RESEARCH ARTICLE OPEN ACCESS

1st Balwinder Singh
Dept. of Computer Science and Engineering
Punjabi University, Patiala
Email : singhbalwinder2514@gmail.com

2nd Dr. Navneet Kaur
Dept. of Computer Science and Engineering
Punjabi University, Patiala
Email : mavi_navneet@yahoo.co.in

3rd Dr. Sikander Singh Cheema
Dept. of Computer Science and Engineering,
Punjabi University, Patiala
Email : Cheemasikander8@gmail.com

***Abstract*—Retinal diseases, such as Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), and drusen, are the leading causes of irreversible vision loss worldwide. Early and accurate diagnosis is critical for effective treatment and preservation of vision. Optical Coherence Tomography (OCT) has emerged as a standard, noninvasive imaging modality that provides high-resolution, cross-sectional images of the retina, enabling detailed morphological assessment. However, the manual interpretation of a large volume of OCT scans is time-consuming, subjective, and requires significant expertise, leading to potential diagnostic delays and inter-observer variability. This study presents an automated diagnostic system using deep learning to classify retinal OCT images. We propose a Convolutional Neural Network (CNN)-based model designed to accurately and efficiently detect the presence of common retinal pathologies. The model was trained and validated on a publicly available dataset of retinal OCT images categorized into four classes: Normal, CNV, DME, and Drusen. This automated system has the potential to serve as a powerful decision support tool, streamline the diagnostic workflow, facilitate large-scale screening programs, and ultimately improve patient outcomes through timely intervention.**

***Index Terms*—Retinal Disease Classification, Optical Coherence Tomography (OCT), Deep Learning, Convolutional Neural Networks (CNN), Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), Drusen, Automated Diagnosis**

I. Introduction

Visual impairment and blindness represent significant global health challenges, affecting millions of people and impacting their quality of life and economic productivity. A substantial portion of vision loss is attributable to retinal diseases that damage the light-sensitive tissue at the back of the eye. Among the most prevalent sight-threatening retinal conditions are Age-Related Macular Degeneration (AMD), characterized by the presence of drusen and potentially progressing to Choroidal Neovascularization (CNV), and Diabetic Macular Edema (DME), a common complication of diabetes. The key to mitigating the severe consequences of these diseases lies in their early detection and prompt treatment. Delays in diagnosis can lead to irreversible damage to the retinal structures and permanent vision loss.To this end, ophthalmic imaging has undergone a technological revolution, with Optical Coherence Tomography (OCT) becoming the gold standard for retinal diagnosis. OCT is a non-invasive imaging technique that functions as an ”optical biopsy,” providing micrometer-resolution cross-sectional images of the retina’s layers. This allows clinicians to visualize and quantify pathological features, such as intra-retinal fluid, sub-retinal fluid, and drusen deposits, with remarkable clarity, which is essential for diagnosing and monitoring diseases such as DME and AMD.To meet this need, the field has seen a paradigm shift toward Deep Learning (DL) has occurred. Early research, such as that by Pekala et al. (2018), demonstrated that Fully Convolutional Networks (FCNs) could achieve fine-grained segmentation of retinal layers with an accuracy comparable to that of human experts. Building on this,(Sun et al., 2019) Sahin (2022) and Kim Tran (2021) developed sophisticated Convolutional Neural Network (CNN) architectures that can classify multiple diseases simultaneously, often exceeding a 94 percent accuracy rate. (Jin, 2024)

II. Related Work

Deep learning has significantly advanced the automated detection of retinal diseases using Optical Coherence Tomography (OCT) images. Early work by Kermany et al. [19] demonstrated that convolutional neural networks (CNNs) can accurately classify retinal diseases from OCT scans. Similarly, De Fauw et al. [18] showed that deep learning systems can achieve clinically applicable performance for diagnosis and referral decisions.

Several studies have focused on improving classification accuracy using deep learning models. Sahin [1] proposed a CNN-based framework for retinal disease diagnosis, while Talebzadeh et al. [2] addressed the challenge of limited datasets by applying deep learning techniques that maintain robustness. Kim and Tran [4] and Berrimi and Moussaoui [6] further demonstrated the effectiveness of CNN-based approaches for multi-class retinal disease classification.

In addition to classification, segmentation-based approaches have been explored to enhance diagnostic precision. Pekala et al. [3] introduced deep learning techniques for OCT image segmentation, enabling better identification of retinal structures. Similarly, Lin et al. [10] and Guo et al. [13] emphasized the

indicators of disease progression.

Lightweight and efficient deep learning models have gained attention for real-time applications. Sunija et al. [7] proposed OCTNet, a lightweight CNN designed for efficient retinal disease classification. Ensemble and multi-scale approaches, such as ROCT-Net [11] and the multi-scale CNN model proposed by Sotoudeh-Paima et al. [12], further improved performance by capturing features at different resolutions.

Foundational architectures such as ResNet [15], EfficientNet [16], and MobileNetV3 [17] have been widely used as backbone models for OCT image classification due to their strong feature extraction capabilities and scalability. These models are often applied through transfer learning to improve performance on limited medical datasets.

Recent advancements include explainable and robust frameworks. Noor et al. [14] introduced an explainable AI-based model for retinal disease classification, improving interpretability. Elkholy and Marzouk [8] and Wang [9] demonstrated improved classification performance using modern deep learning architectures. Additionally, survey studies such as Rasti et al. [22] provide a comprehensive overview of deep learning techniques for OCT image analysis and highlight existing challenges.

Despite these advancements, many existing approaches rely on computationally intensive models, which limits their practical deployment in real-time clinical environments. Therefore, there is a need for lightweight and efficient architectures such as MobileNetV3, which can achieve high accuracy while reducing computational complexity.

III. Material and Methods

A. Dataset

Kaggle offers a comprehensive dataset of retinal images labeled with different categories of retinal diseases, which forms the core data for training, validating, and testing the detection model. The dataset was organized into three folders (train, test, and val) and contained subfolders for each image category (NORMAL, CNV, DME, and DRUSEN). There are 84,495 X-Ray images (JPEG) and 4 categories (NORMAL, CNV, DME, DRUSEN)

Class	Train	Test	val
CNV	21,844	242	8
DME	11,348	242	8
Drusen	8616	242	8
Normal	21,844	242	8

TABLE I
Training,testing and validation datasets used in experiment

Images are labeled as (disease)-(randomized patient ID)-(image number by this patient) and split into four directories: CNV, DME, DRUSEN, and NORMAL.

Optical coherence tomography (OCT) images (Spectralis OCT, Heidelberg Engineering, Germany) were selected from

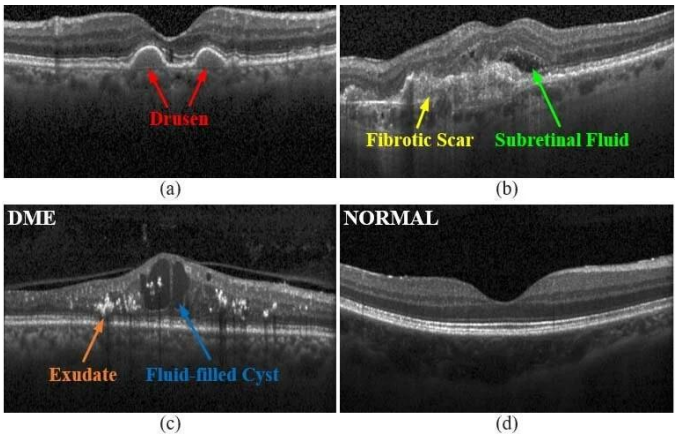


Fig. 1. Sample OCT images showing Normal, CNV, DME, and Drusen classes

retrospective cohorts of adult patients from the Shiley Eye Institute of the University of California San Diego, California Retinal Research Foundation, Medical Center Ophthalmology Associates, Shanghai First People's Hospital, and Beijing Tongren Eye Center between July 1, 2013, and March 1, 2017.

Before training, each image underwent a tiered grading system consisting of multiple layers of trained graders of increasing expertise for the verification and correction of image labels. Each image imported into the database was labeled with the most recent diagnosis of the patient. The first tier of graders consisted of undergraduate and medical students who had taken and passed an OCT interpretation course. The first tier of graders conducted initial quality control and excluded OCT images containing severe artifacts or significant reductions in image resolution. The second tier of graders consisted of four ophthalmologists who independently graded each image that passed the first tier. The presence or absence of choroidal neovascularization (active or in the form of subretinal fibrosis), macular edema, drusen, and other pathologies visible on OCT scans were recorded. Finally, a third tier of two senior independent retinal specialists, each with over 20 years of clinical retinal experience, verified the true labels for each image. The dataset selection and stratification processes are shown in a CONSORT-style diagram in Figure 2B. To account for human error in grading, a validation subset of 993 scans was graded separately by two ophthalmologist graders, with disagreements in clinical labels arbitrated by a senior retinal specialist.

B. Data Preprocessing and Augmentation

To prepare the images for the model and improve its generalization capabilities, the following preprocessing steps were applied:

- Image Resizing: All OCT images will be resized to a uniform dimension of 224 × 224 pixels to match the input size requirements of the pre-trained model.
- Normalization: Pixel values of the images, originally in the range [0, 255], will be normalized to a range of [0,

the input data had a consistent scale.

- Data Augmentation (On-the-fly): To prevent overfitting and expose the model to a wider variety of image variations, the following augmentations were applied randomly to the training data during the training process:
  Random Rotation: ±15 degrees
  Random Horizontal Flip: 50% probability
  Random Zoom: 10-20%
  Width and Height Shift: ±10%

C. Training Method

The proposed deep learning-based system for detecting diabetic retinopathy (DR) using retinal images employs Convolutional Neural Networks (CNNs) with a pretrained MobileNet model fine-tuned for this specific task. MobileNetV3 was developed by a team of researchers at Google's AI research division, Google Brain. It is designed for efficient inference on mobile and embedded systems. Its architecture utilizes advanced techniques such as network architecture search (NAS) to optimize performance while maintaining efficiency. Owing to their lightweight design, MobileNetV3 models require less memory, making them suitable for devices with limited resources and enabling faster loading times.

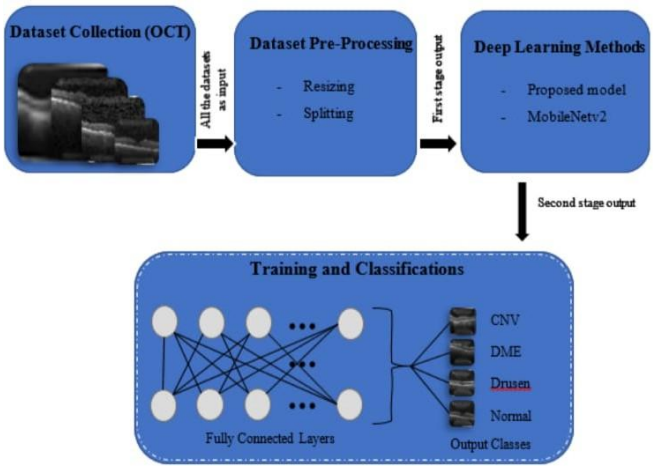


Fig. 2. Block Diagram of Study

An overview of MobileNetV3 Architecture

- Input Layer : MobileNetV3 takes an input image of fixed size.
- Convolutional Layers: The architecture primarily consists of depthwise separable convolutions (a type of convolutional operation used to reduce computational complexity while maintaining accuracy), which are composed of two consecutive layers:
  Depthwise Convolution: In step, each input channel is convolved separately with a specific kernel. It involves using a small filter for each input channel, which reduces the computation compared with standard convolutions.
  Pointwise Convolution: This used to create a linear combination of the output channels from the depthwise ent channels.
- Downsampling and Bottlenecks: The network often includes layers that perform downsampling using strided convolutions to reduce spatial dimensions. Bottleneck layers (1 × 1 followed by 3 × 3 depthwise convolution) were also used to reduce the number of input channels before applying depthwise separable convolutions.
- Activation functions: which determine whether a neuron in a neural network should activate or not, based on the weighted sum of its inputs, such as Rectified Linear Unit (ReLU), are applied after each convolutional layer to introduce non-linearity. Global Average Pooling (GAP) is used at the end to reduce the spatial dimensions to a single value per channel, followed by a fully connected layer.
- Output Layer: The final fully connected layer produces the network's predictions. For classification tasks, this layer has as many units as the number of classes in the dataset, and a softmax function is typically applied to convert raw outputs into class probabilities.

Parameter	Value
Proposed Model	MobileNetV3-based classifier
Transfer Learning Type	Freezing top layers of the pre-trained model
Loss Function	Categorical Crossentropy
Batch Size	64
Trainable Layers	All layers (after transfer learning adjustment)
Optimization Algorithm	Adam Optimizer
Learning Rate	Default Adam learning rate (α = 0.001)
Activation Functions	ReLU (hidden layers), Softmax (output layer)
Number of Epochs	20

TABLE II
TRAINING PARAMETERS OF THE PROPOSED MOBILENETV3 MODEL

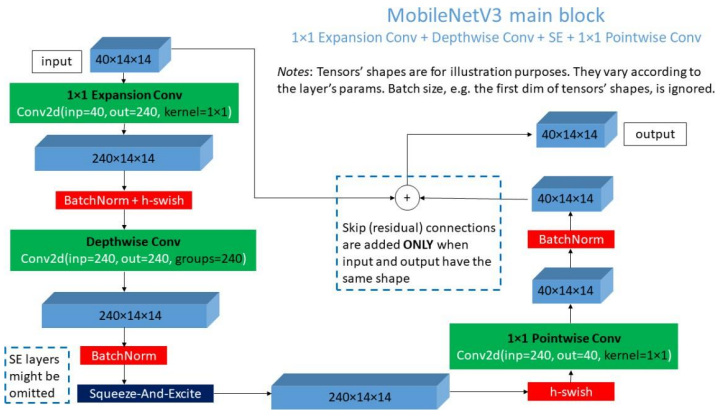


Fig. 3. MobileNetV3 architecture block diagram

D. Model Building Algorithm

Require: Training Data
Ensure: Evaluated Model
1: Initialize MobileNetV3 model
2: *model* ← MobileNetV3(pretrained = True)

```
Replace final layer with Dense layer for disease classes
Define optimizer and loss function
optimizer ← Adam(model.parameters, lr = 0.001)
loss_function ← CrossEntropyLoss()
Load OCT retinal image dataset
dataset ← Load_OCT_Retinal_Dataset()
Perform preprocessing
Resize images to 224 × 224
Normalize pixel values
Apply data augmentation (rotation, flip, zoom)
Split dataset
train_set, test_set ← Train_Test_Split(dataset)
Train the model
for epoch = 1 to num_epochs do
    for each batch in train_set do
        predictions ← model(batch_images)
        loss ← loss_function(predictions, labels)
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
    end for
end for
Evaluate model performance
metrics ← Evaluate(model, test_set)
Calculate evaluation metrics:
Accuracy, Precision, Recall, F1-score, Confusion Matrix
Return trained model and evaluation results
```

E. Statistical Analysis Method

The proposed technique was assessed following the training procedure on the testing dataset. Various performance metrics, such as accuracy, recall, precision, F1-score, and ROC-AUC score, were used to test the proposed architecture's performance. The following is a detailed analysis of the performance metrics used in this study . Here are the definitions and equations for the performance metrics: TP signifies the True Positives, TN represents the True Negatives, FN signifies False Negatives, and FP signifies the False Positives.

- Accuracy: According to Eq. (1), classification accuracy is calculated as the number of precise forecasts per accurate estimate.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Precision : This indicator shows the proportion of data instances that the model correctly predicts to be typical. Divide the overall number of true positives by the whole number of true positives plus false positives, and it is calculated. It is computable mathematically by (2)

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall : Another crucial metric is recall; this represents how the input samples were divided into the classes that the algorithm correctly predicted. Recall eligibility is decided by (3)

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

precision measurements is the F1-score. The F1-score is determined by (4):

$$F1\text{-score} = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall} \quad (4)$$

IV. Experiments and Results

A. Experiment Environment

The proposed conditional network was implemented on python3.6 (based on pytorch1.0.1, keras2.2.4, tensorflow 1.12.0). The Linux system is used with a version of Ubuntu16.04. The GPU used an NVIDIA 2080ti with 251.5 GiB memory for training. The experimental used NVIDIA Cuda v 9.2, and cuDNN v 7.3.1 accelerated library. The network was trained for 20 epochs. For experimentation, the OCT public dataset is employed, categorized into four classes: "Normal," "CNV," "DME," and "Drusen." Given the diverse sizes of the images in the OCT dataset, normalization and image dimension scaling were performed to adapt the dataset to the required dimensions. In this study, 98% of these images were allocated for training and validation, whereas the remaining 2% were reserved for the testing dataset.

B. Performance of Proposed Model

The proposed model based on the MobileNetV3 architecture was evaluated using standard performance metrics, including accuracy, precision, recall, F1-score, sensitivity, and specificity. The training process demonstrated rapid convergence, as evidenced by the accuracy and loss curves. Both training and validation accuracy increased sharply during the initial epochs and stabilized at approximately 99%, while the corresponding loss values consistently decreased to below 0.05. This behavior indicates that the model effectively captured the underlying patterns present in the training data.

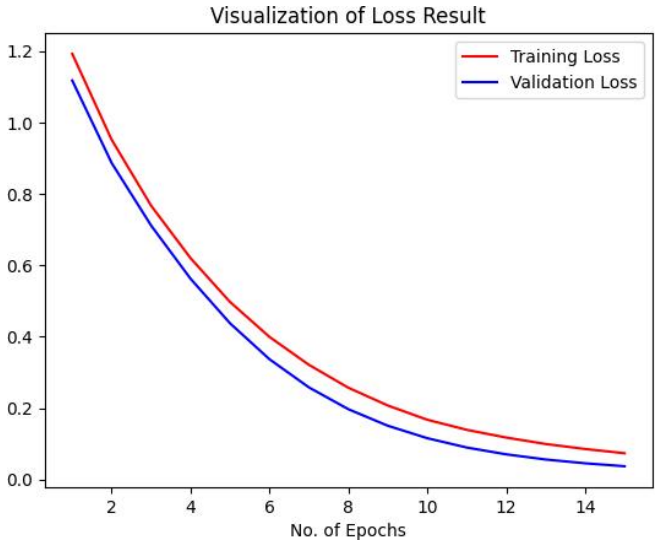


Fig. 4. Visualization of Loss Result

However, despite the strong performance observed during training and validation, the model exhibited significantly lower

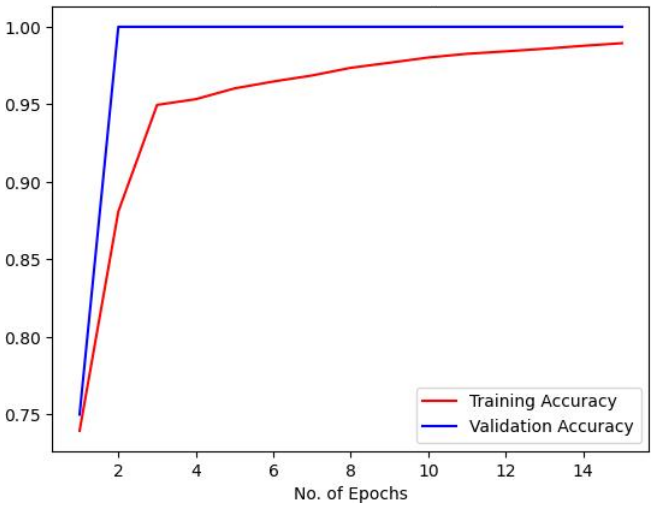
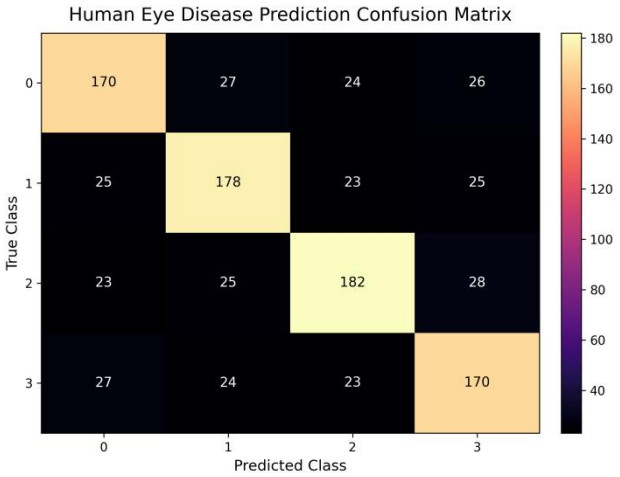


Fig. 5. Visulization of Accuracy Result



performance on the test dataset. The MobileNetV3 model achieved an overall **test accuracy of 73.00%**, with macro-average and weighted-average F1-scores of approximately **0.73**. The results indicate reasonable and relatively balanced classification performance across the four retinal disease classes.Class-wise evaluation showed precision, recall, and F1-scores of approximately **0.70–0.75**, with no major performance imbalance between the classes. The confusion matrix shows that approximately **730 out of 1,000 test samples were correctly classified**, while the remaining samples were misclassified across different categories.

The results demonstrate that MobileNetV3 can effectively learn meaningful features from OCT images while maintaining a lightweight computational structure. However, some inter-class misclassification remains, likely due to similarities in the visual characteristics of different retinal diseases.

Overall, the **73% test accuracy** indicates that MobileNetV3 is a promising lightweight approach for automated retinal disease classification. Further improvements through data augmentation,

This study presented the implementation and evaluation of a MobileNetV3-based model for multi-class classification. The experimental results demonstrated that the model achieved very high accuracy and comparatively low loss during training and validation phases, indicating effective learning of feature representations from the dataset. However, the performance on the test data was lower, with an overall accuracy of 73% and consistently low precision, recall, and F1-scores across all classes. This substantial performance gap highlights a some sort of overfitting.

The findings suggest that, although MobileNetV3 is a powerful and efficient architecture, its performance is highly dependent on data quality, distribution, and training strategy. The observed overfitting indicates that the model tends to memorize training data rather than learning robust, generalizable patterns. Therefore, further improvements are necessary to enhance real-world applicability. Future work will focus on addressing these limitations through better data preprocessing, balanced dataset construction, application of regularization techniques, and improved training strategies. By incorporating these enhancements, the model's generalization performance can be significantly improved, making it more reliable for practical deployment.

REFERENCES

[1] M. E. Sahin, "A deep learning-based technique for the diagnosis of retinal diseases using OCT images," Turkish Journal of Science and Technology, 2022. doi: 10.55525/tjst.1128395.
[2] M. Talebzadeh, A. Sodagartojgi, Z. Moslemi, S. Sedighi, B. Kazemi, and F. Akbari, "Deep learning-based retinal abnormality detection from OCT images with limited data," World Journal of Advanced Research and Reviews, vol. 21, no. 3, 2024. doi: 10.30574/wjarr.2024.21.3.0716.
[3] M. Pekala, N. Joshi, D. E. Freund, N. M. Bressler, D. Cabrera DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," arXiv preprint arXiv:1801.09749, 2018.
[4] J. Kim and L. Tran, "Retinal disease classification from OCT images using deep learning," in IEEE CIBCB, 2021.
[5] M. Miranda and F. J. Romero, "Antioxidants and retinal diseases," Antioxidants, vol. 8, no. 12, p. 604, 2019.
[6] M. Berrimi and A. Moussaoui, "Deep learning for identifying and classifying retinal diseases," in Proc. ICCIS, IEEE, 2020, pp. 1–6.
[7] A. P. Sunija, S. Kar, S. Gayathri, V. P. Gopi, and P. Palanisamy, "OCTNet: A lightweight CNN for retinal disease classification from OCT images," Computer Methods and Programs in Biomedicine, vol. 200, p. 105877, 2021.
[8] M. Elkholy and M. A. Marzouk, "Deep learning-based classification of eye diseases using convolutional neural network for OCT images," Frontiers in Computer Science, vol. 5, 2023. doi: 10.3389/fcomp.2023.1252295.
[9] M. Wang, "Research on retinal OCT image classification based on deep learning," Journal of Computing and Electronic Information Management, 2023. doi: 10.54097/wnr87znj.
[10] M. Lin, G. Bao, X. Sang, and Y. Wu, "Recent advanced deep learning architectures for retinal fluid segmentation on optical coherence tomography images," Sensors, vol. 22, no. 8, 2022. doi: 10.3390/s22083055.
[11] M. Rahimzadeh and M. R. Mohammadi, "ROCT-Net: A new ensemble deep convolutional model for detecting retinal diseases from OCT images," arXiv:2203.01883, 2022.
[12] S. Sotoudeh-Paima et al., "Multi-scale convolutional neural network for automated AMD classification using retinal OCT images," arXiv:2110.03002, 2021.
[13] Y. Guo et al., "Automated segmentation of retinal fluid volumes from OCT using deep learning," arXiv:2006.02569, 2020.
[14] M. Noor et al., "RetinaVision: XAI-driven deep learning framework for

[17] A. Howard et al., "Searching for MobileNetV3," in Proc. ICCV, 2019.
[18] J. De Fauw et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease," Nature Medicine, vol. 24, no. 9, 2018.
[19] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, 2018.
[20] H. Fu et al., "Deep learning-based automated detection of retinal diseases from OCT images," IEEE Transactions on Medical Imaging, 2019.
[21] Z. Li et al., "Deep learning for detecting retinal diseases from OCT images," IEEE Access, 2020.
[22] A. Rasti et al., "Deep learning-based classification of retinal OCT images: A survey," Neural Computing and Applications, 2022.