

Research Article

Diabetic Retinopathy Detection Using Genetic Algorithm-Based CNN Features and Error Correction Output Code SVM Framework Classification Model

Najib Ullah,¹ Muhammad Ismail Mohmand,¹ Kifayat Ullah ,²
Mohammed S. M. Gismalla ,³ Liaqat Ali ,⁴ Shafqat Ullah Khan,⁵ and Niamat Ullah⁶

¹Department of Computer Science, The Brains Institute Peshawar, Pakistan

²Department of Computer and Software Technology, University of Swat, Pakistan

³Department of Electronic and Electrical Engineering, Faculty of Engineering, International University of Africa, Khartoum, Sudan

⁴Department of Electrical Engineering, University of Science and Technology Bannu, Pakistan

⁵Department of Electronics, University of Buner, Pakistan

⁶Department of Computer Science, University of Buner, Pakistan

Correspondence should be addressed to Mohammed S. M. Gismalla; mohsalih@iua.edu.sd

Received 9 March 2022; Revised 25 April 2022; Accepted 22 June 2022; Published 25 July 2022

Academic Editor: Abdul Basit

Copyright © 2022 Najib Ullah et al. 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.

Diabetic retinopathy (DR) is a type of eye disease that may be caused in individuals suffering from diabetes which results in vision loss. DR identification and routine diagnosis is a challenging task and may need several screenings. Early identification of DR has the potential to prevent or delay vision loss. For real-time applications, an automated DR identification approach is required to assist and reduce possible human mistakes. In this research work, we propose a deep neural network and genetic algorithm-based feature selection approach. Five advanced convolutional neural network architectures are used to extract features from the fundus images, i.e., AlexNet, NASNet-Large, VGG-19, Inception V3, and ShuffleNet, followed by genetic algorithm for feature selection and ranking features into high rank (optimal) and lower rank (unsatisfactory). The nonoptimal feature attributes from the training and validation feature vectors are then dropped. Support vector machine- (SVM-) based classification model is used to develop diabetic retinopathy recognition model. The model performance is evaluated using accuracy, precision, recall, and F1 score. The proposed model is tested on three different datasets: the Kaggle dataset, a self-generated custom dataset, and an enhanced custom dataset with 97.9%, 94.76%, and 96.4% accuracy, respectively. In the enhanced custom dataset, data augmentation has been performed due to the smaller size of the dataset and to eliminate the noise in fundus images.

1. Introduction

In telemedicine, automatic classification of ophthalmologic diseases using retinal image analysis has become a routine. Manual segmentation was previously used, but it was difficult, tedious, labor driven, and observer oriented and needed a high skill level. In contrast, computer-assisted detection of ocular disorders is comparatively less expensive, achievable, and purpose oriented and does not require a highly skilled clinician to grade the images. For on time identification and real-time classification of eye illnesses, the advancement

of screening system is required and may be very helpful in the treatment process. There are a number of eye diseases and their causes may also be different. Diabetes is an illness that has become more frequent in recent years, and diabetes may cause eye abnormalities that may result in vision loss. Diabetes is a common disease nowadays, and it may harm the eyes and may result in vision loss [1].

Based on data from the International Diabetes Federation, reported in 9th edition stated that in 2019, diabetes affected around 19 million people in Pakistan which is 148 times higher than their previous report. It is putting them



FIGURE 1: DR classification.

TABLE 1: Stages of diabetic retinopathy [2].

Status	Normal	Nonproliferative diabetic retinopathy			Proliferative diabetic retinopathy
Type of DR	Not seen	Mild DR	Moderate DR	Severe DR	High-risk DR
Retina condition	Healthy	Several little blood vessel swells	The cholesterol is stored as tiny chunk in the veins with visible regions of blood leakage	Blood leakage area may large. Unpredictable beading in the veins. At the optic circle, new blood vessels are formed. Occlusion of a vein	There is a lot of bleeding and fresh blood is forming in the retina. Complete and total blindness

at high risk of life-affecting complications. 8.5 million of these 19 million are still undiagnosed, and they must be vulnerable. Around the globe, approximately 463 million persons are suffering from diabetes. People living with diabetes are always at high risk for diseases associated with diabetes, such as diabetic retinopathy (DR), diabetic macular edema (DME), and glaucoma. Diabetic retinopathy is the most prevalent of all diseases, and it is caused by the damage of blood vessels in the eye retina. Microaneurysms, vitreous hemorrhage, hard exudates, and retinal detachment are only a few of the signs and symptoms of DR, but there might be others.

In Figure 1, we demonstrated different retina images with various diabetic retinopathy levels. The disease begins with minor alterations in the eyes' blood vessels, which is named mild DR. In this case, the patient may be able to overcome the disease and recover completely. If this disease's condition is not managed, it will develop to moderate DR. In the case of moderate DR, blood vessel leakage may start. In the next case, if the disease progresses, it can develop into severe and proliferative DR, which can result in total visual loss.

By 2030, it is estimated that 13 million people in the United States will have DR [1]. If DR is not detected early enough, it can result in a variety of vision problems, including blindness. As a result, a diabetic patient should see an experienced ophthalmologist for an annual or biannual eye checkup and screening.

In addition to identifying DR, its severity degree must be determined in order to be able to cure it. The two primary types of DR are NPDR (nonproliferative DR) and PDR (proliferative DR) [1]. The nonproliferative DR is divided into three phases, which are stated below: The mildest form of DR is mild NPDR, which is followed by moderate and severe NPDR. PDR refers to the most advanced stage of diabetic retinopathy. No DR, mild DR,

moderate DR, severe DR, and proliferating DR are the five stages of DR severity as demonstrated in Table 1. Lesions on fundus photographs that look as little circular shaped red particles at the terminals of the blood vessels are an early indicator of diabetic retinopathy. Microaneurysms, hemorrhages, and/or transudes are signs of moderate DR. In proliferative DR stage, the formation of new blood vessels is occurring, as well as the anomalies mentioned above [3]. Color fundus photographs of a normal retina, as well as varying degrees of DR severity, are shown in Figure 1. A primary challenge with DR recognition is the complication of diagnosing symptoms early on in the course of DR owing to the visual resemblances found among normal fundus photo, moderate DR, and occasionally considerable DR. If diabetic retinopathy developed to an advanced phase, this may cause a loss in vision. To aid medical professionals by identifying DR in real-time, in the literature, several computer-based techniques have been created. Lesions are automatically detected and graded for DR screening, and grading has received a lot of interest from researchers, in these ways to imitate human experts. Blood vessel detection and segmentation in retinal images were created. Since the advent of deep learning algorithms, particularly convolutional neural networks (CNNs), is still a relatively new field of study, several academic communities have already used CNNs for a variety of purposes, including recognizing DR [1]. In the research community, deep learning is widely used for the image classification purpose as it employs neural networks to calculate thousands of mathematical equations with many parameters. Recent DR detection research has generally focused on developing new algorithms for typical fundus images that are predominantly affected by occlusion, refraction, lighting fluctuations, and blur. In this study, we developed a custom dataset and proposed a deep learning approach which works on large dataset and gives better performance.

2. Review of the Previous Work

Various methods to detect DR have been proposed. For multiclass classification, this section focuses on deep learning and neural network algorithms. Fundus images have been divided into two groups by some researchers: diabetic, which encompasses moderate to severe NPDR, and nondiabetic, which signifies the individual does not have DR [4]. Using one primary classifier and backpropagation neural organization processes, the authors suggested a strategy for accurately determining a class where a fundus picture may be classified based on these results. Similarly, a deep learning-based technique for classifying fundus pictures for human ophthalmologist diagnosis has been proposed. Based on Inception V3, the authors developed a Siamese-like CNN binocular model that is able to detect fundus pictures in both the eyes and deliver output from both eyes at the same time [5]. The authors suggested a hybrid strategy for detecting DR [6], in which the deep learning model is aided by histogram (HE) contract limited adaptive histogram (CLACHE). During the diagnosis procedure, the approach uses picture augmentation to bring greater attention and effectiveness. The authors employed five convolutional neural network (CNN) architectures to assess progress indicators for the dataset of diabetic retinopathy patients. Images are divided into three groups depending on the severity of the disease, according to their classification system [7]. The authors presented a novel CNN architecture for diagnosing DR based on ResNet18 in [8]. This method overcomes the issue of significant class imbalance while also generating an area. It is added to the previously mentioned contribution by highlighting semantic areas of the fundus picture to represent the severity degree. The authors developed a strategy for identifying DR based on binary classification, and severity was not a problem. For the datasets under consideration, they used a binary classification system consisting of normal and abnormal photos [2]. The authors of [9] presented a model based on deep learning to categorize a small-scale dataset of diabetic retinopathy images. As an accuracy statistic, they employed Cohen's kappa. In [10], the author designed a CNN-based model and extracted features from a dataset with 30,000 images. While performing preprocessing, they applied denoising techniques and obtained accuracy of 95% and 85% for binary classification and five-class problem of diabetic retinopathy. In [11], a deep learning algorithm has been proposed to grade diabetic retinopathy by modification of GoogLeNet. In the grading process, the authors obtained an accuracy of 81%, while inaccurate prediction achieved 96% accuracy on a custom dataset developed by Jichi Medical University containing 9939 images. In a previous study [12] using 80000 images from the Kaggle dataset, a state-of-the-art DR stage classification technique was constructed. Using complicated DR characteristics such as MAs, HEMs, and exudates on the retina, they employed a CNN architecture with data augmentation to categorize five degrees of DR severity. On a validation set of 5000 samples, a high-end graphical processor unit (GPU) was used, resulting in an accuracy of 75% and a SE of 95%. In the proposed model [13], the authors developed a CNN network for mul-

tistage classification on 128000 fundus images with the technical support of professional ophthalmologists. The developed CNN model obtained an accuracy of 97.5% in the classification of DR stages. The authors in [14] developed a one-of-a-kind approach for classifying fundus photos as normal, moderate, severe, and proliferative DR. The images given as input were preprocessed by morphological operations with disc- and diamond-structuring elements. After this process, six features were obtained considering the perimeter and area of pixels. For the classification, an 8 hidden layer single feedforward neural network was used. 6 units were used as input for each feature value mentioned and 4 units as output, one for each DR level. On the Kaggle dataset and a custom dataset developed by the California Health Care Foundation, the authors employed the Inception V3 architecture for automated classification of diabetic retinopathy in [15]. They got an accuracy of 82% for the Kaggle dataset with a batch size of 64 and 88% for the custom dataset with a batch size of 128. The authors used convolutional neural network-based pretrained Inception V3, which was trained on ImageNet, to perform transfer learning [16] while working on fundus pictures to categorize images into five groups. They got 48.2% accuracy on the EyePACS dataset. The authors of [17] suggested a class coding strategy for predicting and target scores using VGG-D architecture to identify diabetic retinopathy. The design was capable to detect diabetic retinopathy with an accuracy of 82%. The EyePACS dataset was utilized by the authors in [18] to demonstrate the effectiveness of Inception V3 in diagnosing diabetic retinopathy. They also tested AlexNet and VGG16 and got 37.43% and 50.03% accuracy, respectively. Working on a CNN-based architecture, [19] employed 35126 images from the EyePACS dataset to diagnose diabetic retinopathy. Noise removal, normalization, and the use of several hyperparameters resulted in a validation accuracy of 83.68% and a specificity of 93.65%. After using transfer learning and hyperparameter tuning, the authors demonstrated the performance of several pretrained CNN models in the classification of diabetic retinopathy, namely, VGGNet, GoogLeNet, AlexNet, and ResNet. All of these models were tested on 35,126 pictures from the EyePACS dataset. VGG outperforms all of these models in this study. In [20], the authors compared the Inception V4 deep learning system's performance to human graders. They have performed an experiment in Thailand in which 25,326 images were used, resulting in the deep learning approach achieving better accuracy than human graders. This experiment demonstrated that deep learning may be used to identify any. This experiment demonstrated that deep learning may be used to identify any disease. Deep learning lowered the rate of false negatives by 23% while significantly increasing the rate of false positives to 2%. In [21], the authors have developed an ensemble-based model of five CNN models, including Inception V3, ResNet-50, Dense 121, Dense 169, and Xception to classify different diabetic retinopathy severity levels. The model shows performance as 69%, 48%, 65%, 84%, and 51%, respectively. The results show that in this particular study, Dense 169 showed better performance as compared to other models. In [22], the authors presented

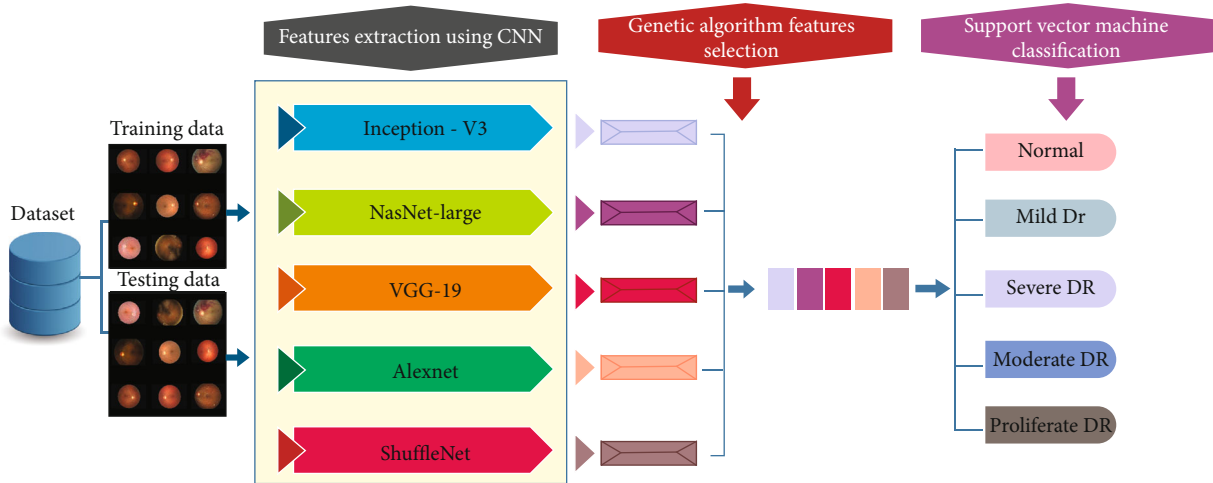


FIGURE 2: Proposed framework.

an AlexNet-based architecture to characterize the severity degree of diabetic retinopathy using appropriate rectified linear activation Units, Pooling, and SoftMax. The team enhanced accuracy across the board by using the Messidor dataset. Researchers often employ the digital images of retina for vessel extraction (DRIVE) and structured analysis of retina (STARE) to segment the vessel network utilizing local and the global vessel features. Different classifiers are also used like support vector machine and K-neighbor, which provide better accuracy, 86% and 55%, respectively, in [23]. They also used pixel clustering to remove the fundus image mask in [23]. Using a Gaussian mixture model (GMM), visual geometry group network (VGGNet), singular value decomposition (SVD), principal component analysis (PCA), and SoftMax, the authors proposed a symmetrically optimal solution for region segmentation, high dimensional feature extraction, feature selection, and fundus image classification in [24]. According to the authors, the VGG-19 model outperformed AlexNet, and the spatial invariant feature transforms in classification accuracy and processing time (SIFT). For people living with diabetes, their body monitoring is a challenging task. In [25], the authors proposed an IoT-based system in which the system uses wearable sensors that will recommend prescriptions and food for patients. In this article, the combination of type-2 fuzzy logic and the fuzzy ontology enhanced the system accuracy. This may be helpful for patients with diabetes. CNN is widely used these days in many computer vision applications. But there are some problems in this system, one of which is the global optimization of CNN training. Training and fast classification play a significant role in the development of CNN. In [26], to increase the convergence and efficiency of CNN training, the authors use the modified resilient backpropagation (MRPROP) technique. To minimize network over-training, a tolerance band is created, which is used with the global best idea for weight update criteria to allow the CNN's training algorithm to optimize its weights more quickly and precisely.

In [27], the author proposed a system that is focused on DR stage classification, but to speed up the training and model convergence, they adopted the lowest learnable parameters. The VGG-16, spatial pyramid pooling layer (SPP), and network-in-network (NiN) are layered to create the VGG-NiN model, which is a highly nonlinear, scale-invariant deep model. Because of the SPP layer's virtue, the suggested VGG-NiN model can process a DR picture at any size. In [28], the authors demonstrated a comparison between different pretrained learning methods and ConvNet, and they also focused on their optimization, for the detection of images and classification. To validate their study, the authors also performed some experiments on a skin detection dataset and a public face dataset that may be able to provide some solutions.

3. Methodology

This section is related to methodology adopted in this study and research framework.

3.1. Framework. The proposed framework represented in Figure 2 defines an architecture of a computer-aided diabetic retinopathy diagnosis system. A benchmark dataset of five classes is loaded to the system, partitioning it into two subdatasets, i.e., trainset and validation. For partitioning the dataset, a hold-out cross-validation mechanism is followed that splits the dataset with a percentage of 70/30, where 70% of the images from five classes are selected randomly for model training along with their labels, while the remaining 30% images with their corresponding labels are selected as the validation set which will be used to evaluate the proposed model's performance. Five advanced convolutional neural network architectures are used to extract robust and noninvariant features from the RGB images, i.e., AlexNet, NASNet-Large, VGG-19, Inception V3, and ShuffleNet, that will perform feature extraction from the training set image and validation set images. A CNN model has two parts, i.e., feature extraction and classification; for the purposed system of feature extraction, an image is input to

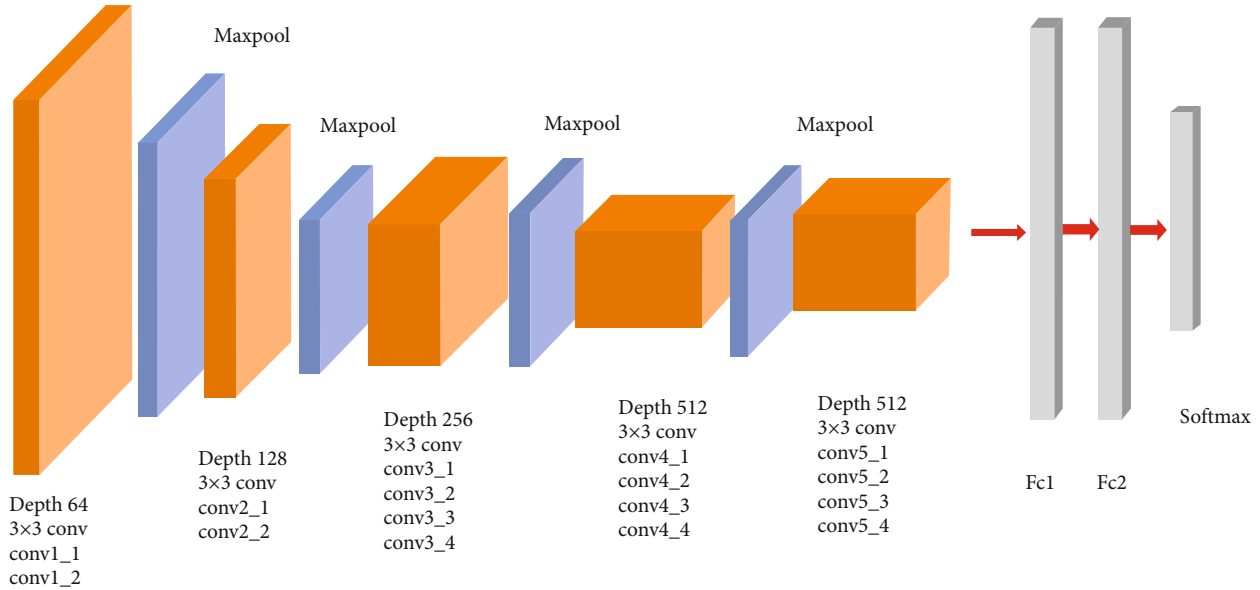


FIGURE 3: The pretrained architecture of VGG 19.

the CNN model, and from the last fully connected layers, a feature vector of 1000 features is extracted. In the proposed feature extraction block, five CNNs are used which will give an output of 5000 images. Feature selection in machine learning is an unsupervised learning approach used to reduce the dimension of the feature vector. The proposed framework genetic algorithm with tournament selection method is used to rank the features into high rank (optimal) and low ranked (unsatisfactory). The nonoptimal feature attributes from the training and validation feature vectors are dropped. Support vector machine classification model is used to develop diabetic retinopathy recognition model. SVM is a binary classifier capable of classifying two classes of data; the benchmark dataset consists of five classes of data, so an error correction output code framework is applied on the SVM classifier to develop a multiclass classifier that can classify more than two classes of data. The SVM model is trained on the training set feature vector and its labels. Upon completion of the model training, the trained model is tested on the validation set features; the model’s predicted labels are compared with the actual labels to create a confusion matrix. Using the information in the confusion matrix, the performance of the model is evaluated using standard classification performance evaluation metrics such as accuracy, precision, recall, and F1 score.

3.2. CNN Architecture. Deep neural networks created on convolutional neural network models are widely used nowadays to deal with challenges in computer vision. In order to distribute the dataset in normal and different categories of diabetic retinopathy patients, CNN-based AlexNet, NASNet-Large, Inception V3, VGG 19, and ShuffleNet models and transfer learning approaches were adopted. The adopted CNN models and the AlexNet, NASNet-Large, Inception V3, VGG 19, and ShuffleNet, along with genetic algorithm for feature selection, are presented systematically in the figure below.

3.2.1. AlexNet. AlexNet is a CNN-based model having approximately 650000 neurons and 60 million parameters. It was introduced by Alex Krizhevsky in 2012. The AlexNet architecture presented in Figure 3 is made up of fully linked layers and one SoftMax tile. Convolutional filters and the ReLU nonlinear transfer function are included in each convolutional layer. A variety of effects may be achieved by blending layers. The input size is changed due to the presence of entirely linked layers. Neural networks include convolutional neural networks, which is a type of neural network. They are made up of neurons that have a weight and bias that can be read. Each neuron receives a number of inputs. It then adds a weight-bearing amount on top of that. Finally, it is transferred by turning, activating, and releasing the device. The figure below demonstrates AlexNet architecture.

3.2.2. Inception V3. Inception, a CNN architecture proposed by Google Inc. in 2014, is used in computer vision and bears a good reputation in this field. The aim of the development of inception deep neural network architecture is to estimate and cover the best local scant formation of a convolutional vision network using complex modules that were already accessible. In December 2015, an improved inception model was proposed by Szegedy that was named Inception V3 presented in Figure 4 which performed better as compared to the benchmark. Inception V3 was used for identification of different objects, human action recognition, classification of video, segmentation, and object tracking. Inception V3 includes smaller convolution factorization, asymmetric convolution factorization, and auxiliary classifiers efficiently reducing the grid size. Inception V3 extends the activate scale of networking filtering before employing maximum or average pooling to avoid a representational bottleneck. Furthermore, deconstruction into smaller convolutions may enhance the network’s accessible space of variations,

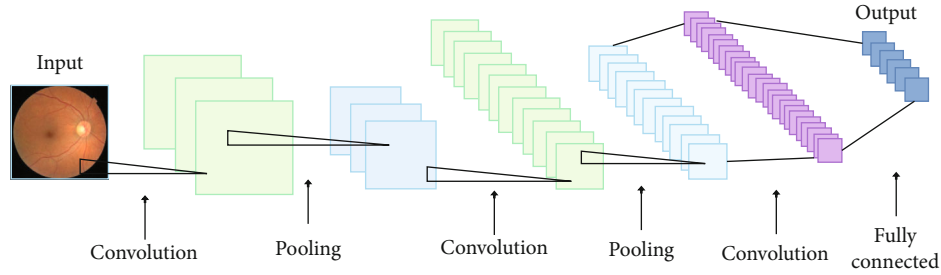


FIGURE 4: The pretrained architecture of Inception V3.

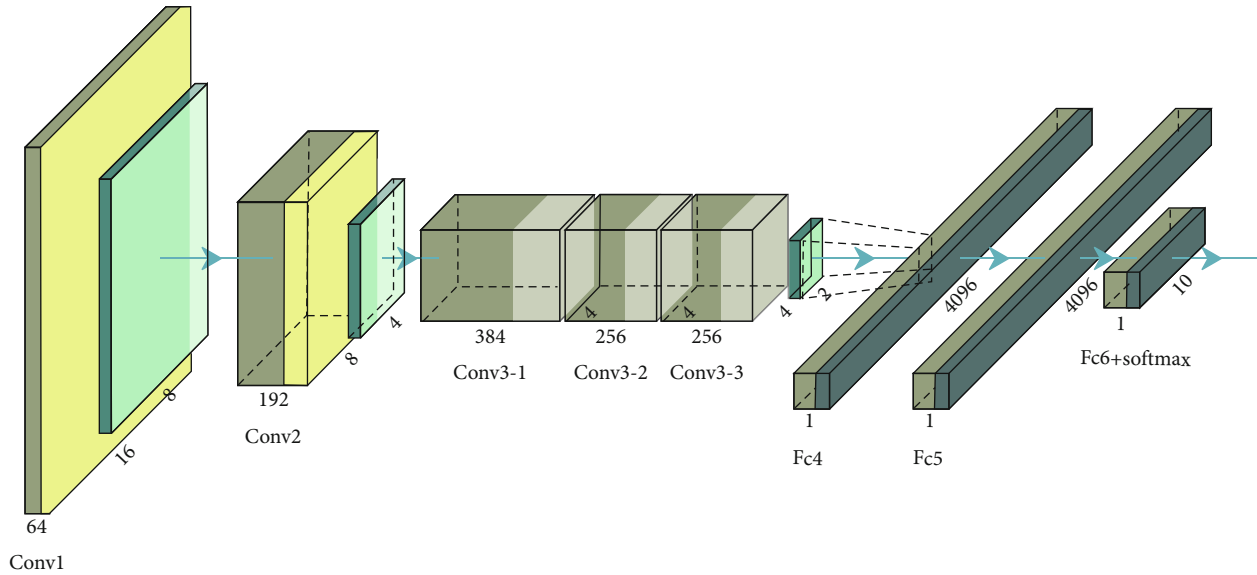


FIGURE 5: The pretrained architecture of AlexNet.

enabling it to provide insight, while the use of auxiliary classifiers allows the network to be accurate. Inception V3 features a more monolithic design and has a lower computation cost. Using the TensorFlow distributed machine learning framework, the Inception V3 model trained networks with a hypostatical pitch.

3.2.3. VGG 19. VGGNet stands for visual geometry group network and multilayered deep neural network. The VGG is a CNN-based network that is applied to the ImageNet dataset. In Figure 5, we presented that VVG 19 is a very simple network and is very useful due to its simple nature. The 3×3 convolutional layers are placed on the top to extend the depth level. In VGG 19, handle max-pooling stages are employed to minimize volume size. VGG 19 is commonly used in medical imaging studies because of its versatility and simplicity.

3.2.4. NASNet-Large. Although previous ANNs have all provided us with unique and significant insights and have impacted the future of ANN design, their particular architectural design was established by humans, influenced by their experiences and biases. Google launched NASNet-Large presented in Figure 6, which framed finding the ideal CNN architecture as a reinforcement learning problem.

NASNet has plenty of computing power. The basic concept was to find the optimal combination of filter sizes, output channels, strides, number of layers, and other characteristics in the specified search space. The accuracy of the searched architecture on the provided dataset was the reward for each search activity in this reinforcement learning environment. NASNet achieves extraordinary better results on ImageNet and also performs well on large datasets.

3.2.5. ShuffleNet. Figure 7 shows the ShuffleNet architecture which is one among computationally efficient CNN architectures that was proposed in 2017 by the member of the research group of Megvii Inc. It was specifically designed for a mobile device with limited computational power, and the ShuffleNet was designed with fewer parameters. ShuffleNet uses two new operations to speed up computations: convolution of points in a group and path shuffling. If we reduce the network size, there will be no loss in the accuracy of the model, but it even performs better.

3.3. Genetic Algorithm-Based Feature Selection. A genetic algorithm is an optimization method used up to the mark in deep learning research. The GA model is an evolutionary search strategy that emulates nature's selection, mutation, and crossover mechanisms. Choosing the most trustworthy

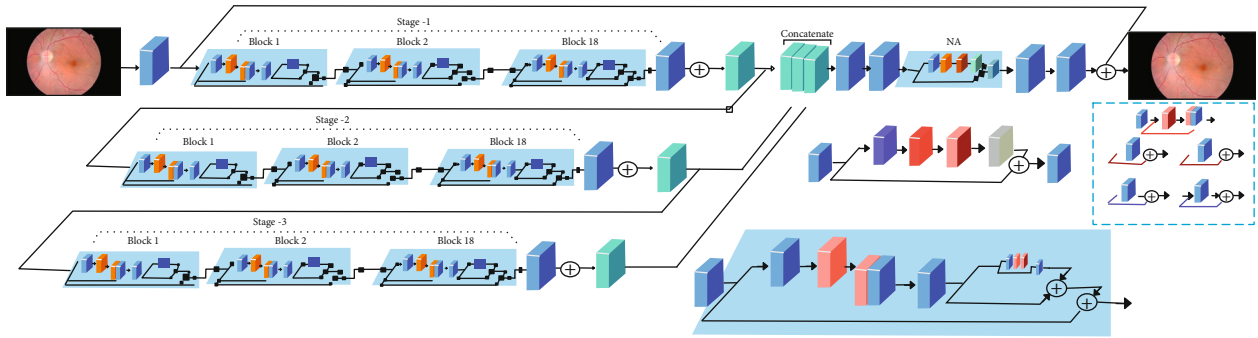


FIGURE 6: Pretrained architecture of NASNet-Large.

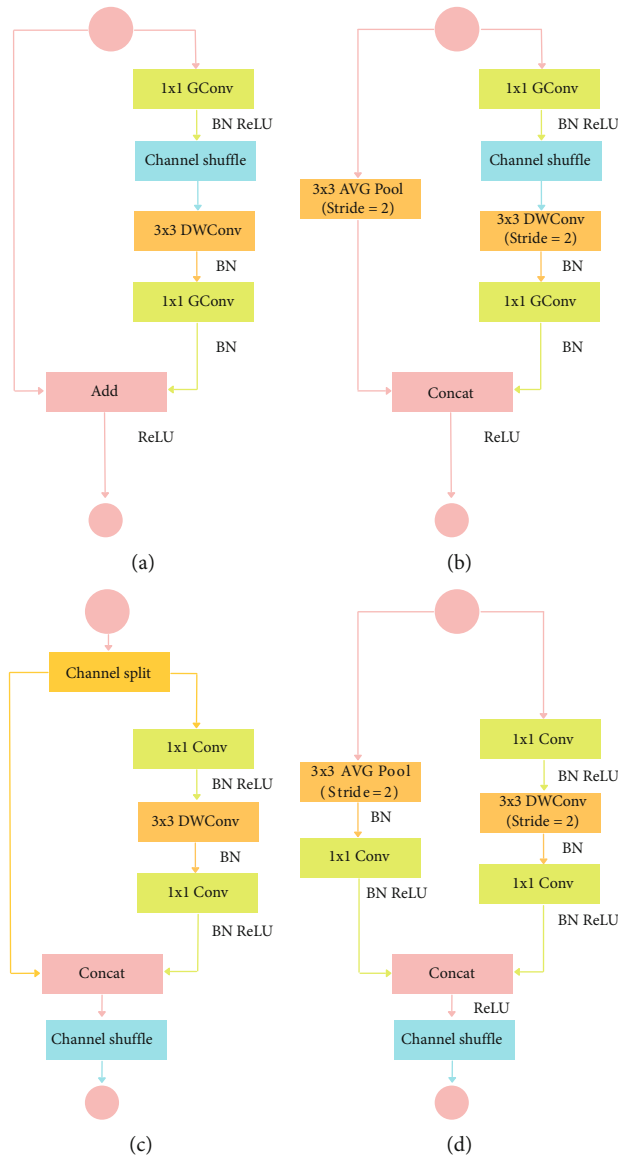


FIGURE 7: The pretrained ShuffleNet.

and discriminative characteristics that decrease the high dimension of feature space to a minimum is feature selection. GA is a metaheuristic feature selection approach that

begins the search and finds many solutions to the problem. GA is an optimizer, which means it will choose the best solution from a list of options. In this particular research case,

```

Z,M Vbest ← ϕ → initialize
P0 ← Gaussian distribution at random with σ =0.3 and μ=0.6
All genes are discretized into binary values.
    {0, Pi < 0.5}
As a result
    {1, otherwise}
While m ≤ M do
m ← m +1.
GA (Pm)
If argmaxm (Pm) ≥ then Vbest ← argmaxm (Pm)
End while
Return best P → Return the most fit one

```

PSEUDOCODE 1: Pseudocode of feature selection method utilizing GA.

5000 features have been extracted in which 800 are optimal, which is indeed a large number that may affect computation cost in terms of module's training time and may affect the classifier accuracy result. It is also possible that due to irrelevant and redundant features, set overfitting maybe found. If that set exists, it is vital to eliminate it. To achieve this goal, we employ an effective genetic algorithm approach as supported by experimental results. In the proposed system, generation of random uniform population is done in the first step. The crossover and probability mutation for every individual generation are 0.5. The model begins by generating a random population. The chromosomes are made up of a number of gene characteristics, each of which has a real number. Below algorithm is the representation of chromosomes. The chromosomal representation is shown in the equation below.

$$Z = \{ \sim \text{Hi} | \sim \text{Hi} \in \{0, 1\} \} \text{ n } i = 1. \quad (1)$$

3.4. Multiclass SVMs. VMs have also been successfully used in practice, particularly for classification problems, and are a hot subject in the ML field right now. FVapnik and his colleagues proposed a unique statistical learning theory-based machine learning method. SVMs use structural risk minimization and are different from other risk models which use empirical risk minimization. This is theoretically sound and may be used in various situations. In SVM, the support vectors (a subset of the training points) must be examined to find the best separating hyperplane. SVMs may achieve high classification accuracy by selecting the optimum separation hyperplane even with minimal training sets.

Given a dataset,

$$\{(x_i, y_i), x_i \in R^n y_i \in \{-1 + 1\}, i = 1, \dots, N\}. \quad (2)$$

As mentioned in the expression, the binary classification

fundamental issue may be asked.

$$\begin{aligned} \min_{m,q,\epsilon} F &= \frac{1}{2} \|m\|^2 + C \sum_{i=1}^N \epsilon_i, \\ \text{s.t. } y_i (M^T \varnothing(x_i) + q) &\geq 1 - \epsilon_i, i = 1, \dots, N, \\ \epsilon_i &\geq 0, i = \dots, N \end{aligned} \quad (3)$$

where $y_i (M^T \varnothing(x_i) + q) \geq 1$ comprises first the given constraints $(M^T \varnothing(x_i) + q) \geq 1$ if $y_i + 1 (M^T \varnothing(x_i) + q) \leq 1$ if $y_i = -1$ and ϵ_i are slack variables in the inequalities set that enable misclassifications; the punishment factor is $C > 0$. Expression (4) states its dual problem.

$$\min_{\partial} \frac{1}{2} \partial^T Q \partial - e^T \partial. \quad (4)$$

Subject to $y^T \partial = 0$,

$$0 \leq \partial_i \leq C, i = 1, \dots, i, \quad (5)$$

where e is vector of all ones, Q is an l by l positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K((x_i, x_j) \equiv \varnothing(x_i)^T \varnothing(x_j)$ The preparation vectors are transferred into upper level Euclidean space by the function, and the input vector x may be described as in this high-ranking dimensional showcase space. Equation (6) provides the decision function.

$$\text{Sgn} \left(\sum_{i=1}^l y_i \partial_i K(x_i, x) + q \right). \quad (6)$$

The aim of the development of support vector machines was a binary classification. It is still a work in progress to figure out how to make it work for multiclass categorization. In most cases, there are a few different techniques to multiclass SVM.

The essential purpose of the development of support vector machines was binary classification. Its extension for multiclass classification is not yet achieved, however still under research. For multiclass SVM, there are a few options.

- (i) One vs. all
- (ii) One vs. one
- (iii) A graph with a direct acyclic path
- (iv) ECO codes

3.5. Error-Correcting Output Codes. Error-correcting output code (ECOC) is a technique for breaking down a multiclass classification issue into a set of binary classifier subproblems. Sejnowski and Rosenberg (1987) presented the approach error-correcting output coding in their well-known system NET talk.

We can solve the problem by linking each of the classes with a row of an $A * B$ "Coding Matrix"; we can provide entries in the range from $\{-1, 0, +1\}$. Each class will be assigned string B that is unique in nature. Here from B , we mean the binary classification problem numbers, we are supposed to construct. In the matrix, each column expresses a difference between two classes "1" and "+1," while class "0" will be ignored. The classes having "-1" and "+1" will communicate with our "li:" and "li+," respectively. These strings will be referred to as "Codewords." Then, B binary functions are learned for each bit position in these binary strings, one for each bit position. The expected outputs of these B binary functions are provided by the codeword for class i during training, for example, from class i . Each n binary function is calculated to yield a B binary sequence, which is then used to classify new X values. The generated strings are compared to the K features, and X is assigned to that class whose features are near to the newly generated strings, according to some distance measure. Coding is a crucial subject to take. Two qualities should be included in a suitable ECO code for a problem that belongs to the K class: row and column separation. The ability of a code to fix mistakes is proportional to the number of rows separated. If d is the least hamming distance, then at least code can be corrected to a single bit error $\lfloor (d-1)/2 \rfloor$.

3.6. Major Contribution. The main contributions of this research study include the following.

- (i) Firstly, a new feature extraction block is developed, using five state-of-the-art CNN models to extract features from the fundus images. Genetic algorithm is used for optimal feature selection extracted through developed block
- (ii) An error correction output code framework is applied on the SVM classifier to develop a multiclass classifier that can classify more than two classes of data
- (iii) A custom dataset is developed by collecting fundus images from a local hospital. The dataset contains five classes of fundus images of healthy persons and those with mild DR, moderate DR, severe DR, and proliferative DR
- (iv) Finally, we carry out experiments on three dataset: Kaggle dataset, custom dataset, and augmented cus-

tom dataset. The proposed model performed well on the enhanced custom dataset as compared to the Kaggle dataset and the custom dataset

3.7. Performance Matrices. All the already discussed CNN models, genetic algorithm, and SVM classifier are used in our experiment. We consider five indicators to measure the performance of our proposed system. All these values were determined using some basic terms from the confusion matrix, and they are true negative (TN), false negative (FN), true positive (TP), and false positive (FP).

Therefore, the following equations give the indicators values:

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{FP} + \text{TN} + \text{TP} + \text{FN}}, \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}, \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}, \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}}, \quad (10)$$

$$\text{F1 Score} = 2 \times \left\{ \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right\}. \quad (11)$$

The accuracy of the classification model is determined by several factors, as shown in Equation (7). Furthermore, as shown in Equation (8), the rate of sensitivity reflects a classifier's capacity to build the focused class correctly. Similarly, as indicated in Equation (9), the specificity rate demonstrates a classifier's separation capabilities. The precision rate evaluates the accuracy with which a class is determined as shown in Equation (10). Equation (11) defines F1 score as the harmonic mean sensitivity (recall) and accuracy value. In this study, as a result, all related evaluation parameters for CNNs were calculated. Based on the parameters mentioned above, the results are presented in the next section.

4. Results and Implementation

This section provides the description of the dataset utilized in this study, experimental setup, and obtained results.

4.1. Dataset. In our research work, the performance of our proposed model is evaluated using three datasets; the first dataset is an eight GB labeled dataset taken from the Kaggle repository, while the second dataset is a custom dataset. The custom dataset contains retina fundus images to classify different levels of diabetic retinopathy. The fundus photos for custom dataset development are obtained from local hospital which are captured on digital fundus camera, i.e., Canon CX-1. Both left and right eye samples of the patients are provided. Each sample is scaled with 0 to 4 for each DR class by the ophthalmologist. A level of 0 indicates the healthy case, while ratings 1, 2, 3, and 4 show mild, moderate, severe, and proliferative DR, respectively. These scales are employed as labels in our proposed model.

The proposed approach differs from the existing system, which distributed fundus images based on the clinical and pathological changes in the retina. Furthermore, we consider the clinical practice, which is, we distributed the fundus image of abnormalities and the treatment methods. Pictures are divided and placed in different folders according to their tags. The process of dividing the images into different categories is done under the guidelines of an expert ophthalmologist. After that, the images are cropped, and the key features are separated. A filtering technique is also used to equalize and contrast the image modification. The third dataset is augmented or enhanced custom dataset. Due to less image in custom dataset, data augmentation is performed. Data augmentation is a technique for increasing the range of data available that has been adopted. After that, flipping, cropping, rotating, zooming, and padding are done.

4.2. Experimental Setup. We developed a novel approach by incorporating CNN models, genetic algorithm, and ECOC multiclass SVM to detect diabetic retinopathy. The architectures used for features learning are AlexNet, Inception V3, VGG-19, NASNet-Large, and ShuffleNet. The features from each CNN model are extracted using the last fully connected layer with a dimension of 1×1000 per image. The genetic algorithm is used to select the optimal features. The SVM model hyperparameters (C value, kernel, filter type, and optimizer) are tuned to select the optimal hyperparameters for classifying DR images into their corresponding classes. Therefore, the parameters are defined using the training results for the improvement of performance. We utilized the value of C as 1, kernel type was RBF, and filter type was feature normalization. We used MATLAB deep learning framework for feature extraction. In contrast, the SVM classifier used is part of the CNN models' statistics and machine learning toolbox. All simulations are performed on HP Z440 work with an Intel Xeon processor consisting of 48 GB of RAM and 8 GB of NVIDIA RTX 2070 SUPER GPU.

When conducting experiments, several performance metrics were used to evaluate the performance of the methodology proposed in this paper, including accuracy, precision, recall, and F1 score. An important GA-based feature selection method is also included, which selects the optimal features while also reducing the dimensions of the feature vector and facilitating efficient model training and development. In the experiment, it is necessary to distinguish between three types of datasets: the benchmark dataset, the customized dataset, and the customized enhanced dataset (all of which are customized). The datasets has been divided into two parts: a trainset and a test set, with a 70/30 split between the two parts to make model validation and parameter tuning more straightforward. To train the classifier, it must be fed with information about the trainset's features as well as the labels that have been assigned to them. The classification model is SVM with an error correction output code intended to be used in error correction. Following the completion of model training, the trained model is evaluated against the validation set, which comprises the reduced feature vector containing only optimal features and the actual validation set, as soon as possible after the training is completed. Confusion matrixes are calculated by comparing predicted labels

from an SVM classifier to the existing labels vector generated by the SVM classifier to calculate the confusion matrixes (see Figure 1). The confusion matrixes contain the rates of four different classifications, namely, true positive (TP), false positive (FP), false negative (FN), and true negative (TN), which are used to calculate the models' accuracy, precision, recall, and F1 score. The experiments are conducted using three sets of data. One is the benchmark Kaggle dataset, the second is a self-developed customized dataset, and the third is an augmented version of the custom dataset. The proposed model shows better results on the Kaggle dataset. On the custom dataset, due to fewer images, the model performance was comparatively lower. To enhance the model's performance in the custom dataset, we performed augmentation for the custom dataset and applied the model on new data that showed improved results. Figure 8 shows the confusion matrices of the experiments we performed on the three different datasets: the Kaggle dataset, the custom dataset, and the augmented custom dataset. The experimental results showed that the proposed method can effectively classify diabetic retinopathy.

For the comparison of the results of the experiments performed on different datasets, accuracy, precision, recall, receiver operating characteristic (ROC) curves, and the area under the ROC are used. Table 2 presents the accuracy, precision, recall, and F1 score of the model for the three datasets. The model achieves a higher accuracy of 97.9% when trained and tested on the Kaggle dataset. On the self-developed custom dataset, the model achieves comparatively lesser accuracy of 94.76% since there are less photos in the dataset.

The researchers performed data augmentation for the custom dataset and developed another version of the custom dataset, the augmented custom dataset. We trained and tested the model on the augmented custom dataset and achieved a comparatively higher accuracy of 96.4%. The model achieves a higher precision rate on the augmented custom dataset and the custom dataset. It is observed that precision is low on the Kaggle dataset.

Figure 9 shows the ROC curves, and from which, it can be appreciated that the performance of model over Kaggle dataset is better than the other two datasets. It is also observed that the augmented custom dataset performed well as compared to custom dataset. The model performance has been shown in Figure 10 in terms of accuracy, precision, recall and F -measure.

4.3. Discussion. The proposed methodology in this paper was evaluated using five-class datasets, a benchmark dataset, a custom dataset, and an enhanced custom dataset. The dataset was split into two sections: training and validation. Parts use a cross-validation mechanism with a percentage of 70/30; 70% of the images and labels were used for training, while the remaining 30% of mages and labels were used for validation. Five pretrained CNN models were used for feature extraction from the images. We used five CNN models for feature extraction, and we got 5000 images. To make the feature vector's dimension smaller, feature selection was used, and for this purpose, the genetic algorithm with tournament selection methods is

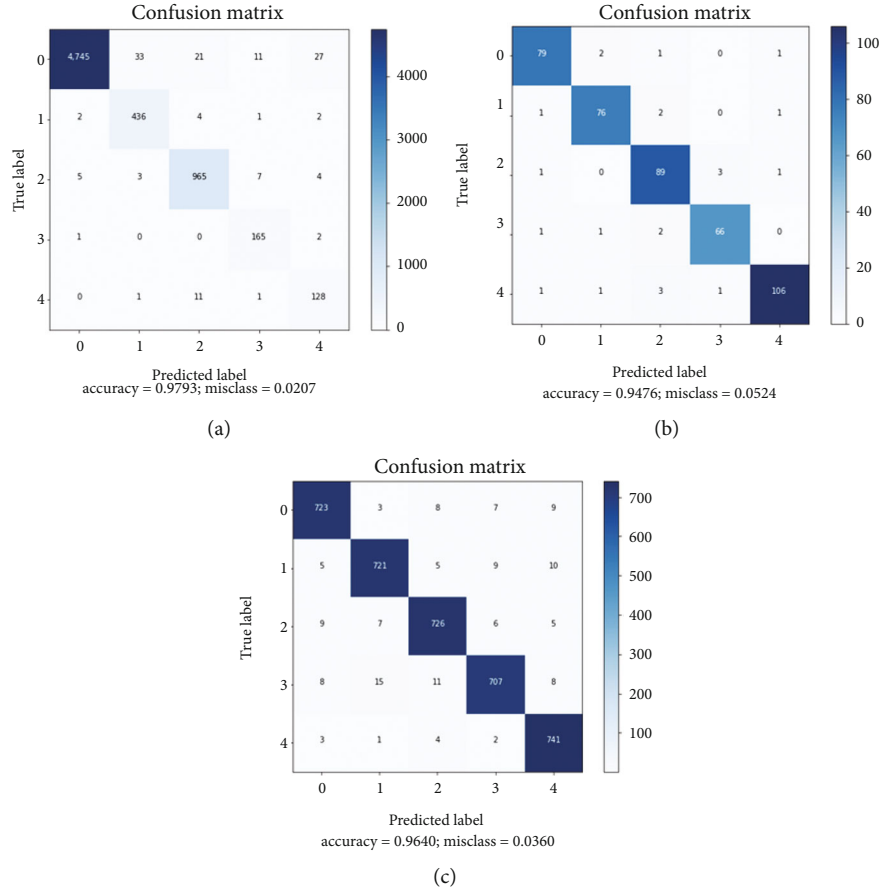


FIGURE 8: Confusion matrix of the proposed model on different datasets.

TABLE 2: Results obtained on the proposed model.

DS	Class	Accuracy	Precision	Recall	F-measure
Kaggle	Normal class		0.9983	0.981	0.9896
	Mild DR class		0.9218	0.9796	0.9499
	Moderate DR class		0.964	0.9807	0.9723
	Severe DR class		0.8919	0.9821	0.9348
	Proliferative DR class	97.9	0.7853	0.9078	0.8421
Custom	Normal class		0.9518	0.9518	0.9518
	Mild DR class		0.95	0.95	0.95
	Moderate DR class		0.9175	0.9468	0.9319
	Severe DR class		0.9429	0.9429	0.9429
Augmented custom	Proliferative DR class	94.76	0.9725	0.9464	0.9593
	Normal class		0.9666	0.964	0.9653
	Mild DR class		0.9652	0.9613	0.9633
	Moderate DR class		0.9629	0.9641	0.9635
	Severe DR class		0.9672	0.9439	0.9554
	Proliferative DR class	96.4	0.9586	0.9867	0.9724

used to rank the features. Support vector machine classification model is used to classify images in classes. All three datasets, the benchmark dataset, the custom dataset, and the enhanced custom dataset, have file classes, i.e., normal,

mild DR, moderate DR, severe DR, and proliferative DR. The method begins with a multiclass dataset being evaluated for accuracy. The average classification accuracy is then evaluated using hold-out cross-validation. Individual

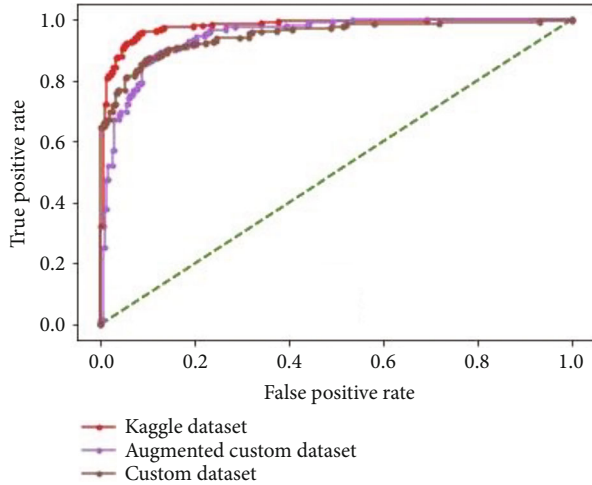


FIGURE 9: Performance demonstration through ROC.

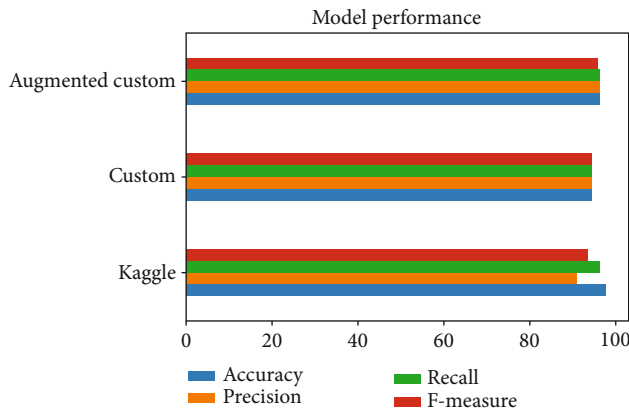


FIGURE 10: Graphical representation of performance.

accuracy data are used to determine the average accuracy value. The findings show how to use the SVM classification approach to explore CNN features using feature extraction on pretrained CNN networks. The SVM classifies the images and returns a confusion matrix for each severity level of the disease. The most accurate dataset is the Benchmark dataset which achieved an accuracy of 97.9% for normalized training feature vector. The model's accuracy increases by incorporating the evolutionary search method. Our unique data package achieves the highest precision of 0.9641 on the augmented custom dataset and the lowest precision of 0.9122 on the Kaggle dataset. In [28], the author gives a thorough examination of the link between ConvNet and various pretrained learning approaches, as well as the consequences of optimization. These hybrid networks improve on state-of-the-art algorithms for image, audio, text, and video identification, classification, and detection. ConvNet has also been used in computer vision for some task-specific applications. To validate the survey, the authors also performed some experiments on a public face and skin detection dataset to provide an authentic solution.

4.4. Conclusion. The recognition of the five stages of severity level of DR using fundus images is a challenging task. Most of the existing studies worked on only the binary classification of DR with significantly good results. This research has presented a deep learning-based model for feature extraction using five pretrained deep architectures and GA for optimal feature selection. The performance of model is tested on three datasets: the Kaggle dataset, a custom dataset, and enhanced custom dataset. The proposed approach achieved an accuracy of 97.9% on the Kaggle dataset and 96.4% accuracy on the enhanced custom dataset with five classes. Our suggested approach performed better on the Kaggle dataset and had poorer outcomes on the custom dataset across all three datasets. The outcomes of this study can help doctors and researchers make better clinical decisions. There are a few flaws in our research that can be addressed in the future studies. A more thorough examination is required, which requires the collection of more patient data. Future research should focus on distinguishing between individuals with normal symptoms and those with nonproliferative symptoms in terms of accuracy. The nonproliferative symptoms may not be seen at all or may not be visible at all on retina imaging. Another possibility is to expand the scope of the presented approach to larger datasets. Other medical issues, such as cancer and tumors, may be addressed.

Data Availability

Data will be available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017.
- [2] M. Chetoui, M. A. Akhloufi, and M. Kardouchi, "Diabetic retinopathy detection using machine learning and texture features," in *Proceedings of the 2018 IEEE Canadian Conference on Electrical Computer Engineering (CCECE)*, vol. 13–16, pp. 1–4, Quebec, QC, Canada, 2018.
- [3] M. Mateen, J. Wen, M. Hassan, N. Nasrullah, S. Sun, and S. Hayat, "Automatic detection of diabetic retinopathy: a review on datasets, methods and evaluation metrics," *IEEE Access*, vol. 8, pp. 48784–48811, 2020.
- [4] D. K. Prasad, L. Vibha, and K. R. Venugopal, "Early detection of diabetic retinopathy from digital retinal fundus images," in *Proceedings of the 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, pp. 240–245, Trivandrum, India, 10–12 December 2015.
- [5] X. Zeng, H. Chen, Y. Luo, and W. Ye, "Automated diabetic retinopathy detection based on binocular Siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744–30753, 2019.

- [6] D. J. Hemanth, O. Deperlioglu, and U. Kose, "An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network," *Neural Computing and Applications*, vol. 32, no. 3, pp. 707–721, 2020.
- [7] Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi, and J. Zhong, "Diagnosis of diabetic retinopathy using deep neural networks," *IEEE Access*, vol. 7, pp. 3360–3370, 2019.
- [8] P. Junjun, Y. Zhifan, S. Dong, and Q. Hong, "Diabetic retinopathy detection based on deep convolutional neural networks for localization of discriminative regions," in *Proceedings of the 2018 International Conference on Virtual Reality and Visualization (ICVRV)*, vol. 22–24, pp. 46–52, Qingdao, China, October 2018.
- [9] A. Samanta, A. Saha, S. C. Satapathy, S. L. Fernandes, and Y. D. Zhang, "Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset," *Pattern Recognition Letters*, vol. 135, pp. 293–298, 2020.
- [10] R. Ghosh, "Automatic detection and classification of diabetic retinopathy stages using CNN," in *Proceedings of the 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, Delhi-NCR, India, 2–3 February 2017.
- [11] H. Takahashi, H. Tampo, Y. Arai, Y. Inoue, and H. Kawashima, "Applying artificial intelligence to disease staging: deep learning for improved staging of diabetic retinopathy," *PLoS One*, vol. 12, no. 6, article e0179790, 2017.
- [12] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia computer science*, vol. 90, pp. 200–205, 2016.
- [13] V. Gulshan, L. Peng, M. Coram 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.
- [14] W. L. Yun, U. Rajendra Acharya, Y. Venkatesh, C. Chee, L. C. Min, and E. Ng, "Identification of different stages of diabetic retinopathy using retinal optical images," *Information Sciences*, vol. 178, no. 1, pp. 106–121, 2008.
- [15] Y. S. Kanungo, "Detecting diabetic retinopathy using deep learning," in *Proceedings of the 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, vol. 19–20, pp. 801–804, Bangalore, India, 2017.
- [16] S. Masood and T. Luthra, "Identification of Diabetic Retinopathy in Eye Images Using Transfer Learning," in *Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA)*, Greater Noida, India, 5–6 May 2017.
- [17] A. Kwasigroch, B. Jarzembinski, and M. Grochowski, "Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy," in *Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Swinoujscie, Poland, 9–12 May 2018.
- [18] X. Wang, Y. Lu, Y. Wang, and W. B. Chen, "Diabetic retinopathy stage classification using convolutional neural networks," in *Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI)*, Salt Lake City, UT, USA, 6–9 July 2018.
- [19] G. Garc, J. Gallardo, A. Mauricio, L. Jorge, and C. D. Carpio, "Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images," in *International Conference on Artificial Neural Networks*; Springer, pp. 1–8, Berlin/Heidelberg, Germany, 2017.
- [20] P. Ruamviboonsuk, J. Krause, P. Chotcomwongse et al., "Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program," *NPJ digital medicine*, vol. 2, no. 1, pp. 1–9, 2019.
- [21] S. Qummar, F. G. Khan, S. Shah et al., "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019.
- [22] I. Ardiyanto, H. A. Nugroho, R. Lestari, and B. Buana, "Deep learning-based diabetic retinopathy assessment on embedded system," in *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1760–1763, Jeju Island, Korea, 11–15 July 2017.
- [23] Grissworld Home Care, "The 4 stages of diabetic retinopathy: what you can expect," 2019. Available online: <https://www.griswoldhomecare.com/blog/2015/january/the-4-stages-of-diabetic-retinopathy-what-you-ca/> (accessed on 13 December 2021).
- [24] M. Mateen, J. Wen, S. Song, and Z. Huang, "Fundus image classification using VGG-19 architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, p. 1, 2019.
- [25] F. Ali, S. R. Islam, D. Kwak et al., "Type-2 fuzzy ontology-aided recommendation systems for IoT-based healthcare," *Computer Communications*, vol. 119, pp. 138–155, 2018.
- [26] S. U. Rehman, S. Tu, O. U. Rehman, Y. Huang, C. M. S. Magurawalage, and C. C. Chang, "Optimization of CNN through novel training strategy for visual classification problems," *Entropy*, vol. 20, no. 4, p. 290, 2018.
- [27] Z. Khan, F. G. Khan, A. Khan et al., "Diabetic retinopathy detection using VGG-NIN a deep learning architecture," *IEEE Access*, vol. 9, pp. 61408–61416, 2021.
- [28] S. Rehman, S. Tu, M. Waqas et al., "Unsupervised pre-trained filter learning approach for efficient convolution neural network," *Neurocomputing*, vol. 365, pp. 171–190, 2019.