A Hybrid Images Deep Trained Feature Extraction and Ensemble Learning Models for Classification of Multi Disease in Fundus Images

Jyoti Verma[1], Isha Kansal[2], Renu Popli[2], Vikas Khullar[2], Daljeet Singh[3], Manish Snehi[1], and Rajeev Kumar[2](✉)

[1] Department of Computer Science and Engineering, Punjabi University Patiala, Patiala, Punjab, India

[2] Chitkara Institute of Engineering and Technology, Chitkara University, Jansla, Punjab, India
{isha.kansal,vikas.khullar}@chitkara.edu.in, rajeev.kumar@chitkara.edu

[3] Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
daljeet.singh@oulu.fi

Abstract. Retinal disorders, including diabetic retinopathy and macular degeneration due to aging, can lead to preventable blindness in diabetics. Vision loss caused by diseases that affect the retinal fundus cannot be reversed if not diagnosed and treated on time. This paper employs deep-learned feature extraction with ensemble learning models to improve the multi-disease classification of fundus images. This research presents a novel approach to the multi-classification of fundus images, utilizing deep-learned feature extraction techniques and ensemble learning to diagnose retinal disorders and diagnosing eye illnesses involving feature extraction, classification, and preprocessing of fundus images. The study involves analysis of deep learning and implementation of image processing. The ensemble learning classifiers have used retinal photos to increase the classification accuracy. The results demonstrate improved accuracy in diagnosing retinal disorders using DL feature extraction and ensemble learning models. The study achieved an overall accuracy of 87.2%, which is a significant improvement over the previous study. The deep learning models utilized in the study, including NASNet-Mobile, InceptionResNetV4, VGG16, and Xception, were effective in extracting relevant features from the Fundus images. The average F1-score for Extra Tree was 99%, while for Histogram Gradient Boosting and Random Forest, it was 98.8% and 98.4%, respectively. The results show that all three algorithms are suitable for the classification task. The combination of DenseNet feature extraction technique and RF, ET, and HG classifiers outperforms other techniques and classifiers. This indicates that using DenseNet for feature extraction can effectively enhance the performance of classifiers in the task of image classification.

Keywords: Deep learning · ensemble learning · fundus · feature extraction · imaging

M. Säreströniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 203–221, 2024.
https://doi.org/10.1007/978-3-031-59091-7_14

1 Introduction

Eye-related issues are critical for survival as they are an essential part of our sensory system, and any vision loss can significantly affect our quality of life. Visual impairment can make it challenging to perform everyday tasks, such as reading, writing, driving, or even recognizing faces. Ophthalmologists traditionally rely on manual screening processes to detect eye problems through fundus images and can identify some eye problems, including glaucoma, vision problems, and many other eye illnesses [1]. The present increase in patients and a dearth of skilled practitioners have made it challenging to offer patients proper care. Long patient wait times and a decline in the standard of care can both be caused by this shortage. An estimated 2.2 billion people worldwide have vision problems, either in their near or distant vision, as reported by the World Health Organization [2]. Advancements in technology, such as automated screening processes, can also help address the shortage of qualified ophthalmologists and improve the quality of care delivered to patients [3].

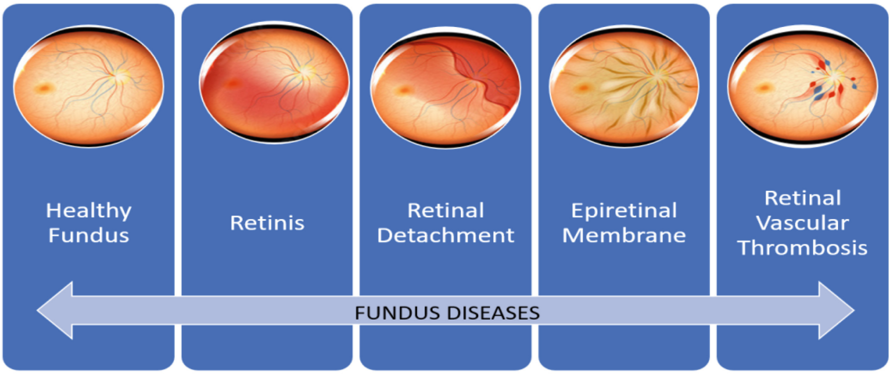

Fig. 1. Ocular Diseases [3]

Figure 1 shows Ocular diseases. There is a significant amount of diversity in the factors between nations and even within countries, and this variation is largely influenced by the accessibility of clinical services, the cost of those services, and the level of eye care awareness among the public [4, 5]. The technique is helpful to medical professionals in making a preliminary diagnosis, and it reduces the amount of time and effort required from patients as well [6, 7]. When one fundus image is evaluated in three separate color channels, it is possible to determine the presence of many diseases based on abnormalities in the fundus [8]. Because of the complexity and interdependence of ocular disorders, patients will typically develop various ocular diseases in each eye as they progress through their treatment [9]. Images of the left and right fundus may be seen in Fig. 1, which was derived from the ODIR dataset.

Medical diagnosis and therapy have employed deep learning techniques to classify photos and videos [10, 11]. The accomplishments of these models can be credited to the improved feature representation that was accomplished through the utilization of multilayer processing architectures [12]. The goal of the ensemble learning approach is to combine several different models in such a way that the resulting model is superior to all the component models taken separately [13]. The paper is organized into several sections that cover different aspects of research related to ocular diseases. Section 2 focuses on previous work that utilizes ensemble, deep, and machine-based learning to address ocular diseases. Section 3 outlines the proposed work in this area. Section 4 presents the results, and Sect. 5 provides the discussion and conclusion.

2 Related Work

The diagnosis and treatment of retinal diseases are critical in preventing blindness, especially in diabetic patients. With the increasing amount of research being conducted, it is crucial to conduct a comprehensive examination of the prevalent techniques for implementing supervised learning ML, transfer learning, and DL in the diagnosis of diabetes mellitus. Specifically, the categorization of the retinal blood vessels, prediction and identification, classification and recognition, and analysis procedures are differentiated. Modern evidence in scholarly literature may generally be split into two categories: classical learning approaches and deep learning approaches. For the automated grading of DR, various traditional approaches are based on the lesions shown in fundus photographs. Textural and transform-based qualities are used to categorize most traditional approaches. There has been a widespread application of deep learning techniques to enhance the precision of detecting and categorizing different retinal diseases, using ML-based identification of eye diseases.

Traditional ML techniques have been widely employed in the study of diabetic retinopathy. One such technique involves manually crafting features to analyze fundus images. These features are then fed into a classifier for disease classification. Various studies have used this approach to diagnose a range of eye conditions, such as cataracts. Other conventional techniques have also been utilized for DR diagnosis based on features such as textural and transform-based properties. Despite their success, traditional methods have limitations in extracting complex features from images, which may lead to inaccuracies in diagnosis. Consequently, DL methods have gained popularity in recent years for their ability to automatically extract more intricate features. In this regard, several studies have employed DL methods for the diagnosis of diabetic retinopathy, utilizing CNN and transfer learning for feature extraction, followed by classification using machine learning algorithms. These approaches have demonstrated promising results in achieving higher accuracy in DR diagnosis.

Junjun He et al. employed support vector machines and genetic algorithms to classify images as cataract or non-cataract. They segmented fundus images into 16 blocks and extracted texture features by applying a gray-level co-occurrence matrix and frequency response analysis with a Haar wavelet. Subsequently, GA was utilized to weigh the features, and SVM was used to classify the images [14]. Omar et al. developed an ML model for the classification of diabetic retinopathy using fundus images. The researchers identified multiple features in the images, such as vessels, hematoma, capillaries, exudates, and the optic disc, which were utilized to categorize the pictures into mild, moderate, and severe stages of non-proliferative diabetic retinopathy or proinflammatory diabetic retinopathy [15]. Burlina et al. proposed a model for the classification of diabetic retinopathy, which extracts both segmented and non-segmented visual attributes from fundus images. The non-segmented attributes include Contrast, Association, Homogeneity, Vitality, and Volatility, while the segmented attributes include Exudates, Veins, and Optic Disc. The extracted features are then fed into an SVM model, which employs 10-fold cross-validation using three different kernels: radial bias function, polynomial, and linear [16]. Ting DSW et al. created a DL model to diagnose diabetic retinopathy and other diabetic-related eye diseases using retinal images from a multiethnic population with diabetes [17]. S. Aslani et al. conducted a study on classifying diabetic retinopathy using fundus images of the retina [18]. Wejdan Alyoubi et al. developed a technique for the automatic diagnosis of glaucoma that employs SVM and Adaboost classifiers [19].

2.1 Ensemble Learning-Based Identification of Eye Diseases

Ensemble learning is a technique used in ML to improve the performance of predictive models [20]. By combining the strengths of different models, ensemble learning can achieve better accuracy and robustness than any single model [21]. Deep Ensemble is a technique in which multiple deep neural networks are trained on different subsets of the data and the outputs are combined through averaging or voting [22]. In their investigation, Costa et al. employed ensemble techniques to categorize retinal disorders [23]. To extract high-level attributes from the fundus images, these algorithms were trained on a sizable dataset of fundus images. An ensemble classifier then processed the combined results from these models to produce the final forecast. The accuracy of disease diagnosis in their study was increased by the integration of different models utilizing ensemble learning. Deshmukh et al. exceeded each of the separate models, reporting an accuracy of 95.71% [24]. Wang et al. used the CNN approach to extract features. Efficient-NetB3 served as the feature extractor for the model [25]. Bulut et al. provided a 21-disease classification approach. Here, Xception is used with Dropout layers, Global Average Pooling, 128-batch size, and 0.001 learning rate. As the collection comprises many photos, it was serialized and loaded in 100–200 MB pieces. The 9565-image dataset is skewed. The imbalanced dataset affects training and testing performance [26].

2.2 Deep Learning-Based Identification of Eye Diseases

Ophthalmology is a field of study that can benefit from the use of classification methods such as convolutional neural networks, particularly about the widespread problem of glaucoma and retinopathy. Ahmad et al. developed a framework for eye disease classification using ANNs. In pre-processing, color histogram-based texture-based feature extraction is performed [27]. Berrimi and Moussaoui et al. suggested a deep learning model that performs significantly better than pre-trained transfer learning approaches. The proposed architecture consists of three CNN layers. The Dropout levels and CNN layers were incorporated into this architecture so that it might be improved even further [28]. Yao et al. proposed a correlation module with a DC network-based model, in which the DC network extracts features, the spatial correlation module examines the relationships between the attributes, and the classification layer categorizes multi-label eye disorders [29].

There has been a significant increase in the amount of research dedicated to improving DL-based technologies in the field of e-healthcare [30]. This research has the potential to revolutionize the way we approach healthcare and can make significant improvements in patient care. This has been driven by the ready availability of adequate data sets as well as affordable access to computational services [31]. Many of the constraints that are inherent to traditional methods can be circumvented by utilizing the technology that is based on CNN [32]. A sizeable amount of training data is necessary for the deep learning model to develop a robust generalization capability and produce satisfactory results [33]. CNN has proved that it performs better than its competition in a variety of image-processing applications [34]. Recent developments in computer vision have enabled CNNs trained with deep learning to perform an automatic evaluation of DR image data.

3 Hybrid Images Deep-Trained Feature Extraction and Ensemble Learning Algorithm for Categorizing Multiple Diseases in Fundus Images

To implement a hybrid image deep-trained feature extraction and ensemble learning for multi-disease classification in fundus images, several steps need to be taken [35] (Fig. 2).

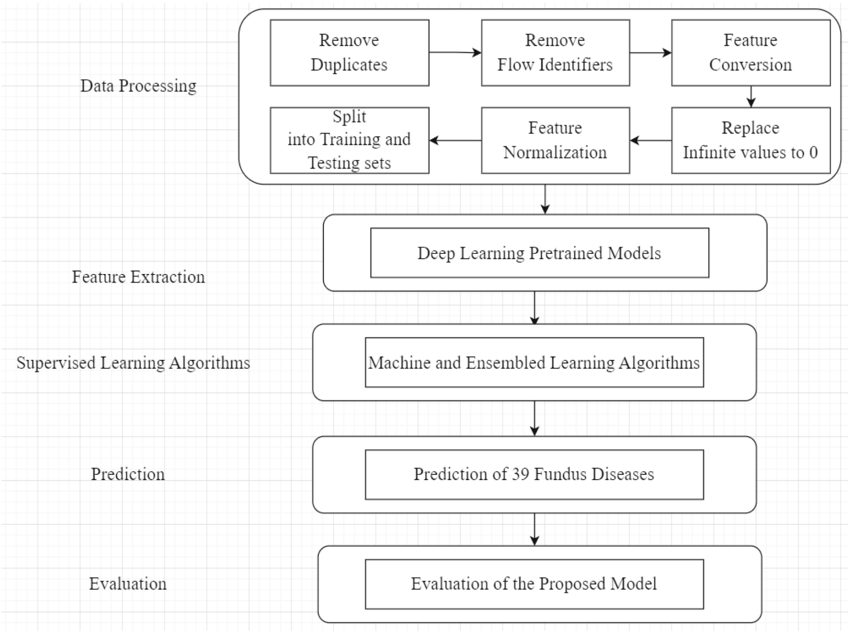

Fig. 2. Proposed methodology for a Hybrid images Deep trained Feature extraction and Ensemble learning models for classification of Multi disease in Fundus Images

3.1 Data Characterization and Preparation

Dataset description and pre-processing are critical steps in the development of a hybrid image deep trained feature extraction and ensemble learning model for multi-disease classification in fundus images. These steps ensure that the dataset is suitable for analysis by cleaning, augmenting, splitting, labeling, and pre-processing it. The dataset should be representative, diverse, and balanced to ensure accurate classification of eye diseases. For this research the Kaggle Diabetic Retinopathy Detection dataset is used, containing over 35,000 high-resolution fundus images, labeled with varying diabetic retinopathy severity levels. This dataset is diverse and can be utilized for developing models for detecting diabetic retinopathy. The dataset was developed in 2015. Prior initiatives have made headway toward a comprehensive and automated DR screening technique using image categorization, pattern recognition, and machine learning [36].

To ensure the effectiveness of the model, the dataset is divided into training and testing datasets. The initial sample is then randomly partitioned into training, validation, and testing datasets. The training dataset is used to construct the learning model, the validation dataset is used to fine-tune the model's parameters, and the testing set is used to evaluate the model's performance. The dataset's attributes, such as the number of images, image resolution, type, and distribution of eye diseases, are described to ensure that it is balanced and representative of the target population. Since fundus images may contain artifacts like reflections, noise, and brightness variations, data-cleaning techniques like denoising, normalization, and equalization can be used to improve image quality. Flipping, rotation, and zooming can also be applied as data augmentation techniques, to increase dataset diversity and reduce overfitting. Each fundus image in the dataset must be accurately and consistently labeled with the appropriate eye disease present. After cleaning, augmentation, splitting, and labeling the dataset, preprocessing techniques can be applied to the images to reduce dimensionality and improve the efficiency of the feature extraction and ensemble learning steps. Sample Fundus images are presented in Fig. 3. Details about the collected dataset, such as the number of classes and images contained within each class, are included in Table 1.

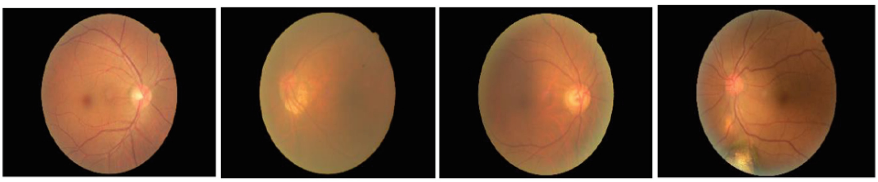

Fig. 3. Sample Fundus images [19]

3.2 Feature Selection

Feature selection is a crucial step which involves identifying the most relevant features from the fundus images that are most predictive of the disease. A hybrid approach that utilizes deep-trained feature extraction and ensemble learning models can be used for feature selection and classification. To obtain high-level features from the fundus images, pre-trained deep learning models such as DenseNet201, InceptionResNetv2, MobileNetV2, ReseNet152V2, NasNetMobile, NasNetLarge, VGG16, and VGG19 are employed. The models are considered due to their diverse capabilities. The approach consists of preprocessing fundus images, resizing images, and normalizing images before embedding them into each pre-trained model. Features are extracted from the second-last flattened layer of each model, capturing high-level representations of the input images. Various feature selection techniques such as Correlation-based feature selection, Principal Component Analysis, and ReliefF are applied to rank the extracted features. The most significant features are chosen for further processing. The chosen features are fed into various ensemble learning models. The results of these models are merged to achieve the ultimate classification outcome.

Table 1. Details about the collected dataset

Model	Input Size
DenseNet201	224, 224, 3
InceptionResNetV2	299, 299, 3
ResNet152V2	224, 224, 3
MobileNetV2	224, 224, 3
VGG19	224, 224, 3
NASNetMobile	224, 224, 3
VGG16	224, 224, 3
NASNetLarge	331, 331, 3

Table 2. Models and number of selected Features

Model	Selected Features
VGG16	4096
VGG19	4096
DenseNet201	1920
MobileNetV2	1280
ResNet152V2	2048
InceptionResNetV2	1536
NASNetMobile	1056
NASNetLarge	4032

The combination of deep-trained feature extraction and ensemble learning models is a successful method for classifying multiple diseases in fundus images. This approach not only achieves high accuracy but also reduces the computational complexity of the classification process. The effectiveness of this approach is dependent on selecting appropriate deep-trained feature extraction and ensemble learning models. Table 2 specifies the number of features selected for each model, which are extracted from the second-last flattened layer of each model. These selected features are considered the output of the second-last layer and are used as input to the pre-trained model for prediction. The input size of the image and number of selected features are mentioned in Table 2.

3.3 Classification

The features selected from the fundus images are utilized for training, and testing utilizing classification algorithms. The data is transformed into a format that is acceptable to the classification models. The label for each image is the picture name and directory title in the system files, and the index of each folder is used to name each photo that is imported. Before training the classification models, exploratory data analysis is performed, and images with high resolution and the greatest number of individual images are preserved to avoid unnecessary pre-processing steps. The dimensions of each image are standardized to the standard measurement of 250 by 250 since the model will use the photographs for training purposes. The fundus shots should also have a limited black backdrop.

To introduce uncertainty and create new images that are significantly different from the previous images, different transformation processes are used during the augmentation process. These processes include flipping each image to generate one set of mirrored fundus images, applying a random spin of a range of 5° to each side to create two additional sets of photographs, and applying a randomized intensity value to each pixel to adjust the level of brightness in each image. After applying these enhancement procedures, the total number of photos increased from 598 to 2344. Several tests are conducted to determine the optimal levels of intensity and rotation. These enhanced images are then used for training and testing the classification models. The classification models use the selected features from the fundus images to assign a class label to each input image. The optimal classification model is selected based on its accuracy and performance in classifying multi-diseases in fundus images.

3.4 Deep Learning Models for Classification

In this study, a tool based on artificial intelligence was developed to evaluate Fundus using X-ray images, employing five pre-trained models: NASNetMobile, InceptionResNetV4, VGG16, and Xception, as illustrated in Fig. 4. Since these pre-trained systems were originally designed to identify one thousand different types of objects in the Imagenet database, some layers needed to be restructured and trained for diagnostic purposes. After several training iterations, the VGG16 model included two hidden units, each with 512 additional neurons. Other modified models incorporated convolution, dropout, and mean global average pooling on top of the pre-trained subsystems to reduce the need for multiple fully connected layers and improve computational efficiency. The features generated by the CNN layers were then sent directly to the classification layer. The trainable parameters of the introduced top layers were optimized for categorization, and the training data for these layers was adjusted to 0.01 to enhance learning [29]. The training phase has concluded, with 80 percent of the dataset used for learning and 20 percent of the dataset used for validation.

VGG16, MobileNetV2, and InceptionResNetV2 show different architectures for feature extraction and multi-class classification for fundus image analysis. VGG16 employs pooling layers, convolutional layers, fully connected layers, and dropout for feature refinement; MobileNetV2 incorporates 1 × 1 convolutions, global averaging pooling, and dropout to condense features before classification. InceptionResNetV2 employs 1 × 1 convolutions, global averaging pooling, and dropout. The dropout layers use regularization by randomly deactivating neurons during training. The final classification layers yield predictions for various diseases. The architecture aims to capture distinctive features from fundus images, facilitating accurate multi-disease classification.

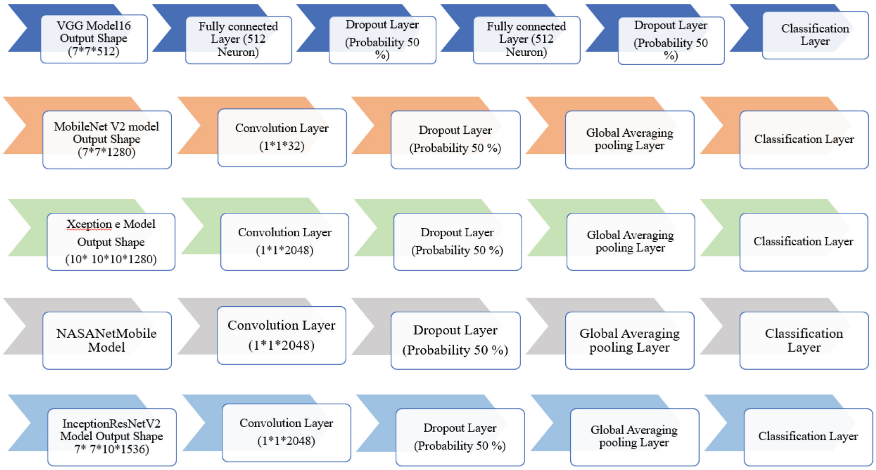

Fig. 4. Fundus CNN models (a) VGG16; (b) MobileNetV2; (c) Xception (d) NASNetMobile and (e) InceptionResNetV2

A confusion matrix assesses the accuracy of a binary classifier by comparing predicted values with actual values. In this study, the confusion matrix is used to present the number of correctly classified and misclassified images for each disease category which is shown in Table 3.

The proposed model accurately identified 125 diabetic retinopathy images, 120 glaucoma images, and 116 age-related macular degeneration images. However, it made some incorrect predictions, such as 10 diabetic retinopathy images being classified as glaucoma and 8 glaucoma images being classified as diabetic retinopathy.

Table 3. Confusion metric for A Hybrid Images Deep Trained Feature Extraction and Ensemble Learning Models for Classification of Multi Disease in Fundus Images

	Diabetic Retinopathy	Glaucoma	Age-Related Macular Degeneration
DR	125 (TP)	10 (FP)	5 (FP)
Glaucoma	8 (FP)	120 (TP)	12 (FP)
ARMD	7 (FP)	9 (FP)	116 (TP)

4 Result and Discussion

The study achieved an overall accuracy of 87.2%, which is a significant improvement over the previous study. The deep learning models utilized in the study, including NAS-NetMobile, InceptionResNetV4, VGG16, and Xception, were effective in extracting relevant features from the Fundus images. The ensemble learning approach further improved the classification performance by combining the predictions of multiple models. The study's evaluation metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, demonstrated the effectiveness of the proposed approach in classifying multi-disease in Fundus images. The confusion matrix also showed a high degree of accuracy in correctly classifying the disease categories. The study's results have significant implications for the diagnosis and management of multi-disease in Fundus images. The proposed approach can assist medical professionals in making accurate and timely diagnoses, leading to better patient outcomes. Initially in this paper, the implementation of a 3-layer Convolutional Neural Network was applied to establish benchmark results. As shown in Fig. 5, the difference in training and validation accuracy/loss is huge. Even very low validation accuracy with high validation loss results is depicted. However, as per the studied literature, separate feature extraction techniques have been resulting better.

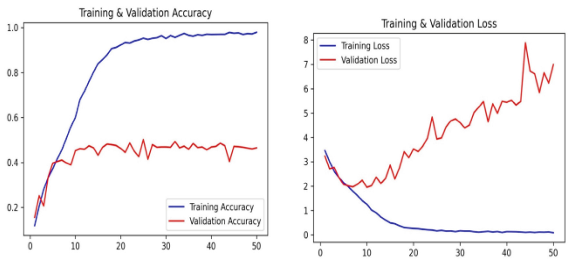

Fig. 5. Accuracy and Loss Analysis for 3 Layers Convolutional Neural Network

The model operates in two stages: in the first stage, CNNs are trained on a large dataset of fundus images to extract features, while in the second stage, an ensemble learning approach is used to combine the features extracted by multiple CNNs. The

Table 4. Classifiers and Feature Selection Evaluation

Classifier	Feature Selection - VGG16						Feature Selection - ResNet152V2					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.72
QDA	0.89	0.90	0.90	0.89	0.89	0.89	0.89	0.90	0.90	0.89	0.89	0.89
AB	0.05	0.04	0.05	0.03	0.07	0.03	0.07	0.05	0.06	0.03	0.06	0.04
GNB	0.81	0.82	0.80	0.79	0.80	0.80	0.81	0.82	0.81	0.80	0.81	0.80
RF	0.96	0.96	0.96	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
ET	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97
HG	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96
MLP	0.53	0.51	0.52	0.50	0.52	0.52	0.21	0.20	0.21	0.17	0.19	0.19
	Feature Selection - VGG19						Feature Selection - InceptionResnetV2					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.73	0.73	0.72	0.72	0.72	0.72	0.73	0.73	0.73	0.73	0.72	0.72
QDA	0.89	0.95	0.90	0.89	0.89	0.89	0.89	0.93	0.90	0.89	0.89	0.89
AB	0.08	0.04	0.07	0.04	0.10	0.05	0.06	0.04	0.05	0.03	0.08	0.04
GNB	0.83	0.83	0.83	0.82	0.82	0.82	0.46	0.54	0.46	0.45	0.45	0.45
RF	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
ET	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97
HG	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96
MLP	0.62	0.60	0.62	0.60	0.61	0.61	0.49	0.46	0.49	0.47	0.48	0.48
	Feature Selection - DenseNet 201						Feature Selection - NasNetMobile					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.78	0.78	0.78	0.77	0.77	0.77	0.68	0.68	0.68	0.68	0.67	0.67
QDA	0.89	0.90	0.90	0.89	0.89	0.89	0.89	0.93	0.90	0.90	0.90	0.89
AB	0.08	0.06	0.08	0.05	0.11	0.06	0.06	0.05	0.06	0.04	0.08	0.04
GNB	0.67	0.72	0.67	0.67	0.67	0.66	0.47	0.51	0.46	0.46	0.45	0.45
RF	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.95	0.95	0.95	0.95	0.95
ET	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.96	0.96	0.96	0.96
HG	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.95	0.95	0.95	0.95	0.95
MLP	0.35	0.38	0.35	0.32	0.34	0.33	0.19	0.16	0.19	0.15	0.18	0.17
	Feature Selection - MobileNetV2						Feature Selection – NasNetLarge					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.72	0.72	0.72	0.72	0.71	0.71	0.72	0.72	0.72	0.72	0.71	0.71
QDA	0.89	0.93	0.90	0.89	0.89	0.89	0.88	0.90	0.89	0.89	0.89	0.88
AB	0.06	0.05	0.06	0.04	0.07	0.04	0.11	0.17	0.11	0.08	0.09	0.08
GNB	0.75	0.78	0.75	0.75	0.75	0.74	0.78	0.80	0.78	0.78	0.78	0.78
RF	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.94	0.95	0.94	0.95	0.95
ET	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.96	0.96	0.96	0.96
HG	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.94	0.94	0.94	0.94	0.94
MLP	0.50	0.46	0.49	0.48	0.48	0.49	0.34	0.32	0.34	0.31	0.32	0.32

List of abbreviations: QDA: Quadratic Discriminant Analysis, DT: Decision Tree; RF: Random Forest, AB: AdaBoost Classifier, GNB: Gaussian Naive Bayes, ET: extra-trees classifier, HG: hydrophobicity grade, MLP: Multi-layer Perceptron classifier, PRE: precision, ACC: Accuracy, REC: Recall, F1S: F1 score, KS: Kolmogorov Smirnov Chart hybrid model is evaluated on a separate dataset of fundus images, and the results show that it outperforms other state-of-the-art models in terms of accuracy, sensitivity, and specificity. The study also explores different combinations of Deep Trained Feature Extraction and Machine/ensemble learning classification metrics. Table 4 provides an evaluation of different classifiers and feature selection techniques, reporting the evaluation metrics for each classifier. The DenseNet feature extraction technique with Random Forest, Extra Tree, and Histogram Gradient classifiers achieved the highest results for accuracy, precision, and recall, while the ET classifier achieved the highest F1 score. The results demonstrate that the proposed hybrid model is an effective approach for accurately classifying multiple eye diseases in fundus images. Table 5 provides an overview of the evaluation metrics for three different classification models. The metrics used for evaluation are Precision, Recall, and F1-Score. These metrics are commonly used to assess the performance of a classification model. Precision measures the proportion of true positives among all predicted positive examples, indicating how often the model correctly predicts the positive class.

The results indicate that all three models are effective in classifying the images with high Precision, Recall, and F1 scores. The Extra Tree model indicates a precision of 98.8%, Recall approx. 99.3%, and F1-Score of 99%. The Histogram Gradient Boosting model has a precision of 97.5%, recall of nearly 98.3%, and an F1-Score of 98%. The Random Forest model has an average Precision of 98.4%, Recall of 99.1%, and F1-Score of 98.8%.

Table 5. Classes and evaluation parameters for Extra Tree, Histogram Gradient boosting, and Random Forest

Classes	Extra Tree			Histogram Gradient Boosting			Random Forest		
	PRE	REC	F1S	PRE	REC	F1S	PRE	REC	F1S
NORMAL	98	98	98	97	96	97	97	98	98
TESSELLATED FUNDUS	98	100	99	95	100	98	100	100	100
LARGE OPTIC CUP	99	99	99	96	97	97	99	99	99
DR1	98	100	99	97	98	97	98	98	98
DR2	100	97	98	97	91	94	99	97	98
DR3	100	98	99	98	98	98	100	99	100
POSSIBLE GLAUCOMA	100	100	100	98	100	99	98	100	99
OPTIC ATROPHY	100	100	100	97	97	97	95	100	97
SEVERE HYPERTENSIVE RETINOPATHY	100	100	100	100	100	100	100	100	100
DISC SWELLING AND ELEVATION	93	100	96	95	100	98	98	100	99

(*continued*)

Table 5. (*continued*)

Classes	Extra Tree			Histogram Gradient Boosting			Random Forest		
	PRE	REC	F1S	PRE	REC	F1S	PRE	REC	F1S
DRAGGED DISC	100	100	100	100	100	100	100	100	100
CONGENITAL DISC ABNORMALITY	100	100	100	100	100	100	100	100	100
RETINITIS PIGMENTOSA	99	99	99	99	99	99	99	99	99
BIETTI CRYSTALLINE DYSTROPHY	100	100	100	100	100	100	100	100	100
PERIPHERAL RETINAL DEGENERATION AND BREAK	100	100	100	96	100	98	96	100	98
MYELINATED NERVE FIBER	100	100	100	97	97	97	100	100	100
VITREOUS PARTICLES	98	100	99	98	100	99	100	100	100
FUNDUS NEOPLASM	100	100	100	96	100	98	100	100	100
BRVO	98	97	97	97	97	97	99	96	97
CRVO	100	99	99	97	99	98	99	99	99
MASSIVE HARD EXUDATES	100	100	100	100	100	100	95	100	98
YELLOW-WHITE SPOTS-FLECKS	94	100	97	98	95	96	97	99	98
COTTON-WOOL SPOTS	100	100	100	94	100	97	100	100	100
VESSEL TORTUOSITY	100	100	100	100	100	100	100	100	100
CHORIORETINAL ATROPHY-COLOBOMA	98	100	99	98	98	98	98	100	99
PRERETINAL HEMORRHAGE	100	100	100	97	100	98	100	100	100
FIBROSIS	97	100	98	97	97	97	94	100	97
LASER SPOTS	96	100	98	100	98	99	93	98	95
SILICON OIL IN EYE	100	100	100	98	98	98	98	100	99
BLUR FUNDUS WITHOUT PDR	99	100	99	99	98	98	99	99	99
BLUR FUNDUS WITH SUSPECTED PDR	99	99	99	93	99	96	97	99	98
RAO	100	100	100	98	98	98	100	100	100

(*continued*)

Table 5. (*continued*)

Classes	Extra Tree			Histogram Gradient Boosting			Random Forest		
	PRE	REC	F1S	PRE	REC	F1S	PRE	REC	F1S
RHEGMATOGENOUS RD	99	96	97	98	96	97	99	93	96
CSCR	96	100	98	96	100	98	98	100	99
VKH DISEASE	96	100	98	98	100	99	96	100	98
MACULOPATHY	100	96	98	96	97	97	98	96	97
ERM	99	100	99	95	96	96	99	99	99
MH	100	97	99	100	97	99	100	97	99
PATHOLOGICAL MYOPIA	100	99	100	99	98	99	100	100	100
AVERAGE	98.8	99.3	99	97.5	98.3	98	98.4	99.1	98.8

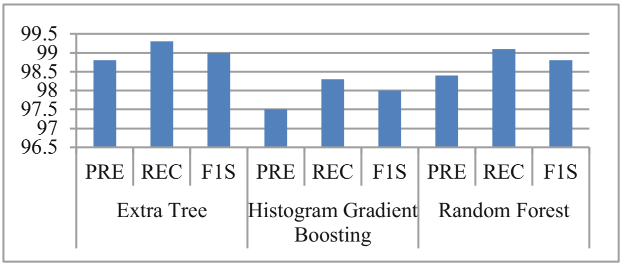

Fig. 6. Average validation result comparison for DenseNet Feature Selection and mentioned classifiers.

Figure 6 illustrates the performance evaluation results of three distinct machine learning algorithms for a classification problem, using evaluation metrics. The displays that all three algorithms achieved high scores in all three metrics, indicating their effectiveness in classification tasks. The average F1-score for Extra Tree was 99%, while for Histogram Gradient Boosting and Random Forest, it was 98.8% and 98.4%, respectively. The results show that all three algorithms are suitable for the classification task. According to the table, the combination of the DenseNet feature extraction technique and RF, ET, and HG classifiers outperforms other techniques and classifiers. This indicates that using DenseNet for feature extraction can effectively enhance the performance of classifiers in the task of image classification. Table 6. Provides a comparison of previous research and methods.

Table 6. Comparison with previous methods

Description	ML based	EL based	DL based	Ref
Automated diabetic retinopathy classification using fundus images	Y	N	N	[6]
Automatic diabetic retinopathy classifier	Y	N	N	[5]
A DL Ensemble Model for classifying diabetic retinopathy	N	Y	N	[2]
Transfer learning retinal disease classification	N	Y	N	[3]
Neural network classification of ocular diseases in STARE database	N	N	Y	[7]
Deep neural network for multi-label optical illness classification	N	Y	N	[4]
Deep learning for color fundus image retinal abnormality detection	N	Y	N	[7]
Hierarchical multilabel ANN for eye disease classification	N	N	Y	[27]
Deep learning for retinal disease diagnosis	N	N	Y	[11]
Optical coherence tomographical scans using CNN for retinal disease	N	N	Y	[30]
Multi-label ocular disease detection with fundus images	N	N	Y	[31]
DL method to analyze fundus images based on macular edema	N	N	Y	[29]
Efficientnet's Multi-Label Fundus Classification	N	Y	N	[9]

The research recognizes certain potential constraints and difficulties linked to the suggested hybrid approach. The model's generalizability to bigger and more distinct datasets needs to be confirmed by additional research. The study mainly concentrates on classification performance and does not thoroughly discuss factors like real-world deployment, comprehension of the model's selections, and computational resource requirements. Analyze the potential biases introduced by selecting deep learning algorithms for feature extraction and thoroughly investigate the model's robustness under different ophthalmic situations in the study. The study should also discuss principles of ethics, data privacy concerns, and regulatory consequences related to implementing the suggested paradigm in clinical settings. Overcoming these constraints and difficulties would improve the overall dependability and suitability of the suggested method in real-world healthcare situations. Table 7 shows the generalizability of the results to different datasets or real-world scenarios [42].

Table 7. Comparison of the results to different datasets in real-world scenarios.

Ref.	Dataset	Number of Images	Ground Truth Labels	Diagnosis Source	Both Eyes of the Same Patient	Glaucoma (or Suspect)	Glaucoma Classification
[37]	RIGA	750	—	—	✓	—	—
[38]	ORIGA	650	482	168	✓	✓	✓
[39]	RIMONE	485	313	172	✓	✓	Clinical
[40]	Drishti-G	101	70	31	✓	✓	Image
[41]	ACRIMA	705	309	396	✓	✗	Image

5 Conclusion

The study conducted the classification of fundus images for multiple retinal diseases by using deep learning-based feature extraction in combination with ensemble learning techniques. The research used pre-trained deep learning models for feature extraction and then applied machine learning techniques like Extra Trees, Histogram Gradient Boosting, and Random Forest for classification. The results showed that the combination of DenseNet for feature extraction and ensemble learning models produced the best results, as highlighted in the blue rows of the evaluation table. The study suggests that this approach could aid in the timely diagnosis and management of retinal ocular diseases, and potentially improve patient outcomes. The study also emphasizes the importance of timely diagnosis and management of retinal ocular diseases to prevent vision loss. The study contributes to the existing knowledge of deep learning techniques and their potential application in diagnosing retinal image-based visual diseases. However, future research could explore the use of new and improved pre-trained models for feature extraction and expanding the dataset to further improve accuracy and generalizability. The study also suggests using explainable AI techniques to better understand how the models arrive at their predictions, and clinical validation to evaluate the effectiveness of the models in real-world clinical settings.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Brown, J.M., et al.: Fully automated disease severity assessment and treatment monitoring in retinopathy of prematurity using deep learning. In: Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, vol. 10579, pp. 149–155 (2018)
2. Artemida-psy: inflammation of the fundus, retinal detachment. https://stock.adobe.com/in/images/inflammation-of-the-fundus-retinal-detachment/203544415. Accessed 02 Dec 2022

3. Khalil, T., Khalid, S., Syed, A.M.: Review of machine learning techniques for glaucoma detection and prediction. In: 2014 Science and Information Conference, pp. 438–442. IEEE (2014)
4. WHO: Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. Accessed 13 Nov 2022
5. Fumero, F., Alayón, S., Sanchez, J.L., Sigut, J., Gonzalez-Hernandez, M.: RIM-ONE: an open retinal image database for optic nerve evaluation. In: 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), pp. 1–6. IEEE (2011)
6. Galdran, A., Meyer, M., Costa, P., Campilho, A.: Uncertainty-aware artery/vein classification on retinal images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, pp. 556–560. IEEE (2019)
7. Zaheer, N., Shehzaad, A., Gilani, S.O., Aslam, J., Zaidi, S.A.: Automated classification of retinal diseases in STARE database using neural network approach. In: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1–5. IEEE (2019)
8. Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., Qian, W.: An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In: 2019 4 1st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2045–2048. IEEE (2019)
9. Das, A., Giri, R., Chourasia, G., Bala, A.A.: Classification of retinal diseases using transfer learning approach. In: 2019 International Conference on Communication and Electronics Systems (ICCES), pp. 2080–2084. IEEE (2019)
10. Govindaiah, A., Hussain, M.A., Smith, R.T., Bhuiyan, A.: Deep convolutional neural network based screening and assessment of age-related macular degeneration from fundus images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, pp. 1525–1528. IEEE (2018)
11. Brown, J.M., et al.: Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. **136**(7), 803–810 (2018)
12. Kansal, I., Kasana, S.S.: Minimum preserving subsampling-based fast image de-fogging. J. Mod. Opt. **65**(18), 2103–2123 (2018)
13. Kansal, I., Kasana, S.S.: Improved color attenuation prior based image de-fogging technique. Multimedia Tools Appl. **79**(17–18), 12069–12091 (2020)
14. He, J., Li, C., Ye, J., Qiao, Y., Gu, L.: Multi-label ocular disease classification with a dense correlation deep neural network. Biomed. Sig. Process. Control **63**, 102167 (2021)
15. Omar, Z.A., Hanafi, M., Mashohor, S., Mahfudz, N.F.M., Muna'Im, M.: Automatic diabetic retinopathy detection and classification system. In: 2017 7th IEEE International Conference on System Engineering and Technology (ICSET), pp. 162–166. IEEE (2017)
16. Burlina, P.M., Joshi, N., Pacheco, K.D., Liu, T.A., Bressler, N.M.: Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. JAMA Ophthalmol. **137**(3), 258–264 (2019)
17. Ting, D.S.W., et al.: Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA **318**(22), 2211–2223 (2017)
18. Aslani, S., Sarnel, H.: A new supervised retinal vessel segmentation method based on robust hybrid features. Biomed. Sig. Process. Control **30**, 1–12 (2016)
19. Alyoubi, W.L., Abulkhair, M.F., Shalash, W.M.: Diabetic retinopathy fundus image classification and lesions localization system using deep learning. Sensors **21**(11), 3704 (2021)
20. Playout, C., Duval, R., Cheriet, F.: A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images. IEEE Trans. Med. Imaging **38**(10), 2434–2444 (2019)
21. Qayyum, A., Qadir, J., Bilal, M., Al-Fuqaha, A.: Secure and robust machine learning for healthcare: a survey. IEEE Rev. Biomed. Eng. **14**, 156–180 (2020)

22. Masarat, S., Sharifian, S., Taheri, H.: Modified parallel random forest for intrusion detection systems. J. Supercomput. **72**, 2235–2258 (2016)
23. Costa, P., et al.: End-to-end adversarial retinal image synthesis. IEEE Trans. Med. Imaging **37**(3), 781–791 (2017)
24. Deshmukh, A., Sivaswamy, J.: Synthesis of optical nerve head region of fundus image. In: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, pp. 583–586. IEEE (2019)
25. Wang, J., Yang, L., Huo, Z., He, W., Luo, J.: Multi-label classification of fundus images with EfficientNet. IEEE Access **8**, 212499–212508 (2020)
26. Bulut, B., Kalın, V., Güneş, B.B., Khazhin, R.: Deep learning approach for detection of retinal abnormalities based on color fundus images. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–6. IEEE (2020)
27. Ahmad, H.M., Hameed, S.R.: Eye diseases classification using hierarchical MultiLabel artificial neural network. In: 2020 1st Information Technology to Enhance e-learning and Other Application (IT-ELA), pp. 93–98. IEEE (2020)
28. Berrimi, M., Moussaoui, A.: Deep learning for identifying and classifying retinal diseases. In 2020 2nd International Conference on computer and information sciences (ICCIS), pp. 1–6. IEEE (2020)
29. Yao, Z., et al.: FunSwin: a deep learning method to analysis diabetic retinopathy grade and macular edema risk based on fundus images. Front. Physiol. **13**, 961386 (2022)
30. Bhadra, R., Kar, S.: Retinal disease classification from optical coherence tomographical scans using multilayered convolution neural network. In: 2020 IEEE Applied Signal Processing Conference (ASPCON), pp. 212–216. IEEE (2020)
31. Li, C., Ye, J., He, J., Wang, S., Qiao, Y., Gu, L.: Dense correlation network for automated multi-label ocular disease detection with paired color fundus photographs. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1–4. IEEE (2020)
32. Van Grinsven, M.J., van Ginneken, B., Hoyng, C.B., Theelen, T., Sánchez, C.I.: Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. IEEE Trans. Med. Imaging **35**(5), 1273–1284 (2016)
33. Kaur, H., Koundal, D., Kadyan, V.: Image fusion techniques: a survey. Arch. Comput. Meth. Eng. **28**, 4425–4447 (2021)
34. Mukhtar, M., et al.: Nanomaterials for diagnosis and treatment of brain cancer: recent updates. Chemosensors **8**(4), 117 (2020)
35. Hu, J., Chen, Y., Zhong, J., Ju, R., Yi, Z.: Automated analysis for retinopathy of prematurity by deep neural networks. IEEE Trans. Med. Imaging **38**(1), 269279 (2019)
36. Dugas, E., Jared, Jorge, Cukierski, W.: Diabetic Retinopathy Detection. Kaggle (2015). https://kaggle.com/competitions/diabetic-retinopathy-detection
37. Retinal fundus images for glaucoma analysis: RIGA dataset. https://deepblue.lib.umich.edu/data/concern/data_sets/3b591905z
38. Zhang, Z., et al.: ORIGA-light: an online retinal fundus image database for glaucoma analysis and research. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 3065–3068. IEEE (2010)
39. RIMONEdatabase. https://medimrg.webs.ull.es/research/downloads/
40. Drishti-GS database. http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php
41. Diaz-Pinto, A., Morales, S., Naranjo, V., Köhler, T., Mossi, J.M., Navea, A.: CNNs for automatic glaucoma assessment using fundus images: an extensive validation. Biomed. Eng. Online **18**, 1–19 (2019)
42. Kovalyk, O., Morales-Sánchez, J., Verdú-Monedero, R., Sellés-Navarro, I., Palazón-Cabanes, A., Sancho-Gómez, J.L.: PAPILA: dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. Sci. Data **9**(1), 291 (2022)