. [10] used Universal Adversarial Perturbations (UAPs) and TL to classify medical images, including OCT data. They categorised the pictures into four categories with 95.3% accuracy using a dataset of 11,200 OCT images for training and testing.

The TL technique causes faster network convergence, while the classification result still needs to be improved. Instead of directly employing features from a pre-trained CNN, feature fusion is used as the image representation for enhancing macular OCT image categorisation. Using the CNN multiscale feature fusion strategy, Sotoudeh-Paima et al. [11] discriminated between normal, drusen, and CNV in OCT images. On a public dataset, their CNN performed with a categorisation accuracy of 93.40%. Elaziz et al. [12] constructed a DL-based ensemble model to recognise four classes of retinal diseases.

They integrated the feature maps from the MobileNet and DenseNet models for the input images. They employed a feature selection technique to filter out irrelevant information and then used machine learning for classification, resulting in an accuracy of 94.31% based on 968 OCT images. Pin et al. [13] proposed an automated approach based on DL and ensemble machine learning classifiers to screen five retinal diseases from OCT images. They used DL to extract features from the images and applied two machine learning algorithms to categorise these features. By merging the outputs of both classifiers, they achieved an accuracy of 97.68%.The attentional process of DL is based on how people pay attention to images. The attention process emphasises vital information while suppressing irrelevant information to increase prediction confidence. Liu et al. [14] used an attention-based retinal OCT image classification method. These features provided categorisation criteria for the normal, CNV, DME, and drusen classes and found an average accuracy of 95.10%.

This paper introduces a CNN-based feature attention fusion network, the FAF-Net, to the automated diagnosis of common macular diseases. FAF-Net combines the feature extraction strengths of ResNet50 and VGG16, two potent deep CNN models, with an attention mechanism contributing to accurately detecting retinal disorders. The key findings of this study are summarised as follows:

- An effective hybrid attention module, incorporating both channel and spatial attention, is provided to enhance CNN feature representation. This module strengthens the representation of vital information while diminishing the emphasis on irrelevant information.
- A fusion operation is implemented to integrate two CNN model architectures with the attention mechanism. This feature fusion enhances the classifier's discrimination capability by augmenting the dimensionality and introducing non-linearity to the fused feature vector.
- The effectiveness of the FAF-Net is assessed using the Gradient-Weighted Class Activation Mapping (Grad-CAM) method, providing insights into the visualisation of attention and the model's performance.

The remaining sections of the paper are arranged as follows. Section 2 describes the suggested FAF-Net network and the details of the associated modules. Section 3 discusses the experimental details, whereas Section 4 presents the outcomes. Section 5 discusses an ablation study, and Section 6 presents the conclusions.

2. Method

OCT scan produces a two-dimensional image of the retinal layers and surrounding tissues. Diagnostic systems must concentrate on the area of the retina with significant changes in morphological features for meaningful feature extraction and superior categorisation. Figure 1 shows the retinal layer deformation due to macular disorders. FAF-Net, an attention-based technique that focuses on deformation zones and captures important morphological features for final classification, is shown in Figure 2. A novel architecture is proposed in which important deep features are effectively chosen, leading to the accurate classification of OCT images. The FAF-Net system comprises four main stages: feature extraction, feature attention, multi-attention fusion, and classification.

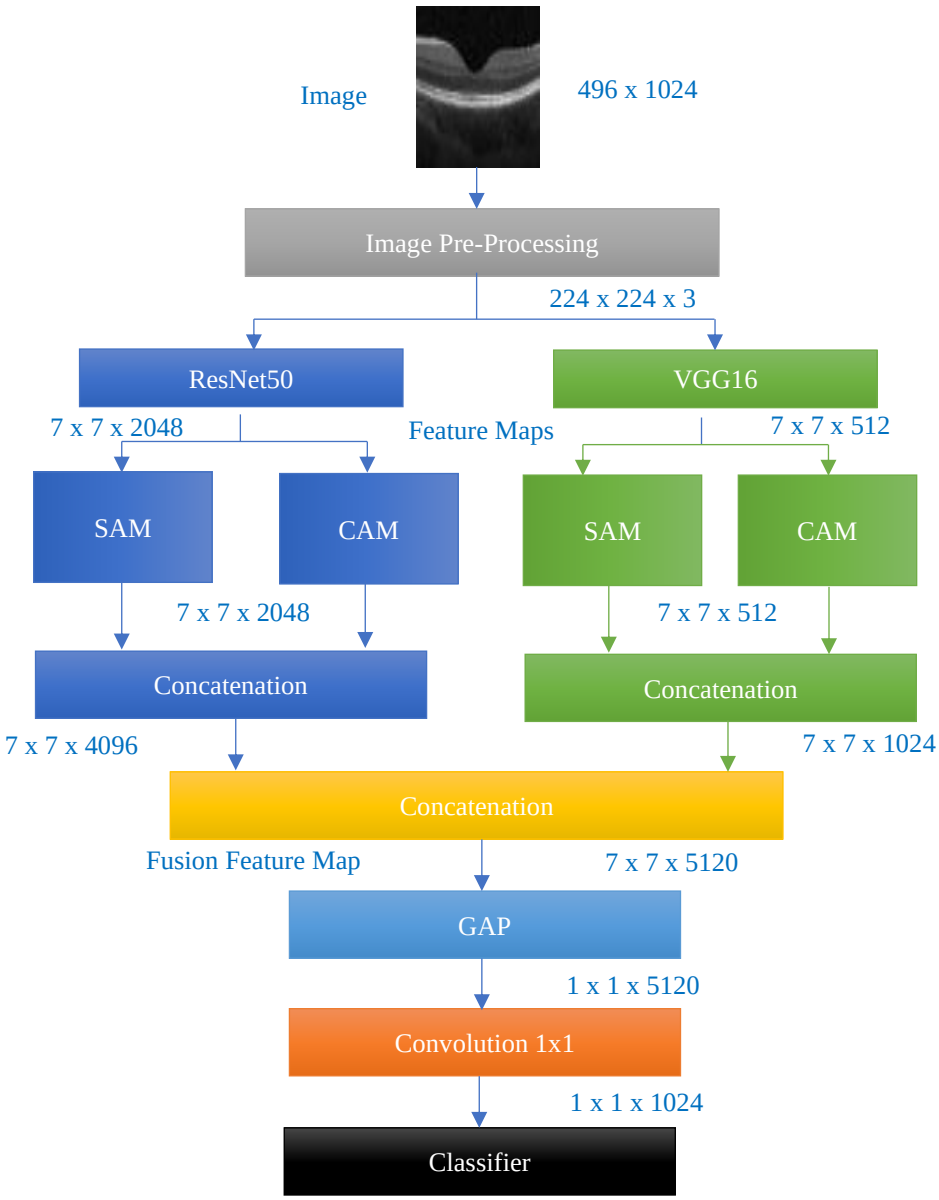


Fig. 2 The architecture of the proposed FAF-Net model

2.1. Feature Extraction

The proposed research used two pre-trained models, VGG16 and ResNet50, to extract features from OCT images. VGG16, a well-known deep learning architecture [15], has 16 layers, 13 of which are convolutional. The input layer of VGG16 accommodates images with dimensions of 224×224×3 pixels. Researchers from various fields have used the VGG16 network to solve problems [16-18]. On the other hand, ResNet50, another widely used transfer-learning model [19], is a deeper model comprising a total of 50 layers, with 48 of them being convolutional layers. As a model's depth increases, challenges related to gradients, such as vanishing and exploding gradients, can emerge. ResNet50 tackles these issues by incorporating skip connections, allowing for the addition of extra layers while preserving essential information throughout data processing. The input dimension for ResNet50 is also 224×224×3. Several researchers have employed this network architecture to address problems in various fields [20 - 23].

2.2. Feature Attention

This section introduces the proposed attention networks, including the Channel Attention Module (CAM) and Spatial Attention Module (SAM), as illustrated in Figure 2. The channel attention module's primary objective is to provide different feature maps with channel coefficients so they can identify correlations across channels. On the other hand, the spatial attention module focuses on assigning spatial coefficients to different feature maps, making the importance of location information across different feature maps simpler.

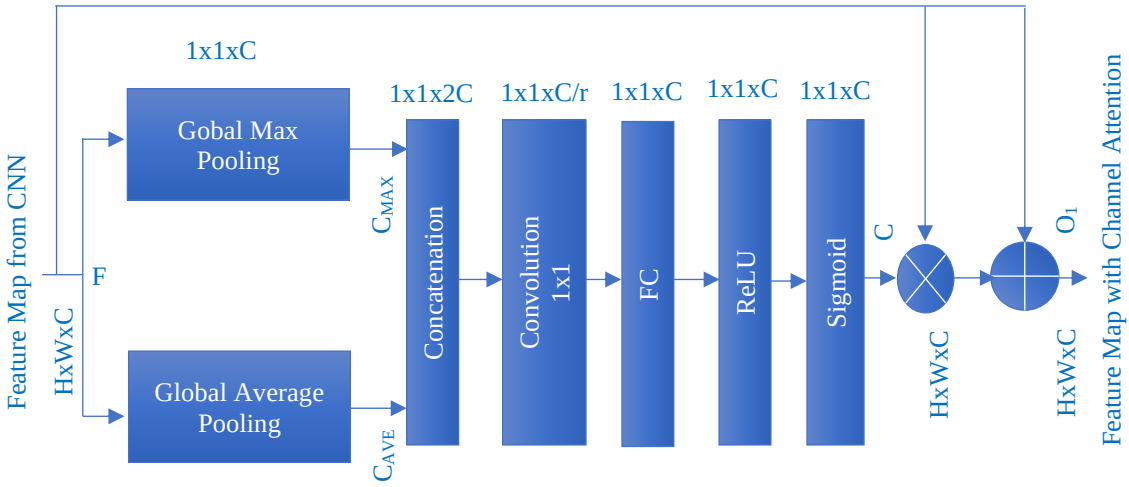


Fig. 3 Channel Attention Module (CAM) architecture

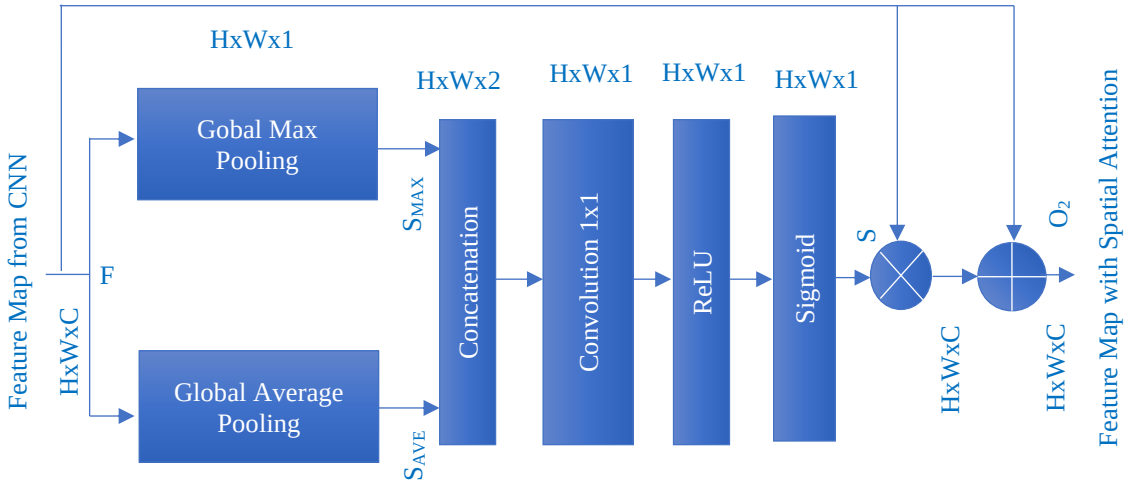


Fig. 4 Spatial Attention Module (SAM) architecture

2.2.1. Channel Attention Module (CAM)

Figure 3 displays the architecture of the CAM. This method uses two global pooling operations as parallel branches to squeeze the feature map channel-wise [24]. The feature map F is subjected to a global max-pooling procedure, resulting in channel-wise maximum value feature maps (C_{MAX}). The parallel global average-pooling technique on F creates channel-wise average value feature maps (C_{AVE}). The next step involves concatenating C_{MAX} and C_{AVE}; the resulting output is reshaped and passed through a convolutional layer for feature map smoothing and compression. A Fully Connected layer (FC) receives the output from the convolutional process and applies the FC result to the sigmoid function to produce the feature map C. Equation (1) illustrates this operation.

$$C = \sigma\left(FC\left(Conv(Concat\left(C_{MAX}, C_{AVE}\right)\right)\right)) \quad \ldots(1)$$

Where σ represents the sigmoid operation, Conv is the convolution, and concat represents the concatenation operation. This attention module can capture each feature map's response across all channels while reducing noise interference. The CNN feature map F is multiplied by feature map C channel-wise, followed by a pixel- wise addition operation with F is performed, as given in Equation (2).

$$O_1 = F \oplus (\alpha F \otimes_{Channel} C) \quad (2)$$

Where $\otimes_{Channel}$ denotes multiplication at the channel level, α is the learnable parameter. It is a non-zero value (used α =1.0), ⊕ represents addition at the pixel level, and O_1 represents the residual module output. With the pixel-level addition of smoothed feature maps and the initial feature map, the multiscale residual attention fusion module produces the output O_1.

2.2.2. Spatial Attention Module (SAM)

Convolutional neural networks frequently use max-pooling and average-pooling operations to reduce feature map sizes while preserving crucial spatial response data in each channel. Due to variability in lesion sizes and forms, they may unintentionally maintain noise [25].

Since noise interference must be suppressed, the spatial attention modules concurrently execute 2D max-pooling and average-pooling to collect the essential spatial response data from all channels, as illustrated in Figure 4.This method produces attention maps S_{MAX} and S_{AVE} by feeding feature map F parallel to the maximum and average map branches. These maps are then concatenated along the channel dimension. The concatenated maps are compressed using a convolutional approach, and a sigmoid function is then used to create the feature map S. Equation (3) illustrates this procedure.

$$S = \sigma(Conv\ (Concat(S_{MAX}, S_{AVE}\ ))) \qquad (3)$$

Where σ represents the sigmoid operation, Conv is the convolution, and concat represents the concatenation operation. This module can effectively reduce noise interference while collecting each feature map's response across all channels. The CNN feature map F is multiplied by feature map S spatially, and then a pixel-wise addition operation with F is performed as given. Equation (4).

$$O_2 = F \oplus (\beta\ F\ \otimes_{Spatial}\ S) \qquad (4)$$

Where $\otimes_{Spatial}$ denotes multiplication at the spatial level, β is a learnable parameter and is a non-zero value (used β= 1.0), ⊕ represents addition at the pixel level, and O2 represents the residual module output. Pixel-wise addition of the initial feature map and the smoothed attention feature map produces the multiscale residual attention fusion module output O2.

2.3. Multi-Attention Fusion Module

The Multi-Attention Fusion Operation (MAFO) transmits the output results of CAM and SAM with the baseline models through Concatenation operations. It performs attention feature fusion of CNNs, as shown in Figure 2. The feature fusion technique output goes through a global average pooling (GAP) to reduce the dimension to 1D and then to the disease identification classifier.

2.4. Classifier Module

In this section, three Fully Connected (FC) layers and a softmax classifier complete the classification process. The concatenation layer creates the feature map to the initial FC layer, a 1024-dimensional feature vector with the ReLU activation function. The second FC layer produces a 512-dimensional feature vector, and the third FC layer performs to create a 4-dimensional feature vector. The Softmax function in the topmost layer determines the output decision classes.

3. Experimental Procedure

This study offers a CNN-based feature extraction and attention fusion network to enhance the discriminative features from pre-trained CNNs like VGG16 and ResNet50.

3.1. Dataset

The training dataset from UCSD[25] contains 108,309 OCT B-scans, divided into 51,140 normal B-scans, 8,616 drusen scans, 11,348 DME scans, and 37,205 CNV scans. In the test set, there are 250 images for each category. The Beijing Tongren Eye Centre, California Retinal Research Foundation, Shanghai First People's Hospital, Medical Centre Ophthalmology Associates, and Shiley Eye Institute collaborated to manage the UCSD retinal OCT dataset collected between July 1, 2013, and March 1, 2017. This dataset is currently the biggest and the finest resource for macular OCT imaging.

3.2. Image Pre-processing

Data pre-treatment techniques were implemented to reduce the hardware resource requirements and enhance the model's reliability. All OCT images were resized to 224x224 pixels to meet the CNN model's requirements for the precise image input size and to speed up processing.

The pixel values were normalised to ensure uniform treatment of each image. Upon completion of the pre-processing steps, the training dataset was partitioned, allocating 20% of the images to the validation set and the remaining 80% to the training set.

3.3. Implementation Details

The model was developed using the Google CoLab development environment, leveraging a potent Graphics Processing Unit (GPU) and deep learning frameworks built on TensorFlow. Training the model involved 50 iterations utilising the Adam optimiser, with a batch size of 16 and a learning rate set at 0.0001. The CNN model with the lowest validation loss was selected as the optimal model, and its performance was evaluated on the test set to assess the training results.

3.4. Evaluation Metrics

The suggested FAF-Net method's classification performance was evaluated using a combination of qualitative and quantitative metrics.

3.4.1. Learning Curves

A learning curve depicts a model's progress as it learns from a training set. Following each training cycle, the model is assessed using the training and separate validation datasets. Examining learning curves during training makes it possible to detect the model learning, correct fitting, underfitting or overfitting, and the representativeness of the training and validation datasets.

3.4.2. ROC Curve

The Receiver Operating Characteristic (ROC) curve is a metric for evaluating a classifier model performance. This curve offers a numerical way of assessing the accuracy and comparing the results of different test or prediction models. The ROC curve demonstrates the balance between sensitivity and specificity. The area under the ROC curve, AUC, determines a test's usefulness. A classifier with an AUC of more than 0.9 is considered outstanding.

3.4.3. Confusion Matrix

A confusion matrix serves as a tool for evaluating the performance of a supervised learning system. It provides a visual and summarised representation of a classification algorithm's performance. A Confusion matrix is a comparison summary of any classification problem's predicted and actual results. The confusion matrix illustrates when one class is explicitly confused with another in machine learning. A confusion matrix can assist in identifying various performance metrics, Sensitivity/Recall, Precision, Accuracy, and F1-score metrics employed to explore and compare the model.

3.4.4. Grad CAM

Grad-CAM (Gradient-weighted Class Activation Mapping) is a method used to reveal where a CNN model concentrates. Since it is class-specific, each input class or image generates a distinct visualisation.Grad-CAM creates a heatmap visualisation for each class label, allowing us to visually confirm the regions of interest that the network has located in an image.

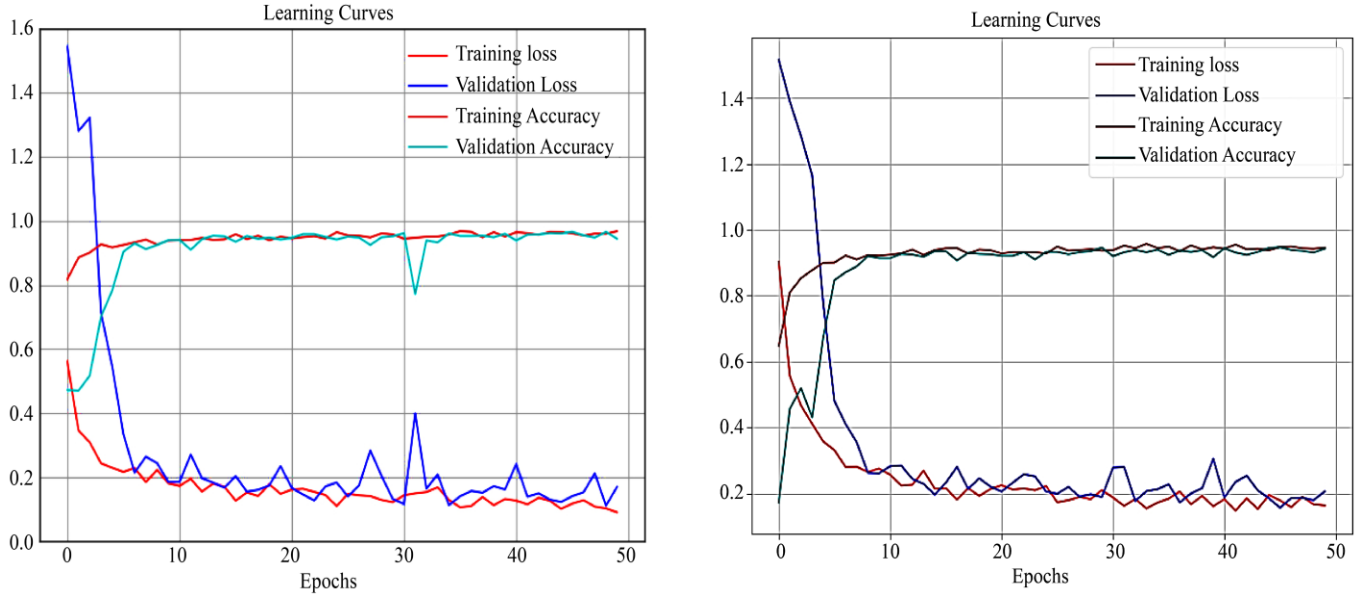
4. Results and Discussion

The proposed FAF-Net model, based on CNN, showed promising results for accurately classifying the OCT images. Learning curves are commonly employed as diagnostic tools in machine learning for algorithms that continuously extract new information from a training dataset.

Figures 5, 6, and 7 display two models: the Complete dataset model, which uses all available datasets for training and prediction across four classes (CNV, DME, drusen, normal), and the Limited dataset model, which chooses 10% of retinal OCT images from each category in the training set to train and predict across the four classes (CNV, DME, drusen, normal). The same test set of 1000 images is assessed on the proposed model, FAF-Net, the Complete dataset model and the Limited dataset model. The confusion matrix data is a quantitative study useful for network monitoring and analysis. The confusion matrix enables the determination of recall, precision, accuracy and f1 score metrics.

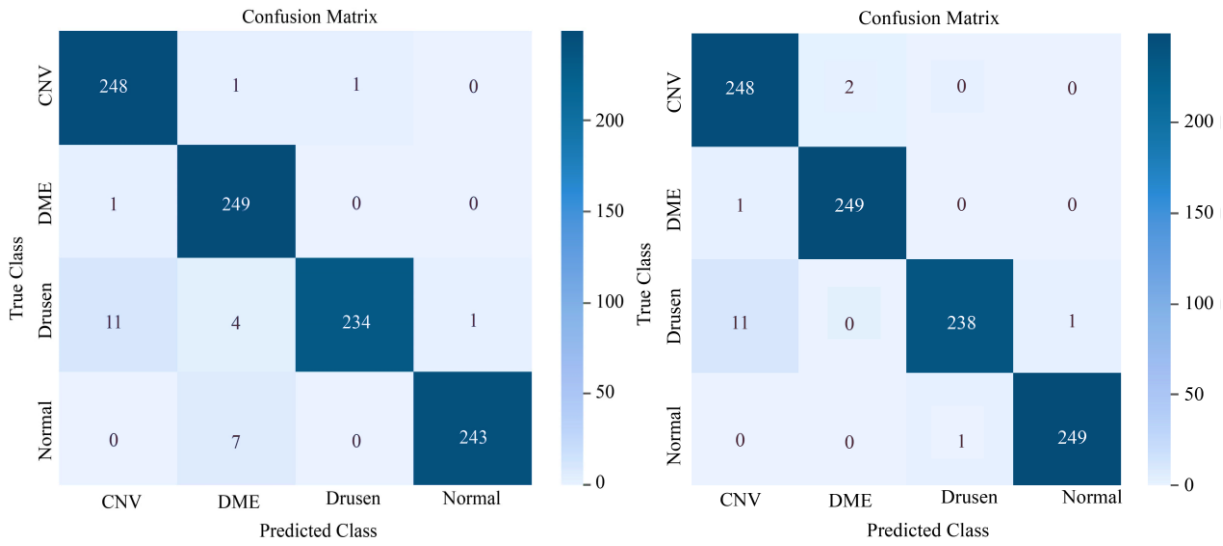
The classification system under test often includes the confusion matrix and associated metrics. The confusion matrix for the proposed FAF-Net model is displayed in Figure 6, while Table 1 presents performance metrics. The evaluation metrics obtained by FAF-Net in the limited data model are inferior to those achieved in the complete data model.

This discrepancy is mainly attributed to the significantly smaller size of the training dataset, resulting in insufficient training. In addition to accuracy measurements, ROC curves are used to assess the accuracy and robustness of the system. ROC curves were generated for every class, corresponding to the CNV, DME, drusen, and normal classes, as shown in Figure 7. The AUC is a performance measure that evaluates the strengths and weaknesses of a model. The AUC in Figure 7 suggests that the classifier accurately distinguishes between the diseased and the normal image.



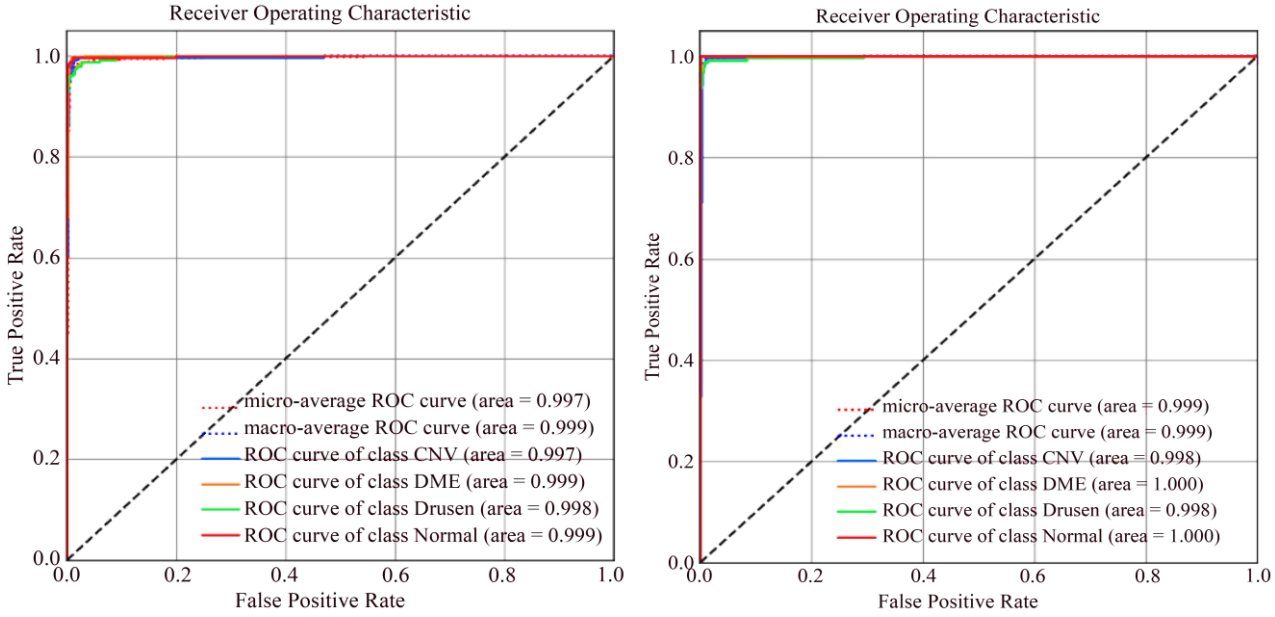
(a) Limited data model **(b) Complete data model**

Fig. 5 FAF-Net model's learning curves.



(a) Limited data model **(b) Complete data model**

Fig. 6 Confusion Matrix of FAF-Net



(a) Limited data model **(b) Complete data model**

Fig. 7 ROC performance of FAF-Net

Table 1. FAF-Net model's categorisation effectiveness

	Class-wise Performance					Overall Performance			
	Class	Recall	Precision	Accuracy	F1-score	Recall	Precision	Accuracy	F1-score
Limited Dataset	CNV	0.992	0.954	0.986	0.973	0.974	0.975	0.974	0.974
	DME	0.996	0.954	0.987	0.975				
	Drusen	0.936	0.996	0.983	0.965				
	Normal	0.972	0.996	0.992	0.984				
Complete Dataset	CNV	0.992	0.954	0.986	0.973	0.984	0.984	0.984	0.984
	DME	0.996	0.992	0.997	0.994				
	Drusen	0.952	0,952	0.987	0.973				
	Normal	0.996	0.996	0.998	0.996				

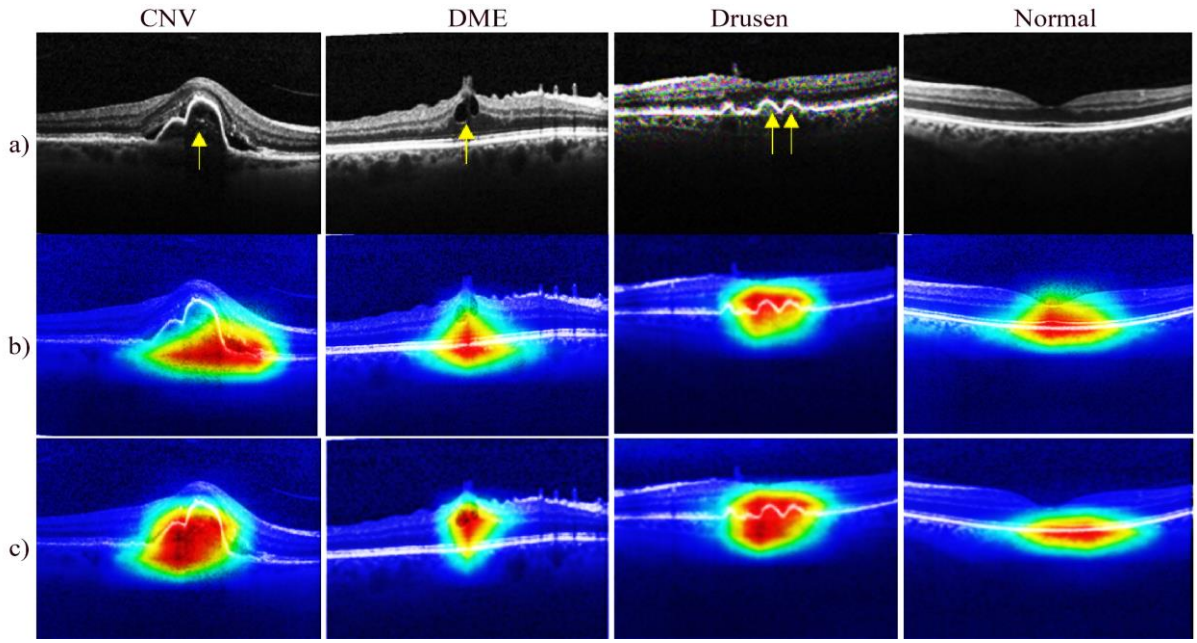


Fig. 8 Samples of a) Input image and Grad CAM representation of feature maps b) Limited data model c) Complete data model

Table 2. Comparison of retinal OCT image classification results

Year	Author	Task	Method	Dataset size	Accuracy
2022	Han et al. [9]	Retinal OCTDisease Classification	TL with CNN models	4749	87.40%
2022	Paima et al. [11]		CNN multiscale feature fusion strategy	120,961	93.40%
2022	Elaziz et al. [12]		Feature selection and Ensemble DL model	84,484	94.32%
2022	Liu et al. [14]		The attention classification model using DL	86,134	95.10%
2022	Minagi et al. [10]		CNN-based TL models	11,200	95.30%
2021	Tayal et al. [8]		CNN-based DL models	84,484	96.50%
2023	Kuntha Pin et al. [13]		Hybrid ensemble DL classifiers.	2,998	97.68%
	Proposed model		CNN-based feature extraction and Attention fusion	108,309	98.40%

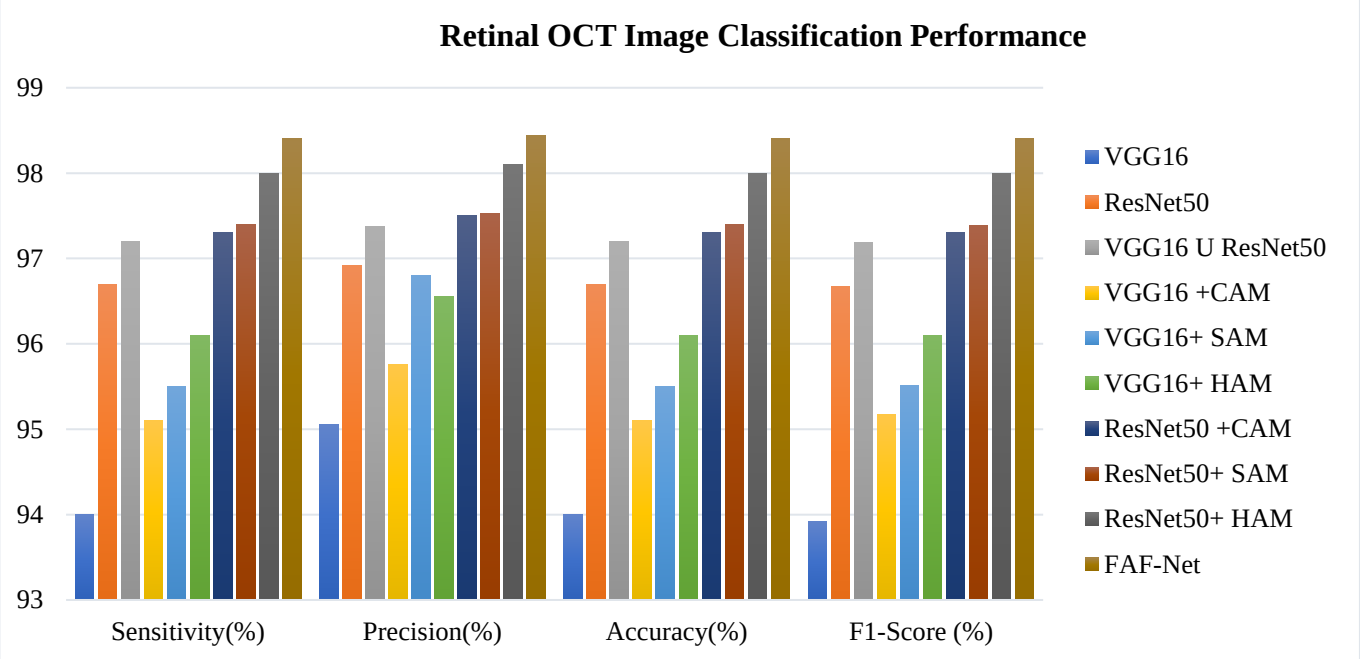


Fig. 9 The effect of additional components on classification performance

Figure 8 displays the visualisation map created by the Grad-CAM algorithm, illustrating how effectively the proposed model performs.This localisation map showcases the classification results of the proposed approach after attention and feature fusion on VGG16 and ResNet50. In OCT images of CNV, irregular retinal pigment epithelium (RPE) elevations can be observed due to damaged choroidal capillaries, interlayer effusion, lipid exudation, and other distinctive features. OCT images of DME depict cystic alterations around the macular fovea and edema of the retinal inner surfaces. As in OCT images, drusen is characterised by isolated RPE protrusions and enhanced choroid reflectance. In contrast, the OCT scans of the normal group reveal a well-organised and multilayered retina. This validation signifies that the GradCAM localisation map confirms the success of the proposed FAF-Net technique.

4.1. Comparison results

Most of the reviewed studies focused on TL for creating new models and integrating basic CNN models with machine learning algorithms. A direct performance comparison is impossible since each research study employed a different dataset to establish its effectiveness. Table 2 compares the FAF-Net OCT image classification approach with the most recent findings given in Section 1.

In contrast to the limitation of relying on a single CNN model, the proposed technique enhances features by combining two pre-trained CNNs with feature attention. The research has outperformed the above-mentioned studies, achieving an impressive accuracy rate of 98.40%. Adding the attention module resulted in a noticeable improvement in classification performance, and the model demonstrated further improvements by incorporating the feature fusion of networks. The limitation of our proposed research is the difficulty of obtaining a wide range of labelled medical image datasets, as the annotation of images demands extensive time and effort from experienced experts.

5. Ablation Studies

The effectiveness of the proposed method depends on three key factors: feature extraction, feature attention, and feature fusion. Our suggested technique improved the multi-CNN feature fusion's spatial and channel-wise discriminative performance. We performed ablation tests on the UCSD Dataset using ResNet50 and VGG16 as the baseline networks to show the performance of each component separately.

The base CNN architecture was methodically incorporated with each element, and the models were trained with the same settings as the proposed model to assess the impact of these added modules. Performance measures were then calculated to assess the impact of these additional modules, as shown in Figure 9. The commonly used image processing techniques produce poor outcomes due to similarities in the retinal OCT-based lesions' images. The basic CNN architecture failed to achieve satisfactory classification performance. However, the integration of the attention module improved the classification performance. The model with the feature fusion showed enhancements in performance, while the models with the feature attention fusion showed further enhancements in image classification.

We compared the models VGG16, ResNet50, VGG16 U ResNet50 (Feature Fusion of VGG16 and ResNet50), VGG16+CAM (VGG16 features modified with channel attention module), VGG16+SAM (VGG16 with spatial attention module), VGG16+(HAM) (VGG16 with channel attention and spatial attention fusion - hybrid attention). Then ResNet50+CAM (ResNet50 features modified with Channel Attention Module), ResNet50+SAM (ResNet50 with spatial Attention Module), ResNet50+(HAM) (ResNet50 with hybrid attention) and the proposed model FAF-Network in this section.

The conclusions from Figure 9 are listed below.

- The feature fusion of two CNN models can boost classification accuracy compared to the baseline models. Deep feature fusion, VGG16 U ResNet50, outperforms the baseline models VGG16 and ResNet50, with an accuracy of 97.20% as opposed to 94.00% and 96.70%, respectively. Since both the baseline networks can learn features and generalise well, feature fusion obtains the information simultaneously, resulting in better performance.
- The attention mechanism has great potential to enhance the performance of CNNs. In addition to providing where to focus, it strengthens the feature representation of interests. VGG16+CAM (accuracy = 95.10), VGG16+SAM(accuracy = 95.50), ResNet50+CAM (accuracy =97.30) and ResNet50+SAM (accuracy =97.40) have better outcomes compared to their baseline models VGG16 and ResNet50. Combining channel attention with spatial attention, VGG16+(HAM) and ResNet50+(HAM) achieved remarkable performance (accuracies of 96.10 and 98.00, respectively) in their tasks.
- The proposed FAF-Net provides the highest classification performance with an accuracy of 98.40%. It offers a fusion technique that uses VGG16+(HAM) and ResNet50+(HAM) feature maps based on hybrid attention, capturing more precise semantic areas.

6. Conclusion

The FAF-Net is presented in this study, serving as a feature extraction and attention fusion network based on CNNs for identifying macular diseases from OCT images. FAF-Net combines deep pre-trained VGG16 and ResNet50 models with attention networks to increase the classification accuracy of macular diseases. According to the findings, the proposed method performs better than standalone classifiers.

The classification effectiveness of the suggested technique was assessed using the UCSD dataset, revealing that macular abnormalities could be detected with a remarkable accuracy of 98.4% (the accuracy increased by 4.4% compared to VGG16 and 1.7% to Resnet50). This approach can significantly improve macular disease screening applications by enabling rapid and accurate disease identification. A limitation of this approach is model training requires large and varied datasets. The OCT picture quality and the types of diseases under study further influence the algorithm's effectiveness. Further research using large and diverse datasets might enhance the accuracy and the generalizability of the proposed methodology, FAF-Net, for applications involving the classification of OCT images.

References

[1] Undurti N. Das, "Diabetic Macular Edema, Retinopathy and Age-Related Macular Degeneration as Inflammatory Conditions," *Archives of Medical Science*, vol. 12, no. 5, pp. 1142-1157, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[2] Zongqing Ma et al., "HCTNet: A Hybrid ConvNet-Transformer Network for Retinal Optical Coherence Tomography Image Classification," *Biosensors*, vol. 12, no. 7, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Nithya Rajagopalan et al., "Retracted Article: Deep CNN Framework for Retinal Disease Diagnosis Using Optical Coherence Tomography Images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 7, pp. 7569-7580, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] David Huang et al., "Optical Coherence Tomography," *Science*, vol. 254, no. 5035, pp. 1178-1181, 1991. [CrossRef] [Google Scholar] [Publisher Link]

[5] Meng Wang et al., "Semi-Supervised Capsule cGAN for Speckle Noise Reduction in Retinal OCT Images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 4, pp. 1168-1183, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Reena Chopra, Siegfried K. Wagner, and Pearse A. Keane, "Optical Coherence Tomography in the 2020s-Outside the Eye Clinic," *Eye*, vol. 35, no. 1, pp. 236-243, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Prakash Kumar Karn, and Waleed H. Abdulla, "On Machine Learning in Clinical Interpretation of Retinal Diseases Using OCT Images," *Bioengineering*, vol. 10, no. 4, pp. 1-24, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Akash Tayal et al., "DL-CNN-Based Approach with Image Processing Techniques for Diagnosis of Retinal Diseases," *Multimedia Systems*, vol. 28, no. 4, pp. 1417-1438, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jinyoung Han et al., "Classifying Neovascular Age-Related Macular Degeneration with a Deep Convolutional Neural Network based on Optical Coherence Tomography Images," *Scientific Reports*, vol. 12, no. 2232, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Saman Sotoudeh-Paima et al., "Multiscale Convolutional Neural Network for Automated AMD Classification Using Retinal OCT Images," *Computers in Biology and Medicine*, vol. 144, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Akinori Minagi, Hokuto Hirano, and Kauzhiro Takemoto, "Natural Images Allow Universal Adversarial Attacks on Medical Image Classification using Deep Neural Networks with Transfer Learning," *Journal of Imaging*, vol. 8, no. 2, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Mohamed Abd Elaziz et al., "Medical Image Classification Utilising Ensemble Learning and Levy Flight-Based Honey Badger Algorithm on 6G-Enabled Internet of Things," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Kuntha Pin, Jung Woo Han, and Yunyoung Nam, "Retinal Diseases Classification based on Hybrid Ensemble Deep Learning and Optical Coherence Tomography Images," *Electronic Research Archive*, vol. 31, no. 8, pp. 4843-4861, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Xiaoming Liu et al., "Joint Disease Classification and Lesion Segmentation Via One-Stage Attention-Based Convolutional Neural Network in OCT Images," *Biomedical Signal Processing and Control*, vol. 71, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Karen Simonyan, and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ArXiv*, pp. 1-14, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[16] Shagun Sharma, and Kalpna Guleria, "A Deep Learning Based Model for the Detection of Pneumonia from Chest X-ray Images using VGG-16 and Neural Networks," *Procedia Computer Science*, vol. 218, pp. 357-366, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Wilson Bakasa, and Serestina Viriri, "VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction," *Journal of Imaging*, vol. 9, no. 7, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Samuel Kumaresan et al., "Deep Learning-Based Weld Defect Classification using VGG16 Transfer Learning Adaptive Fine-Tuning," *International Journal on Interactive Design and Manufacturing*, vol. 17, pp. 2999-3010, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Kaiming He et al., "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[20] Mehdhar S.A.M. Al-Gaashani et al., "Using a Resnet50 with a Kernel Attention Mechanism for Rice Disease Diagnosis," *Life*, vol. 13, no. 6, pp. 1277-1277, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Hareem Kibriya, and Rashid Amin, "A Residual Network-Based Framework for COVID-19 Detection from CXR Images," *Neural Computing and Applications*, vol. 35, pp. 8503-8516, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Muhammad Haris Abid et al., "Multi-Modal Medical Image Classification using Deep Residual Network and Genetic Algorithm," *PLOS ONE*, vol. 18, no. 6, pp. 1-24, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Chun-Ling Lin, and Kun-Chi Wu, "Development of Revised ResNet-50 for Diabetic Retinopathy Detection," *BMC Bioinformatics*, vol. 24, no. 157, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-Excitation Networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7132-7141, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[25] Daniel Kermany, Kang Zhang, and Michael Goldbaum, "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images," *Mendeley Data*, vol. 172, no. 5, pp. 1122-1131, 2018. [CrossRef] [Google Scholar] [Publisher Link]