

Article

Triple-Stream Deep Feature Selection with Metaheuristic Optimization and Machine Learning for Multi-Stage Hypertensive Retinopathy Diagnosis

Süleyman Burçin Şüyun ¹, Mustafa Yurdakul ^{2,*}, Şakir Taşdemir ³ and Serkan Biliş ⁴¹ Computer Engineering Department, Engineering Faculty, Sinop University, Sinop 57000, Turkey; suyun@sinop.edu.tr² Computer Engineering Department, Engineering and Natural Sciences Faculty, Kırıkkale University, Kırıkkale 71450, Turkey³ Computer Engineering Department, Technology Faculty, Selçuk University, Konya 42250, Turkey; stasdemir@selcuk.edu.tr⁴ Eye Diseases Department, Batigoz Medical Group Hospital, Izmir 35200, Turkey; serkanbilis@gmail.com

* Correspondence: mustafayurdakul@kku.edu.tr

Abstract: Hypertensive retinopathy (HR) is a serious eye disease that can lead to permanent vision loss if not diagnosed early. The conventional diagnostic methods are subjective and time-consuming, so there is a need for an automated and reliable system. In this study, a three-stage method that provides high accuracy in HR diagnosis is proposed. In the first stage, 14 well-known Convolutional Neural Network (CNN) models were evaluated, and the top three models were identified. Among these models, DenseNet169 achieved the highest accuracy rate of 87.73%. In the second stage, the deep features obtained from these three models were combined and classified using machine learning (ML) algorithms including Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The SVM with a sigmoid kernel achieved the best performance (92% accuracy). In the third stage, feature selection was performed using metaheuristic optimization techniques including Genetic Algorithm (GA), Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), and Harris Hawk Optimization (HHO). The HHO algorithm increased the classification accuracy to 94.66%, enhancing the model's generalization ability and reducing misclassifications. The proposed method provides superior accuracy in the diagnosis of HR at different severity levels compared to single-model CNN approaches. These results demonstrate that the integration of Deep Learning (DL), ML, and optimization techniques holds significant potential in automated HR diagnosis.



Academic Editor: Keun Ho Ryu

Received: 9 May 2025

Revised: 2 June 2025

Accepted: 6 June 2025

Published: 9 June 2025

Citation: Şüyun, S.B.; Yurdakul, M.; Taşdemir, Ş.; Biliş, S. Triple-Stream Deep Feature Selection with Metaheuristic Optimization and Machine Learning for Multi-Stage Hypertensive Retinopathy Diagnosis. *Appl. Sci.* **2025**, *15*, 6485. <https://doi.org/10.3390/app15126485>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Convolutional Neural Network; eye disease; feature fusion; Harris Hawk Optimization; hypertensive retinopathy

1. Introduction

The eye is one of the most complex and vital organs in the human body. It enables people to perceive their surroundings and process visual information to make sense of it. As the main organ of the sense of sight, the eye functions through the harmonious functioning of sensitive structures such as the cornea, lens, retina and optic nerve [1]. The healthy functioning of these structures directly affects the quality of life of the individual. However, many factors such as genetic factors, aging, infections, chronic diseases such as diabetes, trauma and unhealthy living habits can negatively affect eye health and lead to various diseases [2]. Among eye diseases, serious conditions such as glaucoma, macular

degeneration, diabetic retinopathy (DR) and HR are among the leading causes of vision loss worldwide. These diseases can lead to irreversible vision loss if not diagnosed and treated early. Therefore, early diagnosis of eye diseases is critical to protect the quality of life of individuals and to successfully plan treatment processes [3].

Medical imaging techniques have become an integral part of diagnosis and treatment processes in modern medicine. In particular, methods such as Optical Coherence Tomography (OCT) allow eye diseases to be detected at early stages by imaging sensitive structures of the eye such as the retina, macula and optic nerve with high resolution [4].

However, manual evaluation of these images is time-consuming, subjective and is prone to error. In addition, the lack of specialists, especially in rural areas, makes this process even more difficult. For all these reasons, there is a growing need for technologies that automate the diagnosis of eye diseases and make the process both faster and more accessible. Rapid advances in hardware and software technologies have led to the development of numerous Computer Assisted Diagnosis (CAD) systems in this field [5]. In particular, Artificial Intelligence (AI) methods are emerging as effective tools for the diagnosis of eye diseases. There are a number of studies in the literature on AI-based automatic diagnosis of eye diseases.

Shoukat et al. [6] performed glaucoma detection on gray channel fundus images using ResNet-50 architecture. Fundus images were processed with grayscale technique and optimized for model training by focusing on the optic disk. The performance of the ResNet-50 model was improved by implementing data augmentation techniques and transfer learning methods. The model was evaluated on the G1020 dataset and achieved high performance results such as 98.48% accuracy, 99.30% sensitivity, 96.52% specificity, 97% AUC and 98% F1-score.

Patel et al. [7] proposed a model combining Flexible Analytic Wavelet Transform (FAWT) and Gaussian–Kuzmin distribution-based Gabor filters (GKDG) to develop an automatic diagnostic tool for the diagnosis of glaucoma. They decompose the green channel components of fundus images using FAWT and extract features from these sub-bands using GKDG filters. Neighborhood Component Analysis (NCA) is applied for dimensionality reduction and the features are classified with LS-SVM algorithm. The proposed model achieved 95.84% accuracy, 97.17% specificity and 94.55% sensitivity in the experiments on the RIM-ONE dataset.

Sharma et al. [8] developed a framework for the diagnosis of glaucoma using fundus images. A customized CNN model designed to extract features from the images. In the dimensionality reduction stage, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) techniques were used together. For classification, the Extreme Learning Machine (ELM) method was employed. To enhance the performance of ELM, a Modified Particle Swarm Optimization (MOD-PSO) algorithm was applied to optimize the input weights and biases. The model's performance was evaluated using five-layer cross-validation on the G1020 and ORIGA datasets. The authors obtained success on the G1020 dataset with 97.80% accuracy, 94.92% sensitivity and 98.44% specificity. On the ORIGA dataset, 98.46% accuracy, 97.01% sensitivity and 98.96% specificity values were obtained.

Geetha et al. [9] developed a DL framework for the early detection of glaucoma called DeepGD. For the classification of fundus images, EfficientNetB4 and snapshot ensemble methods were used, and Aquila Optimization (AO) algorithm was applied for hyperparameter optimization. The classified images were segmented with V-Net and 99.35% accuracy, 99.04% precision, 99.19% specificity, 98.89% recall and 98.97% f1-score values were obtained.

Muthukannan [10] proposed a feature extraction-based approach to classify retinal fundus images into age-related macular degeneration (AMD), DR, cataract, glaucoma and normal. They used fundus images from the Ocular Disease Intelligent Recognition (ODIR) dataset. The images were processed using maximum entropy transform to optimize both quality and information content. The feature extraction was performed using a CNN consisting of two convolutional layers and a maximum pooling layer. The hyperparameters of the CNN are optimized with the Flower Pollination Optimization Algorithm (FPOA). After feature extraction, classification is performed using a multi-class SVM. The proposed method achieved 98.30% accuracy, 95.27% recall, 98.28% specificity and 93.3% F1 score.

Kihara et al. [11] developed a ViT-based segmentation model to detect non-exudative macular neovascularization (neMNV) in AMD patients using OCT images. In the study, the model was trained using 125,500 OCT images and the results were compared with human performance. The model achieved 82%, 90%, 79% and 91% sensitivity, specificity, PPV and NPV, respectively. Compared to human evaluations, the ViT-based model provided higher accuracy and performed particularly strongly in patients with late AMD.

Gu et al. [12] developed a model called Ranking-MFCNet for cataract grading from OCT images. By combining multiscale feature calibration and a ranking-based approach, the model aims to classify six different severity levels of cataracts more accurately. The model uses the eaMFC module, which combines the multiscale attention mechanism with external attention layers to provide a better representation of features to distinguish neighboring severity levels. In the experiments performed on the IOLMaster-700 dataset, the model achieved an accuracy of 88.86%, sensitivity 89.08%, precision 90.15% and F1-Score 89.49%.

Liu et al. [13] developed a Swin-TransformerV2-based model for DR diagnosis called STMF-DRNet. The model has a multi-branch structure that allows the extraction of global, local and fine-grained features and aims to improve classification performance by integrating hybrid attention mechanisms and category-based attention mechanisms. In addition, Attention Based Object Localization Module (AOLM) and Attention Based Patch Processing Module (APPM) were implemented to detect lesion regions and reduce noise. The proposed model achieved 77% accuracy, 77% sensitivity, 94.2% specificity, 77% F1-Score and 87.7% kappa value on the clinical dataset.

Kulyabin et al. [14] introduced OCT Dataset for Image Based DL Methods (OCTDL), an open access dataset for DL methods based on OCT imaging techniques. The dataset covers a variety of retinal diseases such as (AMD), diabetic macular edema (DME), epiretinal membrane (ERM), retinal artery occlusion (RAO), retinal vein occlusion (RVO) and vitreomacular interface disease (VID). They also performed experimental studies on OCTDL with ResNet50 and VGG16 models. As a result of these studies, ResNet50 achieved 84.6% accuracy, 89.8% precision, 84.6% sensitivity and 86.6% F1 score, while VGG16 achieved 85.9% accuracy, 88.8% precision, 85.9% sensitivity, 86.9% F1 score and 97.7% AUC.

The studies conducted on HR detection in the literature are as follows.

Irshad et al. [15] developed a method that classifies retinal vessels into arteries and veins to calculate the arterio-venous ratio (AVR) for HR detection. In the study, the region of interest (ROI) is extracted by determining the optic disk (OD) center in fundus images and vessels, classified using intensity differences in different color spaces. The average intensity, contrast and roughness features of the vessels in RGB, LAB and HSV color spaces were extracted, and classification was performed with SVM. In the tests, the mean density, contrast and roughness values for arteries were 124.29 ± 26.09 , 6.40 ± 1.43 and 0.0006 ± 0.0003 , respectively, while for veins these values were 110.61 ± 27.79 , 9.42 ± 2.54 and 0.0013 ± 0.0007 , respectively. The proposed method was tested on 25 fundus images and an accuracy of 81.3% is achieved.

Abbas et al. [16] developed a system called HYPER-RETINO to classify five stages (normal, mild, moderate, severe, malignant) of HR. The model uses semantic and object-based segmentation techniques to detect HR lesions and classifies them using DenseNet architecture. They tested the model on 1400 fundus images and achieved 92.6% accuracy, 90.5% sensitivity and 91.5% specificity.

Suman et al. [17] developed a hybrid DL architecture to classify HR severity levels. In the study, HR was classified into four classes (normal, mild, moderate, severe) by creating an expert-labeled HRSG dataset. The model aims to capture both local and global contextual information by combining transfer learning with the pre-trained ResNet-50 and the improved Vision Transformer (ViT) architecture. With the Decoupled Representation and Classifier (DRC) method, class imbalance is removed and the overall diagnostic accuracy of the model is improved. In the tests, the proposed method outperformed existing HR classification methods, achieving 96.88% accuracy, 94.35% sensitivity, 97.66% specificity and 94.42% F1 score.

Related studies in the literature are summarized in Table 1. According to Table 1, it is seen that many AI-based studies have been carried out in the diagnosis of eye diseases, especially focusing on the detection of diseases such as DR, glaucoma and macular degeneration. However, studies on HR are limited and the literature does not provide a sufficient approach in this field. In addition, existing studies have some important limitations. Most of the previous studies, for example, use feature extraction based on a single CNN model. However, a single model may not be sufficient to capture diverse and discriminative features at different stages of HR. Moreover, although some studies integrate ML based classifiers, they do not examine the effectiveness of feature selection methods, which may lead to redundant information content in the extracted features. Furthermore, although some studies report high accuracy rates, they do not provide a comprehensive analysis of how the model generalizes across different HR severity levels. Whether the model is overfitting, its performance in different patient groups or accuracy rates at specific stages are not examined in detail. These shortcomings indicate the need to develop a more reliable and generalized AI-based system for the diagnosis of hypertensive retinopathy.

The main contributions of this study are as follows:

- A comprehensive literature analysis was conducted to examine existing AI-based studies in the diagnosis of eye diseases.
- Fourteen different CNN models commonly used in the literature were trained on a custom HR dataset and the three best-performing models were determined.
- Deep features extracted from the top three CNN models were combined to create a more robust feature set.
- The combined features were classified by ML algorithms (SVM, RF and XGBoost).
- GA, ABC, PSO and HHO methods were used in the feature selection process.
- The classification performance of the model was analyzed using extensive experiments and different evaluation metrics.

The remaining sections of this paper are organized as follows:

Section 2 describes the materials and methods used in this study, including dataset details, DL and ML techniques. Section 3 covers the experimental setup process, including hardware configurations, training procedures and evaluation metrics. Section 4 presents the experimental results, including the performances of CNN models, feature fusion results and improvements achieved by feature selection. Section 5 discusses the findings and evaluates the strengths and weaknesses of the proposed method. Finally, Section 6 summarizes the main contributions of the work and provides recommendations for future research.

Table 1. Summary of related studies on eye disease diagnosis using AI techniques, including methodology, imaging modality, disease focus, key results, advantages and limitations.

Author(s) and Reference	Year	Methodology	Imaging Technique	Disease	Results	Merits	Limitations
Irshad et al. [15]	2014	SVM-based classification using intensity features	Fundus	HR	Accuracy: 81.3%	Simple and interpretable method	Small dataset, low accuracy
Abbas et al. [16]	2021	DenseNet-based semantic segmentation	Fundus	HR	Accuracy: 92.6% Sensitivity: 90.5% Specificity: 91.5%	High classification performance across stages	Limited scalability to other datasets
Muthukannan [10]	2022	CNN-based feature extraction and SVM	Fundus	AMD, DR, Cataract, Glaucoma, Normal	Accuracy: 98.30% Recall: 95.27% Specificity: 98.28% F1-Score: 93.3%	Very high accuracy and recall	Dataset-specific optimization; may overfit
Kihara et al. [11]	2022	ViT-based segmentation model using encoder–decoder architecture	OCT	neMNV	Sensitivity: 82% Specificity: 90% PPV: 79% NPV: 91% AUC: 0.91	Strong generalization; human-level performance	Focused only on neMNV in AMD
Shoukat et al. [6]	2023	ResNet-50 architecture with transfer learning and data augmentation	Fundus	Glaucoma	Accuracy: 98.48% Sensitivity: 99.30% Specificity: 96.52% AUC: 97% F1-Score: 98%	Excellent performance with transfer learning	Limited generalizability; single disease
Patel et al. [7]	2024	FAWT and GKDG based feature extraction and LS-SVM.	Fundus	Glaucoma	Accuracy: 95.84% Specificity: 97.17% Sensitivity: 94.55%	Effective feature engineering approach	Complex preprocessing pipeline
Gu et al. [12]	2024	A ranking-based multi-scale feature calibration network	OCT	Cataract	Accuracy: 88.86% Sensitivity: 89.08% Precision: 90.15% F1-Score: 89.49%	Fine-grained severity classification	Narrow disease focus
Sharma et al. [8]	2024	Customized CNN, PCA + LDA for dimensionality reduction, ELM optimized with MOD-PSO	Fundus	Glaucoma	G1020 Dataset: Accuracy: 97.80% Sensitivity: 94.92% Specificity: 98.44% ORIGA Dataset: Accuracy: 98.46% Sensitivity: 97.01% Specificity: 98.96%	Optimized feature selection boosted results	Relatively complex architecture
Kulyabin et al. [14]	2024	ResNet50 and VGG16	OCT	AMD, DME, ERM, RAO, RVO, VID, Normal	ResNet50: Accuracy: 84.6% Precision: 89.8% Recall: 84.6% F1-Score: 86.6% VGG16: Accuracy: 85.9% Precision: 88.8% Recall: 85.9% F1-Score: 86.9%	Benchmarking with multiple models and diseases	Accuracy moderate compared to others

Table 1. Cont.

Author(s) and Reference	Year	Methodology	Imaging Technique	Disease	Results	Merits	Limitations
Geetha et al. [9]	2025	EfficientNetB4, Snapshot Ensemble, Aquila Optimization and V-Net	Fundus	Glaucoma	Accuracy: 99.35% Precision: 99.04% Specificity: 99.19% Recall: 98.89% F1-Score: 98.97%	Extremely high accuracy and precision	High computational requirements
Liu et al. [13]	2025	STMF-DRNet	Fundus	DR	Accuracy: 77% Sensitivity: 77% Specificity: 94.2% F1-Score: 77% Kappa: 87.7%	Advanced attention mechanisms for DR	Moderate accuracy; DR-specific only
Suman et al. [17]	2025	Hybrid DL model (ResNet-50 and ViT)	Fundus	HR	Accuracy: 96.88% Sensitivity: 94.35% Specificity: 97.66% F1-Score: 94.42%	Best-in-class hybrid performance	Resource-intensive and complex

2. Materials and Methods

In this study, the first step was to prepare the dataset. Then, 14 well-known CNN models were tested and their performances were evaluated. The three most successful models were used as feature extractors and the features obtained from these models were fused (concatenated) and classified with SVM, RF and XGBoost algorithms. After determining the best-performing ML model, feature selection was performed on the fused features with meta-heuristic algorithms. The details of these methods are explained in the following subsections. Figure 1 shows the schematic diagram of the proposed three-stage approach.

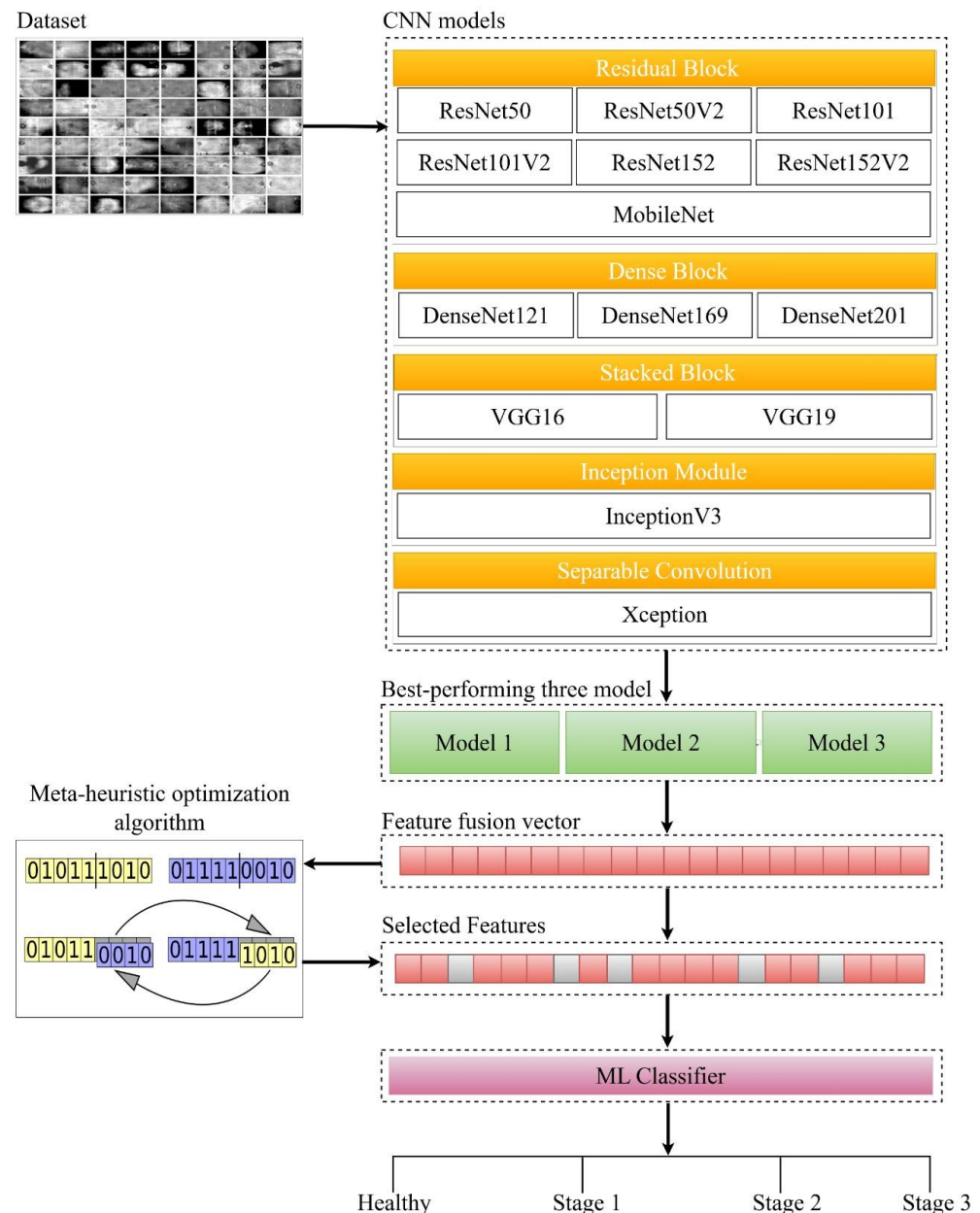


Figure 1. Overview of the proposed three-stage diagnostic framework for HR classification, integrating DL based feature extraction, ML classification and metaheuristic optimization for feature selection.

2.1. Dataset

The HR dataset used in this study consists of specially collected OCT images. The images were obtained from patients from Turkey, and all patient identifiers were anonymized and protected in accordance with ethical rules. The dataset was created using OCT imaging and manually labeled by an expert ophthalmologist. The labeling process was based on clin-

ical criteria that define the different stages of hypertensive retinopathy. The dataset consists of a total of 1875 OCT images (512×512 pixels), and each category is categorized according to a specific stage of the disease. All images have been rescaled to 224×224 pixels to match the input layers of the models.

Figure 2 shows typical OCT images selected for each HR stage. These samples were selected to clearly demonstrate how our model utilizes the distinctive morphological features of each stage. The identification for each disease category is presented in Table 2, and the distribution of images across these categories is visualized in Figure 3. The data distribution is organized to provide a balanced structure to increase the generalization capacity of the model.

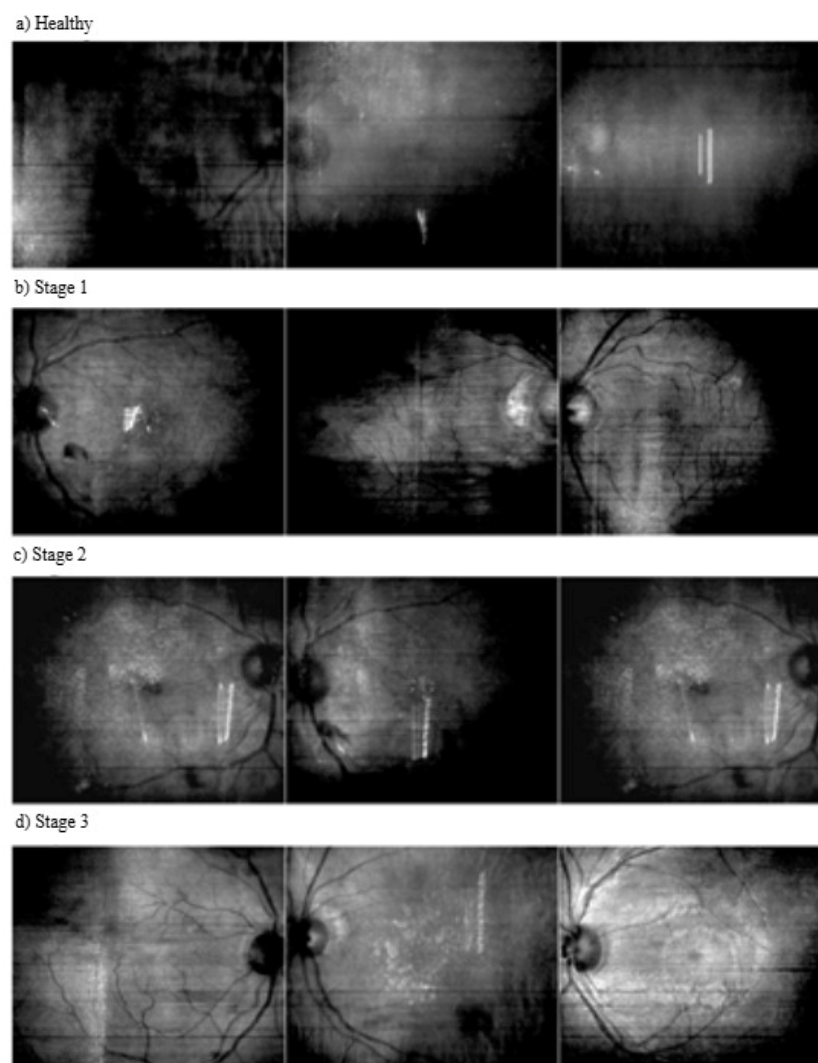


Figure 2. Representative OCT images illustrating stages of hypertensive retinopathy: (a) healthy retina; (b) Stage 1 with mild arterial narrowing; (c) Stage 2 showing prominent arteriovenous crossings and early atherosclerotic changes; (d) Stage 3 with severe vascular damage including hemorrhages and exudates.

Table 2. Clinical identification criteria for each HR stage within the OCT image dataset.

Category	Identification
Normal	No hypertension-related changes or abnormalities in the retina.
Stage 1	A condition with mild arterial narrowing and thickening of the vessel walls.
Stage 2	A condition with more pronounced vasoconstriction, arteriovenous crossing and atherosclerosis.
Stage 3	A condition involving serious vascular disorders with hemorrhages, exudates and cotton wool-like spots in the retina.

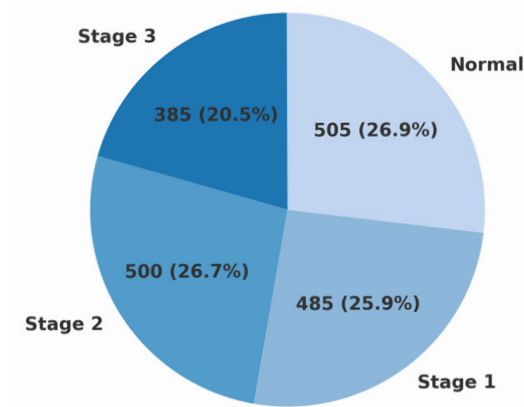


Figure 3. Distribution of the dataset by hypertensive retinopathy stages: Normal (505 images), Stage 1 (485 images), Stage 2 (500 images), and Stage 3 (385 images).

The dataset is divided in a balanced way to train CNN models more efficiently. In this context, 80% of the dataset is used for training purposes, while the remaining 20% is used for testing and validation. During the training phase, the images were randomly rotated between 0 and 30%, horizontally and vertically shifted up to 40% of the image size and diversified with 0–30% shear and zoom transformations for data augmentation.

2.2. CNN

In the field of AI, CNNs are one of the most common and effective methods for analyzing visual data [18]. CNNs are used in complex tasks such as classification, as demonstrated by Pacal et al. [19,20] and Liang et al. [21]; segmentation, as shown in the studies by Zhong et al. [22] and Gao et al. [23]; and object detection and diagnostic modeling, as explored by Shoaib et al. [24], Muthusamy and Palani [25] and Mewada et al. [26]. Their flexible architecture enables high performance in a wide range of applications.

A classic CNN model consists of three basic components: convolution layers, pooling layers and fully connected layers. The Schematic architecture of a classical CNN is shown in Figure 4. Convolution layers extract features from the input data through filters, while pooling layers reduce the computational burden by minimizing the size and preventing overlearning. Fully connected layers use these features to perform classification. Many CNN architectures have been developed using these basic structures. ResNet [27] solves learning problems in deep networks by using residual learning. DenseNet [28] increases the flow of information with dense connections, connecting each layer to all subsequent layers and preventing information loss. VGG [29] achieves balanced performance by stacking small filters on top of each other. Xception [30] reduces computational cost and increases efficiency by using depth separable convolutions. Inception [31] performs robust feature extraction with parallel filters of different sizes. MobileNet [32] is optimized for mobile and embedded devices using depth-separable convolutions, reducing size and computational cost.

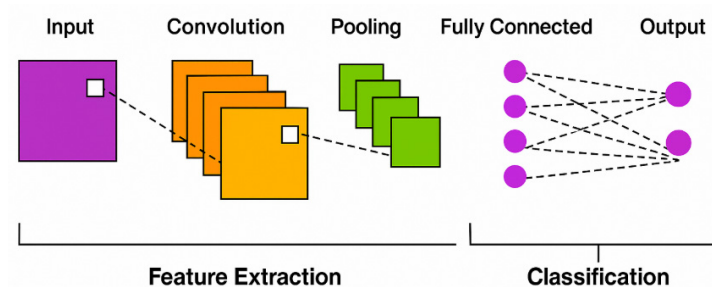


Figure 4. Schematic architecture of a classical CNN highlighting convolutional, pooling and fully connected layers.

2.3. Transfer Learning

Transfer Learning is the process by which an AI model adapts previously learned knowledge to a different but similar problem [33]. DL models are often trained on large and diverse datasets; however, training a model from scratch for each new problem can be both time-consuming and costly. Transfer learning speeds up this process, enabling more efficient results with less data. In this approach, the lower layers of a pre-trained model are usually retained because they have learned general features. The upper layers are retrained specifically for the new problem or completely replaced. In this way, the model can adapt faster to the new dataset and achieve high accuracy rates. It is a widely used technique, especially in areas such as image recognition, natural language processing and audio analysis.

2.4. ML Algorithms

ML algorithms are models that automatically learn from data and perform certain tasks without human intervention [34]. These algorithms make predictions or decisions by learning patterns from data. Some of the most widely used ML algorithms are as follows.

2.4.1. SVM

SVM [35] is a supervised learning algorithm that tries to find the optimal hyperplane that provides the best separation between classes. It can produce particularly effective results on high-dimensional datasets and perform transformations using kernel functions for non-linearly separable data. The linear kernel is the simplest function used when the data can be linearly separated and aims to separate the hyperplane with maximum margin. The polynomial kernel is used to model non-linear relationships and increases in complexity as the degree increases. The radial basis function kernel transforms data into higher dimensional spaces, creating complex decision boundaries and is particularly effective with low-dimensional but non-linearly separable datasets. The sigmoid kernel works similarly to activation functions in neural networks and in some cases can produce similar results to DL methods.

2.4.2. RF

RF [36] is a robust ensemble learning technique that works as an ensemble of decision trees and is supported by bagging. Each tree is trained on different subsets generated by bootstrap sampling of the training data and using randomly selected feature subsets. The process ensures that each tree captures different patterns; as a result, the predictions of all trees are combined by majority voting or averaging. Thus, the variance of the model is reduced, while the risk of overfitting is significantly reduced. RF can effectively model complex data structures, offering high accuracy and generalization capacity in both classification and regression problems.

2.4.3. XGBoost

XGBoost [37] is an enhanced version of the gradient boosting algorithm and is known for its high performance, especially on large datasets. A tree-based model, XGBoost builds successive decision trees to reduce errors, with each new tree attempting to correct errors that the previous trees failed to do. Through its regularization techniques and parallel computing capabilities, XGBoost avoids overlearning and runs fast.

In this study, the three ML algorithms mentioned above were used to classify deep features. They were chosen because they are widely used in the literature, achieve high success rates and have strong generalization capabilities.

2.5. CNN-Based Feature Extraction and Fusion

Machine learning-based image analysis often relies on manually extracted features. However, this method requires domain knowledge and is difficult to apply especially for

complex data patterns. DL methods, on the other hand, eliminate the need for manual feature engineering by learning features automatically. In this study, the best-performing three models among 14 different CNN models are identified and the features extracted from these models over the Global Average Pooling (GAP) layer are fused. The feature fusion method aims to create a more comprehensive feature set that can discriminate between classes by combining the features learned by different models. The feature vectors extracted from the three different CNN models are combined using the mathematical formula in Equation (1):

$$F1 = GAP(f_{CNN1}(X)), F2 = GAP(f_{CNN2}(X)), F3 = GAP(f_{CNN3}(X)) \quad (1)$$

$$F_{fusion} = [F1 \parallel F2 \parallel F3] \quad (2)$$

In Equation (1), X is the input image, $f_{CNNi}(\cdot)$ is the feature extraction function of the corresponding CNN model, $GAP(\cdot)$ is the Global Average Pooling process and F_i is the feature vector extracted from the GAP layer. In Equation (2), the extracted feature vectors are combined. The operator $\parallel$ denotes the merging of the feature vectors on the horizontal axis. F_{fusion} is classified with the ML algorithms mentioned in Section 2.4. Figure 5 illustrates the feature fusion process, where deep features extracted from three CNN models are concatenated to create a more comprehensive feature representation for hypertensive retinopathy classification.

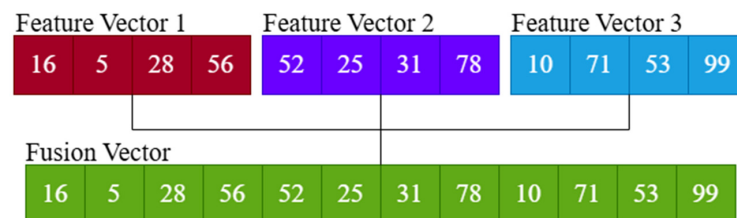


Figure 5. Feature fusion workflow combining deep feature vectors extracted from three distinct CNN architectures to create a comprehensive representation for HR classification.

2.6. Feature Selection with Metaheuristic Optimization Algorithms

Although feature fusion is a robust method that improves classification performance, some features can cause noise and negatively affect the success of the model [38]. Therefore, determining the optimal feature subset and selecting the variables that best represent the dataset is a critical step. Metaheuristic optimization algorithms are widely used to improve the accuracy of the model by identifying the most appropriate features in the large search space and reduce the computational cost by eliminating redundant features.

These algorithms try to maximize a fitness function while identifying the best subset of features, and at the same time preserving the simplicity of the model, avoiding over-learning. Their flexibility allows them to handle non-linear relationships, interactions between features and problem-specific constraints. Feature selection is usually performed using a binary optimization approach, where each feature is expressed as a binary vector. For a dataset with N features, a binary vector of size N is created, where a value of 0 indicates that the feature is excluded and a value of 1 indicates that the feature is included. The optimization process starts with an initial population or initial solution and is iteratively updated to determine the best feature subset. The binary vector-based feature selection process is shown in Figure 6. In all feature selection processes that included metaheuristic optimization algorithms (GA, ABC, PSO and HHO), the F1 score was used as the fitness function. This decision was made to provide a balanced assessment between precision and recall in multi-class classification, especially due to the clinical importance of minimizing false negatives in the diagnosis of hypertensive retinopathy.

In this study, different approaches based on metaheuristic optimization were used to select the most suitable features. GA from the evolutionary algorithms group, ABC and

PSO from the nature-inspired methods, as well as HHO inspired by predator behavior were evaluated. Table 3 provides details of the optimization algorithms used. These algorithms were selected based on their widespread use in the literature and their successful results. Each of them aims to improve classification performance by using different strategies in the feature selection process, to make the model noise-free and more generalizable.

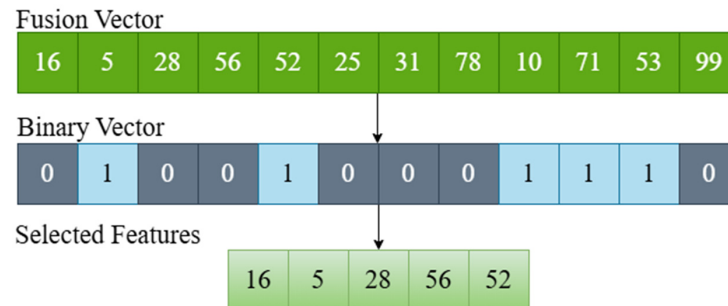


Figure 6. Feature selection process using metaheuristic optimization algorithms to refine the fused feature set by eliminating redundant features and improving classification accuracy.

Table 3. Overview of metaheuristic optimization algorithms used for feature selection: working principle, pseudocode, merits and limitations.

Algorithm	Working Principle	Pseudocode	Merits	Limitations
GA	Based on the principles of natural selection and genetic evolution. Solutions are improved from generation to generation using genetic operators (selection, crossover, mutation). New solutions are generated by selecting the best individuals and the process continues until the optimum solution is reached.	Initialize population X randomly for $t = 1 \dots T$: evaluate fitness $f(X)$ $P \leftarrow$ select parents from X $C \leftarrow$ crossover P with rate α mutate C with rate pm $X \leftarrow$ form new generation from X and C return best individual	Robust global search in complex search spaces	Requires parameter configuration; convergence speed is moderate
ABC	Mimics the behavior of honeybees searching for food sources. Three types of bees (worker, observer and explorer) try to find the best solutions. While developing good solutions, they also keep discovering new ones.	Initialize nectar sources $\{x_i\}$ randomly repeat: // Employed bees for each x_i : $x'_i \leftarrow x_i + \varphi \cdot (x_i - x_k)$ if $f(x'_i) > f(x_i)$: $x_i \leftarrow x'_i$ // Onlooker bees compute $p_i = f(x_i) / \sum f(x_j)$ select sources by p_i and apply employed-bee step // Scout bees if $trial[i] > limit$: $x_i \leftarrow LB + rand \cdot (UB - LB)$ until stopping criterion return best x_i	Good exploration diversity; simple implementation	Convergence speed moderate; exploitation could improve
PSO	Emulates the collective movements of flocks of birds and schools of fish. Each particle (solution candidate) moves in line with its own best position and the swarm's best position. Velocity updates are based on cognitive (personal best) and social (swarm best) components.	Initialize each particle i : position x_i , velocity v_i , personal best $p_i = x_i$ $g \leftarrow$ best of all p_i for $t = 1 \dots T$: for each particle i : $v_i \leftarrow w \cdot v_i + c1 \cdot rand \cdot (p_i - x_i) + c2 \cdot rand \cdot (g - x_i)$ $x_i \leftarrow x_i + v_i$ if $f(x_i) < f(p_i)$: $p_i \leftarrow x_i$ if $f(p_i) < f(g)$: $g \leftarrow p_i$ return g	Fast convergence; few parameters to tune	Risk of local optima; may need restarts
HHO	Inspired by the hunting strategies of Harris's hawks. It works in a balanced way between exploration (searching for prey) and exploitation (capturing prey), both seeking new solutions and improving existing ones.	Initialize hawks x_i randomly within $[LB, UB]$ for $t = 1 \dots T$: $E_0 \leftarrow$ random in $(-1, 1)$ $E \leftarrow 2 \cdot E_0 \cdot (1 - t/T)$ $X_{best} \leftarrow$ hawk with best f if $ E \geq 1$: // exploration update X_i by random exploration formulas else: // exploitation $J \leftarrow 2 \cdot (1 - rand)$ if $ E \geq 0.5$: $X_i \leftarrow X_{best} - E \cdot J \cdot X_{best} - X_i $ else: $\Delta \leftarrow X_{best} - X_i$ $X_i \leftarrow \Delta - E \cdot \Delta $ // optional: apply Lévy flight for further diversification return X_{best}	Balanced exploration–exploitation; strong adaptability	Computationally more intensive; moderate implementation complexity

3. Experimental Setup

In all experiments, models are trained in the same hardware and software environment to ensure a fair comparison. The hyperparameters and experimental settings used are standardized to best evaluate the model performance. The details of the experimental setup are detailed in the subsections.

3.1. Experiment Setting

All experiments in this study were carried out on a system equipped with high-performance hardware and a modern software infrastructure. The details of the experimental environment, including specifications such as the operating system, programming language, framework versions, hardware components and CUDA version, are comprehensively presented in Table 4.

Table 4. Specifications of the hardware and software environment employed for model training and evaluation.

Configuration	Parameter
Operating System	Windows 11, 64 Bit
Programming Language	Python version 3.11.4
Frameworks	Tensorflow 2.14.0, Keras version 2.11.4, Matplotlib 3.7.1
GPU	2 × Nvidia RTX 3090 24 GB
CPU	Intel(R) Core(TM) i9-10920X CPU @ 3.50 GHz
RAM	128 GB
CUDA	v12.7

A standardized set of hyperparameters was used to ensure that all DL models were trained under the same conditions. The chosen hyperparameters and their values are presented in Table 5. The hyperparameter values in Table 5 were determined through a grid search on the validation set, where batch sizes {32, 64, 128}, learning rates $\{1 \times 10^{-4}, 3 \times 10^{-4}, 1 \times 10^{-3}\}$ and optimizers Adam and SGD were compared.

Table 5. Hyperparameters used for training CNN models, including batch size, number of epochs, learning rate, optimizer and weight decay rate.

Batch Size	Epoch	Learning Rate	Optimizer	Weights_Decay_Rate
64	100	0.0003	Adam	0.9

3.2. Evaluation Metrics

The classification performance of DL models can be evaluated by metrics such as accuracy, precision, sensitivity, F1 score and Cohen's kappa coefficient. These metrics are derived from the confusion matrix to analyze the prediction results of the model by class. The classes in the confusion matrix are as follows. True Positive (TP): instances that the model predicts as positive and are actually positive. True Negative (TN): instances that the model predicts as negative and are actually negative. False Positive (FP): instances that the model predicts as positive but are actually negative. False Negative (FN): instances that the model predicts as negative but are actually positive. Table 6 summarizes the classification metrics, including their mathematical formulas and detailed explanations.

Table 6. Performance metrics, equations and detailed descriptions.

Metric	Equation	Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Refers to the proportion of the model's total predictions that are correct.
Precision	$\frac{TP}{TP+FP}$	Refers to the proportion of positive predicted samples that are actually positive.
Recall	$\frac{TP}{TP+FN}$	Measures how many true positive samples are correctly predicted.
F1-Score	$\frac{2 \times (Precision \times Recall)}{Precision + Recall}$	Provides the balance between precision and sensitivity.
Cohen's kappa score	$\frac{p_0 - p_e}{1 - p_e}$	Assesses how well the model performs compared to random guessing and is calculated based on observed accuracy (p_0) and expected accuracy (p_e).

4. Results

This section analyzes the experimental results. First, various CNN models were tested on the HR dataset and the three best models were selected based on their classification performance. These models were then used as feature extractors for deep feature fusion. The fused features were classified using different ML algorithms to determine the best one. To further improve the performance, metaheuristic optimization was applied for feature selection and the final model was built using the best machine learning algorithm. Finally, all results were compared and analyzed.

4.1. CNN Results

In this section, the performance results of 14 different CNN models tested for HR classification are presented and analyzed. The classification performance of CNN models is provided in Table 7. DenseNet169 was the most successful model with 87.73% accuracy, 87.75% precision, 87.73% recall, 87.67% F1-score and 0.8359 kappa value. The high recall value of the model indicates that patients with HR can be successfully detected. Also, the high precision and F1-score values indicate that the overall accuracy of the model is balanced and the false positive rate is low. The Cohen's kappa value is as high as 0.83, which proves that the model performs much better than a random guessing classifier. MobileNet, the second most successful model, performed quite competitively with 86.40% accuracy, 86.60% precision, 86.40% recall, 86.31% F1-score and 0.8180 kappa value. In particular, the precision value of 86.60% indicates that most of the positive samples predicted by the model are indeed positive.

Table 7. Classification performance of CNN models evaluated on the test data.

Model	Accuracy	Precision	Recall	F1-Score	Kappa
DenseNet121	82.93	82.68	82.93	82.67	0.7717
DenseNet169	87.73	87.75	87.73	87.67	0.8359
DenseNet201	85.07	84.91	85.07	84.83	0.8002
InceptionV3	82.13	81.67	82.13	81.60	0.7616
MobileNet	86.40	86.60	86.40	86.31	0.8180
ResNet101	85.07	84.88	85.07	84.92	0.8001
ResNet101V2	85.33	85.25	85.33	85.00	0.8037
ResNet152	85.87	86.01	85.87	85.83	0.8188
ResNet152V2	83.20	83.14	83.20	82.91	0.7754
ResNet50	84.88	84.57	84.88	84.62	0.7965
ResNet50V2	80.27	79.64	80.27	79.42	0.7362
VGG16	85.87	85.38	85.07	85.06	0.8083
VGG19	85.87	85.64	85.87	85.41	0.8108
Xception	84.27	84.18	84.27	84.13	0.7895

The third best performing model, ResNet152, has a very high success rate with 85.87% accuracy, 86.01% precision, 85.87% recall, 85.83% F1-score and 0.8188 kappa value. The precision value of 86.01% shows that the model keeps the false positive rate low. The recall value of 85.87% indicates that most of the HR cases were correctly detected. The kappa coefficient of 0.81 proves that the model has a statistically significant success and provides significantly better results compared to random prediction.

The models that performed less well include DenseNet121, ResNet101, VGG16, VGG19, Xception and InceptionV3. In particular, VGG16 and VGG19 showed acceptable results with an accuracy of 85.87% but exhibited a more uneven distribution compared to models with lower precision, recall and F1-score values. The fact that the VGG architecture is older and requires more parameters compared to advanced CNN architectures

limited the generalization ability of the model and caused it not to be as successful as expected on the HR dataset.

Remarkably, the InceptionV3 and Xception models performed lower than expected with 82.13% and 84.27% accuracy rates. Although these models are known for their ability to extract features at different scales, the low recall and F1-score values suggest that the generalization ability of the model is weakened by overfitting in certain classes. Especially Xception, despite its 84.27% accuracy rate, shows an uneven distribution in precision and recall values, indicating that the model is prone to misclassification.

Figure 7 shows the confusion matrix of the three most successful CNN models. For 0 (Healthy), all models have a high accuracy rate. For 0 (Healthy), DenseNet169 predicted 91 correct predictions, while MobileNet and ResNet152 predicted 92 and 90 correct predictions for 0 (Healthy), respectively. For Stage 1, the errors are more pronounced, with MobileNet and ResNet152 often misclassifying this class as 0 or 2. DenseNet169 made 73 correct predictions, while MobileNet and ResNet152 made 69 correct predictions. For Stage 2, DenseNet169 made 89 correct predictions, while MobileNet and ResNet152 correctly predicted 87 instances. The models made the most errors between classes 1 and 2. In the Stage 3 class, all models correctly predicted 76 instances, while only 1–2 instances were incorrectly predicted. Overall, DenseNet169 gave the most consistent results, while MobileNet and ResNet152 made more errors, especially in classes 1 and 2.

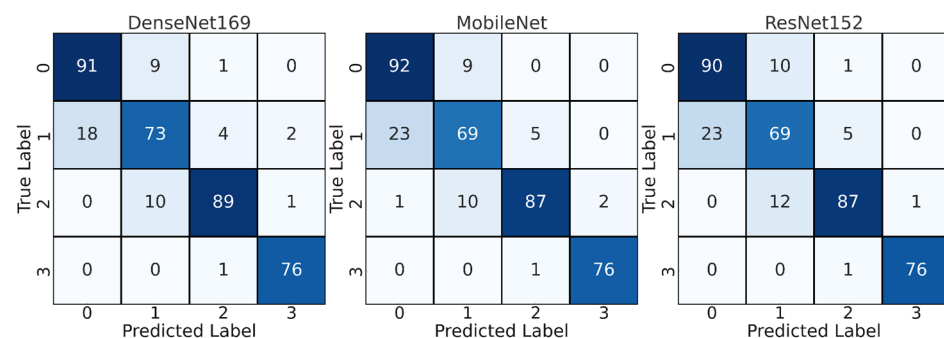


Figure 7. Confusion matrices of three best-performing three CNN models (0: Healthy, 1: Stage 1, 2: Stage 2, 3: Stage 3).

As a result of the overall comparison of CNN models, DenseNet169, MobileNet and ResNet152 were determined as the CNN models with the highest accuracy rate for HR diagnosis. In the next stage, these three models were used as feature extractors to extract deep features and combined with the feature fusion method to further improve the classification performance.

4.2. Feature Fusion Results

In this section, the three most successful models, DenseNet169, MobileNet and ResNet152, were used as feature extractors and the features obtained from the three different models were combined by feature fusion. Then, these fused feature vectors were classified by SVM, RF and XGBoost algorithms. The classification performances of the ML algorithms are presented in Table 8.

The experimental results show that the SVM model achieves the highest performance when the sigmoid kernel function is used. SVM (sigmoid) outperformed other ML models with 92.00% accuracy, 91.93% precision, 92.00% recall, 91.91% F1-score and 0.8930 kappa value. Notably, SVM's polynomial and RBF kernels also provided high accuracy but were inferior to the sigmoid function. The SVM (polynomial) model achieved 91.14% accuracy, while the SVM (rbf) model achieved 90.11% accuracy.

Table 8. Performance results of feature fusion stage on the test dataset.

Model	Accuracy	Precision	Recall	F1-Score	Kappa
SVM (linear)	88.77	88.82	88.77	88.72	0.8497
SVM (polynom)	91.14	91.16	91.14	91.14	0.8818
SVM (rbf)	90.11	90.16	90.11	90.05	0.8677
SVM (sigmoid)	92.00	91.93	92.00	91.91	0.8930
RF	87.97	87.98	87.97	87.91	0.839
XGBoost	88.60	88.52	88.59	88.62	0.8480

When other ML algorithms were analyzed, the RF model showed a relatively low performance with 87.97% accuracy, 87.98% precision, 87.97% recall and 87.91% F1-score. The XGBoost model performed slightly better than RF with an accuracy of 88.60%. In general, the SVM (sigmoid) model achieved the highest performance in terms of both accuracy and other evaluation metrics and was selected as the best ML model.

The results of this stage show that combining features from different CNN models improves classification performance and provides better results compared to traditional single model approaches. Therefore, in the next stage, feature selection was performed using metaheuristic optimization algorithms and classification was performed with the best ML model.

The confusion matrix in Figure 8 shows the classification performance after feature fusion with the SVM (sigmoid) model. Compared to the individual CNN models, the accuracy increased in all classes. In the 0 (Healthy) class, with 95 correct predictions; the success rate was higher than the previous models. In the 1 (Stage 1) class, 80 correct predictions were made, while misclassifications decreased. Especially in the 2 (Stage 2) class, the best result was achieved with 93 correct predictions and a significant improvement compared to the individual CNN models. In the 3 (Stage 3) class, the highest accuracy was achieved with 77 correct predictions. In general, the SVM (sigmoid) model after feature fusion produced better results than the performance of CNNs alone and made the discrimination between classes more successful.

SVM (Sigmoid)

0	95	5	1	0
1	10	80	5	2
2	0	5	93	2
3	0	0	0	77
	0	1	2	3

True Label

Predicted Label

Figure 8. Confusion matrix of SVM model with sigmoid kernel (0: Healthy, 1: Stage 1, 2: Stage 2, 3: Stage 3).

4.3. Feature Selection Results

In this section, metaheuristic optimization algorithms are used to select the most important features from the large feature set obtained after feature fusion and to improve classification performance. GA, PSO, ABC and HHO algorithms were used for feature selection and the best method was determined. The experimental results are shown in Table 9. Experimental results show that the HHO algorithm provides the highest classification accuracy. Feature selection using HHO and SVM (sigmoid) model achieved 94.66% accuracy; this result shows that the selection made after feature fusion increases the

generalization ability of the model and significantly improves the classification accuracy. When the results obtained with other optimization algorithms are compared, GA achieved 93.23% accuracy, ABC algorithm achieved 93.72% accuracy and PSO achieved 93.23% accuracy. The higher accuracy of HHO indicates that it is the most appropriate feature selection strategy for HR classification. In general, feature selection with metaheuristic optimization algorithms improved classification success by removing redundant or low-impact features from the model. The HHO algorithm was determined to be the best method as it provided the highest accuracy and consistency and was used in the final model.

Table 9. Feature selection performance of different metaheuristic optimization techniques.

Model	Accuracy	Precision	Recall	F1-Score	Kappa
GA	93.23	93.27	93.23	93.24	0.909
ABC	93.72	93.73	93.72	93.72	0.916
PSO	92.96	92.98	92.96	92.97	0.9061
HHO	94.66	94.66	94.66	94.64	0.9286

The confusion matrix shown in Figure 9, obtained after feature selection using the HHO algorithm, demonstrates a significant improvement in classification performance compared to the SVM (sigmoid) model employed during the feature fusion stage. Especially in the 0 (Healthy) class, while 95 correct predictions were made in the feature fusion stage, this number increased to 96 with HHO and the number of incorrect predictions decreased. Similarly, in the 1 (Stage 1) class, while 80 correct predictions were made with feature fusion, this number increased to 86 with HHO and incorrect predictions decreased significantly. One of the areas where the model showed the greatest improvement was in the 2 (Stage 2) class. While 93 correct predictions were made in this class in the feature fusion phase, this value increased to 96 with HHO and the rate of incorrect predictions decreased. Especially in the feature fusion stage, the errors made between classes 1 and 2 were minimized by feature selection. In the Stage 3 class, 77 correct predictions were made with feature fusion and this value did not change with HHO, which shows that the model already classifies the most advanced stage diseases with high accuracy. In general, feature selection with the HHO algorithm provided better discrimination, especially in classes 1 and 2, and increased the overall accuracy of the model by reducing false predictions. Although using all features in the feature fusion stage improves the performance of the model, filtering out unnecessary features allows the model to generalize better and the distinction between classes becomes clearer. These results show that optimization with HHO is the best method for HR classification and the model provides higher accuracy.

		HHO			
True Label	0	96	5	0	0
	1	8	86	2	1
	2	0	3	96	1
	3	0	0	0	77
		0	1	2	3
		Predicted Label			

Figure 9. Confusion matrix of SVM classification of features selected with HHO (0: Healthy, 1: Stage 1, 2: Stage 2, 3: Stage 3).

5. Discussion

In this section, the results obtained at different stages of the proposed method are analyzed comparatively. First, the classification results using only CNN models are analyzed. DenseNet169, MobileNet and ResNet152 were identified as the three most successful models, with DenseNet169 having the highest accuracy rate. However, the classification performance of CNN models alone could not fully resolve the ambiguities between classes. First, 14 different CNN models were tested and the three highest performing models were identified. DenseNet169, MobileNet and ResNet152 models were used separately for classification and the accuracy rates were calculated as 87.73%, 86.40% and 85.87%, respectively. Although the DenseNet169 model had the highest accuracy rate, it was observed that the model could not adequately distinguish between different HR stages and made misclassifications at some stages. This indicates that approaches based on a single CNN model may be insufficient for HR diagnosis. Table 10 compares classification performance at different stages of the proposed method.

Table 10. Comparative analysis of classification performance across different stages of the proposed method, including individual CNN models, feature fusion and metaheuristic-based feature selection.

Technique	Accuracy	Precision	Recall	F1-Score	Kappa
DenseNet169	87.73	87.75	87.73	87.67	0.8359
MobileNet	86.40	86.60	86.40	86.31	0.8180
ResNet152	85.87	86.01	85.87	85.83	0.8188
Feature Fusion SVM (Linear)	92.00	91.93	92.00	91.91	0.8930
Feature Selection (HHO)	94.66	94.66	94.66	94.64	0.9286

In the feature fusion stage, the deep features of the three CNN models with the highest accuracy were combined and classified with different ML algorithms. The combined feature set was classified with SVM, RF and XGBoost algorithms. The results showed that the SVM (sigmoid) model achieved the highest accuracy. At this stage, the accuracy rate increased to 92.00%, which is a significant improvement over methods using only individual CNN models. Feature fusion significantly improved the performance of the model compared to individual CNN models. In particular, the misclassifications between stages 1 and 2 of HR decreased and the generalization capacity of the model increased.

In the third stage, feature selection was performed using GA, ABC, PSO and HHO to further improve the accuracy of the model. After feature selection, the highest accuracy rate of 94.66% was obtained using the HHO method. Through the optimization process, the complexity of the model was reduced, redundant features were filtered out and the overall classification accuracy was improved. Especially in classes 1 and 2, misclassifications decreased and the generalization ability of the model improved. In particular, the performance of the model improved by up to 7% in the middle and advanced stages of HR (stages 2 and 3). When the overall comparison is made, there is a significant improvement from the results obtained with CNN models alone to the feature fusion stage, and the greatest success is achieved after feature selection. This process shows that it is critical to use both DL and optimization techniques together to achieve the best performance in HR classification.

As shown in Figure 10, accuracy increases from 87.73% with DenseNet169 to 92.00% (+4.27 points) after feature fusion and reaches 94.66% (+2.66 points) after HHO-based feature selection. Precision, recall and F1 score also show the same upward trend. This significant improvement highlights the cumulative benefit of the pooling and selection steps in improving classification performance.

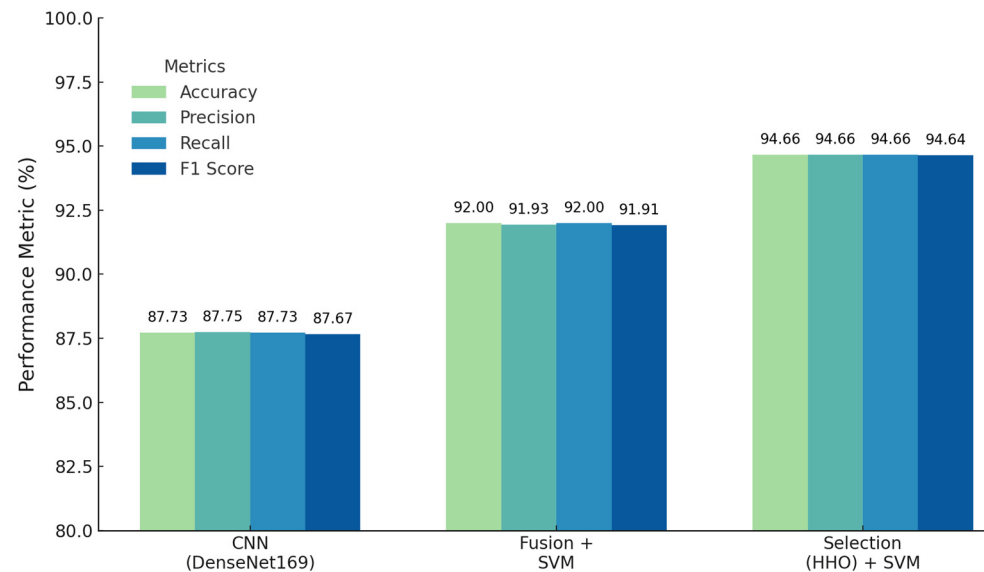


Figure 10. Comparative performance metrics (Accuracy, Precision, Recall, F1 Score) of the proposed hypertensive retinopathy classification approach across three stages: individual CNN model (DenseNet169), feature fusion combined with SVM classification, and metaheuristic-based feature selection using HHO integrated with SVM.

6. Limitations and Future Work

Although this study achieved significant results, there are some limitations: the dataset used was collected from a single hospital, and the sample size and diversity are limited. In the future, it is planned to evaluate the model on broader and more heterogeneous datasets obtained from different centers in order to increase its generalizability. Furthermore, since the model obtained is difficult for expert physicians to interpret in decision-making processes, explainable artificial intelligence techniques will be integrated to make the decision mechanisms more transparent.

7. Conclusions

Early diagnosis of HR is critical to preventing vision loss. In this study, a novel three-stage artificial intelligence method that provides high accuracy and generalization in the multi-stage classification of HR is proposed. In the first stage, the three most successful models were selected from 14 different CNN models (DenseNet169, MobileNet, ResNet152), and the deep features extracted from these models were combined. In the second stage, these combined features were classified using various ML algorithms. The SVM sigmoid kernel achieved the best performance with 92% accuracy. In the third stage, feature selection was performed using HHO, a metaheuristic optimization algorithm, thereby increasing accuracy to 94.66%.

This method combines the information content of different CNN models and removes unnecessary features using HHO, thereby providing a clear advantage over CNN models alone. As a result, classification success and model generalization ability have been significantly improved in the early and advanced stages of HR.

In summary, this study presents an innovative and applicable approach that effectively integrates DL and metaheuristic optimization in the diagnosis of hypertensive retinopathy.

Author Contributions: M.Y. and S.B.Ş.: conceptualization, methodology, review and editing, software, validation, visualization, writing—original draft; Ş.T.: conceptualization, methodology, supervision; S.B.: data collection, data labeling. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was evaluated by the Local Ethics Committee at the meeting dated 6 May 2025 and approved with the decision numbered 2025/281. This study, which was carried out within the scope of the research project titled “Diagnosis of Hypertensive Retinopathy from Fundus Images with Deep Learning”, was carried out in accordance with ethical principles and designed in accordance with scientific and academic rules.

Informed Consent Statement: Patient consent was waived due to the retrospective nature of the study and the use of anonymized data extracted from the hospital database, approved by the institutional ethics committee.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to restrictions imposed by the ethics committee approval.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Shen, C.J.; Kry, S.F.; Buchsbaum, J.C.; Milano, M.T.; Inskip, P.D.; Ulin, K.; Francis, J.H.; Wilson, M.W.; Whelan, K.F.; Mayo, C.S.; et al. Retinopathy, optic neuropathy, and cataract in childhood cancer survivors treated with radiation therapy: A PENTEC comprehensive review. *Int. J. Radiat. Oncol. Biol. Phys.* **2024**, *119*, 431–445. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ba, M.; Li, Z. The impact of lifestyle factors on myopia development: Insights and recommendations. *AJO Int.* **2024**, *1*, 100010. [\[CrossRef\]](#)
- Uyar, K.; Yurdakul, M.; Taşdemir, Ş. Abc-based weighted voting deep ensemble learning model for multiple eye disease detection. *Biomed. Signal Process. Control* **2024**, *96*, 106617. [\[CrossRef\]](#)
- Shin, H.J.; Costello, F. Imaging the optic nerve with optical coherence tomography. *Eye* **2024**, *38*, 2365–2379. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, Y.; Daho, M.E.H.; Conze, P.-H.; Zeghlache, R.; Le Boité, H.; Tadayoni, R.; Cochener, B.; Lamard, M.; Quellec, G. A review of deep learning-based information fusion techniques for multimodal medical image classification. *Comput. Biol. Med.* **2024**, *177*, 108635. [\[CrossRef\]](#)
- Shoukat, A.; Akbar, S.; Hassan, S.A.; Iqbal, S.; Mehmood, A.; Ilyas, Q.M. Automatic diagnosis of glaucoma from retinal images using deep learning approach. *Diagnostics* **2023**, *13*, 1738. [\[CrossRef\]](#)
- Patel, R.K.; Chouhan, S.S.; Lamkuche, H.S.; Pranjal, P. Glaucoma diagnosis from fundus images using modified Gauss-Kuzmin-distribution-based Gabor features in 2D-FAWT. *Comput. Electr. Eng.* **2024**, *119*, 109538. [\[CrossRef\]](#)
- Sharma, S.K.; Muduli, D.; Priyadarshini, R.; Kumar, R.R.; Kumar, A.; Pradhan, J. An evolutionary supply chain management service model based on deep learning features for automated glaucoma detection using fundus images. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107449. [\[CrossRef\]](#)
- Geetha, A.; Sobia, M.C.; Santhi, D.; Ahilan, A. DEEP GD: Deep learning based snapshot ensemble CNN with EfficientNet for glaucoma detection. *Biomed. Signal Process. Control* **2025**, *100*, 106989. [\[CrossRef\]](#)
- Muthukannan, P. Optimized convolution neural network based multiple eye disease detection. *Comput. Biol. Med.* **2022**, *146*, 105648.
- Kihara, Y.; Shen, M.; Shi, Y.; Jiang, X.; Wang, L.; Laiginhas, R.; Lyu, C.; Yang, J.; Liu, J.; Morin, R.; et al. Detection of nonexudative macular neovascularization on structural OCT images using vision transformers. *Ophthalmol. Sci.* **2022**, *2*, 100197. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gu, Y.; Fang, L.; Mou, L.; Ma, S.; Yan, Q.; Zhang, J.; Liu, F.; Liu, J.; Zhao, Y. A ranking-based multi-scale feature calibration network for nuclear cataract grading in AS-OCT images. *Biomed. Signal Process. Control* **2024**, *90*, 105836. [\[CrossRef\]](#)
- Liu, Y.; Yao, D.; Ma, Y.; Wang, H.; Wang, J.; Bai, X.; Zeng, G.; Liu, Y. STMF-DRNet: A multi-branch fine-grained classification model for diabetic retinopathy using Swin-TransformerV2. *Biomed. Signal Process. Control* **2025**, *103*, 107352. [\[CrossRef\]](#)
- Kulyabin, M.; Zhdanov, A.; Nikiforova, A.; Stepichev, A.; Kuznetsova, A.; Ronkin, M.; Borisov, V.; Bogachev, A.; Korotkich, S.; Constable, P.A.; et al. Octdl: Optical coherence tomography dataset for image-based deep learning methods. *Sci. Data* **2024**, *11*, 365. [\[CrossRef\]](#) [\[PubMed\]](#)
- Irshad, S.; Akram, M.U. Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy. In Proceedings of the 2014 Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 11–13 December 2014; IEEE: New York, NY, USA, 2015.
- Abbas, Q.; Qureshi, I.; Ibrahim, M.E. An automatic detection and classification system of five stages for hypertensive retinopathy using semantic and instance segmentation in DenseNet architecture. *Sensors* **2021**, *21*, 6936. [\[CrossRef\]](#)
- Suman, S.; Tiwari, A.K.; Sachan, S.; Singh, K.; Meena, S.; Kumar, S. Severity grading of hypertensive retinopathy using hybrid deep learning architecture. *Comput. Methods Programs Biomed.* **2025**, *261*, 108585. [\[CrossRef\]](#)

18. Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [[CrossRef](#)]
19. Pacal, I.; Ozdemir, B.; Zeynalov, J.; Gasimov, H.; Pacal, N. A novel CNN-ViT-based deep learning model for early skin cancer diagnosis. *Biomed. Signal Process. Control* **2025**, *104*, 107627. [[CrossRef](#)]
20. Pacal, I. Investigating deep learning approaches for cervical cancer diagnosis: A focus on modern image-based models. *Eur. J. Gynaecol. Oncol.* **2025**, *46*, 125–141.
21. Liang, J.; Liang, R.; Wang, D. A novel lightweight model for tea disease classification based on feature reuse and channel focus attention mechanism. *Eng. Sci. Technol. Int. J.* **2025**, *61*, 101940. [[CrossRef](#)]
22. Zhong, J.; Tian, W.; Xie, Y.; Liu, Z.; Ou, J.; Tian, T.; Zhang, L. PMFSNet: Polarized multi-scale feature self-attention network for lightweight medical image segmentation. *Comput. Methods Programs Biomed.* **2025**, *261*, 108611. [[CrossRef](#)] [[PubMed](#)]
23. Gao, Y.; Zhang, J.; Wei, S.; Li, Z. PFormer: An efficient CNN-Transformer hybrid network with content-driven P-attention for 3D medical image segmentation. *Biomed. Signal Process. Control* **2025**, *101*, 107154. [[CrossRef](#)]
24. Shoaib, M.R.; Emara, H.M.; Mubarak, A.S.; Omer, O.A.; El-Samie, F.E.A.; Esmail, H. Revolutionizing diabetic retinopathy diagnosis through advanced deep learning techniques: Harnessing the power of GAN model with transfer learning and the DiaGAN-CNN model. *Biomed. Signal Process. Control* **2025**, *99*, 106790. [[CrossRef](#)]
25. Muthusamy, D.; Palani, P. Deep neural network model for diagnosing diabetic retinopathy detection: An efficient mechanism for diabetic management. *Biomed. Signal Process. Control* **2025**, *100*, 107035. [[CrossRef](#)]
26. Mewada, H.; Pires, I.M.; Engineer, P.; Patel, A.V. Fabric surface defect classification and systematic analysis using a cuckoo search optimized deep residual network. *Eng. Sci. Technol. Int. J.* **2024**, *53*, 101681. [[CrossRef](#)]
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
28. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
30. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
31. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
32. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
33. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl.-Based Syst.* **2015**, *80*, 14–23. [[CrossRef](#)]
34. Alzubi, J.; Nayyar, A.; Kumar, A. Machine learning from theory to algorithms: An overview. *J. Phys. Conf. Ser.* **2018**, *1142*, 012012. [[CrossRef](#)]
35. Cortes, C.; Vapnik, V. Support-Vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
36. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
37. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
38. Yurdakul, M.; Uyar, K.; Tasdemir, S. Enhanced ore classification through optimized CNN ensembles and feature fusion. *Iran J. Comput. Sci.* **2025**, *8*, 491–509. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.