

Research Article

Automatic Cataract Severity Detection and Grading Using Deep Learning

Sunita Yadav  and Jay Kant Pratap Singh Yadav 

Computer Science and Engineering Department, Ajay Kumar Garg Engineering College, Affiliated to Dr. A. P. J. Abdul Kalam Technical University, Ghaziabad, Uttar Pradesh, India

Correspondence should be addressed to Jay Kant Pratap Singh Yadav; er.jaykant@gmail.com

Received 24 December 2022; Revised 11 June 2023; Accepted 12 June 2023; Published 27 June 2023

Academic Editor: Everardo Vargas-Rodriguez

Copyright © 2023 Sunita Yadav and Jay Kant Pratap Singh Yadav. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cataracts are an eye condition that causes the eye's lens to become cloudy and is a significant cause of vision loss worldwide. Accurate and timely detection and diagnosis of cataracts can prevent vision loss. However, poor medical care and expensive treatments prevented cataract patients from receiving appropriate treatment on time. Therefore, an inexpensive system that diagnoses cataracts at an early stage needs to be developed. This study proposes an automatic method for detecting and classifying cataracts in their earliest stages by combining a deep learning (DL) model with the 2D-discrete Fourier transform (DFT) spectrum of fundus images. The proposed method calculates the spectrogram of fundus images using a 2D-DFT and uses this calculated spectrogram as an input to the DL model for feature extraction. After feature extraction, the classification task is performed by a softmax classifier. This study collected fundus images from various open-source databases that are freely available on the Internet and classified them into four classes based on an ophthalmologist's assessment. All the collected fundus images from various datasets with open access are unsuitable for cataract diagnosis. Consequently, a module for identifying the fundus images of good and poor quality is also incorporated into this method. The experimental results show that the proposed system can outperform previous state-of-the-art works by a significant margin compared to a benchmark of four-class accuracy and achieves the four-class accuracy of 93.10%.

1. Introduction

Cataract formation is defined by the clouding of the eye's lens, which occurs when an excessive amount of protein accumulates in the lens and causes opacity that alters the lens's natural shape. This opacity issue may result in the patient's eyesight being blurred. According to the Vision 2020 report, cataracts account for 62.5 million occurrences of visual impairment and blindness worldwide [1]. Due to an aging population, these numbers are anticipated to reach 71.5 million by 2022. Due to the inadequate health infrastructure and a lack of ophthalmologists [2], a considerable proportion of these cataract patients remain undetected. Consequently, cataract diagnosis remains a significant public health concern [3].

Cataracts can be caused by a variety of reasons, including aging, smoking, radiation, diabetes, and so on. Early detection

and diagnosis of cataracts are critical for achieving significant advantages in a short amount of time; otherwise, it may result in permanent vision loss [4]. Therefore, this presents a challenge to detect and grade the cataract at an early stage.

To date, cataracts are clinically detected and graded by ophthalmologists using slit-lamp microscopy and pre-established clinical standards such as the lens opacities classification system (LOCS) III [5]. This manual process necessitates clinical skills and thus poses a significant challenge, particularly in developing countries and rural areas where qualified ophthalmologists are scarce [6]. Furthermore, cataract classification results are subjective and affected by inter-examiner variability [7]. Overall, the manual approach to cataract detection and grading, which requires the expertise of an ophthalmologist, has significantly limited screening access. As a result of the growing health burden associated with cataracts,

new methods to overcome existing limitations and revolutionize the approach to cataract screening are required.

In recent years, the rise of artificial intelligence (AI) and its application in medical science has yielded promising results in specific tasks such as the early detection of breast cancer, lung cancer, and fatal blood diseases [8, 9]. AI-based systems have also shown promising results in the automated identification of age-related disorders such as glaucoma, diabetic retinopathy, age-related macular degeneration, and others in ophthalmology, where a large volume of images and patient data are available [10]. With the rapid expansion of computational infrastructure and power, deep learning (DL)-based AI models are becoming increasingly popular due to their remarkable ability to extract high-level features and previously unknown patterns from massive amounts of data. This piqued the interest of the research community in developing an AI-based automatic cataract detection and classification system.

In recent years, various researchers proposed various methods for automatic cataract detection based on different imaging modalities, such as slit lamp, ultrasound, retro-illumination, or fundus images [11]. Among these imaging modalities, fundus imaging has brought considerable attention in this field, as technologists or even patients themselves can efficiently operate the fundus camera. In contrast, other imaging modalities require experienced ophthalmologists to capture the images [12]. Thus, in order to simplify the process of early cataract screening, an automated fundus image-based cataract detection system is essential. However, fundus image-based cataract detection and grading systems utilize retinal features, such as highly complex tinny blood vessels, which can interfere with other blood vessels and are challenging to extract. Therefore, there is a need for a technique that can easily extract information regarding these minute blood vessels, which are necessary for diagnosing cataracts.

To alleviate the aforementioned problems, a method is used to calculate two-dimensional discrete Fourier transform (2D-DFT) spectrograms of fundus images, and then subsequently, these 2D-DFT spectrograms are used for feature extraction. The advantage of using a retinal 2D-DFT spectrogram is that it preserves all original information used for further diagnosis. 2D DFT has the ability to fully transform the cataract fundus image into the frequency domain, where the low-frequency spectrum contains most of the image information, while the high-frequency spectrum provides more details such as edges, curves, and contours present in the images [13]. Hence, if 2D DFT is applied to the fundus images with various degrees of cataracts, then it produces different types of frequency spectrograms. For instance, a severe cataract has more opacification in the lens, which causes more light to be scattered or absorbed by the lens, resulting in less light reaching the retina. It indicates that the fundus image will be more distorted, that is, a greater proportion of low-frequency components will be visible. In contrast, if the eye is normal, then there is no opacification in the lens, and more light will reach the retina without scattering and absorption, resulting in a clear fundus image, that is, a significant increase in the proportion of high-frequency components. As cataract degrees increase, the shape of the

spectrogram becomes more regular, and the high-frequency components occupy a smaller proportion of the spectrogram, while the low-frequency components occupy a larger proportion [14]. Therefore, high-frequency components present in the 2D-DFT spectrogram of fundus images work as discriminating features to detect and grade cataracts.

DL is an emerging field of AI that uses artificial neural networks made up of numerous layers of artificial neurons to replicate the physiological behavior of the human brain [15]. DL systems are capable of extracting and processing information from images, text, and signals. Convolutional neural networks (CNNs) are popularly used DL models for automatic feature extraction from images [16]. The major advantage of CNN over conventional feature extraction methods is that it does not require human intervention to extract features from the image. The advantages of using CNNs are less prone to error and speed up the feature extraction. However, the major problem associated with CNN networks is the overfitting problem which can be solved by data augmentation.

As a result of the preceding considerations, this work focuses primarily on an automatic approach for cataract detection and grading based on frequency spectrograms of fundus images, which provides a low-cost and timely diagnosis. To validate our method, fundus retinal images with varying degrees of cataracts are acquired from open-source databases and categorized by a professional ophthalmologist as normal, mild, moderate, and severe. Figure 1 shows the sample fundus image of each cataract class. Figure 1(a) shows the normal fundus image without having a cataract, in which a large number of tinny blood vessels can be seen very clearly. Figures 1(b) and 1(c) depict mild and moderate cataract fundus images in which details of tinny blood vessels and the optic disc may be seen to some extent, whereas mild fundus images have more details of tiny blood vessels as compared to moderate images. Figure 1(d) displays severe cataract images in which no details of tiny blood vessels are visible. Hence, the details of tinny blood vessels play a dominant role in cataract classification from fundus images. But, the extraction of these blood vessels as features is a complex and tedious process. To overcome the above limitation, the proposed method uses 2D-DFT spectrograms of fundus images for feature extraction purposes. The advantage of 2D-DFT spectrogram images is that they contain greater details regarding the tinny blood vessels in terms of high-frequency components. Figure 2 shows the 2D-DFT spectrogram images for different severity levels of cataracts that are used for feature extraction and classification.

The proposed method used the CNN model for extracting the features from the 2D-DFT spectrogram of fundus images. These features are then classified using the softmax function into four stages according to the severity of the cataract. The performance of the suggested method is evaluated by using a number of different evaluation matrices, including accuracy, precision, specificity, sensitivity, and F1 score.

The novelty of this study is in its application of 2D-DFT spectrograms of fundus images for automatic cataract detection and grading, which streamlines the process of finer

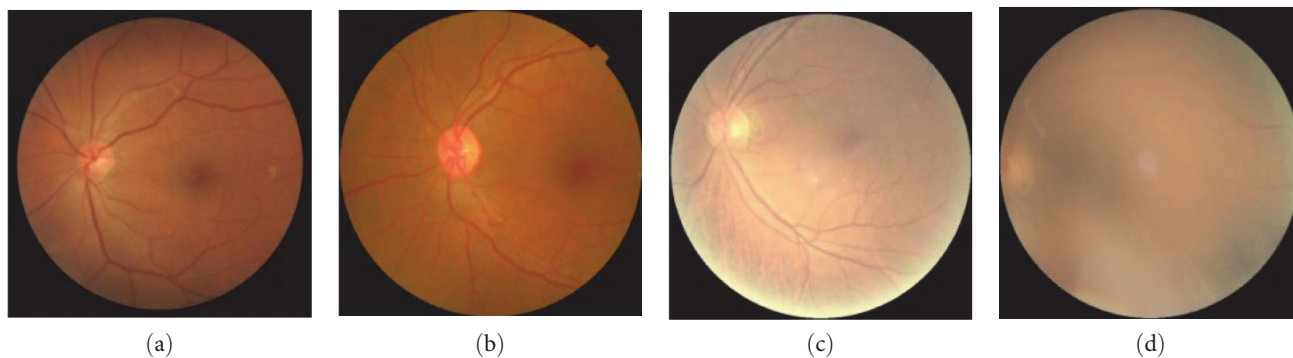


FIGURE 1: Cataract fundus images showing different levels of severity: (a) no; (b) mild; (c) moderate; (d) severe.

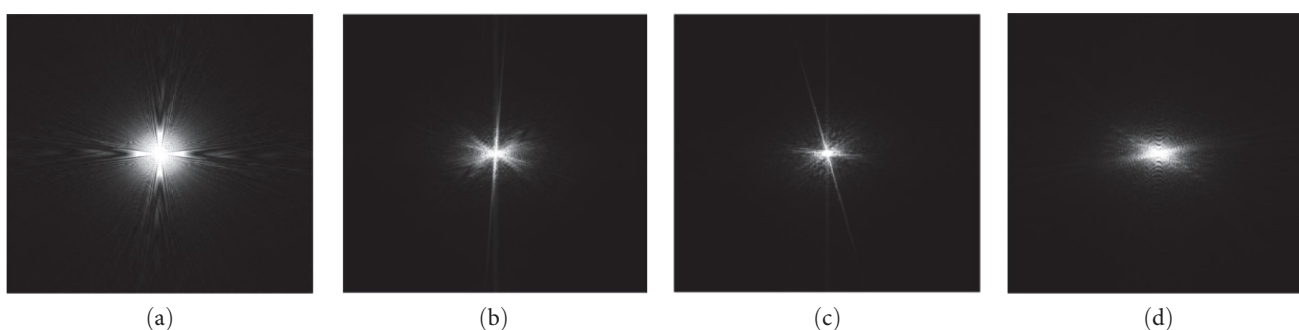


FIGURE 2: Cataract images after applying Fourier transform: (a) no; (b) mild; (c) moderate; (d) severe.

details extraction using CNN to achieve high accuracy. This study makes the following significant contributions:

- (1) To develop a novel 2D-DFT spectrograms-based CNN model for the classification of fundus images into four classes: normal, mild, moderate, and severe.
- (2) Use the quality estimate module to identify fundus images of good quality among the collected images.
- (3) Use data augmentation to expand the dataset by artificially generating the samples, which helps in improving the performance of our proposed method.
- (4) Use the softmax function for the classification of images that maps every output in the range between $[0,1]$ and gives the sum always 1. It is used to classify multiclass images based on multinomial probability distribution.

The following is the rest of the paper: Section 2 outlines the literature reviews and emphasizes the most influential works that inspired this investigation. Section 3 is the core of this study and discusses the proposed methodology's design and explains every part in greater detail. Section 4 provides a discussion of the results obtained in various experiments, and finally, Section 5 concludes this study, highlights the contribution of this research, and outlines our plans for future work.

2. Literature Work

In recent years, the development of computer-aided (CAD) systems for cataract diagnosis and grading from retinal images has achieved widespread popularity in the medical imaging industry. According to numerous studies published in this field, automatic cataract detection and grading systems consist primarily of three steps: preprocessing, feature extraction, and classification, as shown in Figure 3.

The preprocessing step is used to improve the fundus image quality through various stages such as resizing, G-filtering, and normalization. Fundus images collected from different datasets are of different sizes. Therefore, resizing step converts them to the same size to make them suitable for the sequential processing of CNN. G-filtering extracts the G-channel from RGB fundus images and is used to eliminate uneven illumination. The G-channel images preserve essential features of the original images and are the clearest ones among all channels. In the literature, the feature extraction step is performed with two distinct kinds of approaches manual approaches (or hand-crafted feature extraction methods) and automatic feature extraction approaches. Manual or hand-crafted feature extraction approaches use image processing-based techniques to extract hand-crafted features and then classify cataracts using conventional machine learning algorithms. Automatic feature extraction approaches are based on DL algorithms that automatically extract features

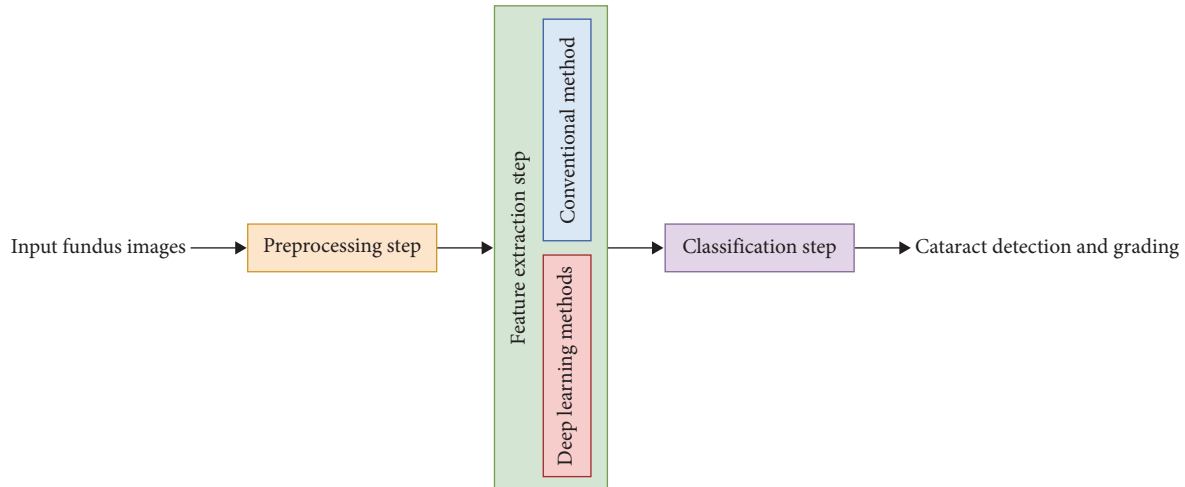


FIGURE 3: The primary steps of cataract detection and classification systems.

without human intervention and then utilize them to train and test machine learning classifiers for cataract classification. Therefore, the literature review is separated into two sections based on feature extraction approaches, namely, (1) automatic cataract detection using conventional methods and (2) automatic cataract detection using DL methods. A detailed review of both types of methods is given below.

2.1. Automatic Cataract Detection Using Conventional Methods.

Conventional methods for the automatic detection and grading of cataracts rely on the manual extraction of information utilizing various image processing algorithms and filters. Many researchers have worked with hand-crafted feature-based methods, like Cao et al. [17] that used improved Haar-wavelet transform to extract appropriate features for automatically grading cataracts. This technique also divides four-class classification issues into three two-class classification problems using a hierarchical strategy and trains neural network-based classifiers on these three sets of two-class problems independently. The result of four-class classification is achieved by integrating all two-class classifiers and archived two- and four-class accuracy of 94.83% and 85.98%, respectively. Yang et al. [18] proposed a method in which an enhanced top-bottom Hat transformation is utilized to enhance the contrast between the foreground and background of fundus images, and a trilateral filter is employed to remove image noise. The images' luminance and texture are then retrieved as classification features, and a backpropagation neural network is used to identify cataracts based on these features. This categorization approach achieves an accuracy of 82.9%. Guo et al. [19] trained and tested multiclass discriminant analysis algorithms for detecting and grading cataracts from fundus images by using wavelet transform and sketch-based features. The two- and four-class accuracy obtained by this method for the wavelet transform-based features are 90.9% and 77.1%, and 86.1% and 74.0% for sketch-based features. Yang et al. [20] introduced an ensemble learning-based cataract identification and grading system that extracts an independent set of wavelet, sketch, and texture-based features from fundus

images. This method uses an ensemble of a backpropagation neural network and support vector machine (SVM) classifiers with the majority of voting and stacking to detect and grade cataracts with 93.2% and 84.4% accuracy, respectively.

Zheng et al. [14] proposed a method that utilized 2D-DFT spectrograms of fundus images as classifying features to classify cataracts. This method reduces the dimensionality of feature vectors using principle component analysis (PCA), and the Adaboost algorithm is used to train and evaluate the linear discriminant analysis classifier in order to accomplish classification. The classification accuracy of this approach is 81.52%. Fan et al. [21] proposed a method that employs PCA to reduce the dimensionality of retrieved wavelet and sketched-based features from fundus images. This approach utilized widely used classification methods such as SVMs, bagging, random forests, gradient boosting, and decision trees in order to classify cataracts. The major advantage of this method is the reduction in computation cost. Song et al. [22] introduced a system for cataract classification that utilized an enhanced semi-supervised learning technique that gained some additional information from unlabelled cataract fundus images in addition to the three primary image features of textures, wavelets, and sketches. This approach combines multiple binary classifiers into a single potent multiclassifier and achieves 88.60% four-class accuracy. Manchalwar et al. [23] studied a system that utilized the histogram of oriented gradients features of fundus images and classified them using minimal distance classifiers in order to find cataracts. Pratap and Kokil [24] employed singular value decomposition to extract features and a SVM as a classifier to detect cataracts in fundus images. This approach yields a two-class classification accuracy of 97.78%.

As per the previous discussion in the literature, the performance of conventional methods is highly dependent on extracted hand-crafted features. Therefore, these methods are subject to the following limitations:

- (1) The manual extraction and selection of hand-crafted features is a time-consuming task that demands professional judgment on feature validity.

- (2) Tiny blood vessel features on retinal images are neglected during feature extraction using traditional approaches, despite the fact that they are essential for cataract identification and grading.

2.2. Automatic Cataract Detection Using DL Methods. The DL-based approaches illustrate the characteristics of autonomous feature extraction and can circumvent the aforementioned drawbacks. A number of researchers have worked hard to come up with DL models for classifying and grading cataracts. For Instance, Zhang et al. [25] came up with an eight-layer deep CNN (DCNN) in which the first five layers are convolution layers, and the last three layers are fully connected layers. The output of the last fully connected layer (FC) of DCNN is used as input by the softmax classifier to make a distribution across the four classes. Ran et al. [26] devised a method that uses the DCNN and random forest classifier to grade the cataract into six classes. The accuracy of this method for grading cataracts into six classes is 90.7%. Imran et al. [27] introduced a method that combines the self-organizing map (SOM) and a radial basis neural network (RBF-NN). This method employs SOM for clustering and determining the starting center and RBF-NN for classifying and grading cataracts. The proposed method has an accuracy of 95.3% for cataract detection and 91.7% for cataract classification into four classes.

Syarifah et al. [28] improved the performance of DCNN using a pretrained CNN architecture termed AlexNet with a lookahead optimizer on Stochastic Gradient Descent (SGD) and Active Design and Analysis Modeling (ADAM). The proposed architecture has a two-class classification accuracy of 97.5%. The limited size of the dataset is the primary limitation of this method. Junayed et al. [29] presented CataractNet, a new DCNN architecture with fewer layers, training parameters, and smaller size kernels designed to save computing time and cost. This approach has a classification accuracy of 99.13% for cataracts. Pratap and Kokil [30] proposed a method for classifying four-class cataracts using a pretrained AlexNet with transfer learning and an SVM classifier. This method used the Naturalness Image Quality Evaluator (NIQE) and the Perceptron Image Quality Evaluator (PIQE), both of which are blind image quality selection parameters, to measure the quality of retinal images taken from open-source datasets. The proposed method has a 92.9% accuracy in grading the cataract into four-classes. Hasan et al. [31] examined the performance of four pretrained CNN models, including DenseNet121, InceptionV3, Xception, and InceptionResNetV2, to diagnose cataracts from retinal images. Moreover, the authors discovered that InceptionResNetV2 outperforms all other models, with a two-class classification accuracy of 98.17%. Weni et al. [32] investigated a CNN-based method for automatically detecting cataracts. The proposed strategy aims to improve diagnostic precision for cataracts while minimizing loss. They were able to achieve an accuracy rate of 95% by using 50 epochs. However, when evaluated directly on ten real-time images, the accuracy of the proposed method is 88%. Varma et al. [33] developed the

fundus images-based automatic cataract classification and grading system that uses a CNN model that contains lesser layers and parameters and small size kernels to obtain better computational cost. The accuracy obtained by this approach is 92.75% for four-class classification of cataracts. Varma et al. [34] proposed the custom-CNN architecture for feature extraction and softmax classifier for classification of cataractous images into four classes according to severity. The accuracy achieved by this method is 92.7% for grading cataracts.

It can be observed from the aforementioned literature that automatic cataract detection using DL methods (particularly CNNs) is more accurate and efficient than existing methods. However, challenges still need to be addressed in the applications of DL models, such as the availability of large labeled datasets and the difficulty of extracting retinal features such as tiny blood vessels, which may affect other blood vessels and depend on the image quality. Although, the first challenge of the lack of a labeled dataset is alleviated to a certain extent in recent research using data augmentation methods. However, the second challenge of extracting tiny blood vessels from retinal images is still alive and needs to be resolved to improve the system's accuracy. Therefore, this study uses 2D-DFT spectrograms of retinal images in order to investigate this issue. The advantage of employing 2D-DFT spectrograms is that they carry details of tiny blood vessels as high-frequency components, which are easier to extract and serve as the discriminating features to detect and grade the cataracts. Besides this, this method employs an image quality selection module that retains retinal images of high quality for further processing and discards those of poor quality.

3. Proposed Work

This section discusses the methodology employed by the proposed study for cataract detection and grading. The complete methodology consists of five major components: image acquisition (dataset construction), image quality selection, preprocessing and data augmentation, feature extraction, and classification. The outline of the proposed study for the automatic detection and classification of cataracts is depicted in Figure 4. Now, let us examine the working of each part in greater depth.

3.1. Image Acquisition (Dataset Construction). The most challenging aspect of this problem is the need for a benchmark dataset containing a large number of labeled retinal fundus images. As a result, this study uses randomly selected retinal images from various publically available datasets, including HRF [35], STARE [36], MESSIDOR [37], DRIVE [38], DRIONS_DB [39], and IDRiD [40] datasets, as well as images obtained from the Internet. The comprehensive description of the datasets is presented in Table 1.

In the dataset, a total of 1,835 fundus images are compiled in such a way that each class has more than 400 images. The class label assignment for fundus images is overseen by Dr. P. K. Gupta, a professional ophthalmologist of a private eye care center, Ghaziabad (UP). The ophthalmologist

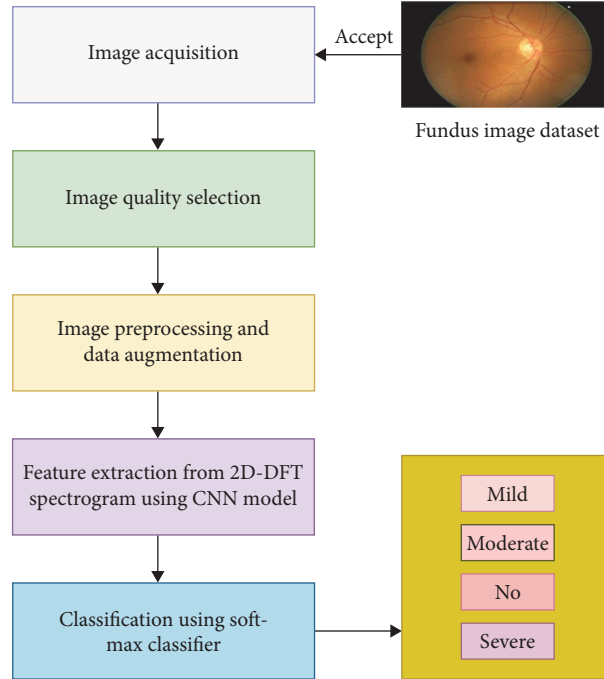


FIGURE 4: Proposed model for automatic cataract detection and classification.

TABLE 1: A comprehensive description of various datasets of fundus images.

S. no.	Dataset name	Dataset size	Image size (pixels)
1.	HRF (high-resolution fundus) [35]	45 images	$3,304 \times 2,336$
2.	STARE (structured analysis of the retina) [36]	397 images	700×605
3.	MESSIDOR [37]	1,200 images (three sets)	Set 1 : $1,440 \times 960$ Set 2 : $2,240 \times 1,488$ Set 3 : $2,304 \times 1,536$
4.	DRIVE (digital retinal images for vessel extraction) [38]	40 images (two sets)	565×584 (both sets)
5.	DRIONS-DB [39]	110 images	600×400
6.	IDRiD (Indian diabetic retinopathy image dataset) [40]	516 images	$4,288 \times 2,848$

TABLE 2: A complete description of acquired fundus images.

S. no.	Category	Total fundus images
1.	Mild	465
2.	Moderate	460
3.	No	450
4.	Severe	460
	Total	1,835

carefully examined each fundus image of different datasets and looked for key indicators such as opacity, discoloration, altered fundus reflex, and reflected light abnormalities. He paid close attention to these indicators and compared them with his knowledge and experience for class-label assignments. This was completely qualitative method and totally depends on clinical judgment of ophthalmologist. The detailed descriptions of created labeled fundus images dataset are given in Table 2.

3.2. Selection of Good Quality Fundus Images. It has been observed that image quality is crucial in deep neural networks [41]. The disparity in image quality between deep neural network training and testing degrades the classifier performance. Maintaining the same degree of image quality throughout the training and testing phases is therefore highly desired. The high-quality fundus images are those which show the clear retinal structures, whereas the poor-quality images do not show the clear retinal structures due to distortion, noise, defocus, blur, weak and overexposure, and eyelash shadow. Consequently, this study includes a quality-selection module that selects high-quality fundus images for subsequent diagnostic evaluation in order to improve the performance.

This quality-selection module includes two subjective image quality descriptors, namely NIQE [42] and PIQE [43], to evaluate the quality of fundus images. NIQE is a blind quality estimator that calculates image quality scores based on observable departures from statistical regularities in

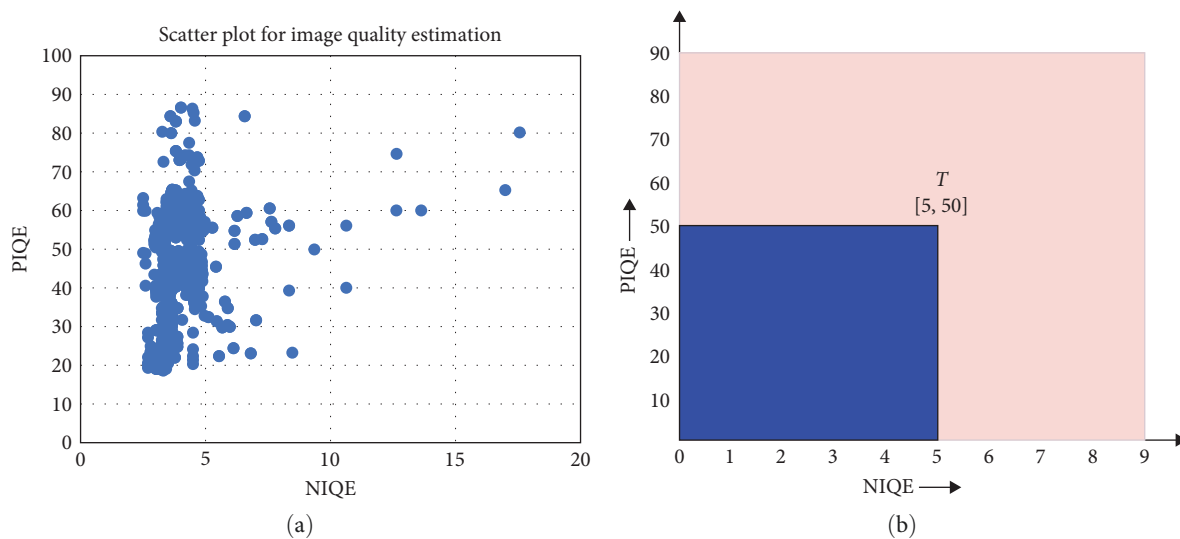


FIGURE 5: (a) Scatter plot corresponding to NIQE and PIQE scores of fundus images obtained from the dataset. (b) Good quality fundus images selection criteria. Blue dot signifies the coordinate point for each fundus image.

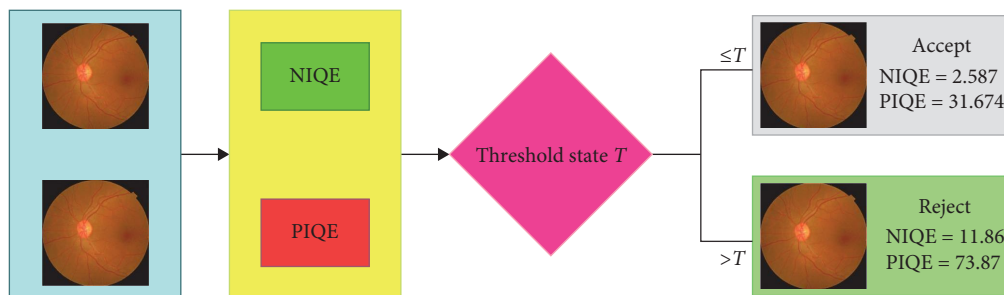


FIGURE 6: Functioning of quality selection module for fundus images.

natural images. It employs a simple and effective spatial domain natural scene statistics model to provide a set of “quality-aware” statistical characteristics. In comparison, PIQE is a nonreference perception-based image quality evaluation method used to assess the quality of images from the real world. The image quality score is calculated using the mean subtraction contrast normalization coefficient and belongs in the range [0–100]. It is given that the lower NIQE and PIQE levels are indicative of better perceptual quality, whereas the higher scores suggest poor perceptual quality. According to the experimental findings presented in references [42, 43], the NIQE and PIQE values for good-quality images do not exceed 5 and 50, respectively.

In this study, NIQE and PIQE scores are calculated for all 1,835 fundus images of the dataset, and the findings are displayed via a scatter plot in Figure 5(a). Retinal fundus images with a NIQE score of less than or equal to 5 and a PIQE score of less than or equal to 50 are considered of good quality, as depicted in Figure 5(b). These carefully selected, good-quality images are then used for training and testing purposes of DL models.

Fundus images whose NIQE and PIQE scores fall below the threshold value (T) are chosen, while the remaining

retinal images are discarded. Figure 6 depicts the operation of the image quality selection module, which employs a threshold point (T) with a value between 5 and 50. The images whose (NIQE score, PIQE score) falls below the threshold (T) are considered for subsequent processing, while the remaining images are excluded due to their poor quality. The selected and rejected retinal images from the collected dataset are described in detail in Table 3.

3.3. Image Preprocessing, 2D-DFT Transformation, and Augmentation. After image acquisition and selection, the fundus images must be preprocessed to increase their quality and achieve greater generalization. The preprocessing steps include resizing the images, green channel extraction, normalization, 2D DFT, and data augmentation. First, resizing operation is performed to unify the fundus images so that they become suitable for series processing. This study used the bicubic interpolation method for resizing the fundus images. Second, the green channel extraction process is employed to extract the green component from the original color fundus images to correct nonuniform illumination. The green component is clearest than the red and blue components. The primary benefit of employing the green

TABLE 3: A complete description of selected and rejected fundus images of the dataset.

S. no.	Category	Number of selected fundus images	Number of rejected fundus images
1.	Mild	400	65
2.	Moderate	400	60
3.	No	400	50
4.	Severe	400	60
	Total	1,600	235

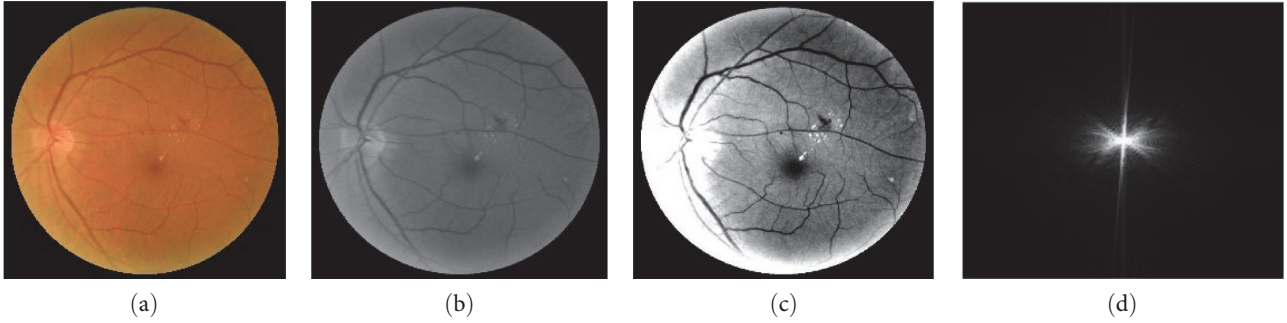


FIGURE 7: Cataract images after (a) resizing, (b) green channel extraction, (c) normalization, and (d) 2D DFT.

channel is that it delivers more contrast and illumination features while preserving all vital information of the original fundus images. In addition, green channel extraction saves computation time by a factor of one-third because it compresses the original image by one-third. Third, normalization removes the interference of background effects and assigns a new intensity range to pixels of fundus images, and the normalized pixel intensity value of each pixel of the fundus image is computed by subtracting the mean intensity value and dividing the result by the standard deviation of the intensity values of all pixels in the fundus image. Finally, 2D DFT is used to convert the fundus images into frequency spectrograms. Figure 7 illustrates the resulting images after applying various preprocessing steps to a fundus image.

Based on the aforementioned discussion, this study used the 2D-DFT spectrogram's features for cataract detection and grading. For an image with dimensions u and v , the 2D DFT is shown as:

$$F(x, y) = \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} f(u, v) e^{-k2\pi(\frac{xu}{U} + \frac{yv}{V})}. \quad (1)$$

Here, $f(u, v)$ represents the picture in the spatial domain, while the exponential term represents the basis function for each point $F(x, y)$ in Fourier space. The base functions are sine and cosine waves with increasing frequencies, so $F(0, 0)$ represents the direct current (DC) component of the image, which corresponds to the average brightness, while $F(M, N)$ represents the maximum frequency. In order to improve DFT spectrogram image representation, The DC-value (also known as the zero-frequency point) $F(0, 0)$ is displayed in the image's center, and as the image point's distance from

the center point grows, its frequency also increases. This task is performed in MATLAB as follows:

$$y_f = \text{abs}(\text{fftshift}(\text{fft2}(\text{image data}))); \quad (2)$$

$$\text{result} = \log(1 + y_f); \quad (3)$$

The idea of using 2D DFT is taken from general science, where light propagates in the lens of the eye of a person with cataracts, it will be scattered or absorbed by opacifications. This means opacifications in the lens can be treated as a low-pass filter that passes only low frequency and stops high frequency. Therefore, fundus photography has no blood vessels on images with severe cataracts because blood vessels are high-frequency components. The appearance of blood vessels in fundus images increases as cataract severity decreases. This idea is practically simulated by 2D DFT in the frequency domain by localizing the blood vessels in the fundus images. Blood vessels are thin edges that can be seen in fundus images and are represented by high-frequency components in frequency domain transformation. For example, if cataract severity is high, then the fundus image is not clear, and no blood vessel is seen; whose result is that the spectrogram of severe cataract fundus image has a very less high-frequency component. The vice versa is true for normal eye fundus images. Hence, it is concluded from the earlier discussion that the structure of the spectrogram becomes more regular as the cataract degree increases, the high-frequency components become less, and the low-frequency components become larger, as illustrated in Figure 8.

Figure 8(a) shows a normal fundus image that contains finer details and its corresponding 2D-DFT spectrogram

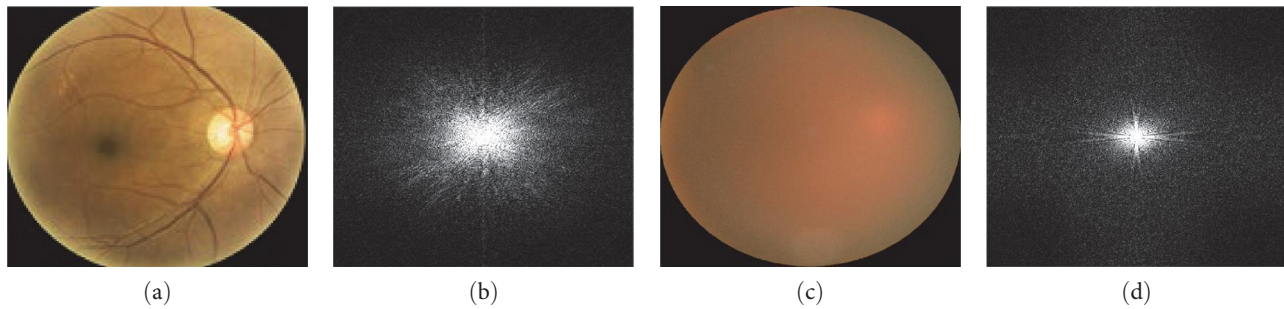


FIGURE 8: Fundus image and corresponding 2D-DFT spectrograms. (a) Normal fundus image, (b) 2D-DFT spectrogram of normal fundus image, (c) severe cataract fundus image, and (d) 2D-DFT spectrogram of a severe cataract fundus image.

image shown in Figure 8(b), where significant amplitudes are away from the zero-frequency center point. Similarly, Figure 8(c) shows a cataractous fundus image in which fundus fine details are missing, and its corresponding 2D-DFT spectrogram image showed in Figure 8(d), where there is a reduction in high spatial frequency contents in the region surrounding the zero-frequency point.

Finally, the data augmentation step is used to increase the number of training samples to address generalization and overfitting issues. The key data augmentation processes, including rotation, flipping (horizontal), cropping, and shifting, are performed on 2D-DFT spectrogram images.

3.4. Feature Extraction and Classification. Feature extraction is a critical step of CAD systems that directly impacts the accuracy of the classification. The application of CNN models to medical image diagnosis is well-recognized worldwide [18]. CNNs are deep neural networks that aim to automatically learn a complex hierarchy of features from medical images in order to diagnose and grade medical disorders. A CNN typically consists of four primary layers, which are the convolution, pooling, fully connected, and classification layers. In CNN, the output of one layer is the input for the next layer, and this output is known as a feature map (or activation map). The convolutional layer is responsible for extracting a number of low-level and high-level features from spectrogram images (also known as activation maps) via a set of linear filters. These features include outlines such as edges, curves, dots, corners, squares, circles, and others. The output of the 3×3 convolution filter is given by:

$$f(l, m, n) = \sum_{k=1}^3 \sum_{i=1}^3 \sum_{j=1}^3 w(l, i, j, k) I(i + m - 1, j + n - 1, k) + b(l), \quad (4)$$

where $I(i, j, k)$ is the image intensity value, $w(l, i, j, k)$ represents weights, and $b(l)$ represents the bias of the convolutional layer.

The pooling layer often reduces the dimensionality of activation maps (i.e., the number of network parameters) via subsampling in order to increase the robustness of retrieved features. The pooling layer may be implemented in one of two ways: (i) using a collection of linear filters to determine

the image's average pixel value below the masked area (average pooling), and (ii) using a set of nonlinear filters to sort the pixel values inside a specific region of the input image, and receiving the pixel value with the highest absolute value as a result (max-pooling). The working of the maximum-pooling layer for a 2×2 grid is defined by the following equation:

$$x(l, m + 1, n + 1) = \max \begin{bmatrix} f(l, 2m + 1, 2n + 1) & f(l, 2m + 1, 2n + 2) \\ f(l, 2m + 2, 2n + 1) & f(l, 2m + 2, 2n + 2) \end{bmatrix}, \quad (5)$$

where $x(l, m, n)$ is the maximum pixel value corresponding to four neighbors in activation map (f).

The role of the batch normalization (BN) layer is to normalize the output of the preceding layer of the network during training in order to boost learning speed and regularize the CNN in order to address the overfitting issue. The BN layer also facilitates other layers of the network to learn independently.

The FC consists of a group of neurons that are linked to all the neurons corresponding to the activation maps present in the preceding layer. The primary role of the FC layer is to produce a compact feature representation of the whole input image. Typically, the outputs of previous FC layers and convolutional layers are processed by a rectified linear unit (ReLU) activation function, which is specified by:

$$y = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (6)$$

where x represents the input to the ReLU activation function and y represents the output created by the ReLU activation function. However, this study used clipped ReLU activation functions instead of simple ReLU to overcome from vanishing gradient problem. The clipped ReLU function is defined as:

$$y = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < c \\ c & x \geq c \end{cases}, \quad (7)$$

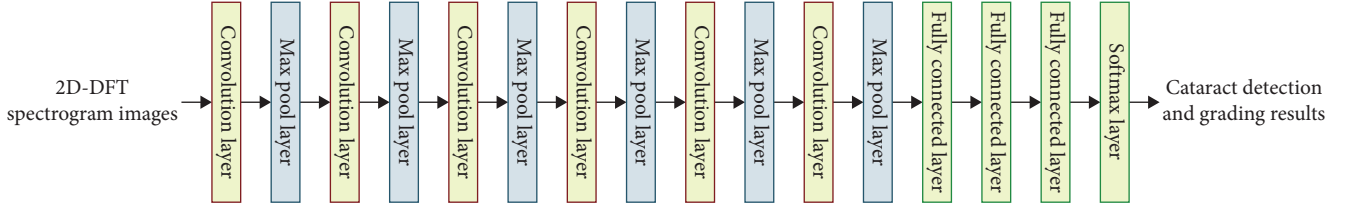


FIGURE 9: A schematic representation of the proposed CNN model.

TABLE 4: The detailed specification of various layers of the proposed CNN model.

S. no.	Layers	Filters	Activation function	Stride	Activations
1.	Convolution layer-1	16	KS: 3×3 , clipped ReLU	–	$256 \times 256 \times 16$
2.	Max-pooling layer-1	–	KS: 2×2	2	$128 \times 128 \times 16$
3.	Convolution layer-2	16	KS: 3×3 , clipped ReLU	–	$128 \times 128 \times 16$
4.	Max-pooling layer-2	–	KS: 2×2	2	$64 \times 64 \times 16$
5.	Convolution layer-3	32	KS: 3×3 , clipped ReLU	–	$64 \times 64 \times 32$
6.	Max-pooling layer-3	–	KS: 2×2	2	$32 \times 32 \times 32$
7.	Convolution layer-4	64	KS: 3×3 , clipped ReLU	–	$32 \times 32 \times 64$
8.	Max-pooling layer-4	–	KS: 2×2	2	$16 \times 16 \times 64$
9.	Convolution layer-5	128	KS: 3×3 , clipped ReLU	–	$16 \times 16 \times 128$
10.	Max-pooling layer-5	–	KS: 2×2	2	$8 \times 8 \times 128$
11.	Convolution layer-6	256	KS: 3×3 , clipped ReLU	–	$8 \times 8 \times 256$
12.	Max-pooling layer-6	–	KS: 2×2	2	$4 \times 4 \times 256$
13.	Fully connected layer-1	–	500; clipped ReLU	–	$1 \times 1 \times 500$
14.	Dropout	–	0.7	–	$1 \times 1 \times 500$
15.	Fully connected layer-2	–	200; clipped ReLU	–	$1 \times 1 \times 200$
16.	Dropout	–	0.6	–	$1 \times 1 \times 200$
17.	Fully connected layer-3	–	40; clipped ReLU	–	$1 \times 1 \times 50$
18.	Dropout	–	0.5	–	$1 \times 1 \times 4$
19.	Fully connected layer-4	–	4; softmax	–	$1 \times 1 \times 4$

where c is the clipping ceiling value that is used for thresholding operations.

However, the softmax activation function is implemented at the end of CNN to calculate the probability distribution for each of the final FC outputs, as stated by:

$$v_i = \frac{e^{u_i}}{\sum_{n=1}^N e^{u_n}}, \quad (8)$$

where u_i is the i^{th} output of the last FC, v_i is the corresponding softmax activation, and N is the number of classes.

The cross-entropy loss function is used to determine the deviation of predicted outputs of the softmax function from actual outputs and is defined as:

$$e = -\sum_{j=1}^N \hat{v}_j \log v_j, \quad (9)$$

where v_j is the actual probability for j^{th} output of the last FC layer to belong particular class. The cross-entropy loss is then minimized using a backpropagation algorithm with optimization

functions, namely, SGD, ADAM, and so on, to update the model parameters necessary for effective image classification.

In this study, a custom-CNN architecture is proposed that contains lesser layers, parameters, and smaller size kernels to achieve better computational cost and accuracy to automatically detect and grade cataracts into four stages: normal, mild, moderate, and severe from fundus (2D-DFT spectrograms) images. Therefore, this architecture contains a set of six consecutive convolutional layers with 2×2 max-pool layers in between them, as shown in Figure 9. The convolutional layers used in this architecture consist of 16, 16, 32, 64, 128, and 256 filters with kernel size 3×3 and padding as same for each of the six layers, respectively. The max-pool layers with kernel size 2×2 and stride of two are employed between convolutional layers to reduce the size of the data representation, which also lowers the number of trainable parameters. The outputs of all six convolutional layers are compiled into a feature map, which is then used to feed a series of FCs. These layers are utilized to identify and classify cataracts. Three sets of FCs and dropout layers are created, with the FCs containing 500, 200, and 50 neurons, respectively, to capture the filtered cataract features. Furthermore, three dropouts are set to 0.7, 0.6, and 0.5 to reduce the risk of model overfitting by removing the output

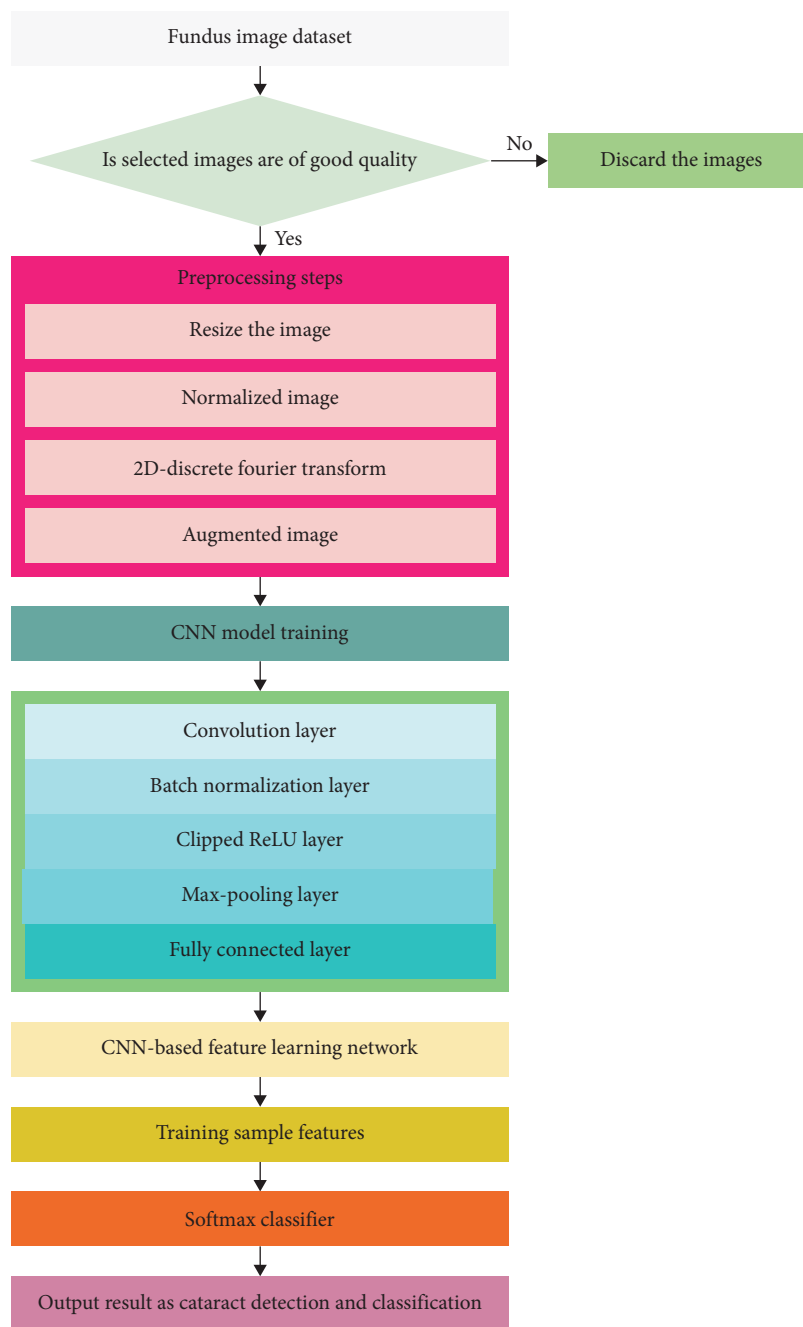


FIGURE 10: A detailed diagram of the proposed method.

of 70%, 60%, and 50% of hidden layer neurons with each update during the training phase. The last FC layer contains four neurons for nonlinear classification. All layers employ the clipped ReLU activation function, with the exception of the classification layer, which uses the softmax function. A detailed description of the underlying architecture of the proposed CNN model is given in Table 4.

The previously described technique is depicted in Figure 10 and can be summed up as follows:

- (1) First, the good-quality fundus images, which are the output of the quality selection module, are preprocessed, and their associated 2D-DFT spectrogram images are obtained.
- (2) Second, the augmented dataset of 2D-DFT spectrogram images is fed into the proposed CNN model for automatic feature extraction.
- (3) Third, in order to facilitate automatic feature extraction, the suggested CNN architecture makes use of layers that are convolutional, batch normalization, clipped ReLU, max-pool, and fully connected.
- (4) The last FCL has four neurons with a softmax activation function that computes the probability distribution for each class in order to classify spectrogram

TABLE 5: Comparative analysis of existing and proposed algorithms (the results are given in percentages).

Methods	Accuracy	Precision	Sensitivity	Specificity	F1-score
Standard SVM	87.34	86.63	86.35	88.57	86.47
AlexNet with softmax	91.32	92.05	91.22	91.44	91.63
ResNet with softmax	92.05	92.02	91.31	91.75	91.65
VGGNet with softmax	91.76	91.63	91.14	92.35	91.38
Proposed method	93.10	93.09	93.13	97.71	93.08

images into four classes such as normal, mild, moderate, and severe.

4. Experimental Results

In this section, the results of various experiments conducted with the proposed method are presented and discussed. All of the experiments in this study have been conducted on a computer equipped with an Intel 7th Generation Core i7-7700 processor, a 4 GB NVIDIA GTX 1050 Ti graphics card, 64 GB of RAM, and a Windows 10 (64-bit) operating system. The simulation programs for the proposed method have been developed and executed on MATLAB R2019a, which includes Image Processing, Neural Network Toolboxes, and Deep Learning Toolboxes.

4.1. Criteria for Performance Evaluation. The method employed in this study has been trained and evaluated on both retinal and their corresponding 2D-DFT spectrogram image datasets. The performance of the method has been evaluated based on the following performance metrics: accuracy, sensitivity, specificity, precision, and F1-score [11].

$$\text{Accuracy} = \frac{\text{T.P.} + \text{T.N.}}{\text{T.N.} + \text{T.P.} + \text{F.N.} + \text{F.P.}}. \quad (10)$$

$$\text{Sensitivity(or recall)} = \frac{\text{T.P.}}{\text{T.P.} + \text{F.N.}}. \quad (11)$$

$$\text{Specificity} = \frac{\text{T.N.}}{\text{T.N.} + \text{F.P.}}. \quad (12)$$

$$\text{Precision} = \frac{\text{T.P.}}{\text{T.P.} + \text{F.P.}}. \quad (13)$$

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}}. \quad (14)$$

Considering both the actual (target) class label and the predicted class label, the input images are divided into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The percentage of accurately predicted classes based on the total number of test images is referred to as accuracy. Precision is the percentage of correctly predicted positive classes based on the total number of positive cases, whereas sensitivity is the percentage of correctly predicted positive classes based on all positive predictions made from test images. In the field of medical

diagnostics, accuracy and sensitivity are the most important performance measures. The F1 score has become a common evaluation metric in the field of medical diagnostics due to the ease with which it can be compared to other metrics. This score condenses information about precision and recall (sensitivity) into a single value. For medical professionals, an FN is more distressing than an FP; hence, sensitivity (recall) is given precedence over precision (accuracy). FPs can be eliminated with more testing, whereas missed conditioned (FN) results can be catastrophic for the patient. In light of this, performance evaluation criteria in this study included precision, recall, specificity, and F1-score in addition to accuracy.

4.2. Performance Measures. This study evaluates the performance of the proposed method as well as other existing methods, including standard SVM, AlexNet-softmax, VGGNet-softmax, and ResNet-softmax, in order to demonstrate the comparison. It is also mentioned that all algorithms are trained and tested on our collected dataset due to the need for more relevant benchmark datasets. In order to illustrate the benefits of the proposed method, all of the algorithms are trained and tested using fundus images, whereas the proposed method is trained and tested using 2D-DFT spectrogram images corresponding to fundus images from the dataset. Table 5 shows the performance comparison of the previously discussed algorithms using various evaluation matrices as discussed above. It can be observed from Table 5 that the proposed method has achieved promising results as compared to all other algorithms. The proposed method obtained 93.10%, 93.13%, 97.71%, 93.09%, and 93.08% of accuracy, sensitivity, specificity, precision, and F1-score.

In addition, Table 5 makes it abundantly clear that the CNN-based algorithm outperforms hand-crafted feature extraction-based methods; this is why the standard SVM method, which relies on the hand-crafted feature extraction method, has the lowest performance. The capacity of CNN-based algorithms to automatically extract features that provide various semantic representations of fundus images is the main factor contributing to their superior accuracy compared to other image processing methods. However, all CNN models continue to struggle with the accurate extraction of retinal features, such as tiny blood vessels. 2D-DFT spectrograms corresponding to the fundus images contain details of tiny blood vessels in the form of high-frequency components, which is easier to extract and work as discriminating features for cataract detection and grading. Consequently, the proposed method used 2D-DFT spectrograms of fundus images as input to the CNN model to extract features successfully and classify them into distinct classes.

TABLE 6: Test accuracy of the proposed method using different quality fundus images.

Image quality of training dataset	Image quality of test dataset	Accuracy of classification (%)
Good images	Good images	93.10
Good images	Poor images	82.62
Poor images	Good images	88.16
Poor images	Poor images	90.54

TABLE 7: A detailed description of the computation time of each step of the proposed method.

Process	Total computation time (s)
Training	160
Validation	35
Single image feature extraction and classification	0.75

4.3. Significance of Image Quality. The image quality has a substantial impact on the testing accuracy results of any DL model. Table 6 shows the results of testing accuracies of the proposed method when the model is trained and tested with 2D-DFT spectrograms of different image quality datasets. It is decided after reviewing Table 6 that the testing accuracy of the proposed model is adequate when 2D-DFT spectrograms from datasets of the same quality are used for training and testing the model. However, a considerable deterioration in testing accuracy is observed when the proposed model is trained and tested using 2D-DFT spectrograms from different image quality datasets. This deteriorated testing accuracy may also be the result of the classifier's poor performance. Although, it is more difficult to determine the correct cause for the deterioration in testing accuracy [11]. Therefore, this study incorporates a quality-selection module that segregates images into good-quality and poor-quality images in order to reach a correct decision regarding testing accuracy.

4.4. Computation Time. The computation time required to complete each stage of the proposed method is outlined in Table 7. It is observed from Table 7 that the proposed method requires 160 s for model training and 35 s for model validation. Table 7 also depicts that the proposed method requires 0.75 s for the feature extraction and classification of a 2D-DFT spectrogram image corresponding to a fundus image. Figure 11 illustrates a comparison chart that compares the computation time of various methods. It is observed from the comparison chart that the computation time of the proposed method is much shorter than that of other methods, such as traditional SVM, AlexNet, VGG19Net, and ResNet50. As a result, the proposed method is suitable in terms of the amount of time required for computation.

4.5. Results Analysis and Discussion. In the first stage of the proposed method, a quality selection module is used to evaluate the quality of the fundus images and select 1,600 good-quality

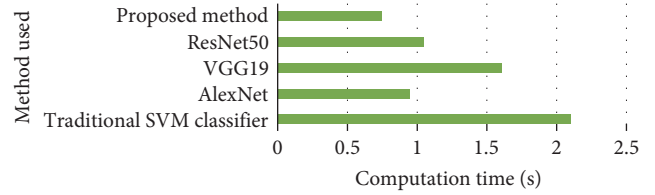


FIGURE 11: Comparison of proposed method computation time (single-image) to other methods.

TABLE 8: A detailed description of hyper-parameters configuration.

Configuration	Value
Learning rate	0.003
Epochs	150
Batch size	128
Optimizer	ADAM
Classification function	Softmax
Network weights	$[-1 \ 1]$

		Confusion matrix				
Output class	Mild	141 22.0%	10 1.6%	5 0.8%	0 0.0%	90.4% 9.6%
	Moderate	7 1.1%	144 22.5%	1 0.2%	2 0.3%	93.5% 6.5%
	No	12 1.9%	0 0.0%	153 23.9%	0 0.0%	92.7% 7.3%
	Severe	0 0.0%	6 0.9%	1 0.2%	158 24.7%	95.8% 4.2%
		88.1% 11.9%	90.0% 10.0%	95.6% 4.4%	98.8% 1.2%	93.1% 6.9%
		Mild	Moderate	No	Severe	Target class

FIGURE 12: Confusion matrix of the proposed method.

images from the dataset that is used in subsequent processing steps. Next, the preprocessing stage improves the quality of fundus images by using resizing, green channel extraction, and normalization operations. Thereafter, 2D-DFT transform is applied to fundus images to obtain spectrogram images corresponding to fundus images, and then augmentation is performed to expand the size of the dataset. Now, this dataset of augmented 2D-DFT spectrogram images is randomly partitioned into an 80:20 ratio to train and test the proposed CNN model. During the training phase, the proposed CNN model is trained on the training dataset in batch mode with a batch size of 128, and network weights are adjusted inside the interval $[-1 \ 1]$. In addition, this CNN model is trained with a total of 150 epochs and optimized using an ADAM optimizer

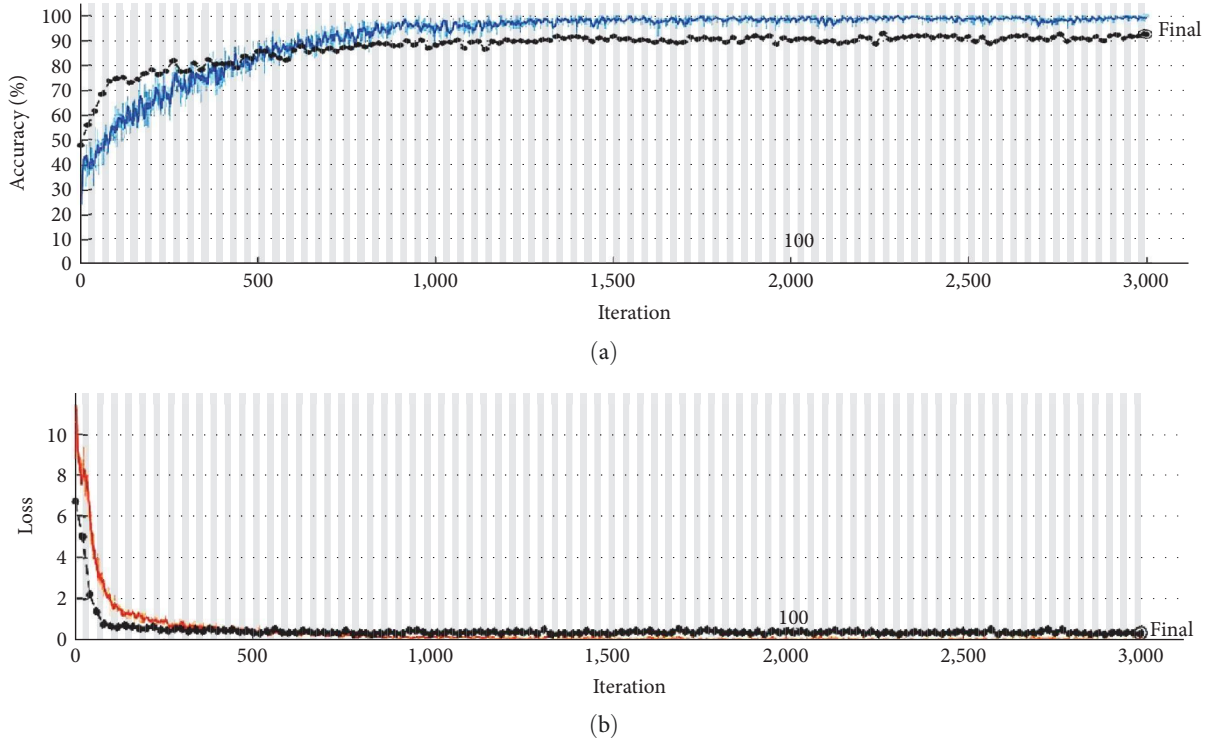


FIGURE 13: (a) The curve of training and validation accuracy of the CNN of the proposed method. (b) The curve of training and validation loss of the CNN of the proposed method.

TABLE 9: Comparison summary between the proposed method and existing state-of-the-art methods.

S. no.	Authors	Year	Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
1.	Pratap and Kokil [30]	2019	Alexnet-SVM	92.87	92.88	93.04	93.04
2.	Cao et al. [17]	2020	Improved Haar wavelet	85.98	84.65	95.29	86.01
3.	Pratap and Kokil [44]	2021	DCNN-SVM	92.90	92.90	93.06	93.06
4.	Varma et al. [34]	2023	D-CNN-softmax	92.7	93.5	92.75	92.75
5.	Proposed method	2022	CNN with 2D DFT	93.10	93.13	97.71	93.09

with a learning rate of 0.003. A detailed description of other fine-tuned hyperparameters is given in Table 8. Multiclass classification task in CNN is performed by softmax activation function that is used as a loss function aligned with the cross-entropy function. It is observed from Table 5 that the proposed method performs better in terms of classification accuracy than the conventional SVM classifier by 5.79%, AlexNet-softmax by 1.79%, VGG19Net-softmax by 1.35%, and ResNet50-softmax by 0.99%. The confusion matrix of the proposed method, as well as its accuracy and loss curves during training and validation, are depicted in Figures 12 and 13, respectively.

The total computation time includes both the time required for extracting features and the time required for classifying images. Figure 11 presents a comparison of the amount of computation time required by various methods. It can be seen from Table 7 that the amount of time required to classify a test sample by a conventional SVM classifier, AlexNet, VGG19, ResNet50, and the suggested technique is 2.1,

0.95, 1.61, 1.05, and 0.75 s, respectively. Therefore, it is concluded that the classification of 2D-DFT spectrogram images using the proposed method takes less time than the conventional SVM classifier-based method due to the automatic feature extraction used in the proposed method. It also takes less computation time than other pretrained networks due to its significantly minimized size and trainable parameters.

The effectiveness of the proposed method is demonstrated by comparing it to the state of art methods. Table 9 presents a comparison of the various state-of-the-art cataract classification methods with the proposed method using commonly available evaluation matrices. It is important to note that these state of art methods are developed using the researchers' private datasets, which are unavailable to the general public. Therefore, these methods are implemented and tested on our dataset. Pratap and Kokil [30] employed a pretrained AlexNet with transfer learning to extract features from fundus images, followed by an SVM classifier for classification. The given method attained an accuracy of

92.87% for detecting and grading four-class cataracts. Cao et al. [17] used an enhanced Haar-wavelet transform to extract features and developed a hierarchical technique for dividing a four-class classification problem into three two-class classification problems. This approach attained a four-class cataract grading accuracy of 85.98%. Pratap and Kokil [44] suggested a novel method for investigating the performance of cataract detection and grading systems in noisy environments. This approach employs a collection of independent SVMs that are trained locally and globally on features extracted by different pretrained CNNs in the presence of varying noise levels. The choice of a CNN network is highly determined by the noise level present in fundus images. The results of this method show that it is robust to noise and has a maximum accuracy of 92.90% for four-class cataract detection and grading. Varma et al. [34] proposed custom-CNN architecture for feature extraction and softmax classifier for classification of cataractous images into four classes according to severity. The accuracy achieved by this method is 92.7% for grading cataracts. Overall results shown in Table 9 indicate that the proposed technique consistently outperforms state-of-the-art alternatives.

5. Conclusion

This study introduced the novel concept of employing 2D-DFT spectrograms of fundus images rather than the original images to detect and grade cataracts automatically using a convolutional neural network. The benefit of using 2D-DFT spectrograms is that they contain details of tiny blood vessels in the form of high-frequency components, which are easier to extract and serve as the distinguishing features to detect and grade cataracts. The cataract dataset is initially compiled from various open-source datasets, followed by quality selection and preprocessing steps to improve the quality of fundus images. The 2D DFT is then utilized to convert fundus images into spectrogram images, which are then fed into the convolutional neural network for feature extraction. The softmax classifier is then used to classify cataracts based on the extracted features. The proposed method demonstrated better results in terms of both automatic feature extraction and classification accuracy when compared with pretrained CNN models, including AlexNet, VGGNet, and ResNet50 for cataract detection and classification. The experimental results revealed that the proposed method surpassed existing methods in terms of accuracy (93.10%), sensitivity (93.13%), specificity (97.71%), precision (93.09%), and F1-score (93.09%). It is also worthy here to mention here that the proposed method takes lesser computation time due to lesser layers, parameters, and smaller size kernels for training and testing of CNN model as compared to other DL methods.

The proposed method can reduce costs and simplify the process of cataract detection and grading, which is advantageous for rural residents who lack access to qualified ophthalmologists. The future work of this study will involve installing this application in rural areas for inexpensive cataract diagnosis using Internet of things-based techniques. In addition, the major area for improvement of the proposed

method is its training and testing on a limited dataset. Therefore, it will be necessary to evaluate the proposed method with real-time and larger datasets in the near future to evaluate its effectiveness in real-time scenarios.

Data Availability

Data supporting the findings of this study is available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors were involved from the initial ideation to the final draught, and each is equally responsible for the final output.

Acknowledgments

This study is supported by Dr. A.P.J. Kalam Technical University, Lucknow, India, under Visvesvaraya Research Promotion Scheme (AKTU/Dean-PGSR/VRPS-2020/5751).

References

- [1] F. F. A. Foster, "Vision 2020: the cataract challenge," *Community Eye Health*, vol. 13, no. 34, pp. 17–19, 2000.
- [2] WHO, *Global Data on Visual Impairments 2010*, World Health Organization, 2012.
- [3] P. Vashist, S. S. Senjam, V. Gupta, N. Gupta, and A. Kumar, "Definition of blindness under National Programme for Control of Blindness: do we need to revise it?" *Indian Journal of Ophthalmology*, vol. 65, no. 2, pp. 92–96, 2017.
- [4] Y.-C. Liu, M. Wilkins, T. Kim, B. Malyugin, and J. S. Mehta, "Cataracts," *The Lancet*, vol. 390, no. 10094, pp. 600–612, 2017.
- [5] M. Karbassi, P. M. Khu, D. M. Singer, and L. T. Chylack Jr., "Evaluation of lens opacities classification system III applied at the slitlamp," *Optometry and Vision Science*, vol. 70, no. 11, pp. 923–928, 1993.
- [6] S. Resnikoff, V. C. Lansingh, L. Washburn et al., "Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs?" *British Journal of Ophthalmology*, vol. 104, no. 4, pp. 588–592, 2020.
- [7] I. L. Bailey, M. A. Bullimore, T. W. Raasch, and H. R. Taylor, "Clinical grading and the effects of scaling," *Investigative Ophthalmology & Visual Science*, vol. 32, no. 2, pp. 422–432, 1991.
- [8] D. Ardila, A. P. Kiraly, S. Bharadwaj et al., "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Medicine*, vol. 25, pp. 954–961, 2019.
- [9] C. G. Yedjou, S. S. Tchounwou, R. A. Aló, R. Elhag, B. Mochona, and L. Latinwo, "Application of machine learning algorithms in breast cancer diagnosis and classification," *International Journal of Science Academic Research*, vol. 2, no. 1, pp. 3081–3086, 2021.
- [10] J. H. L. Goh, Z. W. Lim, X. Fang et al., "Artificial intelligence for cataract detection and management," *Asia-Pacific Journal of Ophthalmology*, vol. 9, no. 2, pp. 88–95, 2020.

- [11] 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.
- [12] A. Güven, "Automatic detection of age-related macular degeneration pathologies in retinal fundus images," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 16, no. 4, pp. 425–434, 2013.
- [13] A. M. Abdul-Rahman, T. Molteno, and A. C. B. Molteno, "Fourier analysis of digital retinal images in estimation of cataract severity," *Clinical & Experimental Ophthalmology*, vol. 36, no. 7, pp. 637–645, 2008.
- [14] J. Zheng, L. Guo, L. Peng, J. Li, J. Yang, and Q. Liang, "Fundus image based cataract classification," in *2014 IEEE International Conference on Imaging Systems and Techniques*, pp. 90–94, IEEE, Santorini, Greece, 2014.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097–1105, Curran Associates, Inc., Lake Tahoe Nevada, 2012.
- [16] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.
- [17] L. Cao, H. Li, Y. Zhang, L. Zhang, and L. Xu, "Hierarchical method for cataract grading based on retinal images using improved Haar wavelet," *Information Fusion*, vol. 53, pp. 196–208, 2020.
- [18] M. Yang, J.-J. Yang, Q. Zhang, Y. Niu, and J. Li, "Classification of retinal image for automatic cataract detection," in *2013 IEEE 15th International Conference on e-Health Networking, Applications & Services (Healthcom 2013)*, pp. 674–679, IEEE, Lisbon, 2013.
- [19] L. Guo, J.-J. Yang, L. Peng, J. Li, and Q. Liang, "A computer-aided healthcare system for cataract classification and grading based on fundus image analysis," *Computers in Industry*, vol. 69, pp. 72–80, 2015.
- [20] J.-J. Yang, J. Li, R. Shen et al., "Exploiting ensemble learning for automatic cataract detection and grading," *Computer Methods and Programs in Biomedicine*, vol. 124, pp. 45–57, 2016.
- [21] W. Fan, R. Shen, Q. Zhang, J.-J. Yang, and J. Li, "Principal component analysis based cataract grading and classification," in *2015 17th International Conference on E-Health Networking, Application & Services (HealthCom)*, pp. 459–462, IEEE, Boston, MA, 2015.
- [22] W. Song, Y. Cao, Z. Qiao, Q. Wang, and J.-J. Yang, "An improved semi-supervised learning method on cataract fundus image classification," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, pp. 362–367, IEEE, Milwaukee, WI, USA, 2019.
- [23] M. D. Manchalwar and K. K. Warhade, "Histogram of oriented gradient based automatic detection of eye diseases," in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, pp. 1–5, IEEE, Pune, India, 2017.
- [24] T. Pratap and P. Kokil, "Automatic cataract detection in fundus retinal images using singular value decomposition," in *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pp. 373–377, IEEE, Chennai, India, 2019.
- [25] L. Zhang, J. Li, I. Zhang et al., "Automatic cataract detection and grading using deep convolutional neural network," in *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, pp. 60–65, IEEE, Calabria, Ital, 2017.
- [26] J. Ran, K. Niu, Z. He, H. Zhang, and H. Song, "Cataract detection and grading based on combination of deep convolutional neural network and random forests," in *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pp. 155–159, IEEE, Guiyang, China, 2018.
- [27] A. Imran, J. Li, Y. Pei, F. Akhtar, J.-J. Yang, and Q. Wang, "Cataract detection and grading with retinal images using SOM-RBF neural network," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 2626–2632, IEEE, Xiamen, China, 2019.
- [28] M. A. Syarifah, A. Bustamam, and P. P. Tampubolon, "Cataract classification based on fundus image using an optimized convolution neural network with lookahead optimizer," *AIP Conference Proceedings*, vol. 2296, Article ID 2020034, 2020.
- [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.
- [30] T. Pratap and P. Kokil, "Computer-aided diagnosis of cataract using deep transfer learning," *Biomedical Signal Processing and Control*, vol. 53, Article ID 101533, 2019.
- [31] M. K. Hasan, T. Tanha, M. R. Amin et al., "Cataract disease detection by using transfer learning-based intelligent methods," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 7666365, 11 pages, 2021.
- [32] I. Weni, P. E. P. Utomo, B. F. Hutabarat, and M. Alfalah, "Detection of cataract based on image features using convolutional neural networks," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, pp. 75–86, 2021.
- [33] N. Varma, S. Yadav, and J. K. P. S. Yadav, "Fundus image-based automatic cataract detection and grading system," *AIP Conference Proceedings*, vol. 2724, no. 1, Article ID 030001, 2023.
- [34] N. Varma, S. Yadav, and J. K. P. S. Yadav, "A reliable automatic cataract detection using deep learning," *International Journal of System Assurance Engineering and Management*, vol. 14, pp. 1089–1102, 2023.
- [35] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International Journal of Biomedical Imaging*, vol. 2013, Article ID 154860, 11 pages, 2013.
- [36] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transaction on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.
- [37] E. Decenci ere, X. Zhang, G. Cazuguel et al., "Feedback on a publicly distributed image database: the messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.
- [38] J. Staal, M. D. Abramoff, 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.
- [39] E. J. Carmona, M. Rinc on, J. Garc ia-Feijo o, and J. M. Mart inez-de-la-Casa, "Identification of the optic nerve head with genetic algorithms," *Artificial Intelligence in Medicine*, vol. 43, no. 3, pp. 243–259, 2008.
- [40] P. Porwal, S. Pachade, R. Kamble et al., *Indian Diabetic Retinopathy Image Dataset (IDRID)*, IEEE Dataport, 2018.
- [41] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, IEEE, Lisbon, Portugal, 2016.

- [42] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.
- [43] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC)*, pp. 1–6, IEEE, Mumbai, India, 2015.
- [44] T. Pratap and P. Kokil, "Efficient network selection for computer-aided cataract diagnosis under noisy environment," *Computer Methods and Programs in Biomedicine*, vol. 200, Article ID 105927, 2021.