Advancing Eye Disease Assessment through Deep Learning: A Comparative Study with Pre-Trained Models

Zamil S. Alzamil
Department of Computer Science; College of Computer and Information Sciences; Majmaah University, Al-Majmaah, 11952, Saudi Arabia
z.alzamil@mu.edu.sa (corresponding author)

Received: 19 March 2024 | Revised: 16 April 2024 | Accepted: 24 April 2024

 DOI: https://doi.org/10.48084/etasr.7294

ABSTRACT

The significant global challenges in eye care are treatment, preventive quality, rehabilitation services for eye patients, and the shortage of qualified eye care professionals. Early detection and diagnosis of eye diseases could allow vision impairment to be avoided. One barrier to ophthalmologists when adopting computer-aided diagnosis tools is the prevalence of sight-threatening uncommon diseases that are often overlooked. Earlier studies have classified eye diseases into two or a small number of classes, focusing on glaucoma, and diabetes-related and age-related vision issues. This study employed three well-established and publicly available datasets to address these limitations and enable automatic classification of a wide range of eye disorders. A Deep Neural Network for Retinal Fundus Disease Classification (DNNRFDC) model was developed, evaluated based on various performance metrics, and compared with four established pre-trained models (EfficientNetB7, EfficientNetB0, UNet, and ResNet152) utilizing transfer learning techniques. The results showed that the proposed DNNRFDC model outperformed these pre-trained models in terms of overall accuracy across all three datasets, achieving an impressive accuracy of 94.10%. Furthermore, the DNNRFDC model has fewer parameters and lower computational requirements, making it more efficient for real-time applications. This innovative model represents a promising avenue for further advancements in the field of ophthalmological diagnosis and care. Despite these promising results, it is essential to acknowledge the limitations of this study, namely the evaluation conducted by using publicly available datasets that may not fully represent the diversity and complexity of real-world clinical scenarios. Future research could incorporate more diverse datasets and explore the integration of additional diagnostic modalities to further enhance the model's robustness and clinical applicability.

Keywords-deep learning; retina; eye disease; transfer learning; ophthalmological diagnosis; DNNRFDC

I. INTRODUCTION

The eyes are the most vital sensory organs of people, as they are responsible for up to 80% of all impressions. The status of eye-related diseases constitutes an enormous problem. There are 2.2 billion visually impaired people worldwide, of which at least 1 billion cases could have been prevented or addressed [1-2]. The main issues on the global problem of eye diseases are reachability, availability, acceptability, and the associated cost of eye care services. One of the major obstacles in resolving these matters is the lack of available skilled human resources. Artificial intelligence and big data analytics are two emerging eye-care technologies that can help improve the access to quality healthcare for disadvantaged populations [1]. These technologies can use color fundus pictures to provide a non-invasive analysis of systemic microvasculature in the retinal image. These retinal image analyses provide eye health-related information and serve as an early indicator of long-term health issues, such as diabetes, stroke, hypertension, arteriosclerosis, cardiorenal, neurodegenerative, and renal diseases [2-5]. Therefore, regular eye screening, consultation, and therapy are trusted worldwide to prevent vision loss and overall health. Early detection and diagnosis of ocular diseases would minimize visual impairment, alleviating its socioeconomic cost. Previous studies generally focused on Diabetic Retinopathy (DR), glaucoma, and Age-related Macular Degeneration (AMD) [6-8]. However, it is critical to detect and diagnose uncommon diseases, such as ischemic optic neuropathy and retinal arterial blockage. These uncommon diseases are generally missed in standard clinical examinations, limiting the use of automated screening technologies by ophthalmologists. Furthermore, current studies emphasize the necessity of multi-disease detection [9-10].

Machine Learning (ML) is now an integral part of addressing complicated issues in most sciences, opening enormous opportunities in medicine. ML approaches can accept a variety of data configurations, apply contextual weighing, and determine the predictive potential of all possible

combinations of variables to assess various diagnostic aspects [11]. In [12], different ML models, including Decision Tree, Random Forest, Naive Bayes, and neural network algorithms, were deployed to classify eye diseases into ten classes. The limitation of this study was the use of single data with only ten types and weak feature selection. In [13], a neural network was employed in conjunction with mathematical morphology to classify eye diseases into two categories, but this study used a low number of images and had a misclassification rate of 52.36%. In [14], a method was proposed utilizing Transfer Learning (TL), based on the VGG-19 model and Random Forest (RF), to perform binary classification on data from 208 patients with glaucoma. In [15], 1200 Optical Coherence Tomography (OCT) images were engaged to classify eye diseases in 23 classes using CNN. This study implemented a small dataset for training and had low performance in the epiretinal membrane class. In [16], Naïve Bayesian, SVM, and J48 models were used to classify 1530 data samples into four categories. This study showed minor parameter tuning and data-splitting strategies. In [17], different ML models were applied, including SVM, NB, RF, and LR on a dataset of 2047 patients. ML-based studies generally struggle with a limited amount of data, parameter optimization, and the issue of overfitting for eye disease classification [11, 18-19].

TABLE I. PREVIOUS ML AND DL-BASED STUDIES FOR EYE CLASSIFICATION

Area	Model	References
Machine Learning	Neural Network	[12, 13]
	Random Forest	[12, 14, 16]
	SVM	[16, 17, 20]
	Naïve Bayes	[11, 12, 16, 17]
	Decision Tree	[16, 18]
	Statistical modeling	[17, 19]
Deep Learning	CNN and TL	[8, 9, 14, 15, 21-26]
	Encoder	[27]

II. TRANSFER LEARNING (TL) FOR EYE DISEASE CLASSIFICATION

ML and Deep Learning (DL) models require a large amount of data to improve their generalizability. Insufficient data can result in overfitting, which has a detrimental effect on the models' performance. In TL, the model is trained on specific data, and its parameters are tuned to the corresponding task at hand. An ML/DL model trained on some datasets can be fine-tuned for an eye disease dataset. Pre-trained models are selected over traditional neural networks by capturing and learning features from images at different hierarchy levels. Local connectedness is a feature of convolution layers that reduces interconnectivity. It is called parameter sharing because the utilization of the weight happens in the pre-trained CNN. A connection between each input or output neuron reduces the number of parameters. Moreover, pre-trained CNNs outperform traditional models. In [22], a TL-based CNN was trained on 490 images and then compared with a clinical procedure. In [23], a CNN was put into service along with a dataset of 32,820 retina fundus images for 1198 patients to detect glaucoma. In [24], a TL-based algorithm was applied on an eye disease dataset from 490 patients to quantify eye vision loss due to glaucoma. In [25], pre-trained models were employed using transfer learning to group eye diseases into four classes, including normal, DME, and early and late AMD. The main drawback of this study was the data split, i.e., 87% for training, 12% for validation, and 0.4% for testing. In [27], an autoencoder was engaged to identify only one eye disease, i.e., glaucoma, using 1426 digital fundus images.

A. UNet

UNet has become the gold standard for biomedical segmentation. UNet is a convolutional autoencoder based on an hourglass architecture [28-29]. The former has been employed in a wide range of biomedical applications, and its classification process is faster than pixel-wise approaches [28]. UNet is a fully Convolutional Neural Network (CNN) that efficiently utilizes data augmentation with the available dataset. It consists of a contracting and an expansive path with 23 convolutional layers. Table II portrays the details of UNet and other pre-trained models.

B. ResNet 152

ResNet is based on microarchitecture modules as opposed to traditional sequential networks, such as AlexNet or VGG. New networks are often built from the ground up utilizing microarchitecture, a collection of tiny building blocks. ResNet-50, a pre-trained model that won in the ILSVRC 2015 competition, has gained significant recognition. Trained on a subset of ImageNet's extensive image database comprising millions of images, this model demonstrates exceptional capabilities in classifying images across over 1,000 object categories. ResNet is renowned for its structural variants, including 18-, 34-, 50-, and 152-layer architectures [30].

TABLE II. CHARACTERISTICS OF VARIOUS PRE-TRAINED MODELS FOR IMAGE DETECTION AND CLASSIFICATION

Architecture	Number of layers	Application	Dataset used for training	Runtime environment
ResNet (2015)	18, 34, 50, 101, 152	Detection, classification	COCO and ImageNet	2GPUs
UNet (2015)	23	Detection, classification	PhC-U373, DIC-HeLa	NVIDIA Titan GPU (6GB)
EfficientNet	237-813	Detection classification	ImageNet CIFAR-100	Google Cloud TPUs

C. EfficientNetB0 and B7

CNNs are primarily designed to optimize resource utilization and can be extended with additional resources to enhance accuracy. In this context, "resolution" pertains to image resolution, "depth" signifies the layer size, and "width" denotes the number of channels in each layer. Increasing depth can process intricate information and yield better generalization, but it may lead to challenges, like gradient vanishing and increased resource demands. However, expanding the network can effectively handle information with reduced depth, making the models more manageable to train. Nevertheless, the expansion may reach a point of diminishing returns in accuracy improvement. Similarly, scaling image resolution can provide finer details, but it may also encounter accuracy saturation issues. EfficientNet addresses these concerns by achieving a balance between resolution, width, and

depth scaling through a compound coefficient Φ, resulting in ameliorated efficiency. [31-32].

Pretrained CNN models were used to classify early and late DME in [33]. The general drawbacks of previous studies were the use of pre-trained instead of developing new models, the small number of images in the dataset, the small number of classes for eye disease types, the lack of rigorous fine-tuning of parameters, cross-validation of the models on different datasets, lack of performance metrics except for AUC, and computational complexity. This study aims to fill these gaps and propose a novel model, called DNNRFDC, to classify eye diseases into various classes. Well-established pre-trained models, such as UNet, EfficientNetB7, EfficientNetB0, ResNet, and ResNet152V2 were fine-tuned and utilized for their performance to be compared with that of DNNRFDC. These pre-trained models employed through setting the top parameter to false so that the eye disease dataset could be fed as an input layer and the results could be classified according to the dataset. The first, second, and third datasets had 45, 4, and 8 classes, respectively.

III. DATASETS

This study deployed three eye disease datasets to evaluate the unbiased performance of the proposed model. The first dataset (Dataset 1) contains 3200 retina fundus images comprising 45 distinct eye diseases [34]. This is a recent and rich dataset collected at the Sushrusha Hospital's Eye Clinic in Nanded (MS), India. Three cameras, Topcon 3D, Topcon TRC, and Kowa, were put into service to collect annotated eye fundus images [2]. The second dataset (Dataset 2) contains 601 images categorized into four classes, entailing normal (300), cataract (100), glaucoma (101), and other retinal diseases (100) [35]. The third dataset (Dataset 3), called ODIR, is a structured ophthalmic dataset comprising data from 5,000 individuals, involving age information, color fundus images of both eyes, and diagnostic assessments by medical professionals [36]. This dataset aims to represent a real-world collection of patient data sourced from various hospitals and medical facilities in China, collected by Shanggong Medical Technology Co., Ltd. The fundus images in this dataset are captured using a variety of cameras available on the market, including Canon, Zeiss, and Kowa, resulting in varying image resolutions [37]. Qualified healthcare professionals assigned labels under the supervision of a quality control manager, classifying eye diseases into eight categories: AMD, cataract, diabetes, glaucoma, hypertension, myopia, normal, and others. To facilitate training, categorical ordinal data were converted into eight trainable labels following a one-hot encoding approach.

IV. METHODOLOGY

A. Data Preprocessing and Analysis

Data preprocessing and analysis were performed to understand the datasets and prepare the DL models. Table III depicts the classes of the first dataset (Dataset 1) and the corresponding number of images [34]. Dataset 1 has well-defined metadata and labeling information in CSV format.

TABLE III. DISEASE TYPES WITH CORRESPONDING NUMBER OF IMAGES

Short Form	Full-Form	No of images
DR	Diabetic retinopathy (DR)	376
ARMD	Age-related macular degeneration	100
MH	Media haze	317
DN	Drusen	138
MYA	Myopia	101
BRVO	Branch retinal vein occlusion	73
TSLN	Tessellation	186
ERM	Epiretinal membrane	14
LS	Laser scars	47
MS	Macular scars	15
CSR	Central serous retinopathy	37
ODC	Optic disc cupping	282
CRVO	Central retinal vein occlusion	28
TV	Tortuous vessels	6
AH	Asteroid hyalosis	16
ODP	Optic disc pallor	65
ODE	Optic disc edema	58
ST	Optociliary shunt	5
AION	Anterior ischemic optic neuropathy	17
PT	Parafoveal telangiectasia	11
RT	Retinal traction	14
RS	Retinitis	43
CRS	Chorioretinitis	32
EDN	Exudation	15
RPEC	Retinal pigment epithelium changes	22
MHL	Macular hole	11
RP	Retinitis pigmentosa	6
CWS	Cotton wool spots	3
CB	Coloboma	1
ODPM	Optic disc pit maculopathy	0
PRH	Preretinal hemorrhage	2
MNF	Myelinated nerve fibers	3
HR	Hemorrhagic retinopathy	0
CRAO	Central retinal artery occlusion	2
TD	Tilted disc	3
CME	Cystoid macular edema	4
PTCR	Post-traumatic choroidal rupture	5
CF	Choroidal folds	3
VH	Vitreous hemorrhage	1
MCA	Macroaneurysm	1
VS	Vasculitis	1
BRAO	Branch retinal artery occlusion	2
PLQ	Plaque	1
HPED	Hemorrhagic pigment epithelial detachment	1
CL	Collateral	1
Disease Risk	Disease Risk	1599

As the ODPM and HR classes do not contain data, they were removed. This study intended to identify different diseases, so the disease risk attribute was also removed. Additionally, as some columns have a smaller number of samples, data augmentation was performed using class weights for the DR class having 376 instances. Since some images in the dataset represent more than one disease, multilabel classification was carried out. Image augmentation is deployed to artificially increase the size of a training dataset by generating modified versions of its images. This study utilized various transformations for data augmentation, including a range of operations, such as shift, flip, and zoom for images. The ImageDataGenerator class in the Keras library enables model fitting engaging augmented image data. For data generation purposes, horizontal and vertical flips were used

with a rotation range of 90°. Data rescaling was performed to convert each pixel's value ranging between 0 and 255 to 0 and 1. In this dataset, some images have high pixel values, whereas others have low values. Since all the images in the same dataset implement the same model, learning rate, and weights, a pixel with a more extensive value range indicates a higher loss, whereas a lower range suggests a low loss. The aggregate of both losses contributes to the weight update of the learning model. Scaling all the images in the same range distributes the loss more evenly. As this investigation aimed at multilabel classification, class_mode was set to raw in flow_from_dataframe during data generation, with a target size of 150×150.

The second dataset (Dataset 2) [35] does not contain any metadata file, but it is well-balanced for all three categories of eye disease. Thus, image augmentation was not applied. The third dataset (Dataset 3) [37] was collected from 6392 eye patients, 2968 and 3424 of which were females and males, respectively, as shown in Figure 1. The age groups 51-61 and 11-21 have the most and least instances. It was observed that diabetes and other classes signify a negative correlation with normal classes, whereas the rest have a little correlation. The correlation heatmap for datasets one and three did not imply significant relationships among them. However, this provides a reasonable estimate of class separability. Dataset 3 also has the issue of class imbalance, as three classes represent 78% of data, whereas the rest five classes represent only 22% of data. Data augmentation was also applied for Dataset 3, corresponding to the class with the highest number of instances, i.e., diabetes.

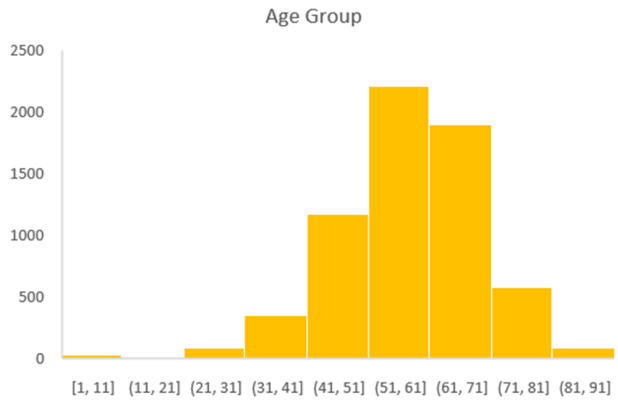

Fig. 1. Age groups of Dataset 2.

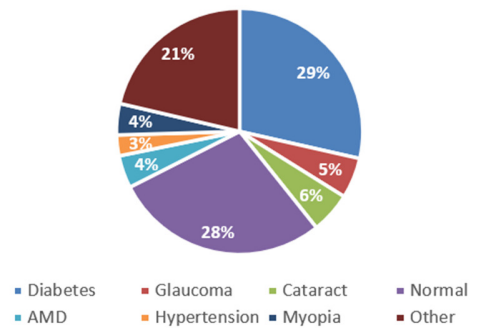

Fig. 2. Classes of Dataset 3.

B. .Pre-Trained Models' Implementation for Eye Disease Classification

Transfer learning leverages a pre-trained model's knowledge to address new problems by applying the insights gained from one task to a variety of others. In this process, the model takes retina fundus images sourced from three different datasets as input and proceeds to group each eye disease based on the dataset. Subsequently, the input, classification, and output layers were replaced by a pre-trained model. In the next step, transfer learning was employed on deep pre-trained CNNs, which were then fine-tuned using experimental data related to eye diseases. This study employed UNet, ResNet 152, EfficientNetB0, and B7 pre-trained models. At first, the U-Net model was implemented. The energy function for UNet was computed based on the softmax function applied pixel-wise, given by:

$$p_e(a) = \frac{\exp(x_e(a))}{\sum_{e'=1}^{E} \exp(x_{e'}(a))} \quad (1)$$

where $x_e(a)$ represents feature channel e for the pixel position $a \in \Omega$, $\Omega \subset \mathbb{Z}^2$, E is the class number, and $p_e(a)$ is a max function. Cross-entropy (C) is given by:

$$C = \sum_{a \in \Omega}^{E} \mathrm{w(a)} \log(p_{l(a)}(a)) \quad (2)$$

where l and w represent the actual pixel label and weight maps, respectively, such that $l, w: \Omega \rightarrow \mathbb{R}$. The weighted map needs to be re-compute to achieve separability of pixels by:

$$m(a) = m_c(a) + m_0 . exp\left(\frac{-(f_1(x)+f_2(x))^2}{2\sigma^2}\right) \quad (3)$$

where m_c represents a weighted map, f_1 and f_2 are the pixel distances from the separable border and the second nearest pixel distance from the border, such that m_c, f_1, and $f_2: \Omega \rightarrow \mathbb{R}$. Weight initialization was obtained based on Gaussian distribution with the standard deviation (SD) of $\sqrt{2} / I$, where I represents the inbound nodes.

This study used Keras, PyTorch, TensorFlow, pandas, NumPy, time, cv2, matplotlib, seaborn, sklearn, tqdm, and mplot3d libraries. The U-Net model contains four convolution blocks. In each convolution block, there are two conv2d layers followed by a max pool layer. Conv2d is a two-dimensional convolution layer in the Keras library that generates a convolution kernel winding with the layers' input, resulting in a tensor of outputs. A max pool 2D layer was utilized to downsample the input dimensions by calculating the maximum value based on a pool size of 2. A stride value of 2 was employed to shift the window. After four convolution blocks, the bottleneck layer was deployed with two conv2d layers. In DL, bottlenecks force the model to learn a compression algorithm for the input data. Four upsampling blocks were added in the U-Net model to add successive layers to a conventional contracting network in place of pooling operations. The final layer was flattened to obtain the desired output size per the dimensions required through the intermediate values received from the previous layers.

The EfficientNet model is based on compound coefficient Φ to uniformly scale the three dimensions of the network: depth (d), width (w), and image resolution (i) [31].

C. DNNRFDC Model Implementation for Eye Disease Classification

The DNNRFDC model has 13 layers, which are fewer than the pre-trained models used. For instance, EfficientNet B7 has more than 800 layers, and ResNet has 152 layers. The first two layers of DNNRFDC were conv2d layers. The parameters of the conv2d layers are the size of units, kernel size, input shape, and activation function. A kernel size of 3 was selected and the softplus activation function was implemented. The third layer was the batch normalization layer, which was employed to standardize the inputs in the model. The fourth and fifth layers were maxpool2d and dropout. Two conv2d layers were added as sixth and seventh layers with a kernel size of 3 and the softplus activation function, followed by the maxpool2d and dropout layers. A flattened layer was also used, followed by a dense layer with a softmax function as a classifier for 4, 8, and 43 output values required for datasets 1, 2, and 3, respectively.

V. EVALUATION

Five-fold cross-validation was performed to repeat the training procedure before determining the performance of the models. The following metrics were applied to gauge the quality of the model: precision, recall, F1-score, F-2 score, accuracy, Area Under the Curve (AUC), Intersection Over Union (IOU), and Matthews Correlation Coefficient (MCC), as exhibited in (4)-(9).

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (4)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (5)$$

$$\text{F1} - \text{Score} = \frac{2*(precision*recall)}{(precision+recall)} \quad (6)$$

$$\text{F2} - \text{Score} = \frac{(5*precision*recall)}{(4*precision+recall)} \quad (7)$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FN + FP + TN)} \quad (8)$$

$$\text{MCC} = \frac{(TP \times TN - FP \times FN)}{(TP + FP)(TP + FN)(TN + FP)(TN + FN)} \quad (9)$$

where *TP*, *TN*, *FP*, and *FN* represent true positives, true negatives, false positives, and false negatives, respectively.

Along with various performance metrics, floating point operations per second (FLOPS) were also used. Arbitrary performance indicators, such as accuracy and recall, might instill illusionary confidence in the output and result. The proposed DNNRFDC and all pre-trained models employed were evaluated for their unbiased performance deploying an averaged Receiver Operator Characteristic (ROC) curve for all classes based on three datasets. The ROC curve is generally utilized to evaluate the performance of the binary classifiers. However, since this study performed multiclass classification, an one-versus-all technique was adopted to construct weighted receiver operating characteristic curves. ROC is an important performance indicator for classifier evaluation, as it does not depend on the class distribution. Comparing different classifiers is possible by condensing their output into a single metric. The area under the ROC curve is known as the AUC metric, which is also used to describe accuracy.

VI. RESULTS AND DISCUSSIONS

Table IV displays the results of the DNNRFDC and the pre-trained models. This study aimed to increase classification accuracy while reducing training time latency and preventing overfitting. Three datasets were implemented, each one with a different number of classification outputs (i.e., 4, 8, and 43). The first dataset's classification accuracy ranged from 77.07% to 84.47% for the pre-trained networks, whereas the proposed DNNRFDC model disclosed better results in all metrics. For Dataset 2, DNNRFDC attained an accuracy of 96.75%, followed by EfficientNetB7 accomplishing 91.02% accuracy. In Dataset 3, the proposed DNNRFDC model achieved an accuracy of 93.06% and all other pre-trained models reached accuracies between 82.43 to 83.73%. It was expected that the deeper architectures should perform well on the large Dataset 1 compared to the smaller Dataset 2. However, the overall highest accuracy was better for Dataset 2, followed by Dataset 3 and Dataset 1. The primary reason is that Dataset 2 contains only four classes, while Datasets 3 and 1 contain 8 and 43 classes. Figure 3 illustrates the accuracy and loss curve per epoch for the proposed DNNRFDC model for eye disease classification on the three datasets.

TABLE IV. PERFORMANCE EVALUATION OF DIFFERENT TRANSFER LEARNING ARCHITECTURES

Datasets	Model	Accuracy	AUC	Precision	Recall	F1-score	F2-score	MCC
Dataset-1	DNNRFDC	92.5	93.17	0.93	0.94	0.93	0.94	0.93
	EfficientNet-B7	84.47	86.52	0.85	0.86	0.85	0.86	0.85
	UNet	80.08	81.97	0.81	0.82	0.81	0.82	0.81
	EfficientNet-B0	77.07	78.56	0.77	0.78	0.77	0.78	0.78
	ResNet 152	82.17	82.99	0.82	0.83	0.82	0.83	0.81
Dataset-2	DNNRFDC	96.75	98.49	0.96	0.97	0.96	0.97	0.97
	EfficientNet-B7	91.02	90.89	0.91	0.92	0.91	0.92	0.91
	UNet	85.97	86.54	0.85	0.86	0.85	0.86	0.85
	EfficientNet-B0	79.39	81.93	0.79	0.8	0.79	0.8	0.78
	ResNet 152	87.02	88.68	0.88	0.89	0.88	0.89	0.88
Dataset-3	DNNRFDC	93.06	94.17	0.93	0.94	0.93	0.94	0.93
	EfficientNet-B7	82.72	83.76	0.83	0.84	0.83	0.84	0.83
	UNet	82.43	84.94	0.85	0.86	0.85	0.86	0.82
	EfficientNet-B0	83.58	83.93	0.84	0.85	0.84	0.85	0.83
	ResNet 152	83.73	85.81	0.86	0.87	0.86	0.87	0.84

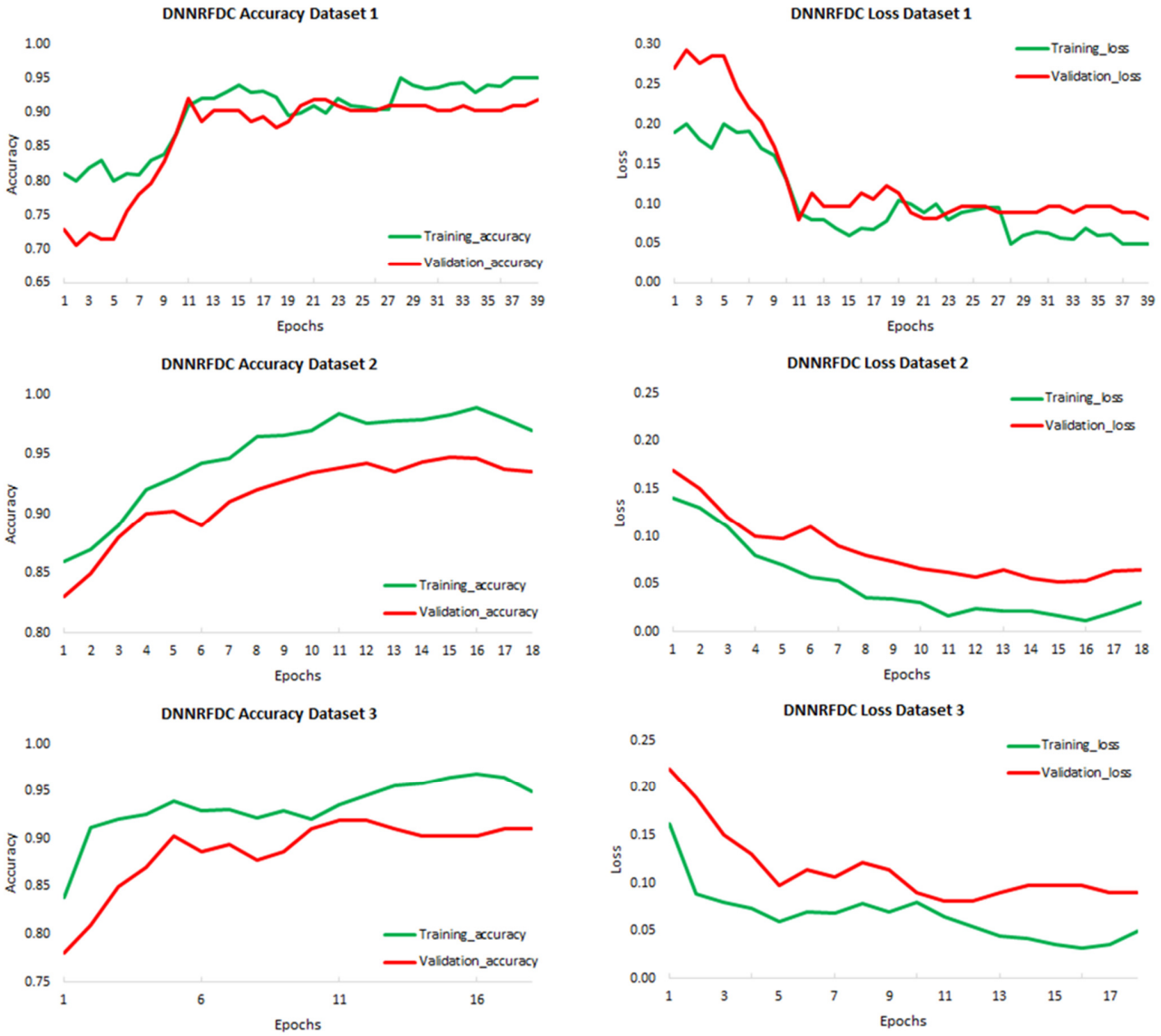

Fig. 3. Accuracy and loss performance of the proposed DNNRFDC model for eye disease classification on all datasets.

The AUC was used to summarize the model performance into a single metric. Figure 3(a) shows that the AUC of the DNNRFDC model was higher for Dataset 2, being 98.49%, followed by 94.17 and 93.17% for Dataset 3 and 1, respectively. It was observed that the AUC obtained was slightly higher than the accuracy. This can occur if the model acquired a high classification for the positive category at the expense of high false negatives. Figure 4 represents the ROC curve for different models on the three datasets. Interestingly, the ROCs for Dataset 1 and Dataset 3 were closer than for Dataset 2 for all models, except for ResNet152, where all three ROCs have different trajectories.

A. Comparison with Other Studies

Table VI manifests a comparison of the results of this study along with previous ones to evaluate the developed model and the tuned pre-trained models. In [38], eye diseases were classified into seven classes, with one normal class, using backpropagation with a linear cyclic learning rate, achieving an accuracy of 89.83%. The accuracies acquired by [12] ranged from 81.53 to 86.63%. In [27], a very high accuracy of 98% was attained, but this study performed binary classification of glaucoma and no glaucoma. In [24], the Duke glaucoma repository was employed to determine the neuroretinal damage caused by glaucoma using the ResNet34 pre-trained model. The results demonstrated a significant AUC of 0.95, but this study has limitations, as it applied only one model for one disease, and the chosen data-splitting strategy could affect the results. This study developed a novel DNNRFDC model. The latest pre-trained models were tuned and trained on three different datasets to appraise the developed and pre-trained models. The issue of data splitting was addressed with k-fold cross-validation to determine the performance based on different subsets of the data. The results revealed an excellent performance for the proposed DNNRFDC model.

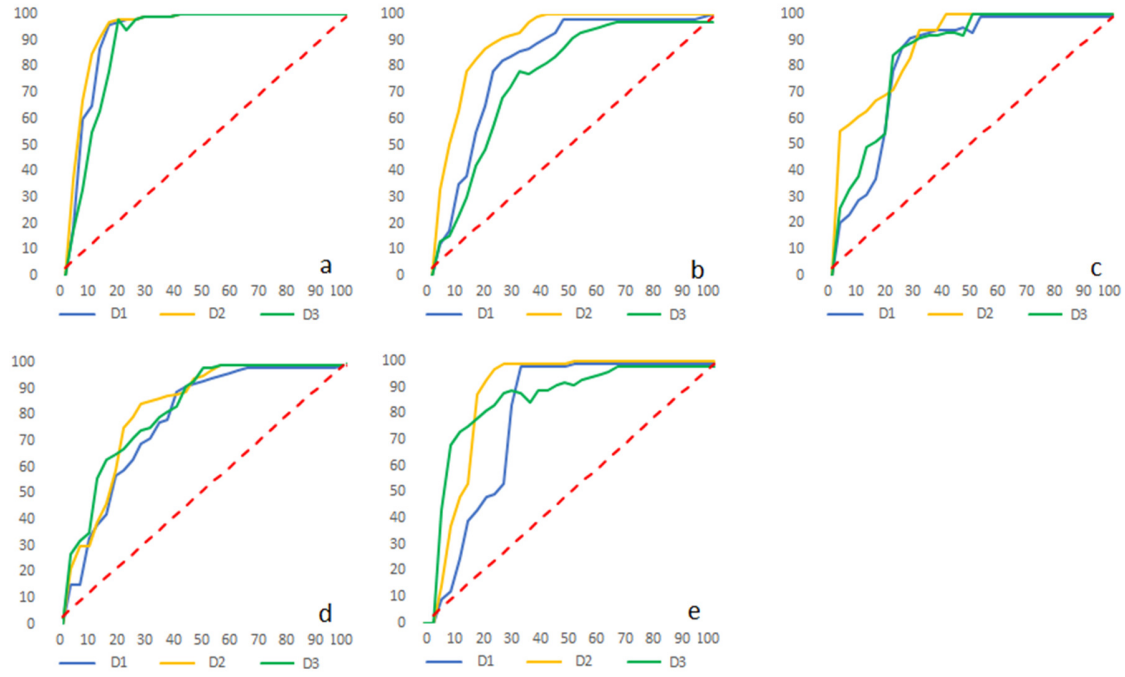

Fig. 4. ROC curves for (a) developed DNNRFDC model, pre-trained models including (b) EfficientNetB7, (c) UNet, (d) EfficientNetB0, and (e) ResNet152.

B. Computational Complexity

The experimental environment for this investigation was an NVIDIA GeForce GTX 1060 GPU on an Intel Core i9-9900K CPU featuring 8 cores and 16 threads. Classification accuracy was not affected by variations in the number of training parameters and the depth of the architectures, as depicted in Table V. Among the four pre-trained networks examined, EfficientNetB7 consistently demonstrated the highest overall accuracy across all three datasets. Notably, the updated version of the pre-trained networks outperformed their predecessor. For example, EfficientNetB7 exhibited superior performance compared to EfficientNetB0. However, it is worth noting that the higher version demanded more resources in terms of depth, parameter count, and inference time per operation, presenting a potential trade-off between performance and resource utilization.

In contrast, the DNNRFDC model boasted a modest 0.2 million parameters and a compact 2.39 MB size, making it the most resource-efficient among the models employed. With only 13 layers of depth, DNNRFDC qualifies as a lightweight CNN model that maintains impressive accuracy while demanding minimal GPU inference time, clocking in at just 1.39 ms. This accelerated processing can be attributed to the efficient utilization of Google Cloud TPU's v2, offering 180 TFLOPS of computing power and 64 GB of high-bandwidth memory. One more significant achievement of DNNRFDC was the Ratio between Trainable and Total parameters (RTT). DNNRFDC tops for RTT with a value of 100%, while UNet, EfficientNetB7, EfficientNetB0, and ResNet152 obtained RTT values of 0.995, 0.995, 0.99, and 0.985, respectively.

TABLE V. COMPLEXITY METRICS OF MODELS

Model	Overall accuracy	Size (MB)	Parameters	Depth	GPU time per inference step (ms)	TFLOPS
ResNet152 V2	84.30	232	60,380,648	152	6.64	11.3
EfficientNetB0	80.01	29	5,330,571	237	4.91	0.39
EfficientNetB7	86.07	256	66,658,687	813	61.62	37
UNet	82.83	386.6	18,839,232	23	17.65	0.52
DNNRFDC	94.10	2.39	208,776	13	1.39	0.0014

C. Limitations

When the classification layer's learned features cannot classify the problem set, transfer learning becomes problematic due to the dissimilarity between the datasets on which it was initially trained and the dataset for the other problem. Additionally, feature transfer fails in this condition. For example, some layers can be removed in a conventional model to improve accuracy. However, doing the same thing with transfer learning reduces the pool of trainable parameters, increasing the risk of overfitting. If the model insists, a lengthy process is required. The proposed DNNRFDC model shows good accuracy on all three datasets with different eye disease classifications. However, its accuracy and stability should be examined on unseen larger datasets. On the other hand, the pre-trained models used in this study are well-established and have displayed remarkable performance on large datasets, such as ImageNet, COCO, and CIFAR. The UNet model was developed for semantic segmentation purposes with the ability of image localization pixel by pixel. However, the UNet model for the classification task disclosed a gradient classification issue on Dataset 3. ResNet152 exhibits architectural complexity due to its significant parameters and size.

TABLE VI. COMPARISON OF DNNRFDC AND EXAMINED PRE-TRAINED MODELS WITH PREVIOUS STUDIES

Ref	Method	Accuracy	F1-score	Sensitivity	Specificity	AUC
[38]	NN	0.90	0.89	0.99	0.89	0.94
[12]	DT	0.86	0.85	0.85	0.87	-
	NB	0.82	0.80	0.81	0.82	-
	RF	0.87	0.86	0.89	0.89	-
	NN	0.86	0.85	0.86	0.86	-
[17]	LR.	-	-	0.96	0.77	0.95
	RF.	-	-	0.94	0.78	0.95
	SVM	-	-	0.93	0.83	0.93
	NB.	-	-	0.96	0.79	0.95
[25]	VGG-16	0.81	0.80	0.80	0.90	-
	VGG-19	0.91	0.78	0.87	0.94	-
	ResNet-50	0.91	0.85	0.89	0.96	-
	DenseNet-121	0.84	0.83	0.80	0.96	-
	DenseNet-169	0.81	0.81	0.90	0.90	-
	DenseNet-201	0.94	0.90	0.50	0.90	-
	Inception-V3	0.53	0.50	0.75	0.70	-
	InceptionResNet-V2	0.84	0.85	0.90	0.90	-
	Xception	0.94	0.88	0.90	0.92	-
[26]	Transfer Learning	0.67	0.52	0.70	0.55	-
[21]	LR1	0.70	-	0.91	0.56	0.72
	LR2	0.69	-	0.70	0.59	0.72
	LR3	0.71	-	0.87	0.62	0.78
	HWG	0.72	-	0.93	0.62	0.80
	CNN	0.80	-	0.91	0.77	0.87
	VGG	0.81	-	0.88	0.79	0.87
	GoogLeNet	0.80	-	0.81	0.74	0.87
	ResNet	0.81	-	0.84	0.74	0.87
	NMD+CNN	0.84	-	0.85	0.83	0.91
	SOD+CNN	0.84	-	0.84	0.81	0.90
	NMD+Attention	0.85	0.00	0.84	0.85	0.91
	TIA-Net	0.86	0.00	0.85	0.87	0.93
[27]	Autoencoder	0.98	0.95	0.98	0.98	-
[22]	DL	-	-	-	-	0.80
[23]	DCNN	-	-	-	-	0.94
[24]	DCNN	-	-	-	-	0.95
This study	DNNRFDC	0.94	0.94	0.95	0.94	0.95
	EfficientNet-B7	0.86	0.86	0.87	0.86	0.87
	UNet	0.83	0.84	0.85	0.84	0.84
	EfficientNet-B0	0.80	0.80	0.81	0.80	0.81
	ResNet 152	0.84	0.85	0.86	0.85	0.86

VII. CONCLUSION

This study described the design and development of an improved eye disease classification model, called DNNRFDC, evaluated on three separate datasets and compared with four well-established pre-trained models. The proposed DNNRFDC model demonstrated superior accuracy in the automatic classification of a diverse range of eye disorders. Compared to existing pre-trained models, such as ResNet152 V2, EfficientNetB0, EfficientNetB7, and UNet, the DNNRFDC model achieved significantly higher overall accuracy metrics across the three datasets. The proposed model introduced several key changes to enhance its performance, namely optimizing the network architecture, refining the training process, and achieving less number of parameters with better computational cost. These modifications contributed to the model's improved accuracy and efficiency, rendering it a more reliable tool for ophthalmological diagnosis and care. The future scope of this research will be to implement various other features that affect eye disease and incorporate more datasets and a variety of advanced models [39-42].

ACKNOWLEDGMENT

Zamil S. Alzamil would like to thank the Deanship of Postgraduate Studies and Scientific Research at Majmaah University for supporting this work under Project No. R-2024-1073.

REFERENCES

[1] "Eye care, vision impairment and blindness." https://www.who.int/health-topics/blindness-and-vision-loss.

[2] S. Pachade *et al.*, "Retinal Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research," *Data*, vol. 6, no. 2, Feb. 2021, Art. no. 14, https://doi.org/10.3390/data6020014.

[3] T. J. MacGillivray, E. Trucco, J. R. Cameron, B. Dhillon, J. G. Houston, and E. J. R. van Beek, "Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions," *British Journal of Radiology*, vol. 87, no. 1040, Aug. 2014, Art. no. 20130832, https://doi.org/10.1259/bjr.20130832.

[4] W. Yang, H. Xu, X. Yu, and Y. Wang, "Association between retinal artery lesions and nonalcoholic fatty liver disease," *Hepatology International*, vol. 9, no. 2, pp. 278–282, Apr. 2015, https://doi.org/10.1007/s12072-015-9607-3.

[5] K. S. Lee, K. H. Nam, D. W. Kim, E. C. Kang, and H. J. Koh, "Risk of Retinal Vein Occlusion in Patients With End-Stage Renal Disease: A 12-Year, Retrospective, Nationwide Cohort Study in South Korea," *Investigative Ophthalmology & Visual Science*, vol. 59, no. 1, pp. 39–44, Jan. 2018, https://doi.org/10.1167/iovs.17-22638.

[6] P. Porwal *et al.*, "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research," *Data*, vol. 3, no. 3, Sep. 2018, Art. no. 25, https://doi.org/10.3390/data3030025.

[7] E. Decencière *et al.*, "Feedback on a Publicly Distributed Image Database: The Messidor Database," *Image Analysis and Stereology*, vol. 33, no. 3, pp. 231–234, Aug. 2014, https://doi.org/10.5566/ias.1155.

[8] J. I. Orlando *et al.*, "REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical Image Analysis*, vol. 59, Jan. 2020, Art. no. 101570, https://doi.org/10.1016/j.media.2019.101570.

[9] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database," *PLOS ONE*, vol. 12, no. 11, 2017, Art. no. e0187336, https://doi.org/10.1371/journal.pone.0187336.

[10] G. Quellec, M. Lamard, P. H. Conze, P. Massin, and B. Cochener, "Automatic detection of rare pathologies in fundus photographs using few-shot learning," *Medical Image Analysis*, vol. 61, Apr. 2020, Art. no. 101660, https://doi.org/10.1016/j.media.2020.101660.

[11] T. D. Stuckey *et al.*, "Cardiac Phase Space Tomography: A novel method of assessing coronary artery disease utilizing machine learning," *PLOS ONE*, vol. 13, no. 8, 2018, Art. no. e0198603, https://doi.org/10.1371/journal.pone.0198603.

[12] S. Malik, N. Kanwal, M. N. Asghar, M. A. A. Sadiq, I. Karamat, and M. Fleury, "Data Driven Approach for Eye Disease Classification with Machine Learning," *Applied Sciences*, vol. 9, no. 14, Jan. 2019, Art. no. 2789, https://doi.org/10.3390/app9142789.

[13] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 720–727, Dec. 2008, https://doi.org/10.1016/j.compmedimag.2008.08.009.

[14] G. An *et al.*, "Glaucoma Diagnosis with Machine Learning Based on Optical Coherence Tomography and Color Fundus Images," *Journal of Healthcare Engineering*, vol. 2019, 2019, Art. no. 4061313, https://doi.org/10.1155/2019/4061313.

[15] S. Kuwayama *et al.*, "Automated Detection of Macular Diseases by Optical Coherence Tomography and Artificial Intelligence Machine Learning of Optical Coherence Tomography Images," *Journal of Ophthalmology*, vol. 2019, Apr. 2019, Art. no. e6319581, https://doi.org/10.1155/2019/6319581.

[16] S. O. Fageeri, S. M. M. Ahmed, S. A. Almubarak, and A. A. Mu'azu, "Eye refractive error classification using machine learning techniques," in *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, Khartoum, Sudan, Jan. 2017, pp. 1–6, https://doi.org/10.1109/ICCCCEE.2017.7867660.

[17] M. E. Nongpiur *et al.*, "Classification Algorithms Based on Anterior Segment Optical Coherence Tomography Measurements for Detection of Angle Closure," *Ophthalmology*, vol. 120, no. 1, pp. 48–54, Jan. 2013, https://doi.org/10.1016/j.ophtha.2012.07.005.

[18] S. P. Imberman, I. Ludwig, and S. Zelikovitz, "Using Decision Trees To Find Patterns in an Ophthalmology Dataset," in *Proceedings of the 24th International Florida Artificial Intelligence Research Society*, 2011, pp. 95–96.

[19] M. H. Brilliant *et al.*, "Mining Retrospective Data for Virtual Prospective Drug Repurposing: L-DOPA and Age-related Macular Degeneration," *The American Journal of Medicine*, vol. 129, no. 3, pp. 292–298, Mar. 2016, https://doi.org/10.1016/j.amjmed.2015.10.015.

[20] U. Farooq and N. Y. Sattar, "Improved automatic localization of optic disc in Retinal Fundus using image enhancement techniques and SVM," in *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, Penang, Malaysia, Nov. 2015, pp. 532–537, https://doi.org/10.1109/ICCSCE.2015.7482242.

[21] X. Xu, Y. Guan, J. Li, Z. Ma, L. Zhang, and L. Li, "Automatic glaucoma detection based on transfer induced attention network," *BioMedical Engineering OnLine*, vol. 20, no. 1, Apr. 2021, Art. no. 39, https://doi.org/10.1186/s12938-021-00877-5.

[22] A. A. Jammal *et al.*, "Human Versus Machine: Comparing a Deep Learning Algorithm to Human Gradings for Detecting Glaucoma on Fundus Photographs," *American Journal of Ophthalmology*, vol. 211, pp. 123–131, Mar. 2020, https://doi.org/10.1016/j.ajo.2019.11.006.

[23] F. A. Medeiros, A. A. Jammal, and A. C. Thompson, "From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs," *Ophthalmology*, vol. 126, no. 4, pp. 513–521, Apr. 2019, https://doi.org/10.1016/j.ophtha.2018.12.033.

[24] A. C. Thompson, A. A. Jammal, and F. A. Medeiros, "A Deep Learning Algorithm to Quantify Neuroretinal Rim Loss From Optic Disc Photographs," *American Journal of Ophthalmology*, vol. 201, pp. 9–18, May 2019, https://doi.org/10.1016/j.ajo.2019.01.011.

[25] D. S. Jagan, "An Artificial Intelligence Based Tool for Eye Disease Classification," *International Research Journal in Global Engineering and Sciences.*, vol. 4, no. 3, pp. 87–107, 2019.

[26] Y. Bao *et al.*, "Self-adaptive Transfer Learning for Multicenter Glaucoma Classification in Fundus Retina Images," in *Ophthalmic Medical Image Analysis*, Strasbourg, France, 2021, pp. 129–138, https://doi.org/10.1007/978-3-030-87000-3_14.

[27] U. Raghavendra, A. Gudigar, S. V. Bhandary, T. N. Rao, E. J. Ciaccio, and U. R. Acharya, "A Two Layer Sparse Autoencoder for Glaucoma Identification with Fundus Images," *Journal of Medical Systems*, vol. 43, no. 9, Jul. 2019, Art. no. 299, https://doi.org/10.1007/s10916-019-1427-x.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Munich, Germany, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

[29] P. K. Gadosey *et al.*, "SD-UNet: Stripping down U-Net for Segmentation of Biomedical Images on Platforms with Low Computational Budgets," *Diagnostics*, vol. 10, no. 2, Feb. 2020, Art. no. 110, https://doi.org/10.3390/diagnostics10020110.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

[31] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, USA, May 2019, pp. 6105–6114.

[32] M. Tan and Q. V. Le, "EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling," May 2019. http://research.google/blog/efficientnet-improving-accuracy-and-efficiency-through-automl-and-model-scaling/.

[33] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomedical Optics Express*, vol. 8, no. 2, pp. 579–592, Feb. 2017, https://doi.org/10.1364/BOE.8.000579.

[34] S. Pachade *et. al.*, "Retinal Fundus Multi-disease Image Dataset (RFMiD)." IEEE, Nov. 25, 2020, [Online]. Available: https://ieee-dataport.org/open-access/retinal-fundus-multi-disease-image-dataset-rfmid.

[35] "Retina dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/jr2ngb/cataractdataset.

[36] "ODIR-2019 Dataset - Grand Challenge." [Online]. Available: https://odir2019.grand-challenge.org/dataset/.

[37] S. M. Khan *et al.*, "A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability," *The Lancet Digital Health*, vol. 3, no. 1, pp. e51–e66, Jan. 2021, https://doi.org/10.1016/S2589-7500(20)30240-5.

[38] H. M. Ahmed and S. R. Hameed, "Eye Diseases Classification Using Back Propagation Artificial Neural Network," *Engineering and Technology Journal*, vol. 39, no. 1B, pp. 11–20, Mar. 2021, https://doi.org/10.30684/etj.v39i1B.1363.

[39] P. Chakraborty and C. Tharini, "Pneumonia and Eye Disease Detection using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5769–5774, Jun. 2020, https://doi.org/10.48084/etasr.3503.

[40] A. N. Saeed, "A Machine Learning based Approach for Segmenting Retinal Nerve Images using Artificial Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5986–5991, Aug. 2020, https://doi.org/10.48084/etasr.3666.

[41] A. Abbas, U. Maqsood, S. U. Rehman, K. Mahmood, T. AlSaedi, and M. Kundi, "An Artificial Intelligence Framework for Disease Detection in Potato Plants," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12628–12635, Feb. 2024, https://doi.org/10.48084/etasr.6456.

[42] S. Alqethami, B. Almtanni, W. Alzhrani, and M. Alghamdi, "Disease Detection in Apple Leaves Using Image Processing Techniques," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8335–8341, Apr. 2022, https://doi.org/10.48084/etasr.4721.