IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 October 2023, accepted 22 October 2023, date of publication 26 October 2023, date of current version 2 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327900

RESEARCH ARTICLE

Ensembled Deep Convolutional Generative Adversarial Network for Grading Imbalanced Diabetic Retinopathy Recognition

HUMA NAZ[1,4], RAHUL NIJHAWAN[2], NEELU JYOTHI AHUJA[1], (Senior Member, IEEE), SHAHA AL-OTAIBI[3], (Member, IEEE), TANZILA SABA[4], (Senior Member, IEEE), SAEED ALI BAHAJ[5,6], AND AMJAD REHMAN[4], (Senior Member, IEEE)

[1]Department of Computer Science, University of Petroleum and Energy Studies, Dehradun 248007, India
[2]Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India
[3]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[4]Artificial Intelligence and Data Analytics Laboratory, Prince Sultan University, Riyadh 11586, Saudi Arabia
[5]MIS Department, College of Business Administration, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[6]Department of Computer Engineering, College of Engineering and Petroleum, Hadhramaut University, Mukalla, Hadhramout, Yemen

Corresponding author: Saeed Ali Bahaj (bahajsaeedali@gmail.com) and Amjad Rehman (drrehman70@gmail.com)

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are also thankful to AIDA Lab CCIS Prince Sultan University, Riyadh Saudi Arabia for support.

ABSTRACT Diabetic Retinopathy (DR) is one of the leading causes of blindness and vision loss worldwide. According to the International Diabetes Federation (IDF), approximately one-third of individuals with diabetes, equivalent to 32.2%, are affected by some form of DR. Due to the uneven data distribution, intra-class variance, and a dearth of ophthalmologists, DR diagnosis is considered challenging. In recent years, Convolutional Neural Networks (CNN) and supervised learning techniques have been potentially useful in computer vision applications. However, unsupervised CNN has received less attention. Moreover, it is more manageable to use synthetic images for model training with the recent advancements in graphics. Therefore, the proposed method combines the actual and augmented views using the Deep Convolutional Generative Adversarial Network (DCGAN) algorithm. The generated views are implemented to balance the minority class in the imbalanced dataset. Furthermore, a novel ensemble convolutional neural network algorithm named Different View Ensemble (DVE) that merges the weighted average prediction of CNN, CNN-i, and CNN+i algorithms has been proposed. The proposed algorithm is evaluated on the DDR and EyePACS datasets, and its performance is compared with K-Means, Fuzzy C-Means (FCM), and Autoencoder-based Deep Embedded Clustering Techniques (DEC). The results demonstrate the superiority of the proposed algorithm, achieving an accuracy rate of 97.4%, specificity of 99.6%, and sensitivity of 92.3%. The promising results underscore the potential impact of this methodology in enhancing the accuracy and reliability of automated diagnostic systems in the field of ophthalmology. Notably, the evaluation considers imbalanced data and a DCGAN-balanced dataset, where the proposed approach exhibits even better performance with balanced classes.

INDEX TERMS Diabetic retinopathy detection, imbalance data, ensembled GAN, healthcare, health risks.

I. INTRODUCTION

DR is a global epidemic and a complication of diabetes that affects the eyes. It is projected that by 2045, the number of people with diabetes will reach 642 million, and approximately 35% (242 million) of them will experience DR. Additionally, it is estimated that by 2045, around 11% (70 million) will have severe sight-threatening retinopathy [1]. The prevalence of DR has shown a significant increase of 25% between 1991 and 2015, surpassing other

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

causes of visual disability [2]. It is the most common symptom among one in three individuals with diabetes, impacting blood flow in the retina.

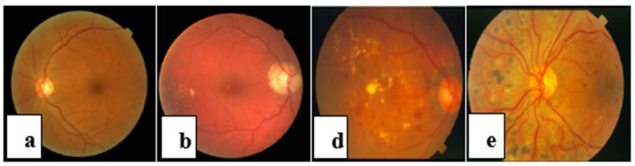

FIGURE 1. Fundus images and abnormal DR grades (a) Mild NPDR (b) Moderate NPDR (c) Severity NPDR (d) PDR [1].

The major pathological features of DR include Microaneurysm (MA), Exudates (EX), and Hemorrhages (HM). DR is broadly categorized into five stages based on the presence of these lesions in the eye fundus: No DR, mild DR, moderate DR, severe DR, and Proliferative DR (PDR), as depicted in Figure 1.

The challenges of the limited availability of ophthalmologists in diagnosing DR patients and the need for automation through computer-aided detection have led to the exploration of deep learning techniques in the medical domain [3], [4]. However, the application of deep learning in this context is hindered by the availability of small or insufficiently annotated sample datasets [5]. This challenge is particularly relevant in the context of imbalanced datasets, where the unequal distribution of training examples can result in biased results favoring the majority class and high misclassification rates for the minority class [6]. Therefore, addressing the issue of imbalanced datasets becomes crucial for achieving accurate and reliable automated DR diagnosis. Imbalanced datasets are prevalent in real-world applications such as disease detection [2], identification of high malignancy diseases [3], rare class emotion prediction, and software quality prediction [4]. Traditional classifiers tend to have accuracy variations, favoring the majority class while performing poorly for the minority class, leading to biased results and high misclassification rates for the minority class.

Addressing the issue of imbalanced class distribution in classification tasks has gained significant attention in recent years. Researchers have explored various approaches and techniques to mitigate the impact of skewed distributions and improve classification performance, as highlighted in the comprehensive review by Kumar et al. [10]. Approaches such as oversampling, under-sampling, two-phase, and cost-sensitive learning can be utilized to address the issue of imbalance classes in DR datasets. Additionally, augmenting images is an effective solution to improve Machine Learning (ML) model training by increasing dataset size and diversity [5], [6], [7].

Image synthesis techniques such as resizing, adding noise, cropping, and color manipulation can enhance model performance, but introducing new data for augmentation is crucial to improve model generalizability in the context of DR. GAN, a data augmentation technique, use a two-model architecture consisting of a discriminator and generator to generate synthetic data points that mimic real-image-like feature distribution and achieve NASH equilibrium (Fig. 2). The generator produces the new probability distribution (PD) $P_G(x)$ based on prior PD $P(x)$ and synthetic images generated from the training dataset. The minimax optimization problem is given by (1).

$$\min_G \max_D V(D, G) = E_{x\sim P_{data(x)}}\left[\log D(x)\right] + E_{x\sim p_z}\left[\log(1-D(z))\right] \quad (1)$$

DCGAN, an extension of the GAN model, employs convolutional and convolutional-transpose layers as the discriminator and generator [11], respectively (Fig. 3). By continuously combining and recombining individual image samples, DCGAN act as an ''interpolating mechanism,'' generating diverse ''views'' on the picture manifold. Leveraging convolutional layers facilitates effective feature extraction and enhances the network's performance in image-related tasks.

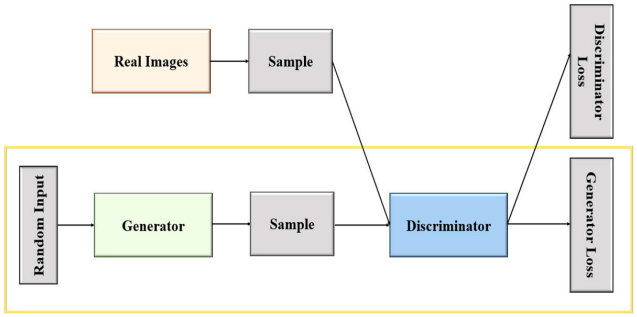

FIGURE 2. Generative adversarial network architecture.

DCGAN utilizes transposed convolutions to filter image details in the opposite manner to convolution. The discriminator includes leaky ReLU, stride convolutional layers, and batch normalization, while the generator produces an RGB image from a $3\times64\times64$ input using transposed convolutions.

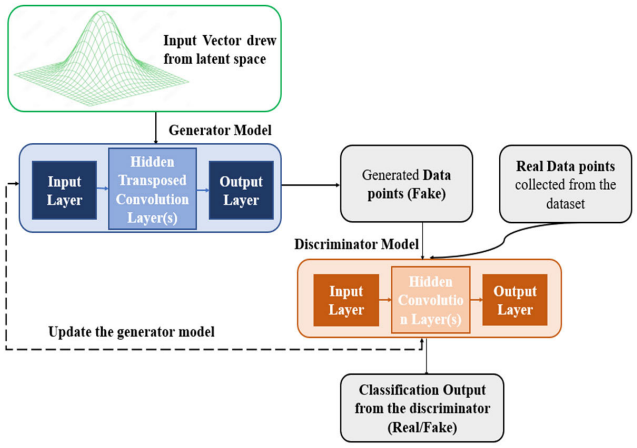

FIGURE 3. Deep convolutional generative adversarial network architecture.

ML researchers have been dedicated to achieving higher classification accuracy in supervised and unsupervised learning tasks [12]. To enhance generalization accuracy, ensembles have proven to be effective by combining constituent classifiers [13]. The capability of ensemble models to integrate weak learners and generate strong learners has motivated researchers to develop and design ensemble techniques [14].

Based on the aforementioned observations, this paper introduces a novel DCGAN ensemble model for fine grading of DR. The proposed model leverages the DCGAN generator to generate different views of fundus images from the Latent Space, as illustrated in Fig. 4. Additionally, a novel ensemble approach using CNN is employed for the classification of fundus images.

The key contributions of the proposed work include-

- The proposed method combines actual and augmented views using DCGAN algorithm, as depicted in Fig. 4.
- The generated views are utilized to address the imbalanced class distribution in the dataset.
- A novel ensemble algorithm called DVE is introduced, which combines the weighted average predictions of CNN+i, CNN, and CNN-i models.
- The proposed algorithm is evaluated on the DDR and EyePACS datasets, comparing it with K-Means, FCM, and Autoencoder-based DEC techniques to demonstrate the model's generalizability.

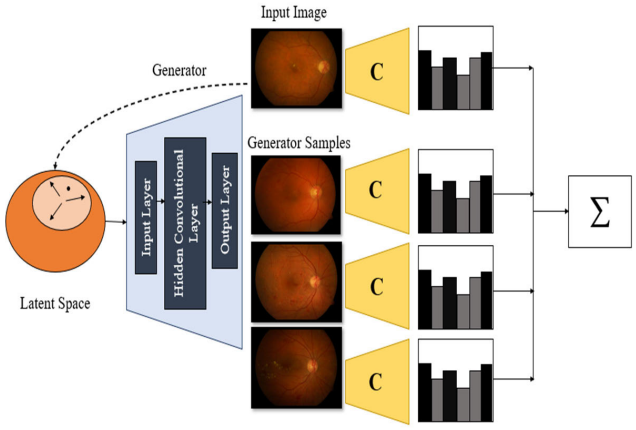

FIGURE 4. DCGAN generated an alternative input image as a sample, and different views of the input image are processed into an ensembled model at test time together with the original image.

The remaining part of paper is arranged in the following manner; The second section includes recent work on DCGAN, ensembled GAN and CNN ensemble architecture. The Third Section presents the methodology of a DCGAN and CNN Ensemble for fine grading of DR. The results and discussions are presented using the fourth section. The paper concludes with the fifth section.

II. LITERATURE REVIEW

This section presents the recent literature work on DCGAN, ensemble CNN and fine grading of DR.

Deep learning technique with small datasets leads to the low performance of the model and overfitting issues. The imbalanced data distribution in DR poses challenges for classification (Fig. 5). Moreover, training models remain a crucial challenge for intra-class variation and class-imbalance data. The GAN framework was first presented by Ian Goodfellow [8] to produce visually realistic images for improving the model's presentation trained on class-imbalance datapoints. Moreover, Unsupervised learning algorithm holds a high potential for medical domain application. Supervised learning methods are generally bounded by biases that specify specific rules, constraining the thinking of other possibilities. Researchers have recently investigated the supervised algorithm with the implementation of CNN. Therefore, Radford et al. [9] proposed DCGAN with certain architectural constraints and represented it as a convincing unsupervised learning model. Since then, different interesting applications and modifications have been proposed by several researchers for GAN and DCGAN [16], [17].

The latest manipulations in GANs have allowed researchers to create interesting realistic images [11]. The generated outputs mimic the actual image variations from training data. Likewise, Wang and Gupta [18] implemented structured GAN to train the model for normal surface scene and style GAN to generate realistic visual variations for indoor scenes. Moreover, researchers [14] illustrated the application of linear separability in latent space and modifying attributes in GAN-separated samples. Furthermore, the applications of linear separability have been explored in many research, including camera attributes [15], [16], face detection and quantifying disentanglement of the latent space [17].

In addition, the recently proposed recurrent GANs are trained using adversarial training [18] and allow users to change images with natural image processing interactively. Liu et al. [19] proposed CoGAN using coupled GAN with multiple modalities to implement joint distribution over images. This process doesn't need data from corresponding image tuples and works with the weight-sharing constraints favoring the common distribution centres.

Many efforts have been made for improved model performance with synthetic data in various prediction tasks, including medical imaging prediction and text classification in RGB images [15], [20], gaze estimation [21], object detection [16], [22] and hand pose detection [23], [24]. In order to categorize clear plastic bottles, Subhajit et al. [30] proposed a lightweight GAN-based model with auto encoding to detect the input image's important features.

The model is further trained on synthetic data with ensembled pre-trained deep learning models to classify transparent plastic bottles. The proposed work is an extension of these approaches, where synthesis data is trained using ensembled DCGAN to improve the model's performance with unlabeled medical imaging data. With misclassified negative samples, Grinsven et al. [31] suggested a selective discriminatory sampling strategy to for quick training of DR characteristics like

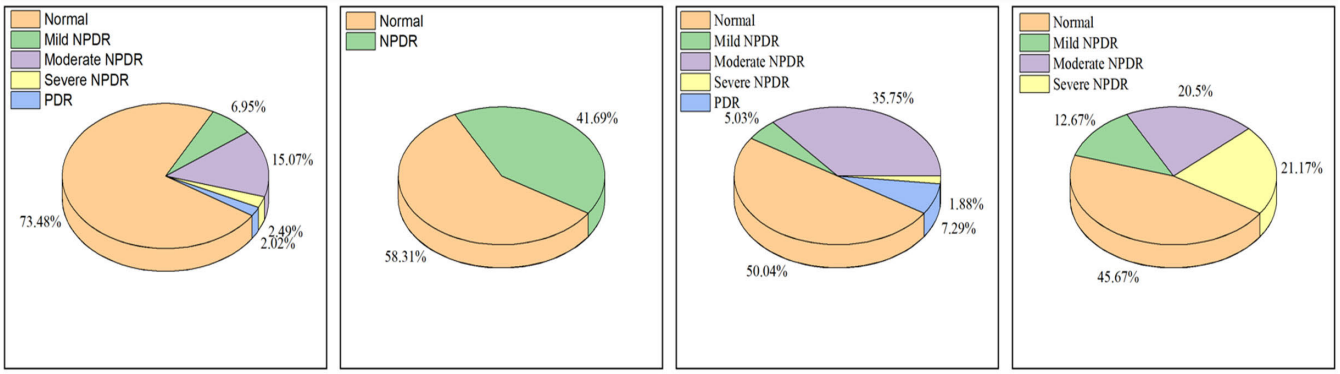

FIGURE 5. Graphical representation of class imbalance in the DR dataset.

HM in fundus images. However, Authors [32] proposed a unique deep network using several filter sizes to train DR features on two open datasets and produce better results.

In the last few years, some studies have been proposed to diagnose lung and Retinal diseases such as Covid using CT [25], [26]. The availability of large databases and growth in deep learning have emerged the automated medical diagnosis to assist experts in better prediction [27]. In comparison to the two CNN models, VGG16 and Xception addition to that, Ayan et al. [28] presented that VGG16 performed better in the detection of pneumonia. The author stated that the performance of the combined model would be better. In continuation, Srivastava et al. [29] categorize X-ray images to detect pneumonia. Firstly, DCGAN trained to synthesize the images and oversample the minority class for better performance. Then, VGG16 is applied as the base model for image classification with transfer learning and CNN. An accuracy rate of 94.5% is achieved on the validation set. The model outperforms the existing model performance with a better accuracy rate. Applying CNN models with radiologist imaging performs better in accurately identifying diseases [25].

In line with recent developments, DR emerges as a severe complication of diabetes, causing irreversible blindness and damage to retinal blood vessels. To address this challenge, the paper [30] introduces an innovative approach. It combines deep CNN with VGG16 and VGG19 models to detect and classify DR based on the visual risk associated with retinal ischemia severity, achieving an impressive accuracy of 90.60%, recall of 95%, and F1 score of 94% on a dataset of 5584 images. The primary objective is to develop a robust system for the automatic detection and classification of DR, contributing to the ongoing efforts in this critical area.

Researchers [31] propose FP-GANs, a novel method for enhancing magnetic resonance image resolution. It divides images into global and texture components in the wavelet domain, using sub-band GANs and sub-band attention for super-resolution. The model also incorporates inverse discrete wavelet transformation for comprehensive image reconstruction. Experiments on the MultiRes 7T dataset demonstrate superior quantitative and qualitative performance compared to other methods.

In Continuation of medical imaging applications, authors [32] introduces BMGAN, a novel 3D synthesis network for merging MR and PET medical images, overcoming limitations like cost and radiation. BMGAN employs bidirectional mapping to embed PET semantic information into a high-dimensional latent space and utilizes a 3D Dense U-Net generator and diverse objective functions to enhance image quality. Importantly, it generates realistic PET images while preserving distinct brain structures across subjects, outperforming other cross-modality synthesis methods in quantitative, qualitative, and classification evaluations. To explore the application of GAN in medical image analysis, the authors [33] propose a multichannel-based generative adversarial network (MGAN) with semi-supervision. It aims to generate sub-fundus images corresponding to scattering DR features, reducing the dependence on labeled data and addressing the challenge of learning inconspicuous DR features in high-resolution fundus images.

Moreover, the shortcoming of trained radiologists and physicians in remote communities is the need to develop an automated Computer Added Diagnosis system for more accurate prediction and detection of disease. The authors proposed a novel ensemble CNN algorithm for the classification of X-ray images into four classes. Transfer learning, followed by data augmentation, is applied to x-ray images for better performance. The results showed an improved accuracy, with 98% on the stacking ensemble and 99% on the voting ensemble.

The authors [34] introduce a Cost-Sensitive classification approach to enhance DR severity prediction from fundus images. By incorporating a regularizer term and addressing label noise through Atomic Sub-Task modelling, the proposed method achieves a 3-5% improvement in quadratic-weighted kappa scores on public datasets within a standard CNN framework. This enhancement is achieved with

negligible computational cost. In continuation of the research advancement in the area of DR grading, the hierarchical structure presented in this research [35] establishes causal relationships between Diabetic Retinopathy (DR) features and severity levels, outperforming traditional deep learning methods. Evaluation on independent sets reveals performance comparable to ophthalmologists, particularly achieving results akin to those with five years of experience in DR severity diagnosis and ten years in referable DR detection.

Furthermore, another study [36] in this area presented to tackle DR) and DME, prevalent causes of blindness. Existing methods separately grade DR or DME, overlooking their correlation. The proposed Cross-Disease Attention Network (CANet) uses image-level supervision to jointly grade DR and DME, achieving superior results on benchmark datasets. CANet's incorporation of disease-specific and disease-dependent attention modules demonstrates effectiveness in capturing internal relationships for comprehensive diagnosis.

An ample amount of research has been conducted to detect DR, but an imbalance of class doesn't yield better outcomes. GAN emerged as a model for generating synthesized images that maintain the class imbalance problem. Moreover, DCGAN produces better-quality images, and an ensemble CNN provides more accurate DR detection results.The research pioneers a novel approach to address imbalanced data in medical imaging datasets, filling a gap in existing literature.

- Utilizing DCGAN, the research employs advanced data augmentation techniques to balance the dataset, potentially boosting model performance.
- Introduces DVE, an ensemble learning technique that combines predictions from multiple models, contributing to enhanced accuracy and robustness in classification tasks.
- The proposed approach achieves metrics—97.4% accuracy, 99.6% specificity, and 92.3% sensitivity—demonstrating its effectiveness, especially in handling imbalanced data scenarios.

III. METHODOLOGY

Classifier ensembles improve generalization ability by combining multiple learners. The ensemble construction involves generating base learners in the learning phase and combining their predictions in the prediction phase using methods like averaging, voting, or stacking. Ensemble pruning selects a subset of base learners to enhance generalization while reducing storage and computational costs. Evaluating the effectiveness of ensemble pruning involves assessing performance gain and ensemble size reduction.

Deep learning has been demonstrated to achieve remarkable results in eye detection tasks, including glaucoma and DR, in recent years [10], [37], [38], [39]. Therefore, our proposed model leverages DCGAN to synthesize high-quality images and utilizes the CNN ensemble method for classification. We introduce a novel algorithm named DVE for generating ensembles based on DCGAN. The proposed architecture consists of two main parts. The first part focuses on the DCGAN algorithm, where the generated diverse views are accumulated and used as representatives of the DCGAN. In the second part, the proposed CNN Ensemble Algorithm is employed for the classification of fundus images. To ensure a balanced representation of errors, under-sampling is applied.

The DCGAN model is composed of dense, reshape, up sampling, and convolutional layers, as detailed in Tables 1 and 2. As described, the input image is fed into the latent space of the DCGAN, where it undergoes transformation through various layers to generate augmented images, as depicted in Fig. 6. The training images, depicted in Fig. 7, are passed through the latent space, resulting in the generation of augmented views using the DCGAN, as shown in Fig. 8. The goal is to find the optimal latent code w that minimizes a composite loss function which is a combination of two terms: $l_1(G(w), x)$ and $\mathcal{L}_{percep}(G(w), x)$. $l_1(G(w), x)$ presents the $L1$ distance (pixel-wise absolute difference) between the generated image ($G(w)$ and the real image x.

$\mathcal{L}_{percep}(G(w), x)$ represents a perceptual loss, quantifying the perceptual difference between the generated and real images. This loss is often calculated using a pre-trained neural network like VGG or a similar architecture. The objective is to find the value of w that minimizes this combined loss, effectively finding the latent code that generates an image closest to the real image.

Mathematically, it's an optimization problem, shown using (2).

$$w^* = arg \min_{w} \mathcal{L}_{img}(\mathrm{G}(\mathrm{w}), \mathrm{x}) + \mathcal{L}_{percep}(G(w), x) \quad (2)$$

The optimization problem can be challenging due to the complex landscape of the loss function. It can heavily depend on the initialization of w, making it computationally expensive and slow to converge. To address these challenges, the approach follows a two-phase to project images into the DCGAN latent space.

First, the w is initialized randomly and gradient descent is used to iteratively refine w to minimize the objective function. This step often involves multiple iterations to approach an optimal w. The key here is to balance the trade-off between reconstruction quality. Once an optimal w is found, it is fed into the pretrained DCGAN generator network. The generator processes w and produces a new image as the output.

The generator in the DCGAN reconstructs features from an n-dimensional latent space. It utilizes a dense layer to generate the feature vector, which is then reshaped into an $N \times N$ feature vector using a reshape layer. The output of the reshape layer is passed through an up-sampling layer and a deconvolutional layer. The Adam optimizer is employed as a gradient descent optimizer to mitigate the issue of exploding gradients. The kernel of 4×4 is used for the DCGAN training. The algorithm is trained for two thousand epochs. The loss of the generator and discriminator during the DCGAN training is depicted in Fig. 9. This DCGAN-based approach

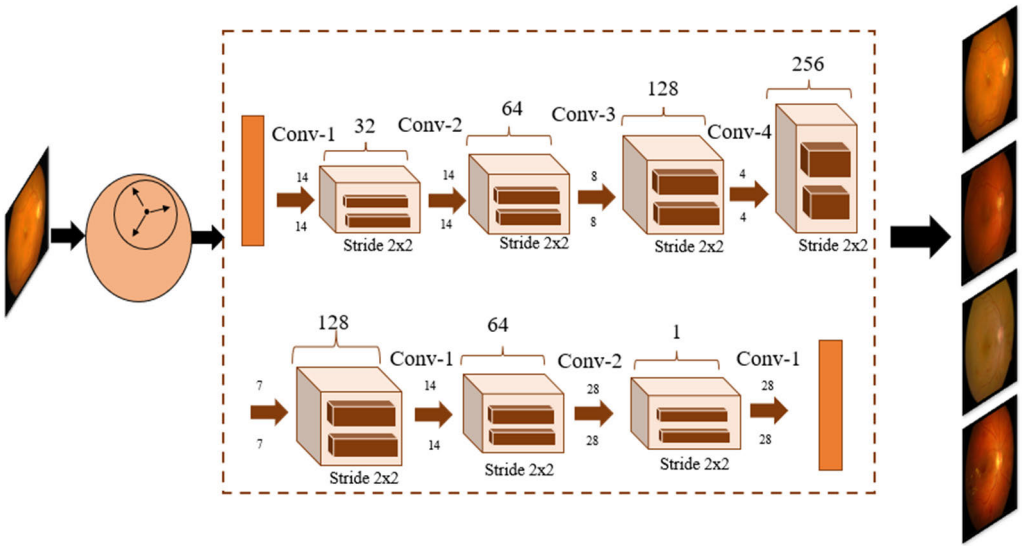

FIGURE 6. System architecture of the DCGAN for augmentation.

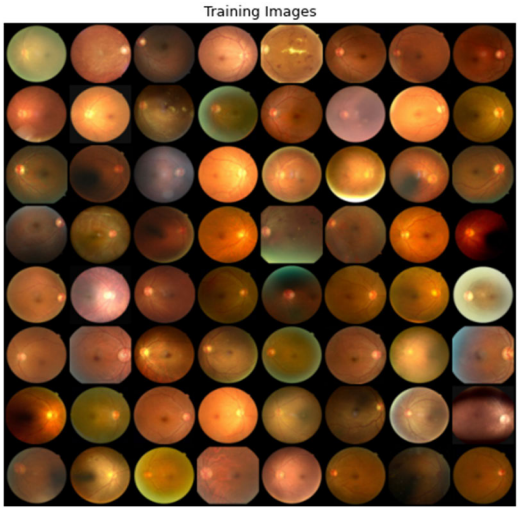

FIGURE 7. Training images.

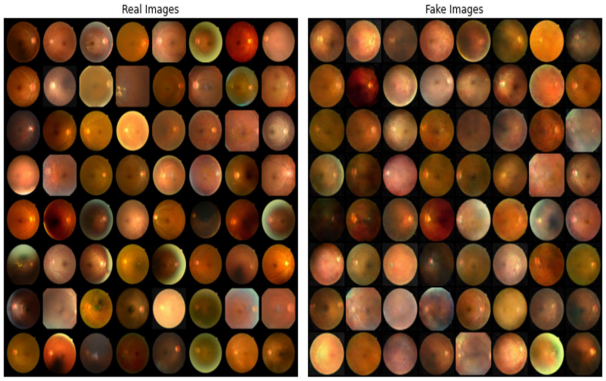

FIGURE 8. Real and Fake Images generated with DCGAN.

is combined with a CNN ensemble method for enhanced classification performance [40]. The DCGAN structure includes a batch normalization stage to normalize intensity values.

TABLE 1. Generator layers and parameters.

Layer - type	Param	Output Shape
conv2d 21 - Conv2D	320	(None, 14, 14, 32)
leaky-re-lu-12 - Leaky RELU	0	(None, 14, 14, 32)
dropout-12 - Dropout	0	(None, 14, 14, 32)
conv2d-22 - Conv2D	18496	(None, 7, 7, 64)
zero-padding2d-3 - Zero Padding2D	0	(None, 8, 8, 64)
batch-normalization-15 - Batch Normalization	256	(None, 8, 8, 64)
leaky-re-lu-13 - LeakyRELU	0	(None, 8, 8, 64)
dropout-13 - Dropout	0	(None, 8, 8, 64)
conv2d-23 - Conv2D	73856	(None, 4, 4, 128)
batch-normalization 16 - Batch Normalization	512	(None, 4, 4, 128)
leaky-re-lu-14 - LeakyRELU	0	(None, 4, 4, 128)
dropout-14 - Dropout	0	(None, 4, 4, 128)
conv2d-24 - Conv2D	295168	(None, 4, 4, 256)
batch-normalization 17 - Batch Normalization	1024	(None, 4, 4, 256)
leaky-re-lu-15 - LeakyRELU	0	(None, 4, 4, 256)
dropout-15 - Dropout	0	(None, 4, 4, 256)
flattern-3 – Flatten	0	(None, 4096)
dense-6 - Dense	4097	(None,1)

As it is commonly known, inhomogeneous intensity values and noise in the images can affect the performance of the automated methods. Although various normalization and denoising algorithms have been applied with different image types to obtain high performance [41], [42], [43], they may lead to loss of information and increase computational costs. The efficiency of the proposed approach has been provided without using any separate steps for normalization and denoising.

Moreover, this pretraining of DCGAN is carried out using a Kaggle dataset consisting of 33,000 fundus images. Specifically, 70% of the dataset (23,100 images) is allocated for

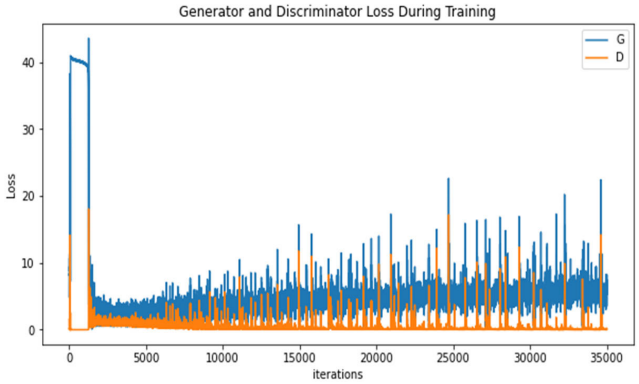

FIGURE 9. Losses for the generator and discriminator during training.

TABLE 2. Discriminator layers and parameters.

Layer - type	Param	Output Shape
dense7 - Dense	633472	None, 6272
reshape3 - Reshape	0	None, 7,7,128
up-sampling2d6 - Upsampling 2D	0	None, 14, 14, 128
conv2d25 - Conv2D	147584	None, 14, 14, 128
batch-normalization18 - Batch Normalization	512	None, 14, 14, 128
activation9 - Activation	0	None, 14, 14, 128
up-sampling2d7 - Upsampling 2D	0	None, 28, 28, 128
conv2d26 - Conv2D	73792	None, 28, 28, 64
batch-normalization19 - Batch Normalization	256	None, 28, 28, 64
activation10- Activation	0	None, 28, 28, 64
conv2d27 - Conv2D	577	None, 28, 28, 1
activation11 - Activation	0	None, 28, 28, 1

model training, 20% (6,600 images) for model validation, and 10% (3,300 images) for model testing. The pretraining is done as a binary classification task where the discriminator classifies images as ''real'' or ''generated.'' The discriminator is pretrained first backpropagation and optimization techniques like stochastic gradient descent (SGD) or Adam and a binary cross-entropy loss function is implemented to compute the classification loss.

Following the discriminator pretraining, the generator network is initialized with random weights. During the generator pretraining phase, synthetic retinal images are generated by feeding random noise as input into the generator. The primary objective of the generator at this stage is to produce synthetic images that are so convincing that the discriminator wrongly classifies them as authentic ''real'' retinal images. The generator's loss is computed based on the discriminator's classification, specifically aiming to minimize the likelihood of the discriminator correctly identifying the images as ''generated.'' Backpropagation, combined with optimization techniques such as stochastic gradient descent (SGD) or Adam, is employed to update the generator's weights and parameters. The utilized parameters, layer types, and output shapes for the generator and discriminator are presented in Table 1 and Table 2, respectively.

CNN classification models are constructed from scratch, and the weights of each classifier are combined to compute the average weight for making the final prediction. The complete methodology is illustrated in Fig. 10. The five-layer CNN classifier is presented in Figure 11. The proposed CNN models consist of input layers followed by a convolutional layer for feature extraction. Max-pooling layers are utilized to reduce the feature size and enhance the robustness of important features. The model also incorporates fully connected layers to aggregate the feature matrix, global average pooling, and an output layer. The SoftMax function is employed for the classification of DR and non-DR cases.

The proposed model can be represented by 5-layers of units, with the input followed by a convolutional layer. When using local filters of size m, the output can be calculated as (N - m + 1). The computation of the convolutional layers is given below using (3).

$$x_i^{l,j} = f\left(\sum_{a=1}^{m} w_a^j x_{i+a-1}^{l-1,j} + b_j\right) \tag{3}$$

The convolutional matrix is represented by b_j as the bias for feature maps. The weights are determined by convolving them with the output of the previous layer's feature map. The nonlinear mapping is determined by the activation function f, and in the proposed model, the SoftMax activation function is used, which is computed using (4).

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{4}$$

K represents the number of classes in the multi-class problem, while e^{z_j} denotes the output matrix computed with the standard exponential function. The max pooling mechanism is employed to enhance the recognition capability, enabling the extraction of more robust features by reducing the feature size. The activation function of the max pooling layer in CNN is represented by (5).

$$x_i^{l,j} = max_{i,j=1}^{r}\left(x_{i,j}\right) \tag{5}$$

Global Average Pooling (GAP) is applied to reduce feature maps and computational time while eliminating overfitting at a specific layer. Each classifier is fine-tuned for classifying DR and non-DR fundus images, employing GAP and setting a dropout of 0.5 to avoid overfitting.

The AdamW optimizer, an Adaptive Moment Estimation algorithm, is utilized to optimize the performance of the CNN algorithm [44]. The main difference between the Adam and AdamW optimizers lies in the way they handle weight decay or regularization. Adam is efficient for large problems with many parameters, requiring less memory, working faster, and demanding less learning rate tuning compared to stochastic gradient descent (SGD). In AdamW weight decay, learning rates are implicitly constrained, allowing for the determination of optimal weight decay individually for each learning rate, thus optimizing learning rates. AdamW

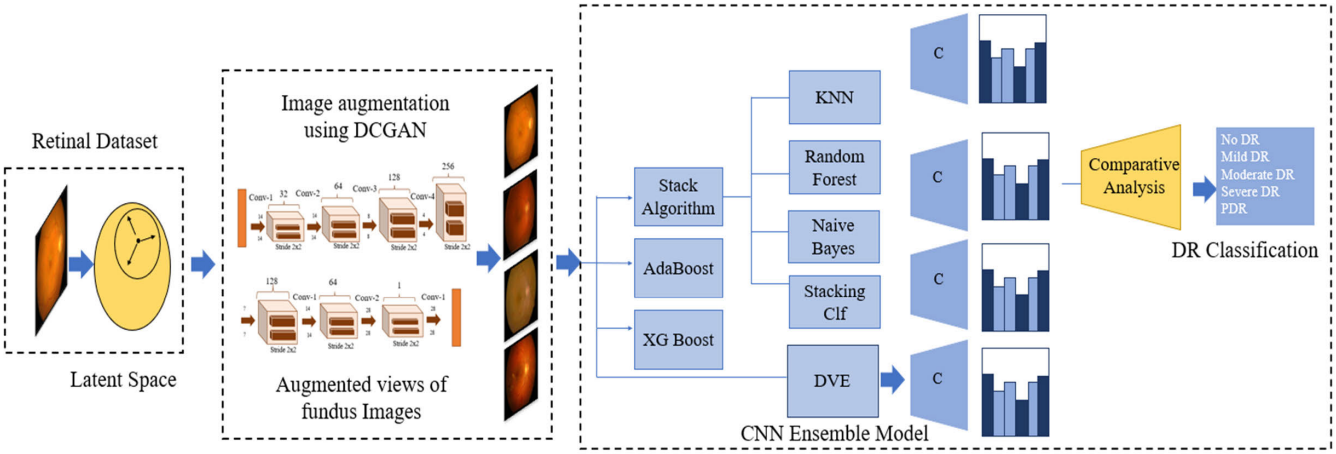

FIGURE 10. Architecture of Preliminary Model: The overall structure of DCGAN-based DVE consists of three parts, Latent Space, DCGAN and CNN-based Ensemble Model. The first and second part process the retinal images through latent space to pretrained DCGAN. The third parts process the augmented images through the DVE, AdaBoost, XG Boost and stack algorithms for comparative analysis, and classification is done for DR fine grading at final stage.

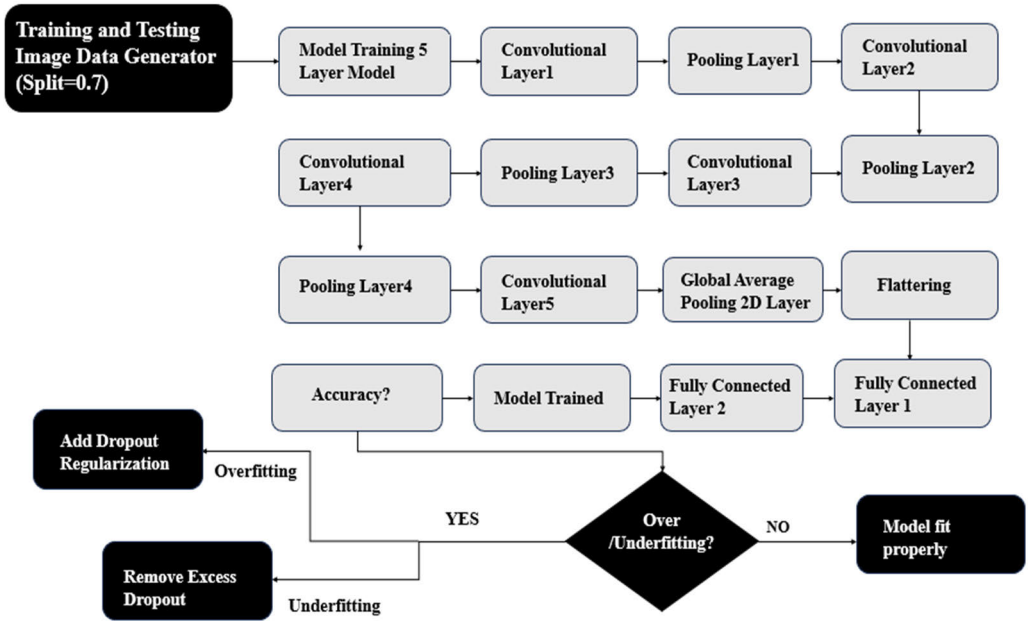

FIGURE 11. 5-Layer CNN architecture for DR classification.

optimizer decouples the optimization learning rate, enhancing performance through separate adjustments [45]. while both Adam and AdamW are variants of the Adam optimizer, AdamW was specifically designed to address issues related to weight decay and provide more reliable and consistent performance in deep learning tasks, especially when weight decay plays a crucial role in regularization. Thus, the proposed work achieves better outcomes with AdamW optimizer. The DVE algorithms follow a finite step to generate a better outcome than the existing approaches as shown below. The DVE ensemble model architecture, depicted in Fig. 12, combines weights from multiple classifiers.

IV. RESULTS AND DISCUSSION

This section provides an overview of the datasets used, implementation details, and performance metrics. The proposed method is evaluated quantitatively and qualitatively using the DDR and EyePACS datasets, showcasing the effectiveness of the approach.

A. DATASETS

The experiments in this study involve two distinct datasets: DDR and EyePACS. The DDR dataset consists of 13,673 fundus images collected from 147 hospitals in China. These images are categorized into five DR grades ranging

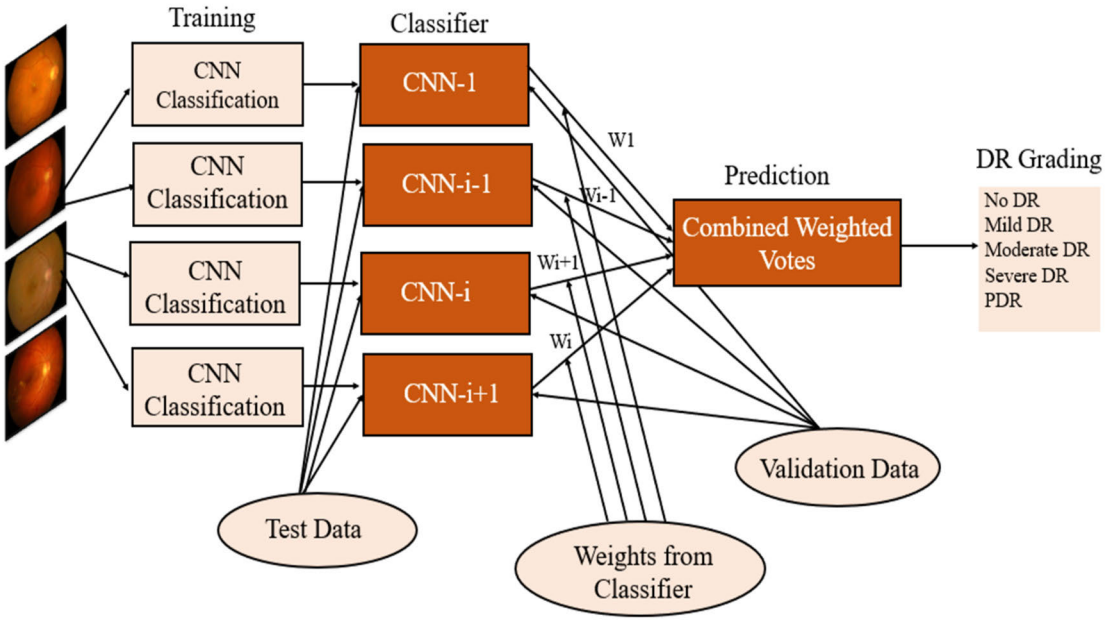

FIGURE 12. DVE architecture for DR grading.

from 0 to 4, representing different levels of severity. Similarly, the EyePACS dataset comprises 88,702 fundus images captured at various care sites in California. The dataset is divided into training (35,126 samples) and testing (53,576 samples) subsets. The images in both datasets are labeled on a scale of 0 to 4, indicating different stages of DR. The diversity and size of these datasets enable deep learning algorithms to effectively recognize and train the model. The EyePACS dataset is widely used by researchers worldwide, and a competition on the Kaggle website has been organized to achieve the best results for DR detection. However, the Kaggle dataset is imbalanced, with a larger number of no-DR images compared to other labeled images.

B. IMPLEMENTATION DETAILS

In this study, the DCGAN model is trained with 1000 epochs and a batch size of 128. A learning rate of 0.002 is used for data augmentation and balancing the minority class. DVE is trained for 100 epochs on the EyePACS dataset and 200 epochs on the DDR dataset using the ADAM optimizer. The cross-entropy function is used as the loss function for all models. The experiments are conducted on a GPU-enabled system with 40 threads, 32 GB of support from a Tesla GPU with 5120 cores, and 256 GB of RAM. The implementation is done using Jupiter Notebook, and Python 3.8 is used for coding the DCGAN and DVE algorithms. In addition to the proposed algorithms, other methods including K-Means, FCM, and autoencoder-based DEC are implemented individually, built, and evaluated. These methods are compared with the proposed algorithms for performance comparison.

C. EVALUATION METRICS

To assess the performance of the proposed DVE algorithm thoroughly, a range of performance measures is employed on both balanced and imbalanced datasets. These measures include commonly used metrics such as accuracy, sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV). Accuracy provides an overall measure of correctness, while SE and SP measure the algorithm's ability to identify positive and negative instances correctly. PPV determines the proportion of correctly classified positive instances, and NPV determines the proportion of correctly classified negative instances. These metrics allow for a comprehensive evaluation of the algorithm's effectiveness on both balanced and imbalanced datasets.

TABLE 3. Ensemble learning algorithm results with balance data on DDR dataset.

Algorithms	Accuracy	SE	SP	PPV	NPV
DVE	96.1	91.2	98.6	96.9	95.8
FCM	83.8	86.9	82.9	58.8	86.9
DEC	88.8	82.4	91.3	82.4	91.3

This study aims to evaluate the performance of unsupervised algorithms, including DEC, K-Means, and FCM, on both balanced and imbalanced datasets. These algorithms are compared with the proposed DVE technique. The evaluation is conducted using the same volume of data for fair comparison. The results, presented in Table 3 and Table 4 demonstrate the effectiveness of the proposed DVE algorithm. It achieves the highest accuracy score of 96.1%

Algorithm 1 Proposed Algorithm for DVE

Input: Training Data-$X_0^{Max-sample}$ Testing Data-$X_0^{Max-sample}$
Validation Data-$X_0^{Max-sample}$
Output: No DR, Mild DR, Moderate DR, Severe DR, PDR
Algorithmic Steps:
For i to max-sample
  First, 5-layer CNN generates the prediction value
  if predicted value is diabetic or non-diabetic
    Probability will be normalized for outcomes
    $NP_u^1\ NP_v^1\ NP_w^1$
    CNN 2 model gives another prediction
    $NP_u^2\ NP_v^2\ NP_w^2$
    CNN 3 model gives another prediction
    $NP_u^3\ NP_v^3\ NP_w^3$
    Classifier weight: $\propto = \frac{NP_u^1}{NP_u^1+NP_u^2+NP_u^3}$
    $\emptyset = \frac{NP_v^1}{NP_v^1+NP_v^2+NP_v^3}$ $\delta = \frac{NP_w^1}{NP_w^1+NP_w^2+NP_w^3}$
    if prediction is diabetic, then
      $NP_u = \emptyset \times NP_u^1 + (1-\emptyset) \times NP_u^2 + (1-\delta) \times NP_u^3$
      $NP_v = \emptyset \times NP_v^1 + (1-\emptyset) \times NP_v^2 + (1-\delta) \times NP_v^3$
      $NP_u = \emptyset \times NP_w^1 + (1-\emptyset) \times NP_w^2 + (1-\delta) \times NP_w^3$
    else:
      $NP_u = \alpha \times NP_u^1 + (1-\alpha) \times NP_u^2 + (1-\delta) \times NP_u^3$
      $NP_v = \alpha \times NP_v^1 + (1-\alpha) \times NP_v^2 + (1-\delta) \times NP_v^3$
      $NP_u = \alpha \times NP_w^1 + (1-\alpha) \times NP_w^2 + (1-\delta) \times NP_w^3$
    return the weighted votes for the predicted values
    weighted
    average ($NP_u NP_v NP_u$)
    End if
  Else
    return the predicted outcome from 5-layer CNN
  End if
End for

on the balanced dataset, as shown in Fig. 13. However, all algorithms experience a decrease in performance on the imbalanced dataset, as observed in Fig. 14.

TABLE 4. Ensemble learning algorithm results with imbalance data on DDR dataset.

Algorithms	Accuracy	SE	SP	PPV	NPV
DVE	91.2	93.1	90.5	79.4	93.1
FCM	89.3	88.2	89.9	81.1	93.9
DEC	85.4	78.7	88.5	76.4	78.4

To assess the training progress and model performance, Fig. 15 and 16 depict the training and validation losses of the DVE model for the DDR and EyePACS datasets, respectively. The decreasing trends of both training and validation loss curves indicate that the model is not overfitting and is effectively learning from the data. This behavior is consistent with the DDR dataset. The convergence of validation and training loss values further indicates that the proposed model fits well

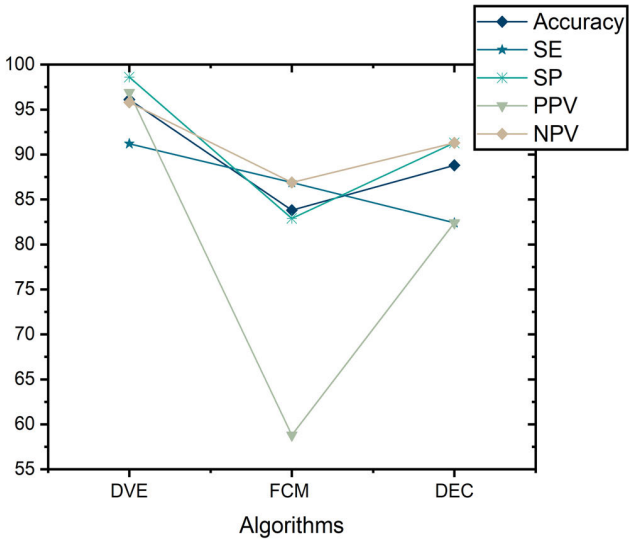

FIGURE 13. Graphical representation of proposed algorithm on balanced data.

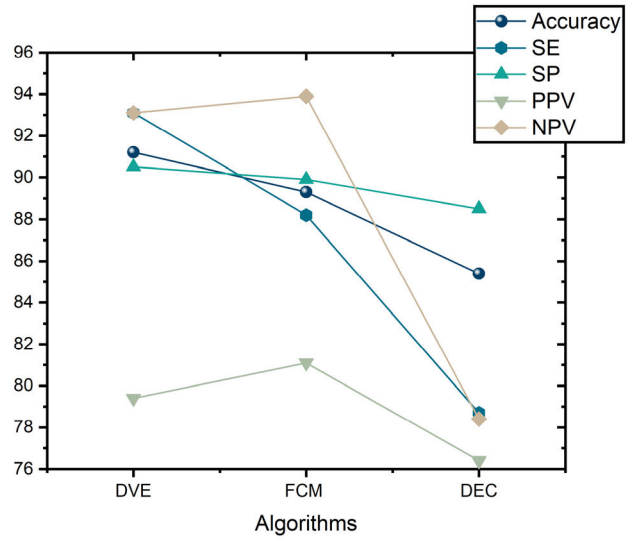

FIGURE 14. Graphical representation of proposed algorithm on imbalanced data.

TABLE 5. DVE algorithm results on DDR Dataset.

Model	Accuracy	SE	SP	PPV	NPV
DVE	96.1	91.2	98.6	96.9	95.8
FCM	83.8	86.9	82.9	58.8	86.9
K-Means	87.4	82.1	89.9	80	91.2
DEC	88.8	82.4	91.3	82.4	91.3

with the given input features and demonstrates its ability to generalize to new data.

Table 5 and 6 present the performance evaluation of the proposed DVE method on the DDR and EyePACS datasets. The algorithm is compared with K-Means, FCM, and autoencoder-based DEC to assess its generalizability. The results demonstrate that the DVE algorithm achieves a

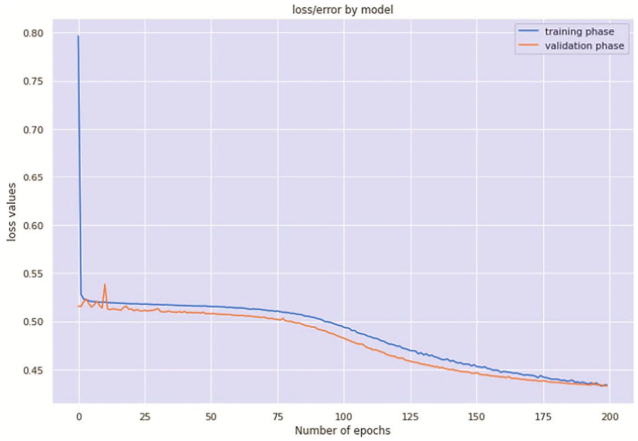

FIGURE 15. Training and validation loss of proposed CNN ensemble model on DDR dataset.

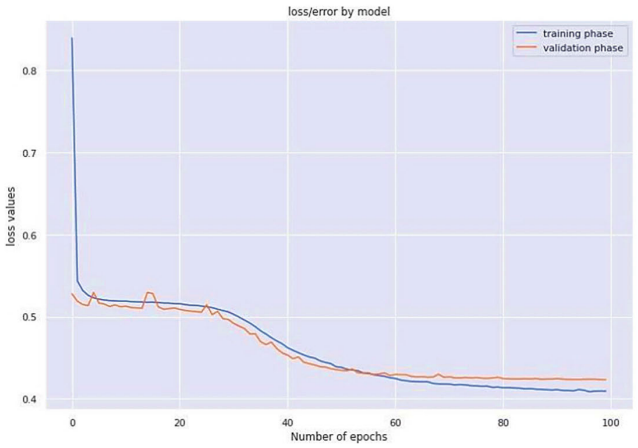

FIGURE 16. Training and validation loss of proposed CNN ensemble model on EyePACS dataset.

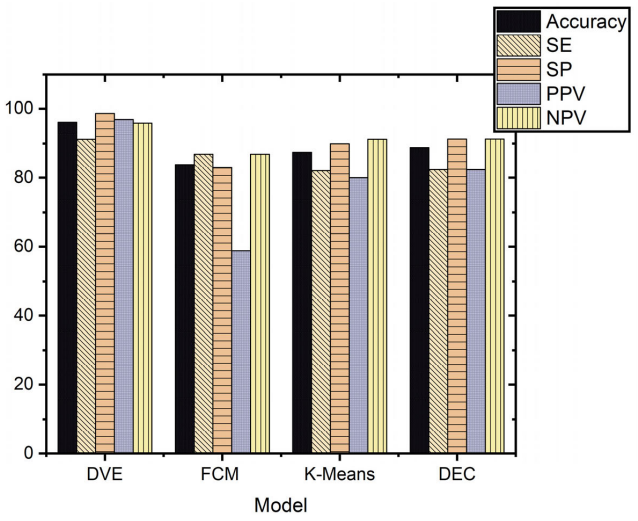

FIGURE 17. Graphical representation of DVE results on the DDR dataset.

superior accuracy of 96.1% on the DDR dataset, surpassing the individual models, as shown in Fig. 17. Moreover, the model's generalizability is confirmed through the results

TABLE 6. DVE algorithm results on EyePACS Dataset.

Model	Accuracy	SE	SP	PPV	NPV
DVE	97.4	92.3	99.6	97.9	96.7
FCM	82.5	75	85.9	70.5	75
K-Means	81.5	77.7	82.5	82.3	80
DEC	96.5	95.2	94.5	91.1	90.3

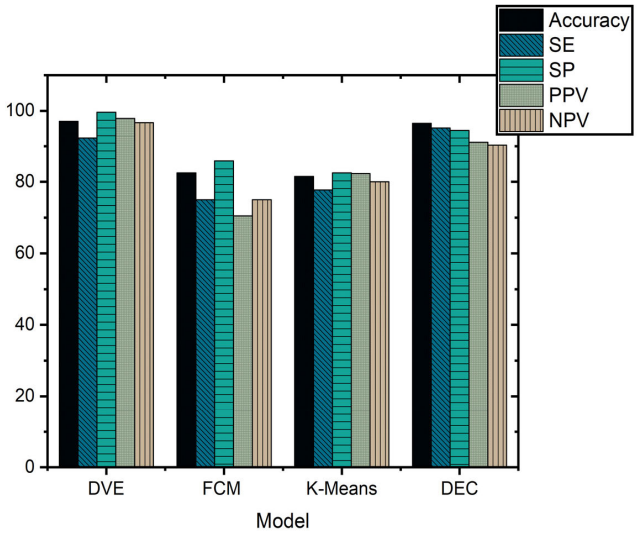

FIGURE 18. Graphical representation of DVE results on the EyePACS dataset.

obtained on the EyePACS dataset. Fig. 18 illustrates that the DVE achieves the highest accuracy rate of 97.1% on the EyePACS dataset, further validating its effectiveness the study highlights the superior performance of the proposed DVE algorithm in comparison with other unsupervised algorithms. It achieves high accuracy scores on balanced datasets and demonstrates generalizability. The results indicate that the DVE algorithm outperforms other algorithms, achieving the highest accuracy scores 96.1% on DDR and 97.3% on EYEPACS dataset. However, it is observed that all algorithms experience a decrease in performance when applied to imbalanced datasets. The training and validation loss curves for the DVE model show a decreasing trend, indicating effective learning and no overfitting. Comparative analysis with K-Means, FCM, and DEC further strengthens the claim of the DVE algorithm's effectiveness. It outperforms these algorithms and demonstrates its generalizability in different dataset settings. The DVE algorithm achieves superior accuracy rates on both the DDR and EyePACS datasets, validating its effectiveness in fine grading of DR images. As a future work, the performance of the proposed model can be compared with the performance of a capsule network-based approach because capsule networks can preserve spatial relationships of learned features and have been used recently for recognitions and classifications [37], [38], [39].

The recent work on DR detection, as demonstrated in Table 7, highlights the utilization of deep learning techniques

TABLE 7. Related work on DR detection.

DL Method	Lesion detection	Dataset (Size)	Performance measure	Ref
CNN (Inception-v3)	×	Messidor-2 (1748)	SN (96.1) SP (93.9)	[45]
CNN-ResNet34	×	Kaggle (35000)	ACC (85) SN (86)	[46]
CNN (AlexNet)	×	Kaggle (22,700) IDRiD (516)	ACC (90.7)	[47]
CNN (Inception V3, InceptionResnet-V2 and Resnet152)	×	Primary dataset (30244)	ACC (88.21) SN (88.51) SP (90.85)	[48]
ResNet-50 and DenseNet-121	×	EYEPACS (88,702) Messidor-2 (1748)	ACC (95.58)	[53]
Multiresolution based CNN (ResNet-50, DenseNet, Xception, and EfficientNet)	√	IDRiD (516) DDR (13,673) EyePACS (88,702)	ACC (90.07 96.20 93.53)	[54]
VGG16, VGG19, ResNet50, ResNet101, **Inception-V3, Xception, ResNets, DenseNet201, DarkNet53, EfficientNetB0**	×	Balanced DDR dataset (31330)	ACC (94.5)	[55]
multi-task deep learning framework, Hierarchical multi-task neural network	√	Independent two testing set with 4502 and 1589 fundus images	AUC Values (Val-94.02, Test1-93, Text2-95.8)	[35]
Novel cross-disease attention network (CANet)	√	ISBI 2018 IDRiD challenge dataset and Messidor dataset	Joint Accuracy metric (65.1)	[36]
Standard CNN (atomic sub task modelling)	×	EyePACS database (88,000) Messidor2 (1748)	Kappa-Score EyePACS (78.71±0.28±0.46) Messidor2 (79.79±1.03/ 63.41±1.99)	[34]
Proposed Work	×	DDR (13,673) EYEPACS (88,702)	**Accuracy No DR (97.72) Mild DR (97.4) Moderate DR (97.2) Severity DR (98.6) PDR (97.3)**	-

such as Inception v3, ResNet34, and Alex Net models [49]. These models were trained using diverse datasets; however, it is important to note that most of the training was conducted using imbalanced open-access datasets. Consequently, the models' performance suffered when it came to generalizing to new data [50]. In light of this, the purpose of the present study is to address the limitations observed in previous works. The justification for this focus on balancing the dataset and utilizing the DVE algorithm lies in the fact that imbalanced training data can lead to skewed model performance, particularly when it comes to generalization. By incorporating a more balanced dataset the study aims to achieve a significant accuracy improvement of 5-7%. As another extension of this work, the performance of the proposed method can be evaluated after integrating a combined loss function, such as the loss function used in [58]. However, the presented performance metric here represents the aggregate performance for all severity classes, showcasing an improvement of 2% compared to existing work [50].

D. APPLICATION OF PROPOSED WORK

There are various applications of the proposed model, as depicted in Fig. 19. The proposed model is planned to be integrated into mobile-based applications for the further evaluation of DR images using a portable fundus camera. It can be used to address DR normal to moderate cases in remote areas.

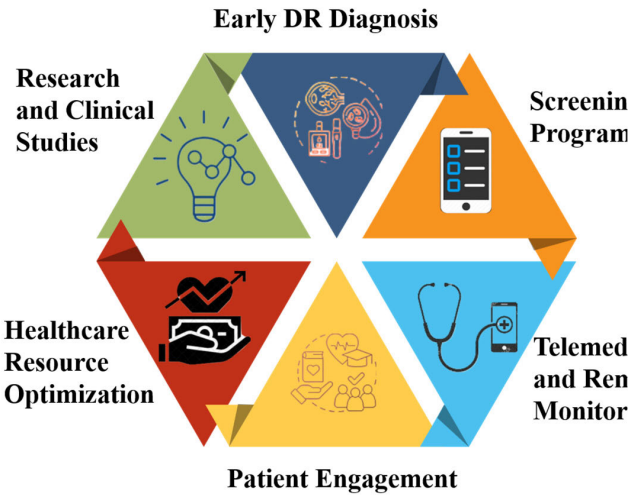

FIGURE 19. Applications of proposed work.

1) EARLY DR DIAGNOSIS

The proposed system can be used in clinical settings to assist ophthalmologists in the early diagnosis of DR. Early detection is crucial for timely intervention and preventing the progression of the disease, potentially reducing the risk of permanent blindness.

2) SCREENING PROGRAMS

It can be integrated into a mobile-based application for large-scale screening programs for DR patients. Automated DR detection can help in identifying individuals who require further evaluation and treatment, streamlining the screening process and reducing the burden on healthcare resources.

3) TELEMEDICINE AND REMOTE MONITORING

In telemedicine applications, patients can capture retinal images using portable devices like MII ret cam portable camera with mobile-based application, and the system can automatically assess the severity of DR.

4) PATIENT ENGAGEMENT

The system can enhance patient engagement by providing visual feedback on the status of their DR with severity levels 0-4. In case on NPDR or PDR patient can visit the nearest ophthalmologist.

5) HEALTHCARE RESOURCE OPTIMIZATION

By automating the initial screening and classification of DR, healthcare resources can be allocated more efficiently. Patients with severe cases can be prioritized for immediate attention.

6) RESEARCH AND CLINICAL STUDIES

The system can be a valuable tool for researchers and clinicians involved in DR studies.

V. CONCLUSION AND FUTURE WORK

Manual screening of DR is a time-consuming and challenging task, often resulting in delayed diagnosis and potential vision loss. To address this issue, researchers have developed automated methods for DR detection. In this study, a novel ensemble learning algorithm called DVE, based on the DCGAN architecture, is proposed for fine grading of DR images. The DVE algorithm demonstrates superior performance compared to state-of-the-art methods in terms of accuracy, sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV). It achieves an overall accuracy rate of 97.4% for the EyePACS dataset and 96.1% for the DDR dataset. Comparative analysis with other algorithms, such as FCM, DEC, and K-means, further validates the effectiveness of the proposed approach. Future research can focus on expanding the dataset to reduce overfitting and optimizing parameters to enhance performance. Additionally, exploring additional data augmentation techniques can be beneficial for further improvement.

DATA AVAILABILITY STATEMENT

The study incorporates the utilization of two datasets: the EYEPACS dataset, accessible at https://www.kaggle.com/datasets/mariaherrerot/eyepacspreprocess, and the DDR dataset, which can be found at https://www.kaggle.com/datasets/mariaherrerot/ddrdataset. These datasets have been strategically selected to facilitate the accomplishment of DR Classification.

AUTHOR'S CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ACKNOWLEDGMENT

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are also thankful to AIDA Lab CCIS Prince Sultan University, Riyadh Saudi Arabia for support.

REFERENCES

[1] D. Nagpal, S. N. Panda, M. Malarvel, P. A. Pattanaik, and M. Zubair Khan, "A review of diabetic retinopathy: Datasets, approaches, evaluation metrics and future trends," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7138–7152, Oct. 2022, doi: 10.1016/j.jksuci.2021.06.006.

[2] B. Krawczyk, M. Galar, Ł. Jeleń, and F. Herrera, "Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy," *Appl. Soft Comput.*, vol. 38, pp. 714–726, Jan. 2016, doi: 10.1016/j.asoc.2015.08.060.

[3] R. Xu, T. Chen, Y. Xia, Q. Lu, B. Liu, and X. Wang, "Word embedding composition for data imbalances in sentiment and emotion classification," *Cognit. Comput.*, vol. 7, no. 2, pp. 226–240, Apr. 2015, doi: 10.1007/s12559-015-9319-y.

[4] T. M. Khoshgoftaar, E. B. Allen, and J. Deng, "Using regression trees to classify fault-prone software modules," *Mach. Learn. Appl. Softw. Eng.*, vol. 51, no. 4, pp. 87–94, 2005.

[5] E. Goceri, *Medical Image Data Augmentation: Techniques, Comparisons and Interpretations*. Amsterdam, The Netherlands: Springer, 2023.

[6] E. Goceri, "Comparison of the impacts of dermoscopy image augmentation methods on skin cancer classification and a new augmentation method with wavelet packets," *Int. J. Imaging Syst. Technol.*, vol. 33, no. 5, pp. 1724–1744, Apr. 2023, doi: 10.1002/ima.22890.

[7] E. Goceri, "Image augmentation for deep learning based lesion classification from skin images," in *Proc. IEEE 4th Int. Conf. Image Process., Appl. Syst. (IPAS)*, Dec. 2020, pp. 144–148, doi: 10.1109/IPAS50080.2020.9334937.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[9] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. 4th Int. Conf. Learn. Represent. (ICLR) Conf. Track*, 2016, pp. 1–16.

[10] A. He, T. Li, N. Li, K. Wang, and H. Fu, "CABNet: Category attention block for imbalanced diabetic retinopathy grading," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 143–153, Jan. 2021, doi: 10.1109/TMI.2020.3023463.

[11] L. Chai, J. Y. Zhu, E. Shechtman, P. Isola, and R. Zhang, "Ensembling with deep generative views," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14992–15002, doi: 10.1109/CVPR46437.2021.01475.

[12] X. Wang and A. Gupta, "Generative image modeling using style and structure adversarial networks," in *Proc. Eur. Conf. Comput. Vis.*, vol. 9908, 2016, pp. 318–335, doi: 10.1007/978-3-319-46493-0_20.

[13] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2242–2251, doi: 10.1109/CVPR.2017.241.

[14] Y. Zhang, Z. Lin, Y. Kang, R. Ning, and Y. Meng, "A feed-forward neural network model for the accurate prediction of diabetes mellitus," *Int. J. Sci. Technol. Res.*, vol. 7, no. 8, pp. 151–155, 2018. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059910862&partnerID=40&md5=40cdc4d37e47645feb76229e7b9c9dfd

[15] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4511–4520, doi: 10.1109/CVPR.2015.7299081.

[16] X. Peng, B. Sun, K. Ali, and K. Saenko, "Learning deep object detectors from 3D models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1278–1286, doi: 10.1109/ICCV.2015.151.

[17] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "StyleGANv2," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Aug. 2020, pp. 8107–8116.

[18] D. J. Im, C. D. Kim, H. Jiang, and R. Memisevic, "Generating images with recurrent adversarial networks," 2016, *arXiv:1602.05110*.

[19] M. Liu, R. Wang, S. Li, S. Shan, Z. Huang, and X. Chen, "Combining multiple kernel methods on Riemannian manifold for emotion recognition in the wild," in *Proc. 16th Int. Conf. Multimodal Interact.*, Nov. 2014, pp. 494–501, doi: 10.1145/2663204.2666274.

[20] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 1–20, Jan. 2016, doi: 10.1007/s11263-015-0823-z.

[21] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and A. Bulling, "Learning an appearance-based gaze estimator from one million synthesised images," in *Proc. 9th Biennial ACM Symp. Eye Tracking Res. Appl.*, Mar. 2016, pp. 131–138, doi: 10.1145/2857491.2857492.

[22] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, "Learning rich features from RGB-D images for object detection and segmentation," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 8695, 2014, pp. 345–360, doi: 10.1007/978-3-319-10584-0_23.

[23] J. Tompson, M. Stein, Y. Lecun, and K. Perlin, "Real-time continuous pose recovery of human hands using convolutional networks," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 1–10, Sep. 2014, doi: 10.1145/2629500.

[24] J. S. Supančič, G. Rogez, Y. Yang, J. Shotton, and D. Ramanan, "Depth-based hand pose estimation: Methods, data, and challenges," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1180–1198, Nov. 2018, doi: 10.1007/s11263-018-1081-7.

[25] L. Visuña, D. Yang, J. Garcia-Blas, and J. Carretero, "Computer-aided diagnostic for classifying chest X-ray images using deep ensemble learning," *BMC Med. Imag.*, vol. 22, no. 1, pp. 1–16, Oct. 2022, doi: 10.1186/s12880-022-00904-4.

[26] S. M. Rezaeijo, M. Ghorvei, and M. Alaei, "A machine learning method based on lesion segmentation for quantitative analysis of CT radiomics to detect COVID-19," in *Proc. 6th Iranian Conf. Signal Process. Intell. Syst. (ICSPIS)*, Dec. 2020, pp. 1–5, doi: 10.1109/ICSPIS51611.2020.9349605.

[27] S. Montani and M. Striani, "Artificial intelligence in clinical decision support: A focused literature survey," *Yearbook Med. Informat.*, vol. 28, no. 1, pp. 120–127, Aug. 2019, doi: 10.1055/s-0039-1677911.

[28] E. Ayan, B. Karabulut, and H. M. Ünver, "Diagnosis of pediatric pneumonia with ensemble of deep convolutional neural networks in chest X-ray images," *Arabian J. Sci. Eng.*, vol. 47, no. 2, pp. 2123–2139, Feb. 2022, doi: 10.1007/s13369-021-06127-z.

[29] G. Srivastava, A. Chauhan, M. Jangid, and S. Chaurasia, "CoviXNet: A novel and efficient deep learning model for detection of COVID-19 using chest X-ray images," *Biomed. Signal Process. Control*, vol. 78, Sep. 2022, Art. no. 103848, doi: 10.1016/j.bspc.2022.103848.

[30] B. Menaouer, Z. Dermane, N. El Houda Kebir, and N. Matta, "Diabetic retinopathy classification using hybrid deep learning approach," *Social Netw. Comput. Sci.*, vol. 3, no. 5, p. 357, Jul. 2022, doi: 10.1007/s42979-022-01240-8.

[31] S. You, "Fine perceptive GANs for brain MR image super-resolution in wavelet domain," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 7, 2022, doi: 10.1109/TNNLS.2022.3153088.

[32] S. Hu, B. Lei, S. Wang, Y. Wang, Z. Feng, and Y. Shen, "Bidirectional mapping generative adversarial networks for brain MR to PET synthesis," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 145–157, Jan. 2022, doi: 10.1109/TMI.2021.3107013.

[33] S. Wang, X. Wang, Y. Hu, Y. Shen, Z. Yang, M. Gan, and B. Lei, "Diabetic retinopathy diagnosis using multichannel generative adversarial network with semisupervision," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 574–585, Apr. 2021, doi: 10.1109/TASE.2020.2981637.

[34] J. D. A. Galdran, H. Chakor, H. Lombaert, and I. Ben Ayed, "Cost-sensitive regularization for diabetic retinopathy grading from eye fundus images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Lima, Peru, 2020, pp. 1–7.

[35] J. Wang, Y. Bai, and B. Xia, "Simultaneous diagnosis of severity and features of diabetic retinopathy in fundus photography using deep learning," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 12, pp. 3397–3407, Dec. 2020, doi: 10.1109/JBHI.2020.3012547.

[36] X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "CANet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1483–1493, May 2020, doi: 10.1109/TMI.2019.2951844.

[37] P. G. Subin and P. Muthukannan, "Optimized convolution neural network based multiple eye disease detection," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105648, doi: 10.1016/j.compbiomed.2022.105648.

[38] P. Ruamviboonsuk, "Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: A prospective interventional cohort study," *Lancet Digit. Heal.*, vol. 4, no. 4, pp. e235–e244, 2022, doi: 10.1016/S2589-7500(22)00017-6.

[39] W. Xiao et al., "Screening and identifying hepatobiliary diseases through deep learning using ocular images: A prospective, multicentre study," *Lancet Digit. Health*, vol. 3, no. 2, pp. e88–e97, Feb. 2021, doi: 10.1016/S2589-7500(20)30288-0.

[40] J. Krause, V. Gulshan, E. Rahimy, P. Karth, K. Widner, G. S. Corrado, L. Peng, and D. R. Webster, "Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy," *Ophthalmology*, vol. 125, no. 8, pp. 1264–1272, Aug. 2018, doi: 10.1016/j.ophtha.2018.01.034.

[41] E. Goceri, "Intensity normalization in brain MR images using spatially varying distribution matching," in *Proc. 11th Int. Conf. Comput. Graph., Visualization, Comput. Vis. Image Process. (CGVCVIP)*, Jul. 2017, pp. 300–304.

[42] E. Goceri, "Fully automated and adaptive intensity normalization using statistical features for brain MR images," *Celal Bayar Üniversitesi Fen Bilimleri Dergisi*, vol. 14, no. 1, pp. 125–134, Mar. 2018, doi: 10.18466/cbayarfbe.384729.

[43] E. Goceri, "Evaluation of denoising techniques to remove speckle and Gaussian noise from dermoscopy images," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106474, doi: 10.1016/j.compbiomed.2022.106474.

[44] T. R. Gadekallu, "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, pp. 1–16, 2020, doi: 10.3390/electronics9020274.

[45] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C.-H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164–118173, 2020, doi: 10.1109/ACCESS.2020.3005152.

[46] E. Goceri, "Classification of skin cancer using adjustable and fully convolutional capsule layers," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104949, doi: 10.1016/j.bspc.2023.104949.

[47] E. Goceri, "Analysis of capsule networks for image classification," in *Proc. Int. Conf. Comput. Graph., Vis., Comput. Vis. Image Process.*, Jul. 2021, pp. 53–60, doi: 10.33965/mccsis2021_202107l007.

[48] E. Goceri, "Capsule neural networks in classification of skin lesions," in *Proc. Int. Conf. Comput. Graph., Vis., Comput. Vis. Image Process.*, Jul. 2021, pp. 29–36, doi: 10.33965/mccsis2021_202107l004.

[49] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, J. Cuadros, and R. Kim, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.

[50] M. T. Esfahani, M. Ghaderi, and R. Kafiyeh, "Classification of diabetic and normal fundus images using new deep learning method," *Leonardo Electron. J. Pract. Technol.*, vol. 17, no. 32, pp. 233–248, 2018.

[51] B. Harangi, J. Toth, A. Baran, and A. Hajdu, "Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Jul. 2019, pp. 2699–2702, doi: 10.1109/EMBC.2019.8857073.

[52] H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma, and W. Qian, "An interpretable ensemble deep learning model for diabetic retinopathy disease classification," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2045–2048, doi: 10.1109/EMBC.2019.8857160.

[53] H. Mustafa, S. F. Ali, M. Bilal, and M. S. Hanif, "Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework," *IEEE Access*, vol. 10, pp. 113172–113183, 2022, doi: 10.1109/ACCESS.2022.3217216.

[54] K. Ashwini and R. Dash, "Grading diabetic retinopathy using multiresolution based CNN," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105210, doi: 10.1016/j.bspc.2023.105210.

[55] F. C. Monteiro and J. Rufino, "Is diabetic retinopathy grading biased by imbalanced datasets?" in *Proc. Int. Conf. Optim., Learn. Algorithms Appl.*, vol. 1754, 2022, pp. 51–64, doi: 10.1007/978-3-031-23236-7_4.

[56] S. H. Abbood, H. N. A. Hamed, M. S. M. Rahim, A. Rehman, T. Saba, and S. A. Bahaj, "Hybrid retinal image enhancement algorithm for diabetic retinopathy diagnostic using deep learning model," *IEEE Access*, vol. 10, pp. 73079–73086, 2022, doi: 10.1109/ACCESS.2022.3189374.

[57] M. Mujahid, A. Rehman, T. Alam, F. S. Alamri, S. M. Fati, and T. Saba, "An efficient ensemble approach for Alzheimer's disease detection using an adaptive synthetic technique and deep learning," Tech. Rep., 2023, pp. 1–20.

[58] E. Goceri, "An application for automated diagnosis of facial dermatological diseases," *İzmir Kâtip Çelebi Üniversitesi Sağlik Bilim. Fakültesi Derg.*, vol. 6, no. 3, pp. 91–99, 2021.

HUMA NAZ received the master's degree from the Department of Computer Science and Engineering, Chitkara University. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, University of Petroleum and Energy Studies (UPES), Dehradun. Her research interests include unsupervised deep learning, data analytics, and machine learning the healthcare applications. She has published more than eight research articles in reputable SCI and Scopus indexed journals. Other than that, she has filed five patents.

RAHUL NIJHAWAN received the M.Tech. and Ph.D. degrees from IIT Roorkee. He is currently an esteemed Assistant Professor with the Thapar Institute of Engineering and Technology, Patiala, Punjab. He has had many prestigious achievements, including winning the Young Scientist Award, in 2018, and the Best Ph.D. Thesis Award. He has published more than 30 research papers in reputable international journals/conferences and guided several research projects. He is a reviewer of several reputed international/national journals, including IEEE TRANSACTIONS. His research interests include medical imaging, remote sensing, computer vision, pattern recognition, artificial intelligence, and machine learning.

NEELU JYOTHI AHUJA (Senior Member, IEEE) received the Ph.D. degree, in 2010. She is currently a Professor and the Head of the Department of Systemics, School of Computer Science, University of Petroleum and Energy Studies, Dehradun. Her Ph.D. degree on developing a prototype rule-based expert system for seismic data interpretation. She has successfully delivered government-sponsored research and development projects and consultancies worth more than 1.5 crores, since 2013. To her credit, she holds successful delivery of half a dozen research and development projects and consultancies funded by the Department of Science and Technology (DST), GOI. Under her supervision, 11 scholars have completed their doctoral work. Currently, six research scholars are undergoing their Ph.D. work under her supervision. Uttarakhand Law Commission and Uttarakhand Human Rights Commission, Government of Uttarakhand have conferred her with the Himalayan Nari Shakti Award, in 2020, and the Institute of Green Engineers have conferred her with the IGEN Women Achievers Award, in 2021, in recognition of her research contributions.

SHAHA AL-OTAIBI (Member, IEEE) received the M.S. degree in computer science and the Ph.D. degree in artificial intelligence from KSU. She is currently an Associate Professor with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Saudi Arabia. She has also the SFHEA and the Senior Fellow Recognition from U.K. Higher Education Academy. She is also a reviewer in some journals and an editorial board member in other journals. Her research interests include data science, artificial intelligence, machine learning, bio-inspired computing, cybersecurity, and information security.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently a Full Professor with the College of Computer and Information Sciences, Prince Sultan University (PSU), Riyadh, Saudi Arabia, and also the Leader of the AIDA Laboratory. She has published over 300 publications in high-ranked journals. Her primary research interests include bioinformatics, data mining, and classification using AI models. She received the Best Student Award from the Faculty of Computing, UTM, in 2012. She received the Best Research of the Year Award from PSU, from 2013 to 2016. Due to her excellent research achievement, she is included in Marquis Who's Who (S & T) 2012. She is an editor of several reputed journals and on a panel of TPC of international conferences.

SAEED ALI BAHAJ received the Ph.D. degree from Pune University, India, in 2006. He is currently an Associate Professor with the Department of Computer Engineering, Hadramout University, and MIS Department COBA, Prince Sattam bin Abdulaziz University, Saudi Arabia. His research interests include information management, forecasting, information engineering, and information security.

AMJAD REHMAN (Senior Member, IEEE) received the Ph.D. degree from the Faculty of Computing, University Teknologi Malaysia (UTM), Malaysia, specializing in information security using image processing techniques, in 2010. He is currently an Associate Professor with the CCIS, Prince Sultan University, Riyadh, Saudi Arabia. He is also a PI in several projects and completed projects funded by MoHE Malaysia, Saudi Arabia. His research interests include bioinformatics, the IoT, information security, and pattern recognition. He received a Rector Award for the 2010 Best Student from UTM Malaysia.