International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I8P109

Volume 74 Issue 8, 136-153, August 2026

Original Article

HybridFONet: Hybrid Feature-based Optimized CNN Algorithm for Precise Ocular Disease Classification from JPG images via CNN and Pre-trained Deep Learning Models

Gurpreet Kaur[1], Amit Kumar Bindal[2]

[1,2]*Department of Computer Science Engineering, M.M. Engineering College, MM(DU), Mullana, Ambala, Haryana, India.*

[1]*Corresponding Author : guri394.kaur@gmail.com*

Received: 02 December 2025 Revised: 10 June 2026 Accepted: 24 June 2026 Published: 29 August 2026

Abstract - *Ocular diseases, including diabetic retinopathy, cataract, glaucoma, age-related macular degeneration, hypertension-related retinal changes, myopia, and other retinal abnormalities, are major contributors to visual impairment and blindness. Manual interpretation of fundus images is reliable but time-consuming and dependent on specialist availability, making automated deep-learning-based screening important for large-scale eye-care support. This study proposes HybridFONet, a hybrid feature-based optimized CNN framework for precise multi-class ocular disease classification from JPG fundus images. The proposed method uses the ODIR-5K fundus-image database and classifies images into eight categories: Normal, Diabetes, Cataract, Glaucoma, Age-related Macular Degeneration, Myopia, Hypertension, and other abnormalities. The workflow includes image preprocessing, resizing, grayscale conversion, Butterworth band-pass filtering, data augmentation, VGG16-based transfer-feature extraction, Crow Search Optimization-based feature selection, and final classification using a one-dimensional CNN with softmax output. Experimental results show that HybridFONet achieved 96.47% accuracy, 97.29% precision, 95.58% recall, 96.32% F1-score, and a 3.53% error rate. Comparative analysis demonstrates that the proposed model outperformed VGG16, Improved AlexNet, VGG19, ResNet50, and ResNet152V2. These findings confirm that optimized deep-feature selection combined with CNN classification improves recognition accuracy, reduces the influence of redundant features, and supports reliable automated ocular-disease screening.*

Keywords *- Ocular disease, Crow Search Optimization, Convolutional Neural Network, Transfer Learning, Retinal imaging, Deep Learning.*

1. Introduction

Diabetic retinopathy, age-related macular degeneration, glaucoma, cataract, and other retinal diseases are the leading causes of low vision and blindness in the world. Early and accurate diagnosis is important, as many eye diseases are asymptomatic and can lead to permanent vision loss if not diagnosed early. The growing number of diabetic eye complications has further added to the need for dependable ocular screening systems. Diabetic retinopathy has been shown to be a global burden, and there is a need for scalable diagnostic methods to enable timely treatment and disease management [1]. Fundus photography and optical coherence tomography are important imaging techniques in ophthalmology. Optical coherence tomography can offer detailed information about the structure of the retina and optic nerve, and has great promise for current and future use in ophthalmology [2]. But, the JPG-based fundus images are still being used in many clinical and screening settings due to their cost-effectiveness, ease of acquisition, and suitability for automated image analysis. These images give valuable visual data for identification of changes in the retina, optic disc, blood vessels and macular area associated with disease. Manual examination by ophthalmologists is considered reliable, but it is time-consuming and requires the presence of experts. Manual diagnosis can be influenced by workload, fatigue, subjective judgment and inter-observer variability in large-scale screening programs. Hence, DL methods, particularly Convolutional Neural Networks (CNNs), have become popular for automated classification of ocular diseases. DL methods have been recently proven to be effective for diabetic retinopathy detection [3] and glaucoma detection and progression prediction [4]. However, the accurate multi-class ocular disease classification from JPEG fundus images is still a challenge because of the presence of image noise, class imbalance, similarity of diseases, and variation in image quality. Despite the several deep learning approaches proposed for ocular disease recognition, there are still some limitations in the current approaches, such as handling class imbalance, optimized feature extraction, generalization of the model, and computational efficiency. Khan et al. [5] presented a deep learning approach for the recognition of ocular disease with inner-class balance, which did not fully exploit hybrid feature optimization for better feature representation of the diseases. Abbas et al. [6] suggested an enhanced transfer learning model with self-attention and dense layers, but these models may require intricate model designs and hyperparameter tuning. Ryan et al. [7] assessed various pre-trained CNN models for ocular disease diagnosis, with the emphasis primarily being on comparative performance and not on an optimized hybrid CNN framework. So, a strong and optimized model is still required to fuse CNN-based feature learning, pre-trained deep features, preprocessing, class balancing and hyperparameter optimization for accurate classification of OD from the JPG fundus images. Although DL has made significant strides in medical image analysis, it is still challenging to accurately classify multiple diseases of the eye from JPG fundus images. Ocular image datasets are frequently characterized by imbalanced classes, noisy samples, variations in illumination, low contrast regions, and high similarity between different disease patterns. This can cause CNN-based models to underperform and potentially overfit, misclassify and fail to generalize to new images.

Many of the models that are available are based on a fixed CNN architecture or direct transfer learning without proper optimization of important parameters like learning rate, kernel size, number of filters, and the representation of features. Therefore, the classification performance can be different depending on data sets quality and training conditions. Hence, the primary issue that is addressed in this research is the growth of an optimized hybrid DL framework that can accurately classify the ocular diseases from the JPG fundus images with minimal effect of class imbalance, redundant features, noise, and suboptimal hyperparameter selection.

The research motivation is to create an automated, accurate and efficient system for ocular disease classification, which can help in initial recognition and aid the ophthalmologist in large-scale screening. Automated deep learning-based systems can be of great importance in improving eye-care services, particularly in regions where specialists are scarce, as timely detection can help minimize the risk of severe vision loss.

JPGs of the eyes can be used to create practical and cost-effective diagnostic models. To achieve high accuracy, however, a model that can extract meaningful disease features, handle class imbalance, reduce irrelevant information and optimize learning parameters is required. This encourages the design of HybridFONet, a hybrid feature-based optimized CNN algorithm that combines CNN learning, pre-trained deep learning models, pre-processing, class balancing and optimization techniques for accurate classification of ocular diseases.

The contributions of this study are as follows:

- The proposed HybridFONet model combines CNN-based feature learning with pre-trained DL models for improved OD classification.
- The model applies optimized feature extraction to improve the representation of disease-specific patterns in ocular images.
- Crow Search Optimization is used to optimize important CNN hyperparameters such as learning rate, kernel size, and number of filters.
- The proposed framework includes preprocessing techniques such as image enhancement, noise filtering, augmentation, and class balancing to improve model robustness.
- The system is designed for multi-class ocular disease classification rather than focusing on a single disease category.
- The proposed model aims to reduce overfitting and improve generalization performance on unseen JPG ocular images.
- The performance of HybridFONet is compared with recent deep learning-based ocular disease classification studies.

The further paper is structured as follows: Section 2 presents a review of prior research on ocular disease classification, CNN-based medical image analysis, transfer learning and optimization-based deep learning methods. Section 3 shows the proposed HybridFONet methodology, which involves the preprocessing of the dataset, extraction of features, CNN architecture, integration of the pre-trained model and Crow Search Optimization. The experimental setup, evaluation metrics, implementation details, simulation results, and comparative analysis are presented in Section 4. The paper is concluded in Section 5, with future directions.

2. Literature Review

This section defines recent hybrid deep learning approaches for ocular disease detection using fundus images. It highlights the models, datasets, key findings, limitations, and future directions reported in existing studies, with emphasis on CNN, Transformer, ensemble, and quantum-based methods.

2.1. Existing Summary of Literature Review

Al-Fahdawi et al. (2024) [8] suggested a multi-label deep learning framework, Fundus-DeepNet, that combines hybrid multi-label learning with deep learning. The proposed method involved cropping the image in a circle, resizing, enhancing the contrast, adding noise, subtracting noise, and finally extracting features using HRNet and an attention block. The left-eye and right-eye features were fused using a SENet, and final classification was done using a Deep Restricted Boltzmann Machine with Softmax. The main data set utilized was OIA-ODIR, which included eight categories of ocular diseases. The key finding was the final score was 92.66% on-site and 92.41% off-site, with good F1, Kappa and AUC values. The limitation was that the data was confidential. This fusion design should be validated on public JPG fundus datasets with explainability in the future. Rieck et al. (2025) [9] suggested an open-access Transformer-CNN hybrid network for general detection of ocular diseases from color fundus images. The approach used a CNN-based feature extraction and a Transformer-based global contextual modeling for the classification of healthy images and nine classes of ocular diseases. The key parameters/tools used were PyTorch 2.5.0, Python 3.11.7, CUDA, AdamW/SGD optimization, transfer learning, fine-tuning, Optuna hyperparameter search, and early stopping and stratified five-fold cross-validation. A total of 5335 fundus images were collected from two hospitals in Bangladesh. The key finding was an average accuracy of 76.40%, balanced accuracy of 81.91% and F1-score of 76.65%. Some limitations were glaucoma, myopia, healthy misclassification, and lack of clinical reliability in isolation. Human-review fallback, uncertainty flags and multimodal clinical integration are future perspectives.

Ejaz et al. (2025) [10] suggested a hybrid fundus image classification method using feature concatenation for the initial diagnosis of a retinal problem. The proposed approach employed two CNN blocks for deep feature extraction for optic disc cupping, macular hole, diabetic retinopathy, and healthy classes. These features were combined using Canonical Correlation Analysis and classified using machine learning classifiers such as gradient boosting, SVM, voting ensemble, KNN, Naive Bayes and random forest. The main dataset used was RFMiD, and the preprocessing involved resizing the image, cropping, augmenting and one-hot encoding. The most important result was that the ensemble classifier had the highest accuracy of 93.39%. One limitation is that the performance was only demonstrated on RFMiD and was only partially validated externally. Future perspectives include end-to-end optimized CNN fusion and testing on independent JPG fundus datasets. Singh et al. (2025) [11] suggested an open-access hybrid CNN-Transformer-ensemble approach. They used deeper CNN backbones, Transformer encoders and ensemble voting strategies to classify 20 retinal diabetic Retinopathy labels in a multi-label classification setting. The key parameters/tools used were ImageNet-pretrained ResNet152d, DenseNet121, DenseNet201, and EfficientNetV2Small, weighted random sampling, LP-ROS balancing, random flipping, rotation, color jitter, ImageNet normalization, AdamW optimization, and cosine annealing with up to 200 training epochs. The main dataset consisted of MuReD, which comprised 2,208 images with 20 labels. The main results obtained were a best model score of 0.9166 and a multi-label F1 of 0.6831. Small data size and imbalanced data, as well as risk of overfitting. Future plans involve extending the dataset validation and Shapley-value interpretability.

Ali et al. (2025) [12] suggested a hybrid DL model for automated diagnosis of eye diseases from fundus images. The approach was to fuse two pre-trained deep networks, DenseNet169 and MobileNetV1, trained on ImageNet to obtain complementary deep features for clinical decision support. The dataset was obtained from the Kaggle platform, which consists of four classes: diabetic retinopathy, cataract, glaucoma and normal. The key finding was an accuracy of 92.99%, precision of 93.02%, recall of 92.85%, F1-score of 92.90% and AUC of 98.77. A limitation is the four-class, single-dataset evaluation. Future plans involve external validation, additional eye categories, easy deployment, and explainable AI integration. Kher et al. (2025) [13] have suggested an explainable ensemble model for the classification of ocular diseases based on feature extraction and Grad-CAM. The proposed method used CLAHE to enhance the image in the LAB color space and used seven pre-trained deep networks to create ensemble models by concatenating features from pairs of deep networks. Classifier fine-tuning and Grad-CAM visualization, followed by DenseNet121 + EfficientNetB5, was the best hybrid combination. The data set included diabetic retinopathy, cataract, glaucoma and normal images of fundus. The accuracy was found to be 95.00%, and the F1-scores for the various diseases were 97% for cataract, 90% for glaucoma and 91% for normal, 100% for diabetic retinopathy. One of the drawbacks was that the preprocessed data was not available for public use. Future prospects involve public reproducible datasets, external validation and better explainability-based clinical evaluation.

Bisht et al. (2026) [14] introduced a hybrid CNN-Transformer network for automated eye disease diagnosis based on color fundus imaging, called EyeFusionNet. The proposed architecture consisted of two tracks: DenseNet169 for local spatial features and a Transformer-in-Transformer module for long-range global Interdependence. The findings were fused and refined through efficient channel attention, and explainable AI visualization was added for clinician interpretation. The EyeDisease Image Dataset, containing 5335 images in 10 classes, was the primary dataset used. The key finding was an accuracy of 89%, and it was shown to outperform some baseline CNN and Transformer models. One drawback is that the public page offers only limited information on the datasets that can be downloaded and how they can be implemented. External clinical validation, robustness testing for JPG compression artifacts, lightweight deployment and better explainability for ophthalmologists are future perspectives. Alqassab et al. (2026) [15] introduced a hybrid quantum CNN for the detection of ocular diseases from fundus images. The method used was anisotropic diffusion filtering, wavelet transformation for hue, contrast and noise enhancement and targeted enhancement to address class imbalance.

The main parameters/tools used were: 8-qubit quantum register, Rx/Rz gates, CNOT operations, quantum convolutional and pooling layers, PennyLane, Qiskit, EstimatorQNN and repeated training runs. OIA-ODIR, a primary dataset of 10,000 images from 8,000 cases with 8 categories of ocular diseases, was used as the main dataset. The major conclusion was that 94% of the time, the accuracy was 94%. One of the difficulties was the lack of data and the need to rely on training volume. Future perspectives involve real testing of quantum hardware, larger public JPG datasets and cost-benefit comparison with classical optimized CNN models.

2.2. Research Gaps

While hybrid deep learning models have demonstrated good performance in ocular disease classification, most studies are evaluated on small or single datasets, limiting their confidence in generalization in real-world settings [8-10]. There is also a lack of external validation, reproducibility, and clinical evaluation of existing work, particularly if private datasets are used, or there is not enough information about implementation [8, 11]. Furthermore, problems like class imbalance, similar disease patterns and limited explainability are still significant [12, 13]. Hence, more efforts are required to create a reproducible, explainable and externally validated model for reliable multi-class ocular disease detection from fundus images.

3. Research Methodology

The methodology adopted in the proposed hybrid ocular disease recognition model using fundus images is described in this section. The pipeline consists of the following components: Dataset preparation from ODIR-5K, image preprocessing, augmentation, transfer-feature extraction using VGG16, feature selection using Crow Search Optimization (CSO), and a Convolutional Neural Network (CNN) classifier. Figure 1 shows the complete workflow designed to reduce noise and redundant features while retaining discriminative retinal patterns for multi-class ocular disease recognition.

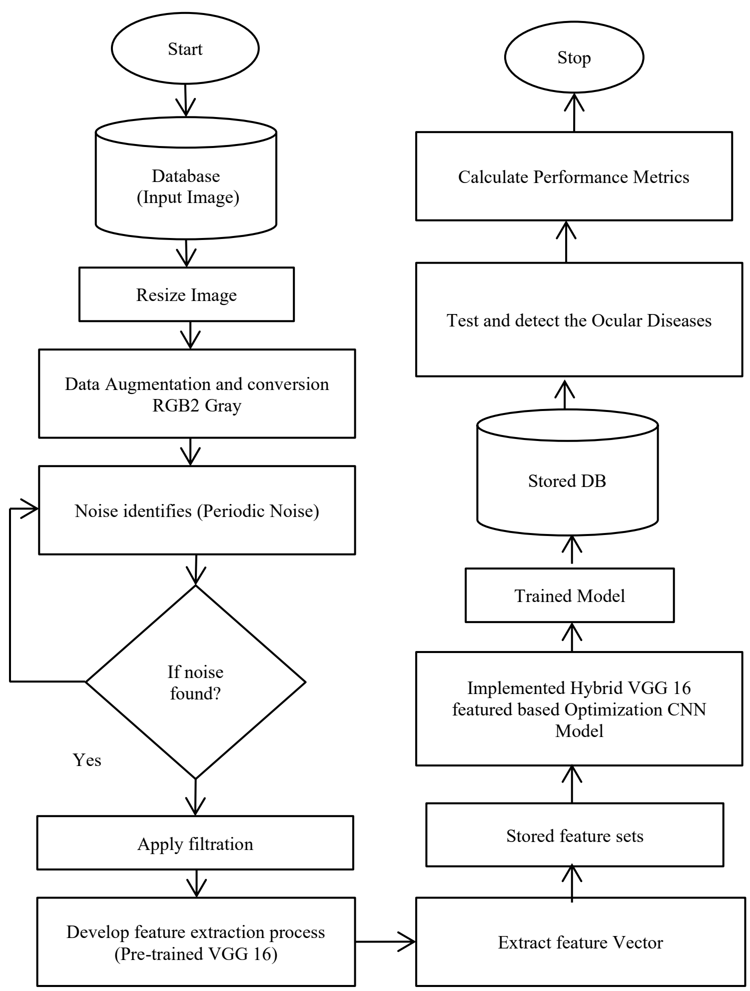

Fig. 1 Overall workflow of the proposed HybridFONet model

3.1. Dataset Collection

The study uses the Ocular Disease Intelligent Recognition dataset (ODIR-5K) [16], an open-access Kaggle dataset containing color fundus photographs and clinical diagnostic keywords for left and right eyes. The diagnostic keyword fields were converted into disease labels representing Normal, Diabetes, Glaucoma, Cataract, Age-related Macular Degeneration, Hypertension, Myopia, and other abnormalities. After label generation, left-eye and right-eye images were filtered by class and combined into a single image-label dataset. To reduce class imbalance, normal images were sampled in a controlled manner and concatenated with disease-class images. Table 1 summarizes the dataset composition and train-test split.

Table 1. Dataset distribution and experimental split for ODIR-5K

Dataset component	Description	Value / Strategy
Dataset	ODIR-5K color fundus images with clinical diagnostic keywords	Left and right eye images
Class	Normal (N), Glaucoma (G), Diabetes (D), Cataract (C), Hypertension (H), AMD (A), Myopia (M), Other (O)	8 categories
Total images used	Images after filtering and preprocessing	7,989
Training set	Used to train the proposed hybrid model	6,391 images
Testing set	Used for model validation and result analysis	1,598 images

3.2. Data Preprocessing and Augmentation

All fundus images are preprocessed prior to feature extraction. The images are read from the dataset directory, and each image is labeled and resized to 240 × 240 × 3 (RGB channel). By standardizing the spatial size, it can be compatible with the input layer of VGG16, and it can make all feature vectors have the same dimensionality.

3.2.1. Upload Image

Fundus images and their respective image paths are uploaded to memory with respective class labels. Unreadable files are ignored, and the valid images are kept for visualization, preprocessing, feature extraction, and model training as shown in Figure 2.

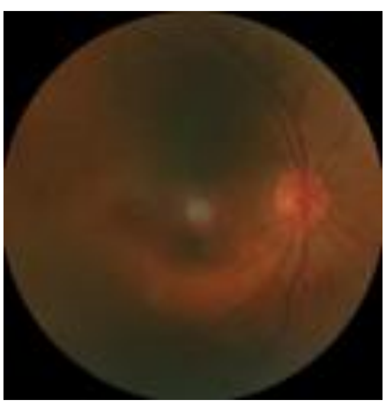

Fig. 2 Uploaded color fundus image

3.2.2. Resize Image

All input images are scaled down to 240 × 240 pixels as illustrated in Figure 3. This size maintains enough retinal structure to recognize disease and also has low computational costs during the feature extraction process using VGG16 and the training of CNN.

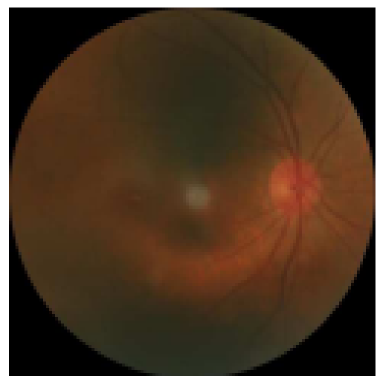

Fig. 3 Resized fundus image

3.2.3. Grayscale Conversion

To simplify the color complexity in the fundus image, while maintaining the important retinal structures, the resized fundus image is converted into grayscale, as shown in Figure 4. This makes preprocessing easier and aids in concentrating on intensity-based features required for the disease recognition.

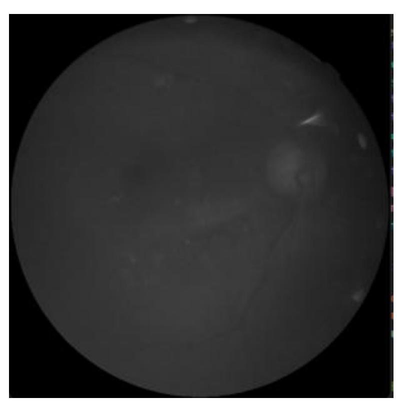

Fig. 4 Grayscale conversion of a sample fundus image

3.2.4. Periodic Noise Addition

To test the robustness of the preprocessing, periodic noise is added to grayscale fundus images (as seen in Figure 5). This mimics acquisition noise that may be structured in medical images due to capture or transmission artifacts.

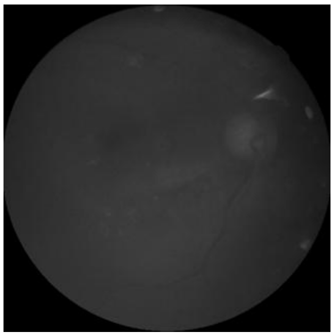

Fig. 5 Sample image after periodic-noise addition

3.2.5. Butterworth Band-Pass Filter

To remove unwanted periodic noise and retain informative retinal structures, a Butterworth band-pass filter is used as shown in Figure 6. The Butterworth response has a smooth frequency transition with no abrupt ringing artifacts, which helps to clean up downstream feature extraction.

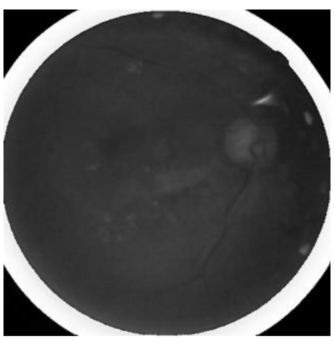

Fig. 6 Output after applying the Butterworth band-pass filter

3.2.6. Data Augmentation

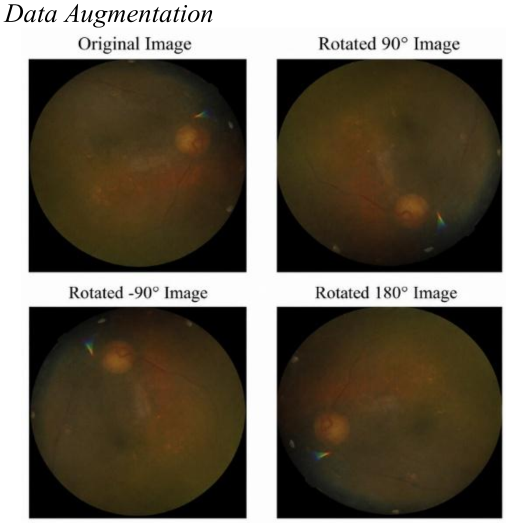

Fig. 7 Image augmentation by rotation at 90°, −90°, and 180°

Data augmentation provides more variety in the training data and helps to prevent overfitting. The reason for using rotational transformations at 90°, −90°, and 180° (as shown in Figure 7) is that fundus images may be taken with a slight difference in orientation. Augmentation allows the model to learn more general features in the retina than the fixed image orientation.

3.3. Pre-Trained Feature Extraction using VGG16 Model

To exploit transfer learning and improve the efficiency of the proposed HybridFONet, feature extraction is performed using the VGG16 model [12].

VGG16 was selected because of its established ability to extract hierarchical visual patterns. The architecture uses 3 × 3 convolutional filters with 1 stride and padding to maintain spatial dimensions, with by 2 × 2 max-pooling layers with stride 2.

Its regular layer design and stacked small filters allow the model to extract complex retinal features efficiently. Figure 8 presents the VGG16 feature-extraction framework.

The VGG16 architecture contains several sequential layers for feature classification and extraction.

- The first and second layers of convolutional use 64 filters of 3 × 3 size. The resulting feature maps are processed by a max-pooling layer to reduce spatial dimensions.
- The third and fourth convolutional layers use 128 filters, followed by another max-pooling layer.
- The fifth, sixth, and seventh convolutional layers use 256 filters, followed by max-pooling to further reduce spatial resolution.
- The 8th-13th layers are convolutional layers, each having 512 filters, which are grouped into two sets to extract high-level features without discarding useful spatial information.
- The fourteenth and fifteenth layers are fully connected dense layers with 4096 units that act as high-level classifiers for the convolutional output.
- The last layer is a softmax classifier that assigns class probabilities according to the learned features [12].

Overall, the VGG16 architecture is well-suited for extracting rich visual representations from retinal images because its stacked 3 × 3 filters capture complex patterns across multiple abstraction levels.

The CSO method is used in this study to refine the extracted feature set, emphasize the most informative retinal regions, and improve classification consistency [17].

CSO is based on the crows' intelligent searching behavior and balances exploration with exploitation during optimization [19]. Figure 9 shows the CSO flowchart.

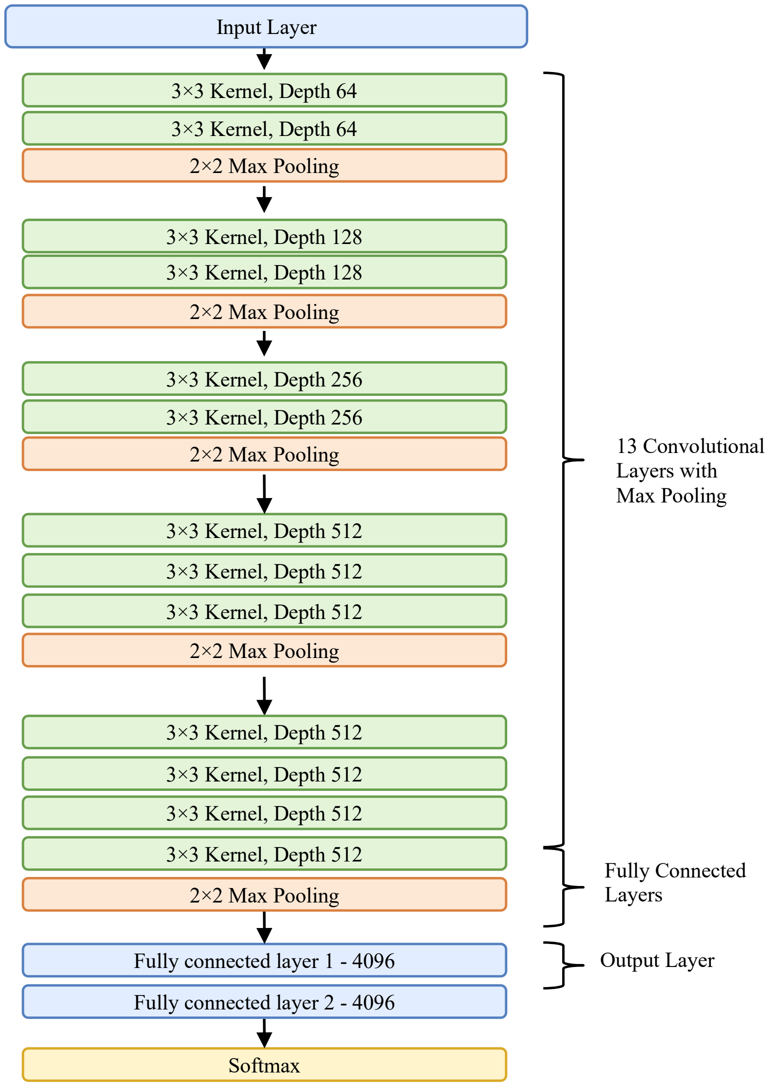

Fig. 8 Framework of the pre-trained VGG16 feature-extraction model

Algorithm 1. Feature Selection Using the CSO algorithm
Input: n = Total no. of crows. $iter_{max}$ = Max number of iterations allowed.
Outcome: Best position of the best-performing crow. 1. Set the positions of all crows randomly within the possible space. 2. Assign each crow a memory variable to store its best-known position. 3. Repeat the further steps while $iter<iter_{max}$: For each $crow_i$ in the population: Choose a random crow to follow. Determine the Awareness Probability (AP) Update the position of $crow_i$ ($x_{i.iter+1}$) using Equation (1) end for loop Updated positions should be within defined boundaries. Evaluate the fitness of each updated crow's position. Update each memory of crow using Equation (2) 4. End while

In ocular disease detection, CSO is utilized to optimize hyperparameters within the model. CSO is able to search a large solution space in the early stages (exploration) and then focus on the best solutions (exploitation) to avoid falling into local minima and to converge more efficiently to the optimal solution. In the Crow Search Optimization algorithm, the crows are randomly distributed in the search space. A fitness function is used to evaluate each crow, and its best position is stored in memory. In each iteration, one of the crows randomly chooses another crow and determines whether to move to the best position that it remembers, with the probability of awareness.

When the selected position is better, the crow updates its memory; otherwise, it keeps the best position. This process is repeated until the stopping condition is satisfied. In this study, CSO is used to optimize the feature selection and model parameters, which reduces the computational complexity and enhances the accuracy of the classification of ocular diseases.

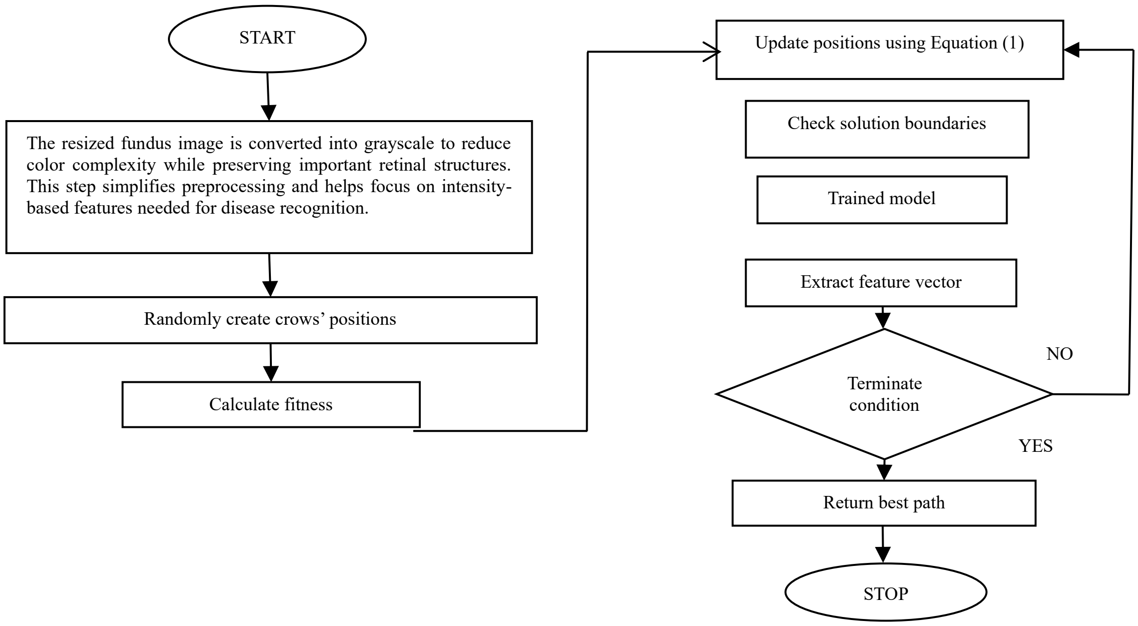

Fig. 9 Flow of the CSO method using feature selection

3.4. Detection Process Using Convolutional Neural Network

CNNs are a key class of DL algorithms in computer vision and have been extensively applied in the ocular disease's detection [18]. CNNs are able to capture the spatial hierarchies of features directly from image data and can detect retinal lesions, vessel patterns, changes in the image due to cataract, and variation of the optic disc due to glaucoma. A CNN has pooling layers, convolutional layers, and fully connected layers, as is shown in Figure 10. The CLs are responsible for applying the convolutional filters to the raw image. These filters, referred to as kernels, traverse the image and perform convolution operations to capture local patterns like edges, textures, or shapes. For ocular diseases, these patterns may allow the model to learn to recognize abnormalities in the eye, such as retinal lesions, cataracts, or glaucoma. The network usually contains pooling layers, which are down-sampling operations, after the convolution operation. Pooling layers are reduced the number of variables and operations by down-sampling. The pooling operation in a typical CNN can be done by max pooling or average pooling. The max pooling is the process of selecting the maximum value in a window (e.g., 2 × 2 or 3 × 3 pixels), while the average pooling is the process of selecting the mean pixel intensity within the window.

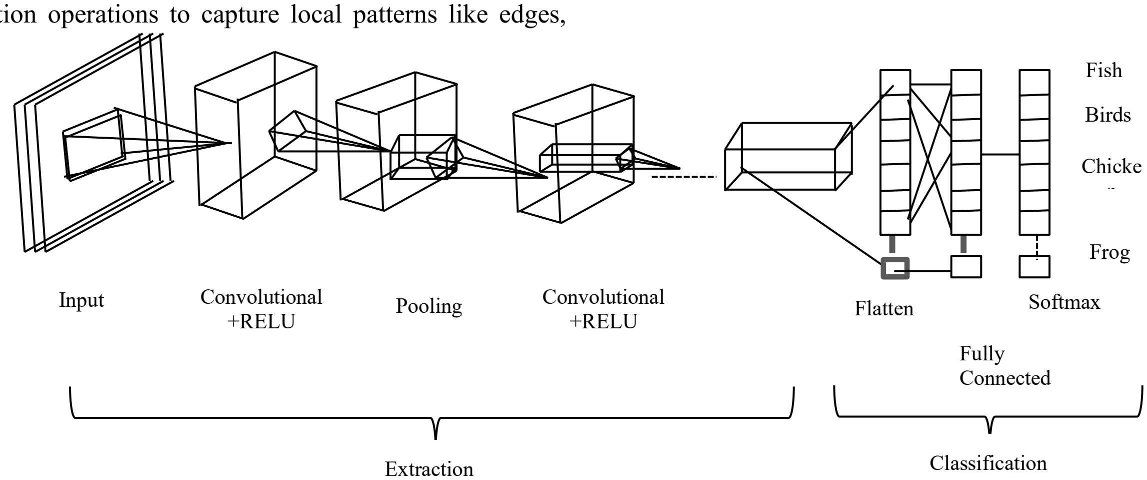

Fig. 10 CNN architecture

Both methods compress the image in space, which decreases the complexity of the calculations and keeps the most important features of the image for further processing. It also helps the network attend to on the most important parts of the image, like the retina or cornea, where disease may be present. After the pooling and convolutional layers, CNNs have fully connected layers that use the attributes learned in the convolutional and pooling layers to make predictions. For the OC diagnosis, the fully connected layers are where the model decides whether a specific condition (DR, AMD or other ocular diseases) exists or not [20].

In CNNs, convolution plays a crucial role in feature extraction. It is a mathematical operation which used two matrices: the input image and the kernel matrix.

This operation involves multiplying the corresponding elements of both matrices and then summing the results. The convolution operation is typically denoted by the * operator. In scientific terms, the convolution operation can be signified as stated in Equation (1).

$$z_{i,j}^{h} = \sum_{m=1}^{M} \sum_{n=1}^{N} A_{(i-1)s+m,(j-1)s+n} * K_{m,n} + b^{h} \quad (1)$$

Here in Equation (1), $z_{i,j}^{h}$ represents the findings of the convolutional operation with the jth column, ith row, and hth channel. b^h represents the bias of the h-th channel. $K_{m,n}$ is an element of the K matrix, while $A_{(i-1)s+m,(j-1)s+n}$ is an element of the input retinal image matrix. The parameters M and N are the kernel sizes, while s is the stride size that shows the shift of the kernel in the convolutional operation.

CNNs also utilize techniques like batch normalization and dropout, basically utilized to enhance training efficiency and reduce over-fitting, respectively. Main strengths of CNNs in the method of OD detection is their ability to process large datasets of images and adapt to them without extensive pre-processing. This can be especially beneficial in medical imaging, where identifying abnormalities can be challenging and require careful analysis. CNNs can be trained on labeled databases of ocular images to learn patterns and help make an early diagnosis and more accurate monitoring of conditions.

Overall, CNNs can be beneficial for detecting fine details in medical images, including the texture of retinal vessels of blood or the slight irregularities in the optic nerve head that may indicate disease. They are able to automatically recognize relevant features from data of the source image, and therefore are very useful in medical diagnostics, especially for the recognition of subtle patterns.

3.5. HybridFONet Model

The model is set up with important hyperparameters to maximize the stability and performance of the training. The batch size is 8, the epochs are 20, and the Learning Rate (LR) is 0.0001. Adaptive learning rates are computed with the Adam optimizer. The convolutional layer has 64 filters with a size 2, max pooling has a pool size of 2, the dense layer has 256 ReLU units, and the output layer has softmax activation for 8 classes. Table 2 summarizes the complete configuration. The complete HybridFONet workflow is explained in pseudo-code 1, starting from loading fundus images, preprocessing, noise filtering and augmentation. It then extracts deep features using VGG16, optimizes the deep features using CSO, trains a CNN classifier and test the final model with accuracy, recall, precision, f1-score and confusion matrix.

Pseudo code 1: HybridFONet Model

Input: Ocular Fundus Image Database (DB)
Output: Model Classification (MC), Metric (Pre, Rec, F1-score, Acc), Confusion Matrix (CM)

```
START
Initialization
Phase I: Data Planning and Data Labeling
Initial categories or classes: [N, D, A, H, G, C, M, O]
 Load fundus images and associate them with ground-truth labels from the DB
Phase II: Image Pre-processing
    FOR each fundus image in DB DO
      Image Resizing:
   Normalize all fundus images to 240 × 240 × 3 pixels
      Normalization:
      Rescale image pixel intensity values to the range (0,1)
      Grayscale Conversion:
      Convert the RGB image into a grayscale image
    END FOR
Phase III: Noise Identification and Filtration
    IF noise_detected THEN
      Apply a Butterworth band-pass filter
      Remove artifacts and improve retinal vessel clearness
    ELSE
      Proceed to Image Augmentation
      Apply image transformation functions Trans:
        • Zooming
```

```
                • Flipping
                • Rotating
                • Scaling
            Balance class distribution
        END IF
Phase IV: Feature Extraction // Pretrained VGG16 Model
        Load pretrained VGG16 model
        (weights trained on ImageNet)
    Freeze initial layers to retain low-level spatial features
        FOR each pre-processed fundus image DO
            Pass the image through the VGG16 convolutional blocks
            Remove High-Dimensional (HD) feature vectors
            from the bottleneck layer
            Store extracted feature sets in Feature DB
        END FOR
Phase V: Optimized Feature Sets
 Input the VGG16 feature vectors into optimized layers
        Apply Crow Search Optimization (CSO)
        to fine-tune hyperparameters:
            • Learning rate
            • Filter sizes
            • Dropout rates
     Phase VI: Dataset Splitting
        Split DB into:
       [1] Train_Set DBtrain (80%)
       [2] Test_Set DBtest (20%)
Phase VII: Model Build (HybridFONet Architecture)
Initialize Convolutional Neural Network
        Define Feature Extraction Layers:
            • Convolutional Layer
            • MaxPooling Layer
        Define Classification Head:
       [1] Fully Connected Layers
                with ReLU activation
       [2] Dropout Layer
                to minimize overfitting
       [3] Softmax Layer
                for an 8-class probability distribution
        Train HybridFONet using feature-rich input
Phase VIII: Model Training
        Set Hyperparameters:
            Learning_Rate = 0.0001
            Optimizer = Adam
            Loss = Categorical Crossentropy
          FOR each iteration (itr) in max_itr DO
            Perform forward propagation on DBtrain
            Calculate loss gradient
            Update weights using backpropagation
            Validate model performance on DBtest
        END FOR
Phase IX: Evaluation
        Generate predictions Y for DBtest
        Calculate the Confusion Matrix (CM)
        by comparing Y with true labels
        Compute performance metrics:
     END
```

3.6. Hyperparameter Configuration

Table 2. Hyperparameter configuration of the proposed HybridFONet model

Method	Strategy / Value	Description	Motive
Preprocessing	240 × 240 × 3	Standardizes all input fundus images	Ensures compatibility with VGG16
Train-test split	80:20	6,391 training and 1,598 testing images	Provides independent validation
Feature extractor	VGG16 (ImageNet; top excluded)	Extracts deep convolutional features	Uses transferable features
Feature selection	CSO, max selected features = 20,000	Selects the best VGG16 feature subset	Reduces redundancy and complexity
Classifier	1D CNN	Conv1D, MaxPooling, Flatten, Dense, Softmax	Learns class boundaries from selected features
Optimizer	Adam	Adaptive gradient optimization	Stable convergence for deep models
Learning rate	0.0001	Step size for weight updates	Balances training speed and stability
Batch size	8	Number of samples per update	Fits memory limitations and stabilizes updates
Epochs	20	Training iterations over the dataset	Allows sufficient learning without excessive overfitting
Loss function	Sparse categorical cross-entropy	Multi-class loss for integer labels	Matches the eight-class classification task

This section presents the key hyperparameters and implementation settings used to train the HybridFONet model. These parameters define the preprocessing, feature extraction, feature selection, classifier training, and optimization strategy used for stable ocular disease classification. Table 2 summarizes the final configuration.

3.7. Model Validation

The held-out test set is used for model validation. Once the best saved model is loaded, predictions are made for the test features and compared to the actual labels. Evaluation metrics are accuracy, recall, precision, F1-score and error rate.

3.8. Comparison Methods

To assess the performance of the proposed HybridFONet model, its performance is compared with established deep-learning architectures: VGG16 [5], Improved AlexNet [6], VGG19 [7], ResNet50 [7], and ResNet152V2 [7]. The comparison focuses on accuracy, recall, precision, F1-score, and error rate so that performance is not judged by accuracy alone.

- VGG16: VGG16 is used as a baseline because it demonstrated that deeper convolutional networks using small 3 × 3 filters can learn strong hierarchical visual representations. In this study, the proposed method extends the VGG16 baseline by adding CSO-based feature selection and a CNN classifier to reduce redundant feature influence [5].
- AlexNet: AlexNet is included because it is a landmark CNN architecture that popularized deep convolutional learning for large-scale image classification. It provides a useful comparison with a shallower CNN design relative to VGG and ResNet families [6].
- VGG19: VGG19 is compared because it follows the same design principle as VGG16 but increases depth to 19 weight layers. This comparison helps determine whether additional depth alone improves ocular disease recognition or whether the hybrid feature-selection approach is more effective [7].
- ResNet50: ResNet50 is included because residual learning uses skip connections to train deeper networks more effectively than plain CNN stacks. Its bottleneck residual blocks provide a strong modern baseline for medical image classification tasks [7].
- ResNet152: ResNet152 is the deepest residual comparator and is used to test whether very high depth improves performance. ResNet152 can learn powerful representations, but it may require more computation and may not always outperform a hybrid optimized-feature approach on limited or imbalanced medical datasets [7].

4. Simulation Results and Analysis

A HybridFONet framework is proposed to classify ocular diseases precisely into multiple classes from JPG fundus images using image preprocessing, data augmentation, transfer-feature extraction using VGG16, feature selection using Crow Search Optimization and a one-dimensional CNN classifier. The proposed workflow consists of the following steps: retinal images from ODIR-5K dataset are standardized and enhanced, deep visual representation of retinal images is extracted using the pre-trained VGG16 model, CSO is used to select the most informative feature subset, and finally, the optimized feature subset is classified into eight classes: Normal, Diabetes, Age-related Macular Degeneration, Cataract, Glaucoma, Hypertension, Myopia, and other abnormalities. The simulation results demonstrate that the hybrid optimized-feature strategy can provide better feature representation and remove redundant information, and can obtain higher classification performance than the compared baseline deep-learning models.

The experimental setup used Python-based image-processing, ML, and DL tools. The implementation environment, model parameters, feature-extraction settings, and evaluation configuration are summarized in Table 3. The final CNN training configuration uses an LR of 0.0001 with the Adam optimizer.

4.1. Tools and Parameters

Table 3. Simulation tools used for HybridFONet implementation

Component / Tool	Setting / Value	Purpose / Description
Programming environment	Python 3.10.11	Model implementation and experiment execution
Deep-learning framework	TensorFlow / Keras	VGG16 loading, CNN construction, model checkpointing, and prediction
Data handling	Pandas, NumPy, Pickle	CSV reading, label generation, feature-matrix handling, and saved test-data loading
Image processing and visualization	OpenCV, Matplotlib	Image resizing, grayscale conversion, noise addition, Butterworth filtering, and result visualization
Machine-learning support	Scikit-learn	Train-test split and metric computation
Operating platform	Windows 10, Intel Core i7 11th Gen, 24 GB RAM	Simulation workstation used for model training and evaluation
Input image size	240 × 240 × 3	Fixed RGB image size used before VGG16 feature extraction
Train-test split	80:20	6,391 training images and 1,598 testing images
Total processed images	7,989	Images used after filtering and class construction
Saved feature-test samples	1,417	Samples available in the saved test-feature matrix used for feature analysis
Feature extractor	VGG16 with ImageNet weights, include_top = False	Transfer feature extraction without the original VGG16 classifier
Extracted feature dimension	25,088	Flattened VGG16 feature vector length
Feature selection method	Crow Search Optimization (CSO)	Selection of optimized deep features
Maximum selected features	20,000	Reduced feature subset used for CNN classification
Classifier	1D CNN	Conv1D, MaxPooling1D, Flatten, Dense, and Softmax layers
Batch size	8	Number of samples used per weight update
Epochs	20	No. of complete passes through the training data
Learning rate	0.0001	Step size for Adam optimizer updates
Optimizer	Adam	Adaptive gradient-based optimization
Loss function	Sparse categorical cross-entropy	Multi-class loss for integer class labels
Conv1D filters	64	Feature filters used in the 1D convolution layer
Conv1D kernel size	2	Kernel size for local feature learning
MaxPooling1D pool size	2	Down-sampling of selected feature maps
Dense layer units	256	High-level learned representation before output
Dense activation	ReLU	Nonlinear activation in the hidden dense layer
Output activation	Softmax	Eight-class probability prediction

4.2. Evaluation Performance Metrics

Performance metrics provide statistical measures for assessing the effectiveness of the proposed classification model. In this study, accuracy, recall, precision, F1-score, and error rate are used to evaluate efficiency.

4.2.1. Accuracy

Accuracy represents the overall percentage of correct predictions made by the model and provides a general indication of classification efficiency, as shown in Equation (2).

$$Accuracy = \frac{(TRUE\_Positive + TRUE\_Negative)}{(TRUE\_Positive + TRUE\_Negative + False\_Positive + False\_Negative)} \quad (2)$$

4.2.2. Precision

Precision is particularly useful when reducing false positives is important, as shown in Equation (3).

$$Precision = \frac{TRUE\_Positive}{(TRUE\_Positive + FALSE\_Positive)} \quad (3)$$

4.2.3. Recall

Recall is important when minimizing false negatives is necessary, as shown in Equation (4).

$$Recall = \frac{TRUE\_Positive}{(TRUE\_Positive + FALSE\_Negative)} \quad (4)$$

4.2.4. F1-Score

The F1-score is useful when precision and recall must be balanced, as shown in Equation (5).

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

4.2.5. Error Rate

The error rate represents the proportion of incorrect predictions and provides a direct measure of misclassification, as shown in Equation (6).

$$Error\ Rate = (FP + FN) / (TP + TN + FP + FN) \quad (6)$$

where FP shows false positives, TP denotes true positives, FN denotes false negatives, and TN denotes true negatives.

Together, these metrics provide a balanced assessment of classification performance by considering both correct predictions and error behavior.

4.3. Database Description

The experiment uses the OD Intelligent Recognition dataset, ODIR-5K, which is an open-access fundus-image database available through Kaggle [16]. The dataset contains color fundus figures for right and left eyes, together with clinical diagnostic keyword fields. In this work, diagnostic keywords are converted into image-level disease labels. The final classification problem contains eight categories:

Normal (N), Diabetes (D), Hypertension (H), Myopia (M), Glaucoma (G), Cataract (C), Age-related Macular Degeneration (A), and other abnormalities (O). Table 4 summarizes the detailed dataset description, and Figure 11 shows representative images from each category. In the experimental implementation, 7,989 images were used. The data were seperated into 6,391 training images and 1,598 testing images using an 80:20 split. Figure 12 shows the class distribution used to understand imbalance and guide evaluation beyond accuracy.

Table 4. Detailed description of the ODIR dataset used in the simulation

Database component	Description / Value	Role in the experiment
Dataset name	ODIR-5K / Ocular Disease Intelligent Recognition	Color fundus image database with clinical diagnostic keywords
Source	Kaggle open-access dataset	Used for ocular disease recognition experiments
Image type	JPG/color fundus photographs	Left-eye and right-eye retinal images
Label source	Clinical diagnostic keyword fields	Converted into disease class labels
Classes	N, D, H, M, G, C, A, O	Normal, Diabetes, Hypertension, Myopia, Glaucoma, Cataract, AMD, and Other
Total images used	7,989	Images retained after filtering and preprocessing
Training set	6,391 images	Used for model fitting
Testing set	1,598 images	Used for validation and result analysis
Split ratio	80:20	Independent training and testing split
Class-imbalance handling	Controlled normal-image sampling and augmentation	Reduces the dominance of majority categories
Preprocessing	Image loading, resizing, grayscale conversion, noise filtering	Standardizes image size and improves robustness
Augmentation	Rotation at 90°, −90°, and 180°	Increases variation and reduces overfitting

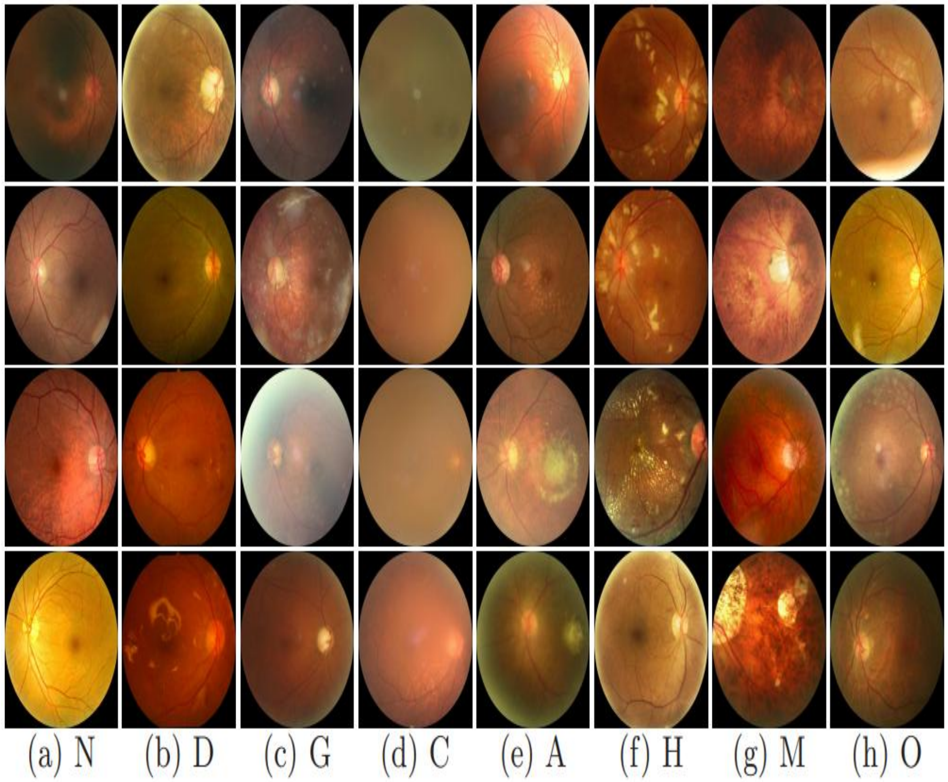

Fig. 11 Representative fundus images grouped by ocular-disease category [19]

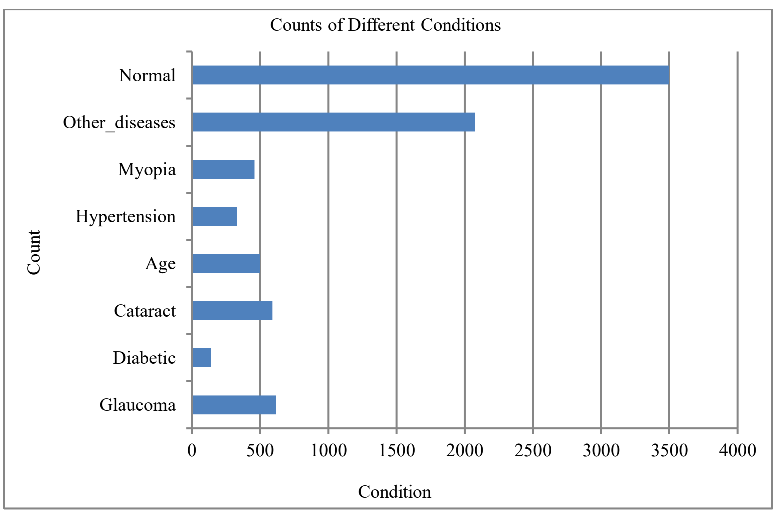

Fig. 12 Class distribution of the ODIR-5K fundus images used in this study

4.4. Proposed HybridFONet Results

The complete HybridFONet result flow starts with image loading and diagnostic-keyword label generation, followed by preprocessing, augmentation, VGG16 feature extraction, CSO feature selection, 1D CNN classification, and final evaluation using the confusion matrix and performance metrics. The pre-trained VGG16 model is used in the feature extraction stage (as illustrated in Figure 13) to extract discriminative retinal features from fundus images. Firstly, all the retinal images are resized to 240 × 240 × 3 pixels and preprocessed as per the requirements of the VGG16. The model is pre-trained with ImageNet weights, and the top fully connected layers are omitted so that the model can learn features in the convolutional layers. Each image is processed using convolutional and max-pooling layers, which capture retinal textures, vessel structures, optic-disc characteristics and abnormalities related to the disease. The output is extracted and flattened to high-dimensional feature vectors of size 25,088 and stored for feature optimization and CNN-based classification.

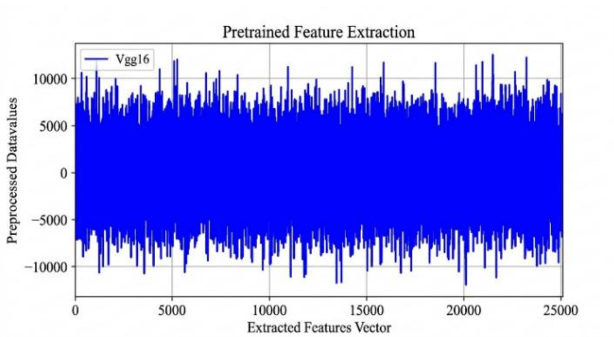

Fig. 13 Distribution of extracted VGG16 features before CSO feature selection

The deep features extracted from VGG16 are reduced using the Feature Selection technique, which is done by the Crow Search Optimization (CSO) algorithm, as illustrated in Figure 14. In this process, each crow is a potential feature subset, with ones and zeros indicating the selected and unselected features, respectively. The algorithm modifies the position of the crows according to the memory and awareness probability, which enables the exploration and exploitation of the best combination of features. Finally, the highest fitness solution is kept, and 20,000 optimized features are selected for efficient CNN-based ocular disease classification.

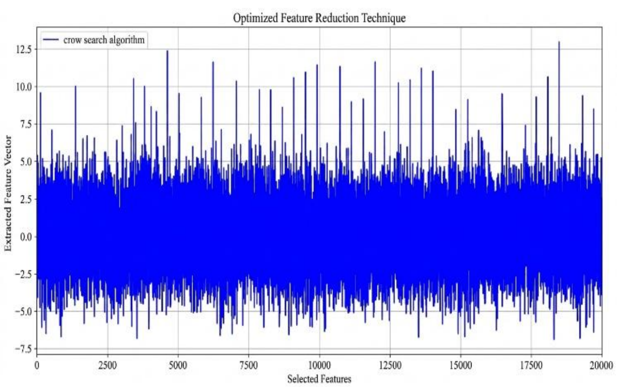

Fig. 14 Optimized feature distribution after CSO feature selection

The confusion matrix in Figure 15 reveals high diagonal dominance, which means that the majority of the samples were classified correctly. Classes with the largest number of correct predictions are Normal (279), Diabetes (280) and other abnormalities (233), and the smaller classes like Hypertension, AMD, Myopia, Glaucoma and Cataract have fewer samples with slightly higher relative misclassification. The overall performance of the optimized hybrid feature pipeline is 96.47% accuracy, 97.29% precision, 95.58% recall, 96.32% F1-score, and 3.53% error rate, which validates the improvement of the overall recognition performance.

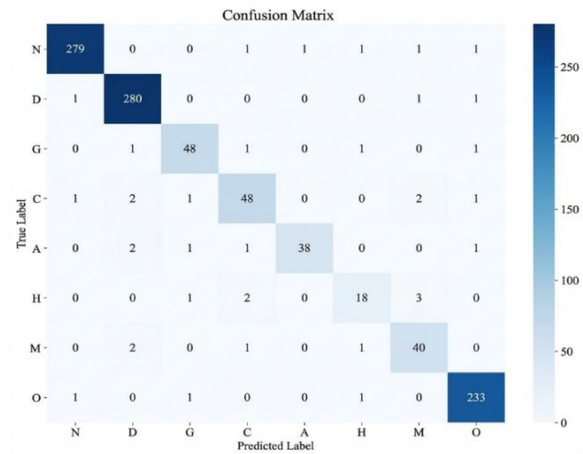

Fig. 15 Confusion matrix obtained for the HybridFONet classifier

The class-wise efficiency of the HybridFONet model is shown in Table 5 in terms of false positive, true positive, false negatives, recall, precision, and F1-score values.

The method was able to get good classification results for Normal, Diabetes and Other abnormalities with F1-scores of 98.59%, 98.25% and 98.31% respectively. Glaucoma also demonstrated a balanced performance with 92.31% precision, recall and F1-score. AMD had high precision of 97.44% and comparatively lower recall of 88.37% which means that there were some missed cases of AMD. Cataract, Hypertension, and Myopia exhibited relatively poor performance because of class imbalance and similarity of retinal features.

The lowest F1-score (78.26%) was obtained for hypertension. In general, the class-wise results validate the effectiveness of HybridFONet in most of the ocular classes, and the need for balancing and validating the dataset in smaller classes.

Table 5. Class-wise metrics extracted from the confusion matrix

Class	TP	FP	FN	Precision (%)	Recall (%)	F1-score (%)
Normal (N)	279	3	5	98.94	98.24	98.59
Diabetes (D)	280	7	3	97.56	98.94	98.25
Glaucoma (G)	48	4	4	92.31	92.31	92.31
Cataract (C)	48	6	7	88.89	87.27	88.07
AMD (A)	38	1	5	97.44	88.37	92.68
Hypertension (H)	18	4	6	81.82	75.00	78.26
Myopia (M)	40	7	4	85.11	90.91	87.91
Other (O)	233	5	3	97.90	98.73	98.31

The classification efficiency of the proposed HybridFONet model is summarized in Table 6. The model was capable to correctly classify the majority of fundus images in each of the eight ocular disease classes with an accuracy of 96.47%. Its precision of 97.29% shows that very few disease cases were falsely predicted, recall of 95.58% shows that most disease cases were predicted, and the F1-score of 96.32% shows that there is a strong balance between precision and recall. The error rate of 3.53% suggests that there is some degree of misclassification.

Table 6. Reported aggregate performance metrics of HybridFONet

Metric	HybridFONet value (%)
Accuracy	96.47%
Precision	97.29%
Recall	95.58%
F1-score	96.32%
Error rate	3.53%

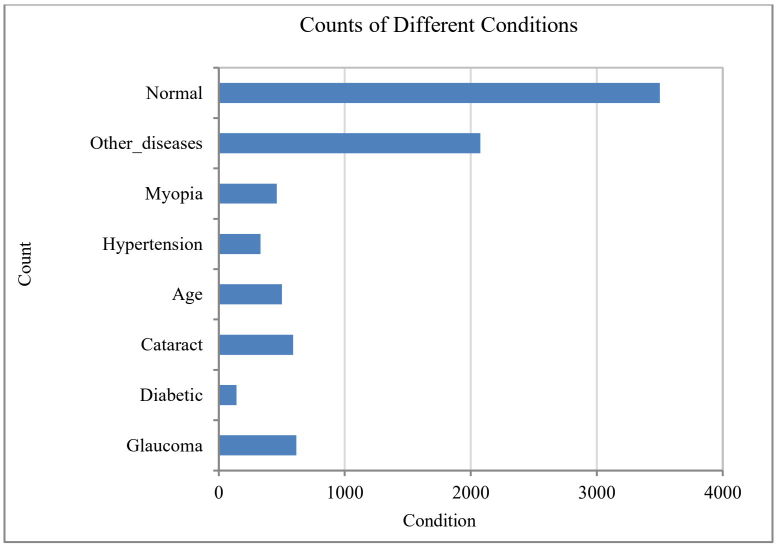

Fig. 16 Visual representation of the reported HybridFONet performance metrics

Figure 16 visually represents the reported efficiency of the proposed HybridFONet model, including accuracy, recall, precision, F1-score, and error rate. The graph shows high values for all major evaluation metrics, with precision reaching 97.29%, accuracy reaching 96.47%, F1-score reaching 96.32%, and recall reaching 95.58%. The low error rate of 3.53% further confirms that only some proportion of fundus images were incorrectly classified.

4.5. Comparative Analysis

The proposed HybridFONet model was compared with VGG16, Improved AlexNet, ResNet50, VGG19, and ResNet152V2. Table 7 and Figure 17 compare the model's using accuracy, recall, precision, F1-score, and error rate so that the proposed method is not evaluated by accuracy alone.

HybridFONet achieves the highest accuracy at 96.47%, which is 10.47 percentage points higher than VGG16 [5], 16.98 percentage points higher than Improved AlexNet [6], 45.04 percentage points higher than VGG19 [7], 25.40 percentage points higher than ResNet50 [7], and 24.27 percentage points higher than ResNet152V2 [7]. It achieved a precision of 97.29% and a recall of 95.58%, which means that it has the ability to detect actual ocular disease cases with fewer false-positive predictions compared to the baseline networks. The F1 score of 96.32% also indicates that the proposed method achieves a balance between recall and precision. The lowest error rate of 3.53% indicates that VGG16 feature extraction, CSO feature selection, and 1D CNN classification is more reliable than using conventional CNN architectures alone.

Table 7. Comparative performance of HybridFONet and existing models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Error rate (%)
VGG16 [5]	86.00	86.75	84.00	86.00	14.00
Improved AlexNet [6]	79.49	78.71	78.00	79.00	20.51
VGG19 [7]	51.43	39.00	40.00	38.00	48.57
ResNet50 [7]	71.07	71.00	71.00	71.00	28.53
ResNet152V2 [7]	72.20	77.00	72.00	72.00	27.80
HybridFONet	96.47	97.29	95.58	96.32	3.53

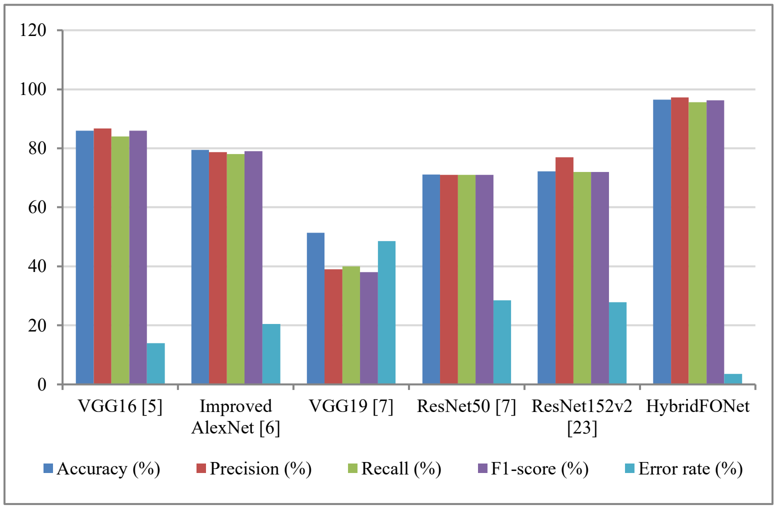

Fig. 17 Visual comparison of HybridFONet with existing deep-learning models

4.6. Discussion

The simulation results demonstrate that the proposed HybridFONet framework is effective to classify the multi-class ocular disease from fundus images. The features of VGG16 are powerful transfer features capturing retinal texture, lesion patterns, vessel structures, optic-disc characteristics, and macular variations. CSO converts the high-dimensional VGG16 representation (25,088 extracted features) to 20,000 selected features, thereby reducing redundant information before the 1D CNN classifies the features. This staged design is the reason behind the superior results obtained by the model when compared to the direct baseline CNN and pre-trained architectures. The confusion matrix also shows that the proposed approach is reliable, as the majority of the predictions are located on the diagonal. The class-wise performance is particularly good for Normal, Diabetes and Other abnormalities, and comparatively low for Hypertension, Cataract, Myopia, Glaucoma and AMD. Such behavior is anticipated in the classification of ocular diseases as smaller classes, and similar retinal abnormalities can lead to higher misclassification. The advantage of HybridFONet is that it combines transfer learning, feature selection and CNN-based classification in a single workflow. Yet, external validation on independent clinical data, better class balancing, and explainable AI techniques like Grad-CAM are still required prior to clinical use.

5. Conclusion

In this research, a hybrid feature-based optimized CNN model, HybridFONet, for the classification of ocular diseases from JPG fundus image was developed and evaluated. To improve the presentation of features and reduce the feature redundancy before the final disease prediction, the proposed framework has included image preprocessing, image augmentation, transfer feature extraction using VGG16, feature selection using Crow Search Optimization, and 1D CNN classification. The model was evaluated on ODIR-5K dataset for eight-class classification: Normal, Age-related Macular Degeneration, Diabetes, Glaucoma, Cataract, Hypertension, Myopia, Other abnormalities. The results showed that the overall performance was good with an accuracy of 96.47%, a precision of 97.29%, a recall of 95.58%, an F1-score of 96.32%, and an error rate of 3.53%. Comparative analysis showed that HybridFONet surpassed VGG16, Improved AlexNet, VGG19, ResNet50 and Res-Net152V2 in terms of key metrics. The results show that the use of optimized feature selection and hybrid CNN classification can enhance the reliability of recognition of ocular diseases. Future research efforts should be directed towards external validation with larger multi-center datasets, better class balancing, and the visualization of explainable AI and the integration of decision-support systems in clinically interpretable ways.

Conflict of Interest

The authors have no conflict of interest.

Author Contribution

Gurpreet Kaur and Amit Kumar Bandal contributed to conceptualization, methodology design, software implementation, validation, original draft preparation, review, editing, proofreading, and supervision. All authors have read and agreed to the final version of the manuscript.

References

[1] Mohammed Ramzi Mohammed, and Ashraf Fathi “Global Prevalence of Diabetic Retinopathy over the Last Decade (2015–2025): A Scoping Review,” *medRxiv*, pp. 1-10, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mario A. Vasilescu, and Mioara L. Macovei, “The Perspective of using Optical Coherence Tomography in Ophthalmology: Present and Future Applications,” *Diagnostics*, vol. 15, no. 4, pp. 1-20, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Feifei Cao et al., "Development and Validation of a Deep Learning Model for Early Detection and Screening of Diabetic Retinopathy," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[4] Xiao Chun Ling et al., "Deep Learning in Glaucoma Detection and Progression Prediction: A Systematic Review and Meta-Analysis," *Biomedicines*, vol. 13, no. 2, pp. 1-26, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kalla Bharath Vardhan et al., "Eye Disease Detection using Deep Learning Models with Transfer Learning Techniques," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 12, no. 1, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Halit Çetiner, "Cataract Disease Classification from Fundus Images with Transfer Learning based Deep Learning Model on Two Ocular Disease Datasets," *Gümüşhane University Journal of Science and Technology*, vol. 13, no. 2, pp. 258-269, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Antonella Santone et al., "A Method for Ocular Disease Diagnosis through Visual Prediction Explainability," *Electronics*, vol. 13, no. 14, pp. 1-33, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Shumoos Al-Fahdawi et al., "Fundus-DeepNet: Multi-label Deep-Learning classification System for Enhanced Detection of Multiple Ocular Diseases through Data Fusion of Fundus Images," *Information Fusion*, vol. 102, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Celina Rieck et al., "A Novel Transformer-CNN Hybrid Deep Learning Architecture for Robust Broad-Coverage Diagnosis of Eye Diseases on Color Fundus Images," *IEEE Access*, vol. 13, pp. 156285-156300, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Sara Ejaz et al., "Fundus Image Classification using Feature Concatenation for Early Diagnosis of Retinal Disease," *Digital Health*, vol. 11, pp. 1-24, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Deependra Singh, Saksham Agarwal, and Subhankar Mishra, "Retinal Fundus Multi-Disease Image Classification using Hybrid CNN-Transformer-Ensemble Architectures," *Proceedings of the International Health Informatics Conference IHIC*, Springer, Singapore, vol. 1113, pp. 103-120, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ismael Abdulkareem Ali, and Sozan Abdulla Mahmood et al., "Enhancing Clinical Decision Support: A Deep Learning Approach for Automated Diagnosis of Eye Diseases from Fundus Images," *UHD Journal of Science and Technology*, vol. 39, no. 2, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Geethanjali Kher et al., "DeB5-XNet: An Explainable Ensemble Model for Ocular Disease Classification using Feature Extraction and Grad-CAM," *Informatics in Medicine Unlocked*, vol. 54, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Akshaj Singh et al., "EyeFusionNet: A Hybrid CNN-Transformer Network for Automated Diansis of Eye Diseases using Colour Fundus Imaging," *EAI Endorsed Transactions on Intelligent Systems and Machine Learning Applications*, vol. 2, pp. 1-13, 2025. [CrossRef] [Publisher Link]

[15] Ans Ibrahim Mahameed Alqassab, M.-Á. Luque-Nieto, and Mazin Abed Mohammed, "Identification of Multiple Ocular Diseases using a Hybrid Quantum Convolutional Neural Network with Fundus Images," *Scientific Reports*, pp. 1-21, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[16] Kaggle, Ocular Disease Recognition, Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k/data

[17] Abdelazim G. Hussien et al., "Crow Search Algorithm: Theory, Recent Advances, and Applications," *IEEE Access*, vol. 8, pp. 173548-173565, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Jiantao Zhao, and Lili Li, "Research on Image Classification Algorithm based on Convolutional Neural Network-Tensorflow," *Journal of Physics: Conference Series*, vol. 2010, no. 1, pp. 1-7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jessica Ryan et al., "Harnessing Deep Learning for Ocular Disease Diagnosis," *Procedia Computer Science*, vol. 245, pp. 914-923, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Nadim Mahmud Dipu, Sifatul Alam Shohan, and K.M.A Salam, "Ocular Disease Detection using Advanced Neural Network based Classification Algorithms," *Asian Journal of Convergence in Technology*, vol. 7, no. 2, pp. 91-99, 2021. [CrossRef] [Google Scholar] [Publisher Link]