

A new technique for cataract eye disease diagnosis in deep learning

Salwa Shakir Mahmood¹, Sihem Chaabouni², Ahmed Fakhfakh³

¹ATISP research unit National School of Electronics and Telecommunications of Sfax (ENET'Com), Sfax University, Tunisia

² Centre De Recherche En Num'Érique De Sfax, Tunisia Enet'com, Sfax Technopole, Sfax University Tunisia

³Centre De Recherche En Num'Érique De Sfax, Tunisia

ABSTRACT

Automated diagnosis of eye diseases using fundus images is challenging because manual analysis is time-consuming, prone to errors, and complicated. Thus, computer-aided tools for automatically detecting various ocular disorders from fundus images are needed. Deep learning algorithms enable improved image classification, making automated targeted ocular disease detection feasible. This study employed state-of-the-art deep learning image classifiers, such as VGG-19, to categorize the highly imbalanced ODIR-5K (Ocular Disease Intelligent Recognition) dataset of 5000 fundus images across eight disease classes, including cataract, glaucoma, diabetic retinopathy, and age-related macular degeneration. To address this imbalance, the multiclass problem is converted into binary classification tasks with equal samples in each category. The dataset was preprocessed and augmented to generate balanced datasets. The binary classifiers were trained on flat data using the VGG-19 (Visual Geometry Group) model. This approach achieved an accuracy of 95% for distinguishing normal versus cataract cases in only 15 epochs, outperforming the previous methods. Precision and recall were high for both classes – Normal and Cataract, with F1 scores of 0.95-0.96. Balancing the dataset and using deep VGG-19 classifiers significantly improved automated eye disease diagnosis accuracy from fundus images. With further research, this approach could lead to deploying AI (Artificial Intelligence)-assisted tools for ophthalmologists to screen patients and support clinical decision-making.

Keywords: Fundus images, Ocular diseases, Deep learning, VGG-19, Imbalanced data, Binary classification, Transfer learning

Corresponding Author:

Salwa Shakir Mahmood

ATISP research unit National School of Electronics and Telecommunications of Sfax (ENET'Com),
Sfax University, Tunisia

E-mail: salwa1982m@gmail.com

1. Introduction

Ocular diseases encompass any abnormalities or impairments that disrupt the proper functioning of the eye or negatively affect visual sharpness. Retinal conditions are leading reasons for blindness worldwide, including glaucoma, diabetic retinopathy, cataracts, and age-related macular degeneration. Studies have estimated that over 400 million people will have diabetic retinopathy by 2030 [1, 2]. The timely detection of these diseases helps avoid vision loss. However, a significant gap exists between ophthalmologists and patients. Manual fundus examination is time-demanding and relies heavily on specialist experience, complicating large-scale screening. Therefore, automated computer-aided diagnosis is critical for detecting eye diseases [3].

The global prevalence of eye diseases varies widely, contingent on age, sex, occupation, socioeconomics, hygiene, customs, climate... Studies have demonstrated that tropical populations have a higher incidence of ocular infections than temperate regions, attributed to environmental factors, including dust, humidity, sunlight, and other factors [4]. World Health Organization quotes that approximately 2.2 billion have near- or far-vision

impairment globally [5]. According to projections, 50% of these cases could have been avoided or treated. Approximately one billion people suffer from moderate to severe distance vision issues or blindness stemming from uncorrected refractive errors, cataracts, glaucoma, corneal opacities, diabetic retinopathy, and trachoma. Over 800 million people also experience near-vision deficits due to uncorrected presbyopia [6]. Affordable eye care access is imperative for underserved groups. Deep learning is gaining traction in medical imaging, showing promise for detection, classification, and diagnosis. Automating disease identification may reduce ophthalmologists' workload.

This study aims to develop a system that can accurately categorize eye diseases into one of two classes - cataract or normal (non-cataract). However, the utilized dataset exhibited substantial class imbalance, rendering disease categorization inadvisable as imbalance engenders training instability. A binary class-balancing approach was employed, whereby equal samples from two classes were extracted and input into a pre-trained VGG-19 (Visual Geometry Group) model. Rather than simultaneous multiclass disease classification on an entirely imbalanced dataset, this tactic focused on balanced two-class – standard and cataract - categorization, thereby enhancing model training and performance. Consequently, this research commenced by balancing the dataset by extracting equivalent samples across classes and training the models using the pre-existing VGG-19 architecture.

The methodology loads and populates the dataset with matched image quantities per class. Transfer learning was used to optimize the VGG-19 model. Balancing the data improved class accuracy for standard and cataract classes. The rest of this article is structured subsequently. Section 2 reviews the relevant prior work. Section 3 comprehensively delineates the tools and techniques used in this study. Section 4 analyzes the experimental outcomes and the performance of the model. In the concluding Section 5, we summarize this research study's essential findings and contributions. We also discuss potential avenues for future work, building on the results presented here.

2. Related work

Various methods have been suggested for ocular disease classification, including a two-stage technique using CNNs (convolutional neural network) for optic disc localization [7], knowledge distillation models with sequential deep network training [8], and ReLayNet (Retinal Layer) - an encoder-decoder fully convolutional network for semantic segmentation of retinal layers and fluids from OCT (Optical coherence tomography) scans [9]. These studies demonstrate deep learning techniques for automated feature extraction and analysis from fundus and OCT ophthalmic imaging. Researchers developed a method to diagnose different retinal diseases using OCT [10]. They performed pixel-wise classification of OCT scans with CNNs using dilated convolution filters, evaluating performance on 400 AMD (Age-related macular degeneration) patient scans. In another study, Hu et al. [11] proposed a CNN approach to detect intraretinal fluid in OCT images. Their CNN model was trained on 1289 OCT scans, achieving a 0.911 Dice score in cross-validation. Distinctly, the authors presented a supervised learning technique employing a novel convolutional multitask structure [12]. This model was trained to concurrently segment bright and red lesions and detect lesions from fundus images, achieving strong performance with a 0.839 AUC (Accuracy). In addition, researchers have proposed a new method for segmenting retinal blood vessels using conditional random fields connected to a convolutional neural network (CRF-CNN) [13]. They tested how well their CRFs-CNN model performed by analyzing its accuracy and effectiveness on color fundus images from two existing datasets called STARE (Structured Analysis of Retina) and DRIVE (Digital Retinal Images for Vessel Extraction) [14, 15]. Similarly, Khan et al. [16] developed an automated deep-learning method to detect diabetic macular edema and retinopathy. They accomplished this by optimizing a neural network image classification model.

Researchers have also proposed using deep-learning methods called GANs (Generative Adversarial Networks) to detect glaucomatous optic neuropathy, including over 8000 color fundus images for model training. They attained an area under the curve (AUC) of 0.98, with 92.2% specificity and 95.6% sensitivity. Another study diagnosed different retinal diseases from OCT images using fine-tuned CNNs, such as GoogLeNet [17]. They were classified into diabetic macular edema, dry age-related macular degeneration, and no pathology. VGG-19 has also been used to detect cataracts in color fundus images [2]. Other studies have investigated evaluation principles for different methods [18], optimization of deep learning models for eye disease detection [19], and benchmarking state-of-the-art deep neural networks [20].

3. Methodology

The methodology used to classify ocular images showing cataract eye disease accurately uses deep learning techniques. First, the dataset was carefully curated to ensure a balanced representation of each ocular condition. The fundus images are then preprocessed and input into deep neural networks, such as VGG19, which can analyze such visual data. The VGG19 model was trained through experiments to optimize the classification performance.

The image dimensions were adjusted to improve accuracy, with convolutional layers extracting relevant features and reducing dimensionality. The final sigmoid layer performs classification. The methodology is summarized in a flowchart (see figure 1) outlining the steps of dataset preparation, image preprocessing, deep network training utilizing techniques and transfer learning from VGG19, and model assessment that enables precise classification of medical images for automated diagnosis of ocular diseases. To create a binary classifier for detecting cataracts versus normal eyes, the examples labeled as diabetic retinopathy, glaucoma, age-related macular degeneration, hypertension, and the two other classes were removed. This simplified the task into distinguishing between just two categories - cataract and regular. As the first step of this project, exploratory data analysis was performed by examining a dataset sample. Specifically, the contents of the first 5 patient records were inspected to gain a preliminary understanding of the data structure and characteristics. Figure 2(b) contains information on patients from the ODIR-5K dataset, including ID, age, sex, left and right fundus image filenames, diagnostic keywords for each eye, labels, and target vectors. For each patient entry listing the patient ID, demographics, left and right fundus image filenames, diagnostic keywords describing pathological findings in each eye, the image file directory path, the overall diagnostic label (N=Normal, D=Diabetic Retinopathy, G=Glaucoma, C=Cataract, A=Age-related Macular Degeneration, H=Hypertension), and a target vector encoding the diagnosis. This structured data provides the input images and diagnostic metadata for training machine learning models to classify ocular diseases based on retinal fundus photographs.

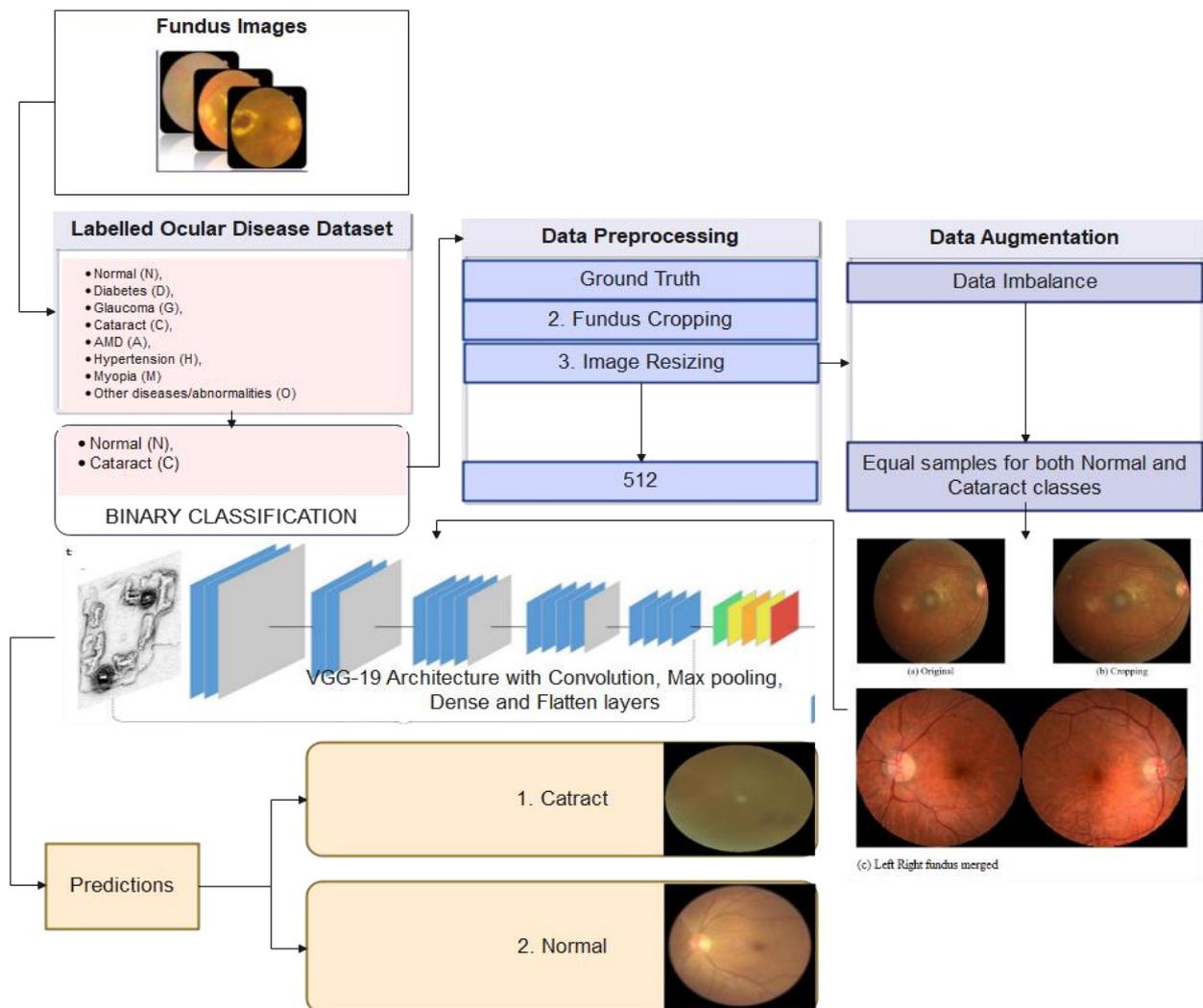


Figure. 1. Proposed System flow

3.1. Dataset

The ODIR-5K dataset contains color fundus photographs of the left and right eyes of 5,000 patients, along with age information and diagnostic keywords provided by doctors [21]. Shang-gong Medical Technology compiled a desensitized dataset of 5000 patients' left and right fundus images from various Chinese hospitals. They added metadata such as age, gender, and diagnostic keywords to determine disease labels for classification. The images were captured using diverse cameras, resulting in varying dimensions. This multiclass, multi-label dataset

encompasses eight ocular disease categories, with individual patients potentially exhibiting multiple conditions. The classes include normal, glaucoma, diabetes, AMD, cataract, hypertension, myopia, and other abnormalities. Figure 2 (A) exhibits fundus images demonstrating various ophthalmological disease pathologies.

Fundus pictures provide a non-invasive way to observe ocular abnormalities indicative of various eye diseases. For example, glaucoma damages the optic nerve, increasing cupping and brightness in the disc (Fig. 3(A)b). Diabetic retinopathy produces microaneurysms, hemorrhages, and exudates visible as red and yellow spots (Fig. 3(A)c). AMD causes neovascularization and geographic atrophy in the retina (Fig. 3(A)d). Cataracts blur anatomical structures like vasculature and fovea (Fig. 3(A)f). Hypertension alters vessel morphology, narrowing arterioles and causing AV (Arteriovenous) nicking (Fig. 3(A)e). Myopia thins the retinal pigment epithelium, creating peri-papillary atrophy (Fig. 3(A) g). Other abnormalities like macular degeneration, pigment proliferation, and epiretinal membranes can also be observed (Fig. 3(A) h). Overall, fundus imaging provides a

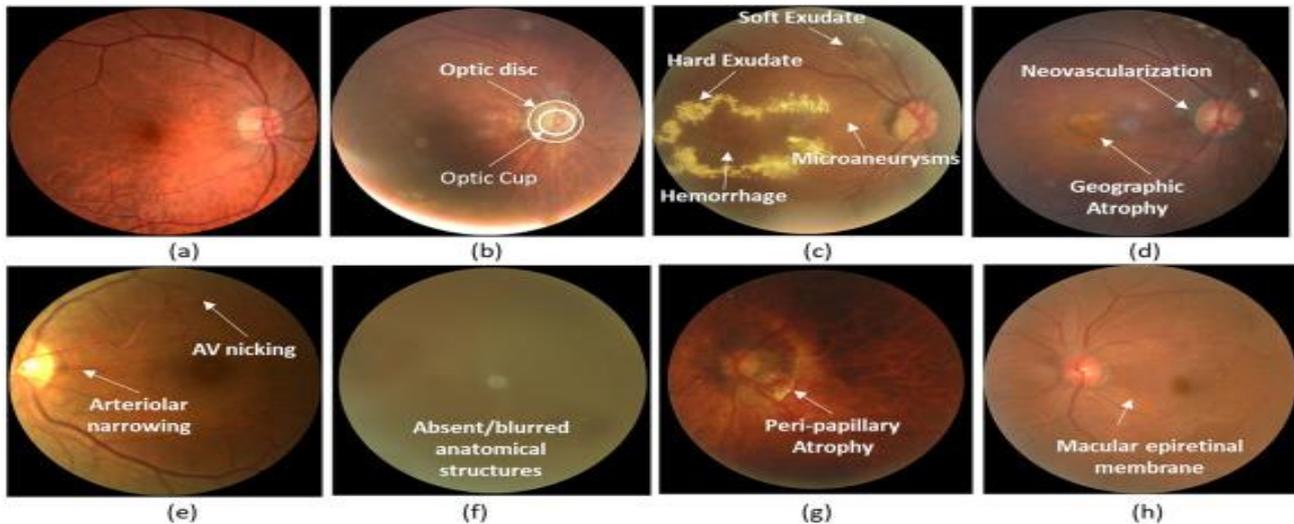


Figure. 2 (A) Fundus images demonstrating various ophthalmological disease pathologies: (a) a normal fundus devoid of abnormalities, (b) an image of glaucoma, (c) diabetic retinopathy lesions, (d) signs of age-related macular degeneration, (e) manifestations of hypertension, (f) cataract opacity, (g) peripapillary changes from myopia, and (h) additional unspecified anomalies. Fundus images demonstrating various ophthalmological disease pathologies. [22]

```
C:\Users\user\AppData\Local\Programs\Python\Python37\python.exe H:/catract/Preprocessing.py
(6392, 19)
```

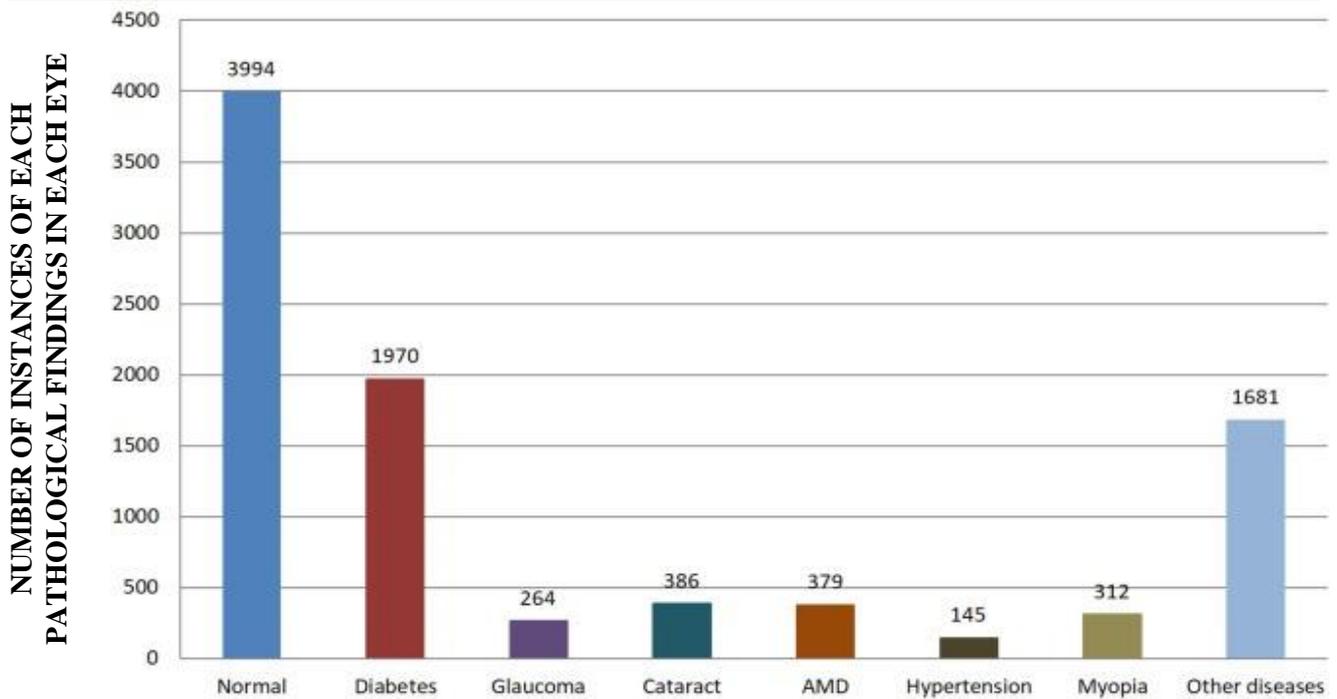
ID	Patient	Age	Patient	Sex	...	Labels	target	filename
0	0	69	Female	...	['N']	[1, 0, 0, 0, 0, 0, 0, 0]	0_right.jpg	
1	1	57	Male	...	['N']	[1, 0, 0, 0, 0, 0, 0, 0]	1_right.jpg	
2	2	42	Male	...	['D']	[0, 1, 0, 0, 0, 0, 0, 0]	2_right.jpg	
3	4	53	Male	...	['D']	[0, 1, 0, 0, 0, 0, 0, 0]	4_right.jpg	
4	5	50	Female	...	['D']	[0, 1, 0, 0, 0, 0, 0, 0]	5_right.jpg	

```
[5 rows x 19 columns]
```

Figure. 2 (B). first five records from the dataset

powerful tool to screen for multiple ocular diseases in a non-invasive manner through characteristic morphological changes in the retina and optic disc.

The image dataset exhibited variability in terms of image dimensions and shapes. Specifically, 114 images were non-rectangular triangles with differing sizes, such as 3456x188 and 5188x250 pixels. This variability is attributed to the images captured by different cameras and settings. Additionally, while 150 images were 1000 pixels, 10 images were 2000 pixels. To normalize this data, black borders were removed, cropping and resizing were applied to make images rectangular, and left and right images merged into a single image. Figure 3 represents a bar graph that illustrates the distribution of pathological observations in individual eyes within the ODIR-5K dataset, which comprises 5000 retinal fundus photographs.



PATHOLOGICAL FINDINGS FOR VARIOUS DISEASES IN EACH EYE

Figure. 3. Overall distribution of multiclass diseases dataset (before Preprocessing)

The graph presents a comprehensive representation of problematic labels and their corresponding frequencies inside the ODIR-5K dataset, which was utilized to develop automated eye disease diagnosis systems using retinal pictures. The dataset exhibits a class imbalance as the number of normal fundus photos surpasses the number of instances of most disorders, thus emphasizing the disparity in distribution.

3.2. Data preprocessing

The ODIR-5K dataset contains variable-sized fundus images captured by diverse cameras. Preprocessing ensured high quality and involved manual screening to exclude invalid images, cropping black borders, augmenting data, and normalizing. After removing poor-quality images per expert annotation, 1104 images across two classes (Normal and Cataract) were selected—594 for the cataract and 510 for the normal.

Moreover, most fundus images contain non-informative black borders; retaining such negligible backgrounds equivalently increases the negative sample proportions, adversely impacting diagnostic lesion detection. Consequently, an automated cropping process is implemented: images undergo segmentation into background and foreground components, with the foreground pathology-containing region localized, and image dimensions are subsequently resized based on the foreground position and dimensions. This cropping normalizes images to 512*512 pixels, removing extraneous information and improving the feature representation.

Moreover, the quantity of data points within each category is disproportionate, resulting in discriminatory categorization outcomes that favor types boasting copious images via increased training weights over categories with limited data points. Thus, attenuating the ramifications of the data imbalance while enhancing categorization efficacy mandated down sampling of the dominant class to mitigate the data asymmetry. Furthermore, implementing normalization to rescale all imagery within intervals of (0, 1) streamlined network optimization by standardizing the input data range.

3.3. Transfer learning (TL) and convolutional neural network (CNN)

LeCun et al. [23] initially used a Convolutional neural network (CNN) for recognizing handwritten digits. CNN uses convolution instead of matrix multiplication. It performs best for medical image segmentation, enhancement, and classification [21]. Different layers comprise CNN architecture comprising convolutional, batch normalization, ReLU (rectified linear unit), pooling, and fully connected layers.

Deep learning models, like various CNN architectures, including GoogleNet [24], AlexNet [25], VGGNet [7], MobileNet [26], and ResNet [27] suggested for ImageNet classification, require immense datasets to tackle multifaceted problems and achieve best performance. Transfer learning circumvents the arduous, time-consuming

task of amassing ample labeled data by exploiting information from pre-trained models trained on massive databases, like ImageNet, irrespective of their original purpose. Therefore, transfer learning enables leveraging existing CNN architectures without collecting extensive medical imaging data [28].

This study implements TL (Transfer Learning) on CNN architectures, specifically VGG19, for fundus image classification. This latter is chosen to analyze the fundus image domain adaptation from the ImageNet database, considering its numerous layers and parameters. The architecture was applied using Keras over TensorFlow and optimized using the Adam optimizer to enable the transfer of learning from natural to medical images.

3.4. Proposed CNN'S VGG19 model via transfer learning (TL)

The ODIR-5K dataset, discussed in section III part A, contains fundus images for every patient's eyes. The final label is either a normal eye or a cataract disease eye. Images were resized to $512 \times 512 \times 3$ to standardize data via bilinear interpolation. For the model, concatenated left and right eyes were input to CNN. CNN feature maps were pooled globally. The resulting feature vector was optimized using sigmoid activation function $f(s)$ -(Eq.(1)) which limits the feature predictions to between (0,1).

$$f(s) = \frac{1}{1+e^{-s}} \quad (1)$$

Sigmoid activation predicts the probability for each label separately. Labels with a chance exceeding 0.5 are classified as cataract disease eyes, while those below 0.5 are normal eyes. The MSLE (Mean Squared Logarithmic Error) loss function calculates the divergence between the predicted and actual values for normal eyes and eyes with cataract disease. The VGG-19 model utilizes a consistent CNN architecture with 3x3 filters, single stride convolutions, identical padding, and 2x2 max-pooling layers with a stride of 2 rather than relying on numerous hyperparameters. The convolution and max-pooling layers follow a uniform pattern. The model contains two fully connected layers. VGG-19 is an extensive network with over 138 million trainable parameters due to its structured design focused on stacked 3x3 filter convolutions. Following the classification layer, comprised of a densely connected classifier and dropout, a sequence of convolutional layers was implemented. In a dense layer, each neuron connects to all neurons in the prior layer, contrasting convolutional layers. Thus, dense layers learn from the preceding layer's features, unlike convolutional layers. The activation function for the densely connected layers must be specified, determining how they process the inputs from the previous layer.

Leveraging transfer learning by utilizing weights that were pre-trained on the ImageNet dataset, the VGG19 model, which is a deep convolutional neural network architecture initially designed by the Visual Geometry Group at Oxford, has over 20 million parameters yet only 238,081 were trainable as a result of freezing the convolutional base layers, with the model being trained to perform binary classification likely on a custom dataset through the use of a sigmoid activation on the dense output layer along with the selection of Adam as the optimizer to update the trainable weights during training as well as the application of binary cross entropy as the loss function, which is commonly used for binary classification tasks, and no early stopping was employed so the model was trained for a fixed number of epochs; overall, the VGG19 model made use of transfer learning on pretrained ImageNet weights, had minimal trainable parameters, and was tuned as a binary classifier for a likely custom dataset utilizing standard practices like the Adam optimizer and binary cross entropy loss. Table 1 provides the model specification for the undertaken research work.

Table 1. Employed model.

NO	Layer name	Layer type
1	vgg19	Functional
2	flatten	Flatten
3	dense	Dense

4. Implementation

By employing a subset of the ODIR-5K dataset containing only normal and cataract images, preliminary discourse elucidates the implementation specifics and assessment metrics, followed by presenting the empirical outcomes to validate the proficiency of the proposed VGG19-CNN architecture. Specifically, 1104 images across two classes (Normal and Cataract) were selected from the full ODIR-5K dataset. This refined dataset contained 594 images labeled as cataracts and 510 images labeled as normal. We propose modifications to the VGG19 architecture by adding custom-flattened and dense layers, creating an enhanced model tailored for our binary classification application using only normal and cataract eye images. The experimental implementation leveraged the Google Colab platform with a Tesla T4 GPU and employed the Jupiter Notebook within the

TensorFlow framework. Utilizing the ODIR-5K dataset, the data was partitioned with a 70/30 split for training and testing sets. Parameterization instantiated input dimensions of 512×512 pixels, Adam optimizer configured at a 0.001 learning rate, MSLE for the loss function calculation, batch size 32, and 15 epochs of iterative training.

Table 2 presents the summary of the employed model. The classification model utilized the VGG19 architecture pre-trained on the ImageNet dataset. Transfer learning was employed in this study by immobilizing the weights of the initial layers and retraining the subsequent fully connected network layers. The final layer consists of a densely connected sigmoid output, which is utilized for binary classification. No data augmentation was performed. The optimization algorithm used in this study was Adam, whereas the loss function employed was the mean-squared log error. The use of early discontinuation still needs to be implemented. The model consisted of 20 million parameters, with 25,089 being trainable in the retrained, fully connected layers. The remaining 20 million parameters in the convolutional layers were maintained constant. Transfer learning in this approach facilitated efficient training that emphasized the adjustment of higher-level features while capitalizing on the broad range of features acquired by VGG19 from its exposure to various image datasets. Implementing a selective retraining technique facilitated the effective acquisition of knowledge from a restricted set of medical images, allowing for the differentiation between normal eyes and cataracts diseases eyes by utilizing diagnostic indicators included in retinal fundus photographs.

Table 2. A summary of the employed model

Training Details	VGG19
Data Augmentation	NO
Transfer Learning	Yes
Weights	Pre-trained on ImageNet
Last Layer	Dense (1, activation = 'sigmoid')
Feature Extraction Enabled	Yes
Classification Enabled	Yes
Optimizer	Adam
Loss Function	MSLE
Early Stopping patience	NO
Number of total Parameters	20,049,473
Number of trainable Parameters	25,089
Number of Non-trainable params	20,024,384

Performances are assessed based on two key metrics: Accuracy (equation (3)) and F1-score (equation (4)).

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (3)$$

While TP represents True Positives, emblemizing exemplars wherein both predicted and factual declensions are affirmative, contrariwise, TN, that is, True Negatives, impersonates contingencies where both indices are dissenting; FP along with FN signify False Positives in addition to False Negatives, correspondingly, mirroring the fallacious qualifications: FP where the prototype speciously prescribes positive mottoes to negative paradigms, plus FN wherein sanguine prototypes are incorrectly emblazoned as pessimistic; the numerator of the proportion, (TP + TN), merges the precise arrangements, while the denominator finds the sum of information dots arranged; ergo, the ratio explicates the section of accurate characterizations out of the gross prognostications, typifying the prototype's integrity.

$$F1\text{-score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}) \quad (4)$$

With:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (5)$$

The precision constitutes a ratio delineating the model's ability to elucidate positive exemplars from the aggregate of pattern accurately points it defines as sanguine; superior precision indicates the prototype's mastery in diminishing fallacious positive arrangements.

$$\text{recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (6)$$

Recall measures the effectiveness of a model in correctly identifying positive samples in a dataset. It is calculated as the ratio of true positives to the sum of true positives and false negatives. True positives are cases correctly identified as positive, while false negatives are instances where the model incorrectly predicts the negative class despite a positive ground truth. A higher recall indicates that the model is better at capturing

actual positives and fewer false negatives. For example, a recall of 0.9 means that 90% of malignant cases were correctly identified, with only 10% mislabeled as benign, which exemplifies a proficient diagnosis. Using recall as a single metric makes model comparisons easier, especially in domains like cancer prediction, where overlooking true positives can have severe consequences. Therefore, models that maximize recall are essential. While accuracy, a ubiquitously utilized metric furnishes an overarching appraisal of preciseness in prognostications and ought to approximate unity as nearly as feasible, in circumstances where unequal class dissemination prevails, accuracy exclusively may not render a consummate depiction of model achievement, in such situations, the F1-score arises as salient, for it meditates on both precision and recall, tendering an equilibrated scale that chronicles false positives and false negatives alike; therefore, by assimilating the F1-score in conjunction with accuracy, we procure a more composite gauging of the model's feat, markedly in predicaments where the class arrangement is asymmetric or when false positives and false negatives necessitate parity in weighting. A confusion matrix (figure 4) evaluated a classification model's performance in distinguishing normal versus cataract fundus images. The valid class represented the rows, while the predicted classes were the columns. Out of 80 typical cases, the model correctly classified 80 as usual but misclassified 3 as cataracts. Of 90 cataract cases, it accurately predicted 85 as cataracts, wrongly predicting 5 as normal. The high diagonal values showed the model correctly classified most cases. Overall accuracy was $172/180 = 95.6\%$, indicating robust discrimination of normal and cataract images. Though a few misclassifications existed, the high accuracy demonstrates effective learning of features to categorize fundus images.

The confusion matrix assessed the model's test performance, revealing strong capabilities for detecting cataracts from fundus photographs.

5. Results and discussion

The proposed VGG19 model demonstrates strong performance for this classification task, achieving an overall accuracy of 94% on the test set. The image data comprised 594 cataracts and 500 normal eye images, split 70/30 into training and test sets. Precision and recall were high for both classes, with F1 scores of 0.95-0.96, indicating effective learning of discriminative features. The confusion matrix revealed slightly better recognition of normal (96% recall) versus cataract eyes (94% recall), though precision was higher for detecting cataracts (0.97). Overall, the vital test accuracy demonstrates this approach can accurately distinguish cataracts from normal eyes, supported by robust precision and recall metrics. The model's effective generalization to new data suggests applicability to real-world screening and diagnosis after further validation across diverse datasets. The combination of high precision and recall underscores the model's reliable identification of positive examples of both classes. Based on these solid classification results on key metrics, the proposed VGG19 model demonstrates suitability and effectiveness for this problem. Table 3 presents a comparative analysis of the relevant studies.

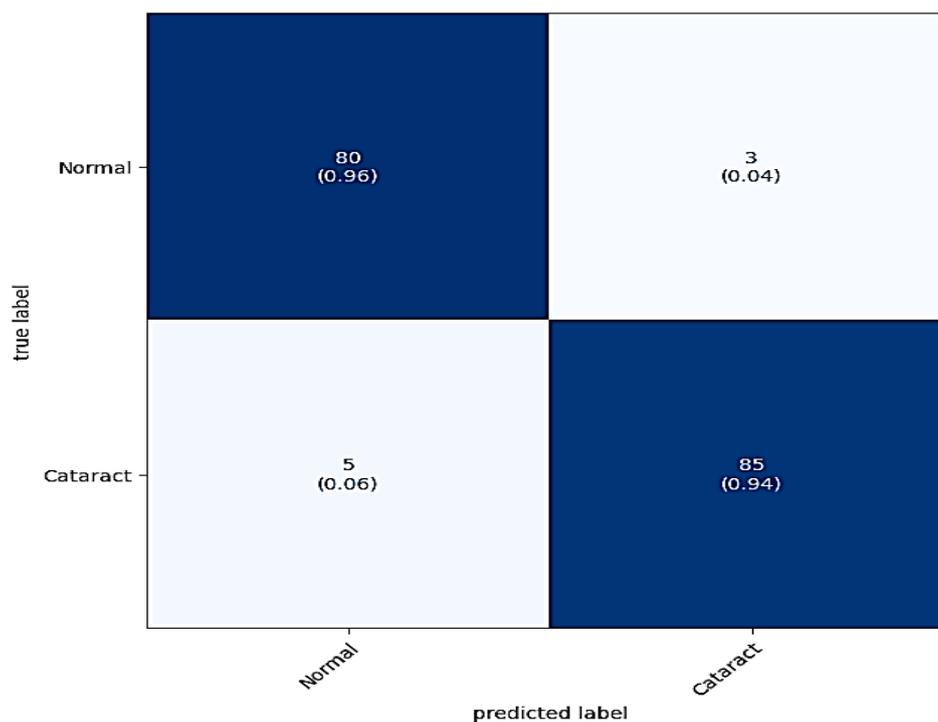


Figure 4. Confusion Matrix

Table 3. A summary of Previous studies

Reference	Method Employed	Result Achieved
CataractNet [29]	Custom CNN architecture with tuned loss and activation functions	99.13% accuracy
Automated Detection of Cataracts Using a Deep Learning Technique [30]	VGG16 pre-trained CNN	92.1% accuracy
Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network [31]	Hybrid CRNN fusing CNN and RNN	97.39% accuracy
Artificial Intelligence for Cataract Detection and Management [32]	Review of AI approaches - slit lamp and fundus imaging	Early promising results, limitations exist
Computer-aided diagnosis of cataract severity using retinal fundus images and deep learning [33]	Ensemble of AlexNet, VGGNet, ResNet CNNs with SVM	96.25% accuracy for 4-class classification
Deep Learning Approach for Automated Detection of Myopic Maculopathy and Pathologic Myopia in Fundus Images [34]	Custom DL models and system	84-87% sensitivity, 87-98% AUC for lesions, 92% accuracy for pathologic myopia detection
Automatic cataract grading methods based on deep learning [35]	ResNet18 and GLCM features with SVM and FCNN	92.66% accuracy for 6-class grading, 94.75% for 4-class
Retinal image blood vessel classification using hybrid deep learning in cataract diseased fundus images [36]	Hybrid DenseNet and ShuffleNet	98-99% accuracy
Automated identification of cataract severity using retinal fundus images [37]	Transfer learning CNNs with SVM	95.65% accuracy for 4-class classification

Despite the limited number of only 594 cases of cataracts in the dataset, sampling and passing the data into the pre-trained VGG-19 model achieved promising classification performance between cataracts and normal subjects. The proposed VGG19 model for cataract classification was trained using the Adam optimizer for 15 epochs. As shown in Table 1, the model achieved strong performance, attaining a training accuracy of 0.948 and a loss of 0.025 after 15 training epochs. The equivalence between the training and validation accuracies indicates minimal overfitting, with the model generalizing well to new data. The high accuracy and low loss after only 15 epochs of training highlight the model's efficiency in learning discriminative features for cataract classification. Given the limited size of the cataract dataset, the model's ability to rapidly learn and generalize demonstrates the strength of using transfer learning from the pre-trained VGG19 weights. Figures 5 (a)(b) show accuracy and loss plots for training and testing. The accuracy plot Figure 5 (a) offers valuable information regarding the classification model's optimization trajectory and generalization capabilities. The training accuracy exhibited a consistent and steady increase, starting from an initial level of approximately 50% and reaching ideal values close to 94% by the 15th session. This pattern indicates the gradual and reliable acquisition of the discriminative characteristics that underlie the dataset. The observed practice of the testing curve closely aligned with the training accuracy, showing a high generalization level with a limited overfitting occurrence. Significantly, the testing accuracy reached 94% at epoch 15, aligned with the training performance.

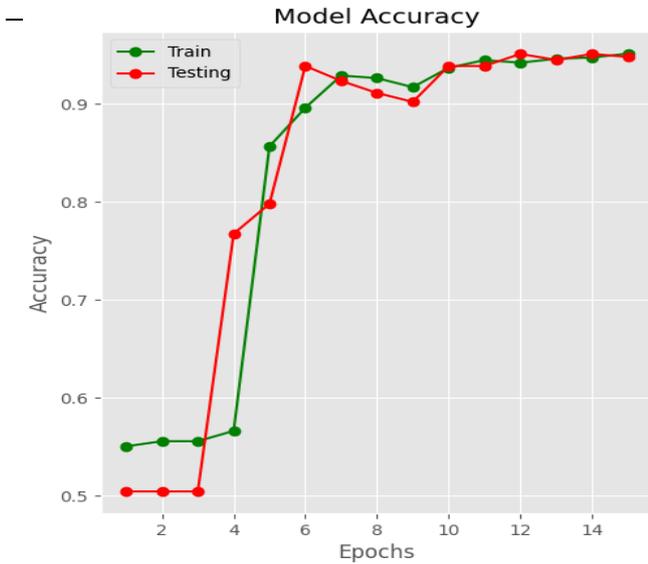


Figure. 5 (a). Accuracy Plot (Training and Testing)

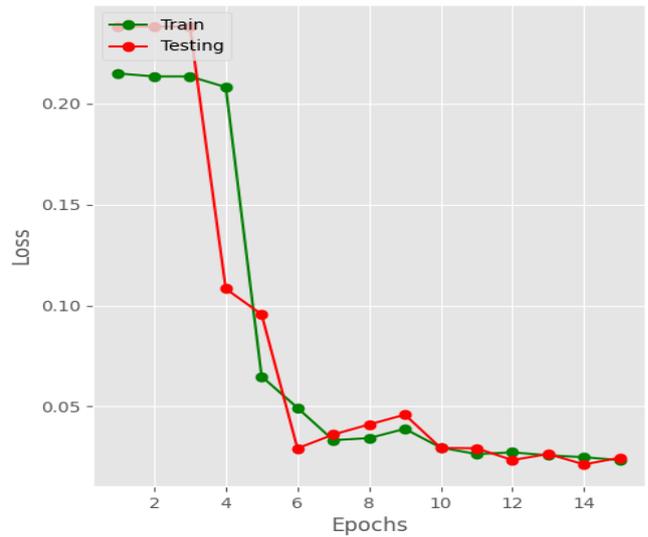


Figure. 5 (b). Loss Plot (Training and Testing)

The prompt suggests that the prompt's early convergence to high accuracy on the holdout data indicates the prompt's efficient extraction of representative characteristics and patterns crucial for differential diagnosis from the fundus images. Several crossings between the training and testing curves suggest that utilizing dropout and data augmentation techniques throughout the training process is a form of regularization, thereby mitigating the risk of overfitting.

The loss plot (figure 5 (b)) depicted training and validation loss over 15 epochs. The training loss curve started around 0.22, indicating a high initial error, but steadily decreased to 0.02 by epoch 15 as the model learned. Similarly, validation loss began around 0.23, declining to converge at 0.02 by epoch 15, mirroring the training curve. The decreasing loss and convergence of the training and validation curves demonstrated the successful minimization of errors and good generalization by the model. Approaching near zero loss by the final epoch signaled accurate cataract and standard fundus image classification.

6. Conclusion and future work

This study, the VGG-19 model is used to classify different eye diseases, predicting whether an eye was normal or had cataracts. The model performed very well, exceeding expectations. An accuracy of 94% was achieved for the normal versus cataract classification task. This proposed strategy surpasses existing CNN models for ocular disease classification in terms of accuracy while requiring less latency. Moreover, it can be easily adapted for other medical image classification tasks. The VGG-19 model shows promise for real-world ocular disease diagnosis systems. A key advantage is the adaptability of this technique to different medical image classifications.

Furthermore, segmentation could be incorporated to improve performance. Generative adversarial networks may help address class imbalance issues by synthesizing realistic pathological images. Overall, this model shows strong potential to aid medical experts and transform ocular disease screening, though additional research is needed to optimize accuracy, particularly for glaucoma cases. However, the results obtained thus far are encouraging, and with more data and experimentation, this has the potential to become a very useful disease classification tool.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

- [1] A. Adio, A. Alikor, and E. Awoyesuku, "Survey of pediatric ophthalmic diagnoses in a teaching hospital in Nigeria," *Nigerian Journal of Medicine: journal of the National Association of Resident Doctors of Nigeria*, vol. 20, no. 1, pp. 105-108, 2011.
- [2] N. Li, T. Li, C. Hu, K. Wang, and H. Kang, "A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection," in *Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Virtual Event, November 15–16, 2020, Revised Selected Papers 3*, 2021: Springer, pp. 177-193.
- [3] Y. Elloumi, M. Akil, and H. Boudegga, "Ocular diseases diagnosis in fundus images using a deep learning: approaches, tools and performance evaluation," in *Real-Time Image Processing and Deep Learning 2019*, 2019, vol. 10996: SPIE, pp. 221-228.
- [4] P. Paudel *et al.*, "Prevalence of vision impairment and refractive error in school children in B a R ia–V ung T au province, V Vietnam," *Clinical & experimental ophthalmology*, vol. 42, no. 3, pp. 217-226, 2014.
- [5] "Blindness and vision impairment." <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment> (accessed 8/12/2023, 2023).
- [6] J. D. Steinmetz *et al.*, "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study," *The Lancet Global Health*, vol. 9, no. 2, pp. e144-e160, 2021.
- [7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [8] X. Meng, X. Xi, L. Yang, G. Zhang, Y. Yin, and X. Chen, "Fast and effective optic disk localization based on convolutional neural network," *Neurocomputing*, vol. 312, pp. 285-295, 2018.
- [9] C. S. Lee, A. J. Tying, N. P. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography," *Biomedical optics express*, vol. 8, no. 7, pp. 3440-3448, 2017.
- [10] C. Payout, R. Duval, and F. Cheriet, "A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2434-2444, 2019.
- [11] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285-1298, 2016.
- [12] K. Hu *et al.*, "Retinal vessel segmentation of color fundus images using a multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179-191, 2018.
- [13] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel Segmentation in color images of the Retina," *IEEE Transactions on medical imaging*, vol. 23, no. 4, pp. 501-509, 2004.
- [14] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402-2410, 2016.
- [15] Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, and M. He, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199-1206, 2018.

-
- [16] M. S. M. Khan, M. Ahmed, R. Z. Rasel, and M. M. Khan, "Cataract detection using convolutional neural network with VGG-19 model," in *2021 IEEE World AI IoT Congress (AIIoT)*, 2021: IEEE, pp. 0209-0212.
- [17] T. Guergueb and M. A. Akhloufi, "Ocular diseases detection using recent deep learning techniques," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021: IEEE, pp. 3336-3339.
- [18] H. S. Gill and B. S. Khehra, "Fruit image classification using deep learning," 2022.
- [19] H. S. Gill, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and F. Alassery, "Multi-Model CNN-RNN-LSTM Based Fruit Recognition and Classification," *Intelligent Automation & Soft Computing*, vol. 33, no. 1, 2022.
- [20] M. Rajalakshmi, V. Saravanan, V. Arunprasad, C. T. Romero, O. I. Khalaf, and C. Karthik, "Machine Learning for Modeling and Control of Industrial Clarifier Process," *Intelligent Automation & Soft Computing*, vol. 32, no. 1, 2022.
- [21] N. Gour and P. Khanna, "Multiclass multi-label ophthalmological disease detection using transfer learning based convolutional neural network," *Biomedical Signal Processing and Control*, vol. 66, p. 102329, 2021.
- [22] N. Gour, M. Tanveer, and P. Khanna, "Challenges for ocular disease identification in the era of artificial intelligence," *Neural Computing and Applications*, pp. 1-23, 2022.
- [23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [24] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [26] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
- [28] A. Ng, "Nuts and bolts of building AI applications using Deep Learning," *NIPS Keynote Talk*, 2016.
- [29] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "CataractNet: An Automated Cataract Detection System Using Deep Learning for Fundus Images," *IEEE Access*, vol. 9, pp. 128799-128808, 2021, doi: 10.1109/ACCESS.2021.3112938.
- [30] R. Angeline, R. Vani, A. Jeshron Sonali, and D. A. Rao, "Automated Detection of Cataract Using a Deep Learning Technique," in *Computational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021*: Springer, 2022, pp. 399-408.
- [31] A. Imran, J. Li, Y. Pei, F. Akhtar, T. Mahmood, and L. Zhang, "Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network," *The visual computer*, vol. 37, pp. 2407-2417, 2021.
- [32] J. H. L. Goh *et al.*, "Artificial intelligence for cataract detection and management," *The Asia-Pacific Journal of Ophthalmology*, vol. 9, no. 2, pp. 88-95, 2020.
- [33] J. K. P. S. Yadav and S. Yadav, "Computer-aided diagnosis of cataract severity using retinal fundus images and deep learning," *Computational Intelligence*, vol. 38, no. 4, pp. 1450-1473, 2022.
- [34] R. Du *et al.*, "Deep learning approach for automated detection of myopic maculopathy and pathologic myopia in fundus images," *Ophthalmology Retina*, vol. 5, no. 12, pp. 1235-1244, 2021.
-

- [35] H. Zhang, K. Niu, Y. Xiong, W. Yang, Z. He, and H. Song, "Automatic cataract grading methods based on deep learning," *Computer methods and programs in biomedicine*, vol. 182, p. 104978, 2019.
- [36] Y. Kumar and B. Gupta, "Retinal image blood vessel classification using hybrid deep learning in cataract diseased fundus images," *Biomedical Signal Processing and Control*, vol. 84, p. 104776, 2023.
- [37] A. Imran, J. Li, Y. Pei, F. Akhtar, J.-J. Yang, and Y. Dang, "Automated identification of cataract severity using retinal fundus images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 8, no. 6, pp. 691-698, 2020.