

Enhancing The Precision Detection and Grading of Diabetic Retinopathy through Digital Retinal Imaging Using 3D Convolutional Neural Networks

Autho Allwine^{*1}, Mutiara S Simanjuntak², Wahyu Aji Pulungan³

^{1,3}Computer Science, Universitas Lampung, Indonesia

²Department of Electrical and Computer Engineering, National Kaohsiung University Of Science and Tecnology, Jiangong Campus. No 415 Jiangong Rd., Sanmin Dist., Kaohsiung City 807, Taiwan

Email: allwine@fmipa.unila.ac.id

Received : Jan 30, 2025; Revised : Jun 3, 2025; Accepted : Jun 22, 2025; Published : Jun 30, 2025

Abstract

Diabetic retinopathy (DR) is a pressing global health issue that affects the retina and is closely linked to diabetes, leading to vision impairment and blindness, particularly in adults. With the rising incidence of diabetes, the need for efficient and accurate DR screening is critical for early intervention and improved patient outcomes. Automated screening solutions can streamline this process, allowing healthcare professionals to focus more on patient care. In this study, we harnessed advanced deep learning techniques, specifically 3D convolutional neural networks (3D-CNNs), to classify DR into binary categories (presence or absence) and five multiclass categories: mild, moderate, no DR, proliferative DR, and severe DR. Our goal was to enhance diagnostic Precision in ophthalmology. To optimize our models, We embraced two methods transformative data augmentation: random shifting and random weak Gaussian blurring, empowering our model to reach new heights, as well as their combination. Our results showed that, for binary classification, the combined augmentation achieved significant success, The multiclass model was trained without any data augmentation excelled in Precision. These findings highlight the importance of large, high-quality research datas in deep learning algorithms. By leveraging advanced methodologies and robust data, we can transform diabetic retinopathy screening, promoting earlier detection and better treatment outcomes for those affected.

Keywords : 3D Convolutional Neural Networks (3D-CNNs), Binary Classification, Data Augmentation, Deep Learning, Diabetic Retinopathy (DR), Diagnostic Precision, Multiclass Classification, Ophthalmology.

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. INTRODUCTION

The disorder of diabetes affects the internal sugar-regulating mechanism of the body. In 2017, over 451 million people suffered from the condition. High blood sugar levels cause serious harm to the body's organs and are responsible for severe complications such as heart attacks, blindness, cataracts, glaucoma, retinal detachments, and dementia. An increasing number of individuals suffering from any irrespective age group of diabetic retinopathy, or DR, are experiencing vision problems [1], [2], [3], [4].

While retinal damage does not show up in gross examination, their diagnosis tends to become quite obvious through a simple clinical examination. One can say, in a sense, that retinal injuries may be classified into four stages: mild, moderate, severe, and proliferative retinopathy. Microaneurysms, which are outpouching-like formations, develop in very small retinal veins in early phases, while the same vein would be found completely occluded in moderate stage. The most severe one may cause retinal detachment [2].

The literature presents an excellent number of descriptors in the extraction of aspects from DR scans [5], [6], [7]. Almost always, these methods are focused on automatic detection of the DR lesions [8], [9], [10]. CNNs were also used in this research [11] and they were able to define images of DRs

trained in data augmentation/an enhancement technique, achieving Precision of 75% on the validation data. Shanthi and Sabeenian [12] used a customized AlexNet architecture to classify ophthalmoscopic images from the Messidor research data.

Khanish et al. in [13] applied knowledge transfer models like the VGGNet, AlexNet, ResNet and GoogleNet and achieved a recognition rate of 95.68% vis-a-vis the publicly accessible Kaggle hub. In [14], A complete patch-based convolutional neural network (CNN) framework is developed utilizing a mere 28 ophthalmoscopic images and reaches An Precision rate of 94.0%. In [2], an attempt was made by the authors to developed a 3D capsule network and confirmed its effectiveness using the Messidor research data. Their model achieved an impressive Precision of 98.64% when tested At stage-3 in ophthalmoscopic images. In [4], the authors developed a very deeply unified network to classify EyePACS and Messidor-2 research datas achieving an Precision rate of 88% for EyePACS and 95% for Messidor-2. In [15], some researchers used CNNs that 87% on the Messidor-2 research data and achieved 90% true positive rate on the EyePACS research data. Sayres et al. showed utilization of DL models which achieved a True positive rate of 91% on the Messidor-2 research data and a true positive rate of 94.5% on the EyePACS Research data. In [16], the authors applied the VGG-19 architecture in knowledge transfer techniques on 9 retinal pathologies and a single normal retina class, achieving a total Precision of 30.5% on confounding factors such as Transformation, orientation, and luminance adjustment alteration notwithstanding a limited sample size. The investigator applied A deep neural network (DNN) with convolutional layers model in 23 to the EyePACS research data and achieved a whopping A detection rate of 98% by using several augmentation techniques like shearing, flipping, rotation, cropping, zooming, translation and Krizhevsky augmentation. Shankar et al. had developed the deep learning model using a collaborative network and histogram analysis-based segmentation that achieved an Precision of 99.28% for the grouping activity against the Messidor research data. Beede et al. examined the clinical characterization surrounding distributed eye screening processes. Multiple algorithms have been suggested to explain scan reports [5], [6], [7] to diagnose diabetic retinopathy. Typically investigators have focused predominantly on the automated detection of lesions relating to diabetic retinopathy.

Knowledge transfer models were used by the authors in [13]: VGGNet, AlexNet, ResNet and GoogleNet-to achieve a recognition Precision of 95.68% with the publicly available research data repository on Kaggle. A whole patch-oriented CNN network is created In [14], with the use of only 28 ophthalmoscopic images to achieve a true positive rate value of 0.940. The authors developed A 3D capsule architecture in [2] and tested it using the Messidor research data, obtaining Precision of up to 98.64% from ophthalmoscopic images from stage 3. In [4], an extensively connected and deep network was built for the grouping of EyePACS and Messidor-2 research datas, achieving a accuracy of 88% for EyePACS and 95% for Messidor-2. In [15], investigators employed CNNs and noted a true positive rate of 87% on the Messidor-2 research data, and of 90% on the EyePACS research data. Sayres et al [17] exploited Deep Learning models that attained a true positive rate of 94.5% on the EyePACS research data and a true positive rate of 91% on the Messidor-2 research data. The research highlighted in [16] used a knowledge transfer method with the VGG-19 framework to detect 9 retinal conditions along with One class of normal retina an Precision of 30.5% over all ten classes while performing brightness modification augmentation, rotation, and translation methods, despite a low sample size. The researcher utilized a deep neural network with convolutional layers framework [18] on the EyePACS research data and achieved a true positive rate of 98% while applying methods such as shearing, rotation, zooming, flipping, cropping, translation to enhance the data, and Krizhevsky augmentation. Shankarel al. [19] presented a deep learning model utilizing Histogram analysis-based segmentation alongside a collaborative network to classify the Messidor research data with 99.28% precision.. Beede et al. [20] investigated the clinical details of eye screening methods. A variety of techniques have been devised for

examining Analyze reports for identify DR [5], [6], [7]. In general, the focus of researchers has been automatic lesion detection related with DR [8], [9], [10]. In [11], a CNN has been used to classify diabetic retinopathy images, obtaining 75% Precision on the validation database using computational data data enhancement techniques. Sabeenian and Shanthi in [12] categorized ophthalmoscopic images using a customized AlexNet architecture framework tested with the Messidor database. The researchers in [13] achieved a recognition rate of 95.68% by employing knowledge transfer paradigms such as VGGNet, GoogleNet, ResNet and AlexNet on the public Kaggle hub. A comprehensive patch-oriented CNN framework created in [14] working with just those 28 ophthalmoscopic images, achieving a true positive rate of 0.940. Referring to [2], the authors created 3D capsules and showed the Messidor-based model research data, attaining 98.64% Precision for phase 3 ophthalmoscopic photographs. In [4], a comprehensive and intricately tied the architecture was established To assign between the Messidor-2 and EyePACS research datas, achieving a precision of 88% for EyePACS and 95% for Messidor-2. In [15], CNNs were employed to yield a true positive rate of 87% on the Messidor-2 research datas and 90% on the EyePACS research datas. Sayres et al. [17] implemented DL models that yielded a high true positive rate of 94.5% on the EyePACS, and a higher true positive rate of 91% on Messidor-2 research datas. The research in [16] applied a knowledge transfer-based VGG-19 architecture method for The grouping of 9 retinal pathologies and a single healthy retina class by using a limited sample size, and reported an Precision of 30.5% across all ten categories by using brightness alteration, translation, and rotation as data enhancement techniques. The researcher implemented An advanced CNN framework [18] on the EyePACS research datas, attaining true positive rate of 98% utilizing shearing, rotation, zooming, flipping, cropping, translation as augmentation, and Krizhevsky data enhancement techniques.

Shankar et al. [19] used ntensity-based segmentation with a Collaborative network attaining 99.28% precision in classifying the Messidor research data. Beede et al. [20] carried out research characterizing workflows for ophthalmic screening. Minimal example learning has become the subject Of great significance because it concerns classifying categories with very few examples, notably categories that do not appear Throughout the training process [21], [22], [23]. Optimization forms the basis of few-shot learning approaches. Such systems rely on gifted inductive bias and need intricate inference techniques. Typical methods for this form of learning include meta-learning, augmentation/generative techniques, knowledge transfer, and semi-supervised approaches. In addition to health-related opportunities, the academic researchers and industrial scholars have explored other scientific, technological, and engineering subjects through Deep learning-based approaches [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]. Though the analyses presented provide all workable means for 2-class and Multiple-class classification of DR, the bulk of them have leveraged data in the humble 2D realm. Higher dimensions, on the other hand, 3D come along with copious scale and geometric information, posing some challenging opportunities for image recognition algorithms [39], [40], [41]. Research, therefore, must bear in mind to exploit higher-dimensional data in the aforementioned tasks.

To implement representation learning techniques with a low sample size [42] with data expansion, we employed a 3D-convolutional neural network design [43] In the spatial realm for both multi-category classification tasks using the DR research datas. Different data augmentation techniques, weak random Gaussian blurring, and random shifting were used in different combinations in order to analyze the various effects these kind of augmentation techniques would pose on a classification activity. This research plays a useful part in contributing to the current literature on DR classification. As far as the authors are aware, very few works available in academic literature have attempted to resolve this problem in three-dimensional space. This research mainly performs what performs and benefits itself from two different aspects of discussions taken simultaneously regarding spatial and temporal. Finally, it is worth elaborating to find out what impact different data data enhancement techniques have more

significantly on 3D classification performance. Another promising area for research is few-shot learning, where only a limited number of images may be available for working with in performing any DL technique. The remaining paper is properly structured as follows: Section 2 has a short description on the research datas, Section 3 is a discussion on the methods applied in the research, Section 4 gives the detailed descriptions with respect to all experiments being conducted, Section 5 includes findings with discussion, and Section 6 proposes conclusions from the work discussed.

1.1. Research data Description

For experimentation and evaluation of retinal image analysis and diabetic retinopathy identification, we used two independent research datas in our study. The first research data, known as TeleOphta, contains an array of ophthalmoscopic images that showcase specific pathological features, namely microaneurysm lesions and exudates. From this rich collection, we meticulously constructed an additional 83 3D volumes representing subjects and 99 3D volumes representing healthy subjects with visible signs of disease, such as exudates and microaneurysms. Most importantly, each of these volumes was organized at the subject level, ensuring that the data reflected each patient conditionally without overlap. In order to have a more robust data-set and improve our results, we used augmentation techniques, namely random shifting and weak Gaussian blurring. These approaches allowed for increased variety and volume of data to analyze. Sample images from the TeleOphta research data can be seen in Figure 1. To explore the retinal structures in a neighbour hood around the measurement area, each volume in this research data is acquired over $210 \times 210 \times 12$ voxels.

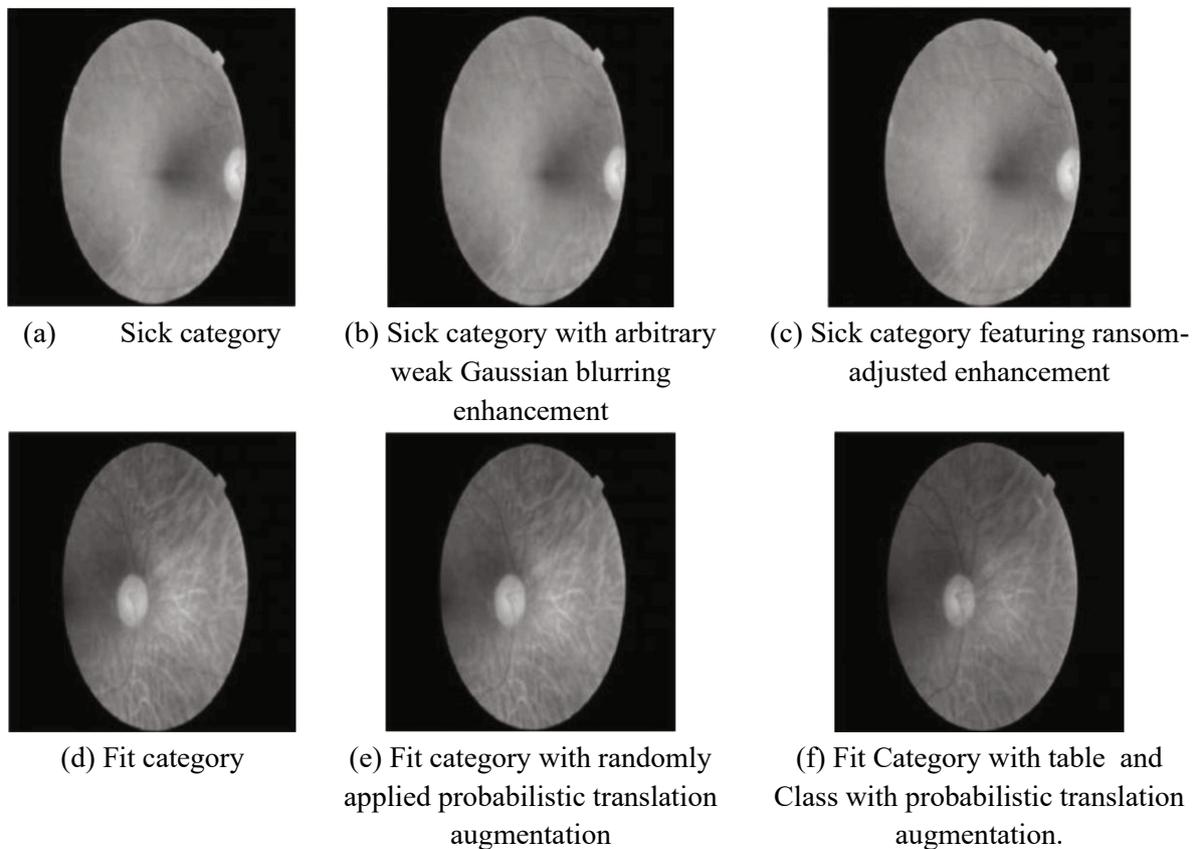


Figure 1. Discover a collection of sample images from the TeleOphta database that showcase innovative augmentation techniques.

The second research data includes images from retina scans filtered with a Gaussian process to assist different computer vision tasks; these images are pivotal in the diagnosis of diabetic retinopathy. This research data has five label levels: (no retinopathy, Minor, Intermediate, Critical and Complicated proliferative retinopathy). In short, we managed to construct 262 3D volumes per category, and to achieve more clarity and avoid subject duplicity, we also split the samples at the subject level. Figures 2–4 show sample images that highlight these varying degrees of diabetic retinopathy. Similarly to the TeleOphta research data, we performed random shifting and weak Gaussian blurring augmentations to this research data, so that the research data variability and capability of training would increase. The research data was obtained from Kaggle, a popular website with many research datas for different machine learning tasks.

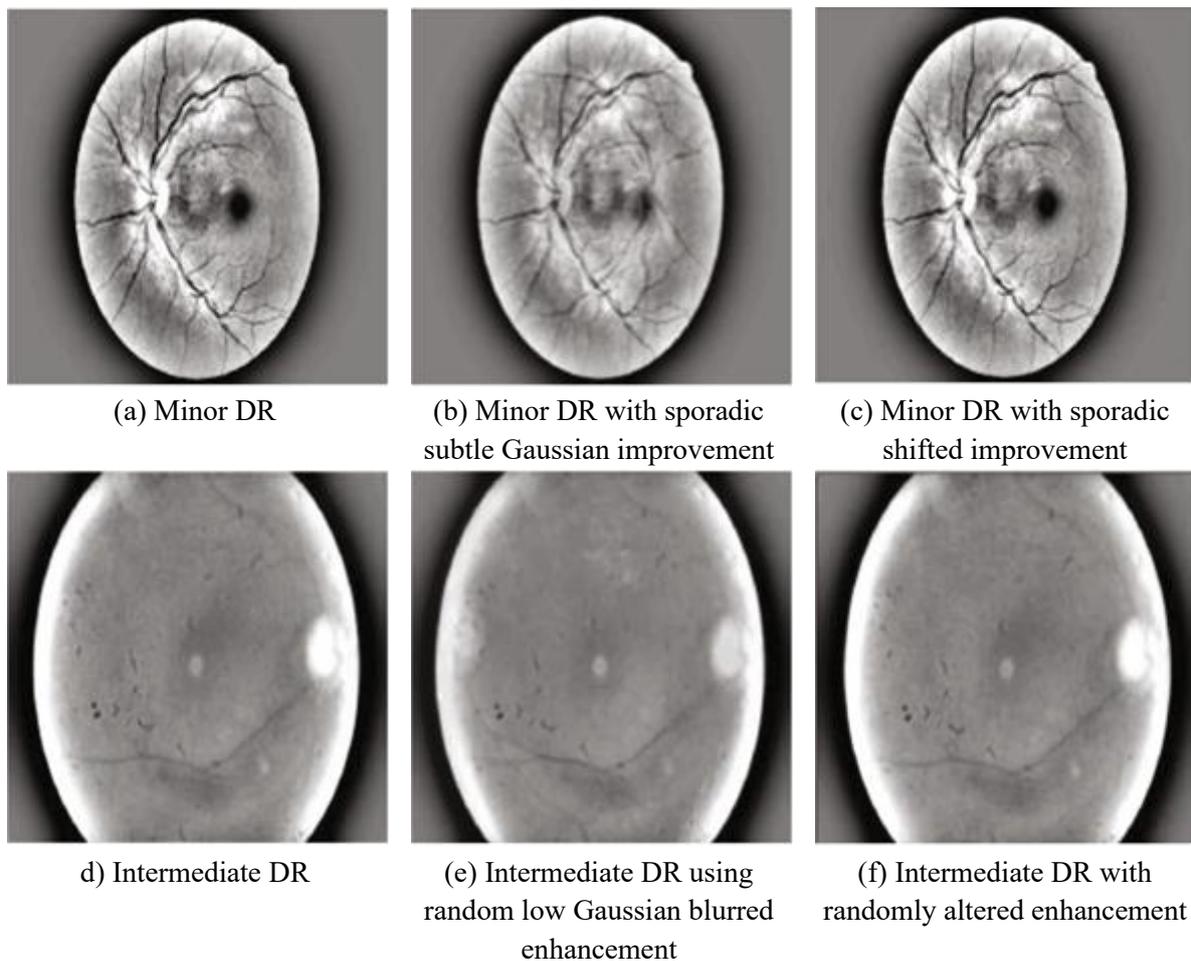


Figure 2. Illustrative examples of Minor and Intermediate classes from the Kaggle DR database, showcasing the effectiveness of randomly altered and lightweight Gaussian blurred enhancement.

3D volumes to values in the common range of 0 to 255 intensity values. This process was integral for structuring the data in a format conducive for the next steps of analysis and data processing, leading to better model training and evaluation

2. METHOD

In this study we utilized a variety of 3D-CNN models to deep dive into 2-class and Multiple-classclassification tasks, with a focus on the complex structures of these models. The application of 3D-

CNN architectures and their configurations is best exhibited through figures 5 and 6, alongside various other experiments, this is the reason why these figures hold critical importance in this research.

Figure 5 displays a brief yet concise analysis of the attributes of the models under consideration – data augmentation, a combination of multiple techniques, as well as shifted weak Gaussian blurring. The number of convolutional layers in feature map is adjusted by these models, all while utilizing a combine augmentation scheme that incorporates 10 convolutional feature maps, which is then enhanced with 8 and 9 feature maps through weakening and shifting algorithms respectively. The difference in feature maps account for the slight differences in architectural layout.

The input cubic data of the volumetric data that is provided to the model has a prime size of 210 x 210 x 12. This data is first subjected to a combination of rescale and zero one normalization which takes all the input values between 0 and 1. A normalization process is done based on the specific minimum and maximum values for each specific channel and ensures that the data is clean enough to be processed and analyzed. We employ the recurrent structural unit, block-A, after the said normalization process has been applied. Such blocks incorporate vital parts appended several times which in case of block-A is five. Block-A has a 3D convolution layer which helps build meaningful and spatial features from the data provided as an input. Also noticeable is the batch normalization layer which normalizes the output of multiple previous layers to complete the process with the aim to stabilize and speed up the entire training time. After that, there is a rectified linear unit layer which brings an appealing feature that enables the network to learn complicated data patterns, that feature chance-level element as non-linearity will be incorporated. Finally, a downsizing 3D max pooling layer is introduced and added which aids in retaining the relevant features and reducing the dimensionality mods even further.

Upon completing Block A, we bring forward block B which we only do once. This block has three fully connected FC layers having neuron counts of 300, 150 and 2. These layers are integrated into the model as well, since they aid in the grouping decision making. To reduce the risk of overfitting, we apply a dropout layer with a 0.1 rate. This layer changes the state of some training units which helps to train a more generalized model. The architecture ends with the a classification layer and softmax function that accomplishes the last stage of the binary classification task. The model defined as the ReLU activation function mathematically as $\max(x, 0)$ enables the model to hardly learn with negative values as in this only positive activations are brought to use.

To boost the effectiveness of the learning procedure even more, we apply batch normalization which minimizes the variation among the fundus volumes. This method improves the rate of training by making the learning environment more uniform thereby enabling a very fast convergence with less dependence on the starting conditions. In addition, the dropout policy is a particular regularization strategy that is useful to address the problem of overfitting to the training research data and consequently helps the model get adapted to new data that has never been seen before. Moreover, L2 regularization is also introduced to the weight and bias values to yield smaller values by penalizing large sizes of weights and biases. The output post-processing with the softmax is performed such that the output values are considered as probabilities making it simple to understand what the outputs of the particular classification are. Discover the crucial hyperparameters of our architecture of the novel 3D-CNN model intended for the outstanding results during a binary classification by the aid of the eight characteristic maps in the convolutional layers which are highlighted in Table 1.

The Architectural Differences Between Models Using Different Data Augmentation Techniques is Further Expounded in Figure 6, Providing Key Insights into Our Approaches. These convolutional layers have different input feature map sizes based on the augmentations that are being used in each of the respective architectures: 12 feature maps through each of the architectures that employed combined augmentations, 10 feature maps for the architectures that were augmented with no augmentations at all,

and 11 appear through classifications With probabilistic translation augmentations and random weak Gaussian blurring. Other than these couple of variants, the core components of the architectures remain the same, providing confidence in the robustness of our design paradigm.

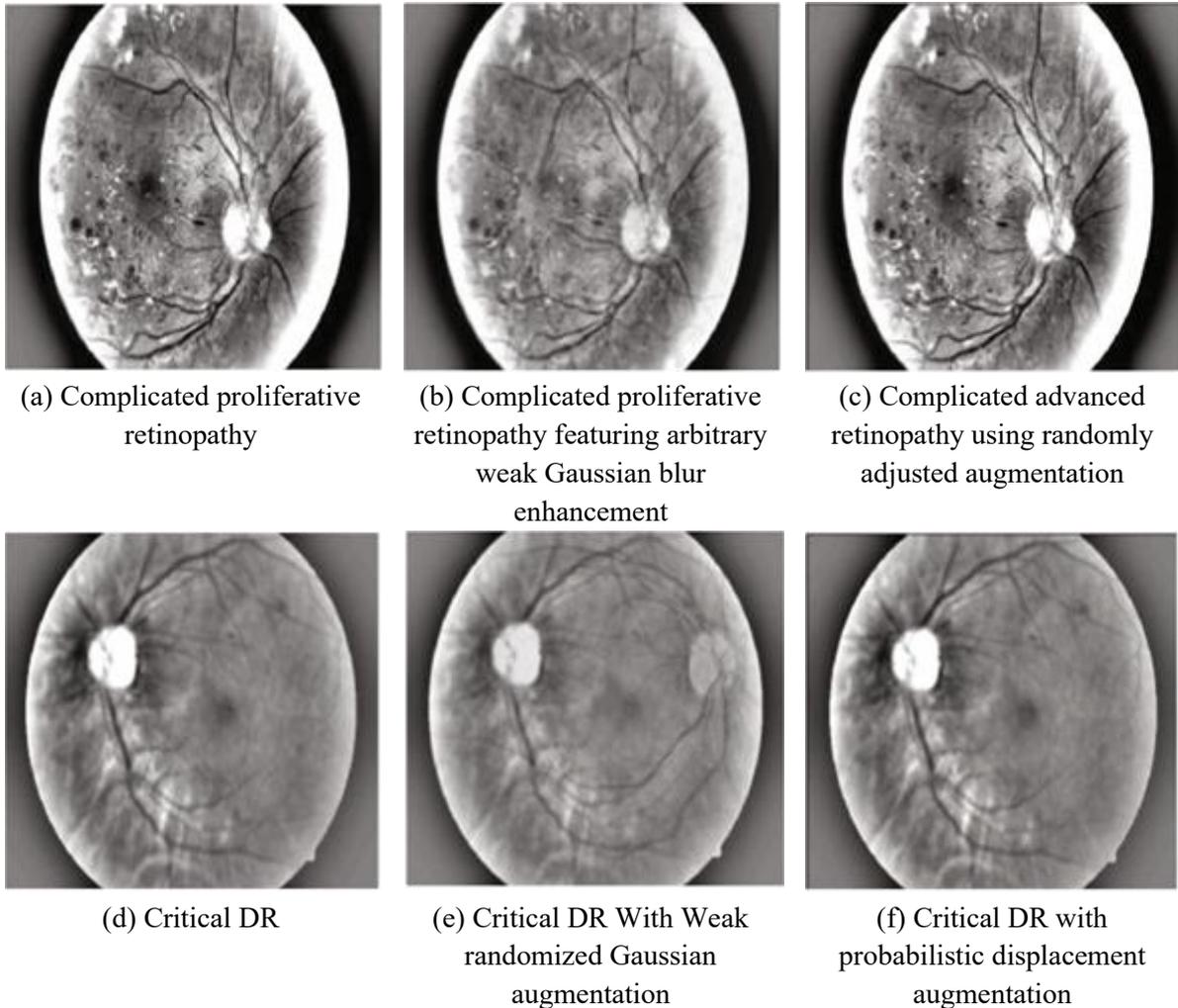


Figure 3. Definitive samples of critical and Complicated advanced retinopathy from the Kaggle DR database, utilizing randomized shifted and incidental mild Gaussian blur augmentations.

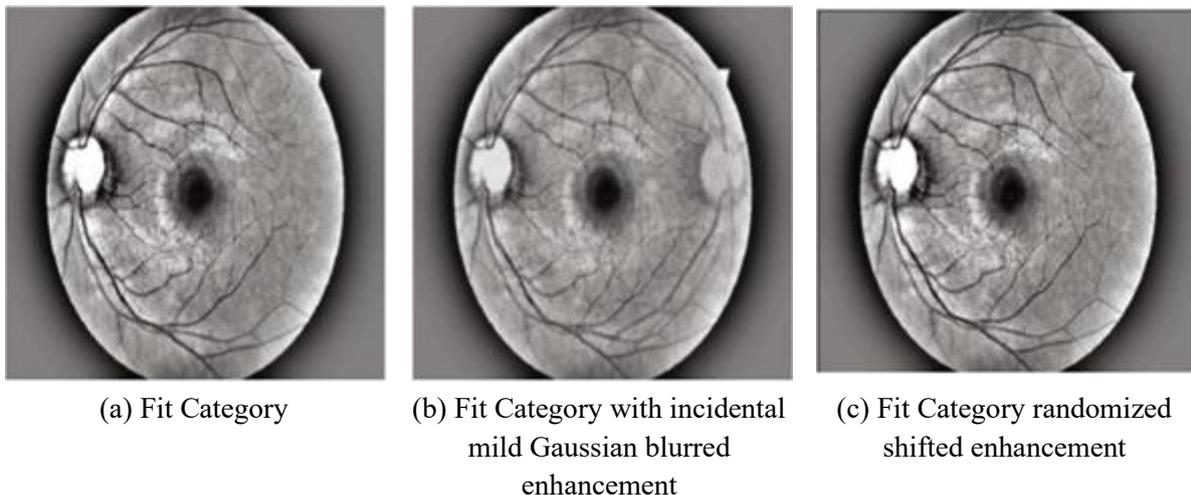


Figure 4. Sample images from the "Fit Category" class in the Kaggle Diabetic Retinopathy database. These images have been altered using with incidental mild and randomized shifted enhancement to enhance the research data and improve model robustness.

Being built around this architecture, an input layer of the network is capable of accepting volumetric data of dimensions $512 \times 512 \times 2$, operating with a zero center normalization onto the 3D input. In this architecture, block-A is yet again repeated six times, same as in the previous model, where this A convolutional layer within the block performs feature extraction, followed by an exponential linear unit (ELU) activation layer and batch normalization layer. This avoids the vanishing gradient problem. Just like before, a max-pooling layer is used to condense and refine the extracted features for further processing.

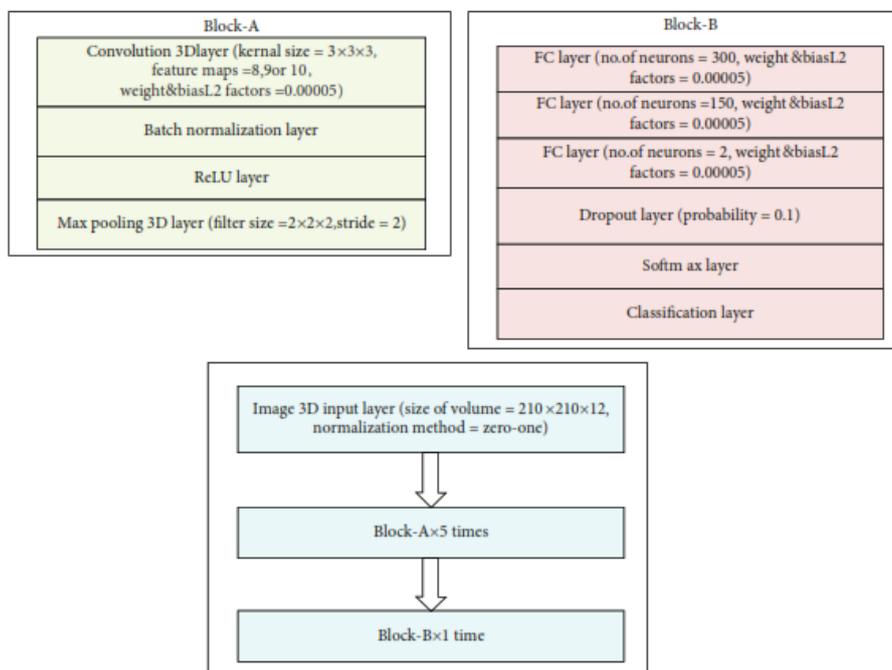


Figure 5. Binary classification (sick/healthy) set-ups with random Gaussian blurred augmentations and random shift applied without any augmentation, along with combined augmentations. Augmenting features will be selected utilizing a tenfold cross-validation method for examining hyperparameter configurations.

After our familiar block-A, we can see that there is another block, let's call it block B, which contains three fully connected layers with number of neurons 500, 300, 5. There is a dropout layer in the same block with a probability of 0.1 that was also implemented to prevent overfitting. The architecture ends with a classification layer and a softmax function designed for the multi-class classification problem of which there are five classes. Using ELUs is especially beneficial, as it enables the model to deal with negative values, thus tackling the issues related to the vanishing gradient problem while simultaneously ensuring a reduction in computational complexity.

$$ELU(0 < \alpha): \begin{cases} \alpha (e^x - 1), & x < 0, \\ x, & x \geq 0. \end{cases} \quad (1)$$

The detailed specification of the hyperparameters to ascertain the architecture used is given in Table 2 concerning the 3D-CNN architecture applied in the multiclass classification problem. Therefore,

this architecture contains 10 feature maps in the 3D convolutional layer, where these maps capture different features from the input data to produce better classification performance.

2.1. Experiments

In a spatial domain study, we conducted experiments on 2-class and Multiple-class classifications for effective discrimination among various classes of Diabetic Retinopathy (DR). We transformed the original research data in terms of quantity using random shifting and random mild Gaussian blurring as our techniques for data enhancement. The applied amount of random weak Gaussian blurring was set at a value of $\sigma = 1.5$, and a random shifting of either 1 or 2 pixels was applied. We then combined training samples produced by both data enhancement techniques.

The experiments were carried out with the following key tasks: 1) Binary classification without augmentation for accurate distinction between healthy and diseased categories; 2) Auxiliary experiments performing random weak Gaussian blurring on the healthy and diseased categories; 3) Binary classification using random shifting for healthy and diseased samples; 4) Normal binary classification testing healthy and diseased categories using both random blurring and shifting augmentations; 5) Simple classification of multiple classes without augmentation; 6) Basic classification of different classes using applied random weak Gaussian blurring; 7) Classification with multiple classes using various translations 8) Finally, simple multiclass classification that includes both random translations and random weak Gaussian blurring augmentations.

The research data For the activity of binary classification thus consisted of 72 samples for each respective class; 8 samples worked as a validation set. Therefore, we kept in the test split 19 healthy and 3 diseased among a total of 140 samples available. During testing experiments carried on this subset, we trained using the whole research data made of a total of 80 samples per class; the validation was made on previously unseen instances of the test partition. Such a meticulous approach allowed to test the efficiency of our methods with highest scrutiny.

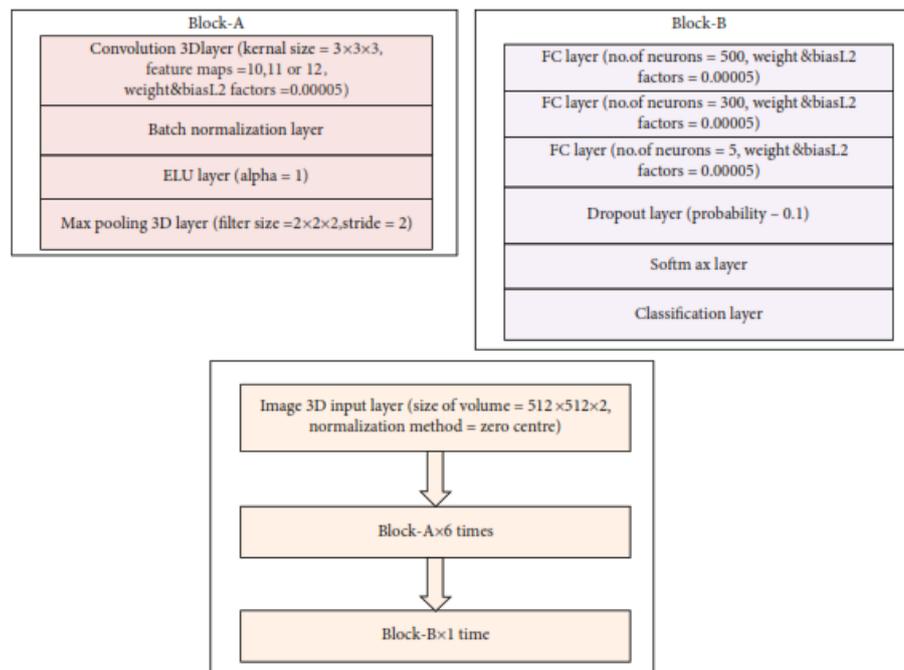


Figure 6. Innovative constructs for multi-class classification tasks capitalize on random shifts and the scarcely discernible, yet highly artistic, random weak Gaussian blur with some dynamic combined augmentation strategy. Cross-validation enabled meticulous consideration of hyperparameters using a ten-times repeated methodology

Table 1. Summary of the hyperparameters for the 3D-CNN model, which is aimed at binary classification and includes 8 maps from the convolutional layers.

Layer Name	Type	Kernel Size	Number of Kernels	Step Size	Padding	Rate of dropout	Input Dimensions	Output Dimensions
Conv1+BN+ReLU	Convolution + BN	3 × 3 × 3	8	1	Same	N/A	210 × 210 × 12	210 × 210 × 8
MaxPool1	Max Pooling	2 × 2	N/A	2	N/A	N/A	210 × 210 × 8	105 × 105 × 8
Conv2+BN+ReLU	Convolution + BN	3 × 3 × 3	8	1	Same	N/A	105 × 105 × 8	105 × 105 × 8
MaxPool2	Max Pooling	2 × 2	N/A	2	N/A	N/A	105 × 105 × 8	53 × 53 × 8
Conv3+BN+ReLU	Convolution + BN	3 × 3 × 3	8	1	Same	N/A	53 × 53 × 8	53 × 53 × 8
MaxPool3	Max Pooling	2 × 2	N/A	2	N/A	N/A	53 × 53 × 8	27 × 27 × 8
Conv4+BN+ReLU	Convolution + BN	3 × 3 × 3	8	1	Same	N/A	27 × 27 × 8	27 × 27 × 8
MaxPool4	Max Pooling	2 × 2	N/A	2	N/A	N/A	27 × 27 × 8	14 × 14 × 8
Conv5+BN+ReLU	Convolution + BN	3 × 3 × 3	8	1	Same	N/A	14 × 14 × 8	14 × 14 × 8
MaxPool5	Max Pooling	2 × 2	N/A	2	N/A	N/A	14 × 14 × 8	7 × 7 × 8
FC 1	Fully Connected	N/A	300	N/A	N/A	N/A	7 × 7 × 8	300
FC 2	Fully Connected	N/A	150	N/A	N/A	N/A	300	150
FC 3	Fully Connected	N/A	2	N/A	N/A	0.1	150	2
Dropout	Regularization	N/A	N/A	N/A	N/A	0.1	N/A	N/A
Softmax	Activation Function	N/A	N/A	N/A	N/A	N/A	2	2

FC (Fully Connected): Involves the connections of every neuron to all others in the next layer. MaxPool: Reduces the size by taking the maximum value from certain areas of an image. BN: Batch Normalization; sets the input values to improve the speed and stability in the training process. ReLU (Rectified Linear Unit): This is an activation function that outputs the same value for positive input or 0 if not. Conv: A convolution operates with a filter applied to detect patterns in the data.

Table 2. Hyperparameters for multi-class classifications in the proposed 3D-CNN model with ten layers of convolution feature maps were established.

Layer Name	Type	Kernel Size	Number of Kernel	Step Size	Rate of dropout	Input Dimensions	Output Dimensions
ELU+ BN +Conv1	Convolution + BN	$3 \times 3 \times 3$	10	1	N/A	$512 \times 512 \times 2$	$512 \times 512 \times 10$
MaxPool1	Max Pooling	2×2	N/A	2	N/A	$512 \times 512 \times 10$	$256 \times 256 \times 10$
MaxPool2	Max Pooling	2×2	N/A	2	N/A	$256 \times 256 \times 10$	$128 \times 128 \times 10$
Conv2+BN+E LU	Convolution + BN	$3 \times 3 \times 3$	10	1	N/A	$256 \times 256 \times 10$	$256 \times 256 \times 10$
Conv3+BN+E LU	Convolution + BN	$3 \times 3 \times 3$	10	1	N/A	$128 \times 128 \times 10$	$128 \times 128 \times 10$
MaxPool3	Max Pooling	2×2	N/A	2	N/A	$128 \times 128 \times 10$	$64 \times 64 \times 10$
Conv4+BN+E LU	Convolution + BN	$3 \times 3 \times 3$	10	1	N/A	$64 \times 64 \times 10$	$64 \times 64 \times 10$
MaxPool4	Max Pooling	2×2	N/A	2	N/A	$64 \times 64 \times 10$	$32 \times 32 \times 10$
Conv5+BN+E LU	Convolution + BN	$3 \times 3 \times 3$	10	1	N/A	$32 \times 32 \times 10$	$32 \times 32 \times 10$
MaxPool5	Max Pooling	2×2	N/A	2	N/A	$32 \times 32 \times 10$	$16 \times 16 \times 10$
Conv6+BN+E LU	Convolution + BN	$3 \times 3 \times 3$	10	1	N/A	$16 \times 16 \times 10$	$16 \times 16 \times 10$
MaxPool6	Max Pooling	2×2	N/A	2	N/A	$16 \times 16 \times 10$	$8 \times 8 \times 10$
FC 1	Fully Coupled	N/A	500	N/A	N/A	$8 \times 8 \times 10$	500
FC 2	Fully Coupled	N/A	300	N/A	N/A	500	300
FC 3	Fully Coupled	N/A	5	N/A	N/A	300	5
Dropout	Regularization	N/A	N/A	N/A	0.1	N/A	5
Softmax	Activation Function	N/A	N/A	N/A	N/A	5	5

BN indicates batch normalization; Conv represents convolutional layers; FC stands for fully connected layers; ELU refers to exponential linear unit; and MaxPool is shorthand for maximum pooling.

For the multi-category classification problem, the research data were divided into training, validation, and test sets for convenience in modeling. The training partition consisted of 225 samples for each category, which is a reasonable number to furnish the model with considerable attributes for learning emanating from several instances. The model was validated using 25 samples per category. This enables the assessment of model performance on unseen data during training. A test split was also

developed that included 12 samples for each class, which would be employed to estimate the final performance of our model once it underwent training and validation.

In attempts to boost performance in binary classification activity, several parameters were adjusted. The minibatch size was set at 2, allowing model parameters to be updated frequently, and perhaps motivate the process of training. The starting value of the learning rate was again selected to be 0.001 in order to balance speed of learning versus stability. We determined a total of epochs to be fixed at 50, which would provide enough time for the model to converge to learn features of training data properly. In the training, we employed a piecewise learning rate schedule with the Adam optimizer, widely employed for its adaptive learning characteristics. Cross-entropy was used as the loss function, suitable to classification problems. Conducting 41 experiments, the entire process took a little less than 642 min (or 10.7 h), further manifesting the hard work and dedication we expended through the refinement and thorough assessment of the performance of our model.

For multi-class classification tasks, we created a similar setting to retain consistency and utilize our previous insights. The configuration involved a minibatch size of 2 and an initial learning rate of 0.001, effective according to our previous attempts. We reduced the total number of epochs to 30 to reduce time consumption while still obtaining satisfactory performance. A piecewise learning rate schedule was applied using the same reliable Adam optimization method, still with categorical cross-entropy for the use of our loss function since it handles the multi-class nature of our activity very appropriately. In total, there were 41 experiments run for this setting, requiring approximately 8448 minutes, or approximately 140 hours. This is a very great deal of time to spend on a task and shines a light on our commitment to persevering and doing all we can for the best possible results in our classification tasks.

3. RESULT

The findings that pertain to our research on dichotomizing states of health into healthy and sick categories are summarized in Tables 3 and 4, with the visual representation of the findings being covered by Figures 7 and 8.

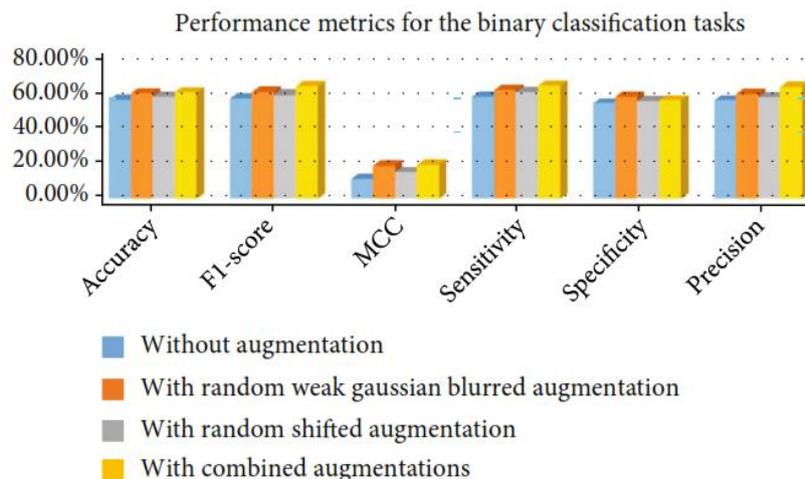


Figure 7. A Compelling Overview Of Metrics Used To Evaluate The Performance Of Binary Classification Models.

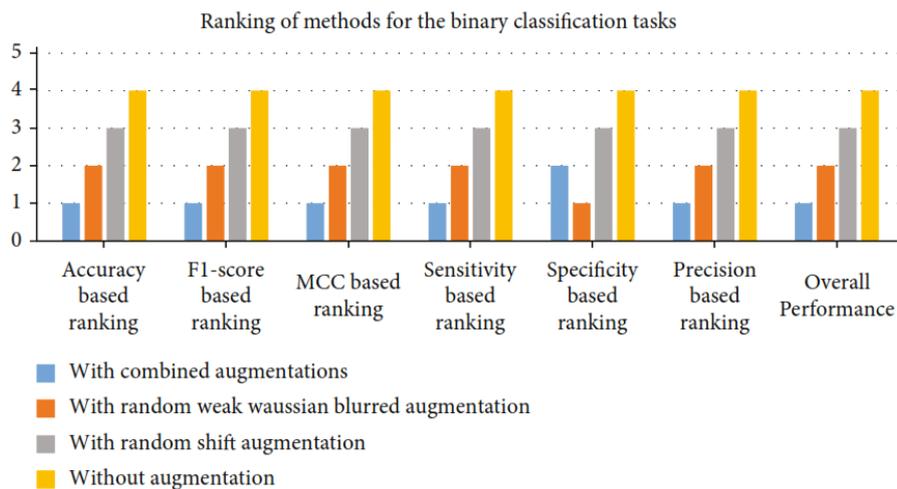


Figure 8. A Comprehensive Ranking Of Binary Classification Methods, Based On The Performance Metrics Outlined

We performed various experiments with different architectures of 3D-CNN trained from scratch and evaluated them from the perspective of performance characteristics, including such measures as Precision, F1-score, (MCC) Matthews correlation coefficient, True positive rate, precision and specificity.

This evaluation can be defined mathematically by:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$F_1 - score = \frac{2TP}{2TP+FP+FN} \quad (3)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (4)$$

$$True\ positive\ rate = \frac{TP}{TP+FN} \quad (5)$$

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Substituting TN for true negative, TP for true positive, FN for false negative samples, and, FP for false positive the ranking of methods for binary classification in relation to metrics is shown in Table 4. We applied the various augmentation techniques in the training of the validation split, as shown in Tables 3 and 4. Now, with regards to the effects, today there are winning combination techniques which outperformed all the others like one for weak Gaussian noise and Shift-based image augmentation. A Random weak Gaussian noise model for augmentation is presented first, followed by the Random Shifted augmentation model. The model with no augmentation was least effective. Hence, all performance measures, with the exception of specificity, gave similar rankings, hence demonstrating a pretty strong correlation among the performance measures.

Table 3. Showcases Quality measures for binary image analysis, comparing results from weak Gaussian blur, no augmentation, random shift, and their combination.

Task	MCC	Precision	F ₁ -score	Specificity	True positive rate	Precision
With Weak randomized Gaussian augmentation	0.1876	59.38%	60.12%	57.5%	61.25%	59.04%
Without augmentation	0.1126	55.63%	56.44%	53.75%	57.5%	55.42%
With blended augmentation operations (plus test split)	0.1908	59.89%	63.32%	55.42%	63.64%	63%
With probabilistic translation augmentation	0.1502	57.5%	58.54%	55%	60%	57.14%

Table 4. Ranks the different methods applied to the Binary image analysis tasks, highlighting their performance based on the same augmentation strategies: no augmentation weak randomized Gaussian blur, random probabilistic translation, and blended augmentation operations.

Performance Metric	Rank 1	Rank 2	Rank 3	Rank 4
Precision	With blended augmentation operations(59.89%)	With Weak randomized Gaussian augmentation(59.38%)	With probabilistic translation augmentation(57.5%)	Without augmentation (55.63%)
F ₁ -score	With blended augmentation operations(63.32%)	With Weak randomized Gaussian augmentation(60.12%)	With probabilistic translation augmentation(58.54%)	Without augmentation (56.44%)
MCC	With blended augmentation operations(0.1908)	With Weak randomized Gaussian augmentation(0.1876)	With probabilistic translation augmentation(0.1502)	Without augmentation (0.1126)
True positive rate	With blended augmentation operations(63.64%)	With Weak randomized Gaussian augmentation(61.25%)	With probabilistic translation augmentation(60%)	Without augmentation (57.5%)
Specificity	With Weak randomized Gaussian augmentation(57.5%)	With blended augmentation operations(55.42%)	With probabilistic translation augmentation(55%)	Without augmentation (53.75%)
Precision	With blended augmentation operations(63%)	With Weak randomized Gaussian augmentation(59.04%)	With probabilistic translation augmentation(57.14%)	Without augmentation (55.42%)
Overall Effectiveness	Using integrated enhancements	Applying a slight Gaussian blur augmentation randomly	By using random shift	Without augmentation

For multiclass classification, we mainly look at overall Precision, MCC, RCI, index of balanced Precision (IBA), index of balanced Precision (IBA), confusion entropy (CEN) and GM. Overall Precision is the ratio of correctly predicted values to the total values. In Tables 5 to 8, one finds the details of the various statistics that were tabulated for multiclass classification. Additional details include class-wise for IBA, CEN, MCC, RCI and GM as well as values of overall Precision across there are four situations: no enhancement, randomly shifted enhancement, randomly weak Gaussian blurred enhancement, and merged enhancement methods. Merged augmentation is defined as an inclusion of

both weak Gaussian noise and Shift-based image augmentation at the same time. We include test subset statistics from a model trained without augmentation.

Table 5. Outlines values for multiclass classification tasks, the overall Precision, (RCI) Relative Class Importance, and class-wise Class Entropy (CEN) assessed across the four augmentation scenarios.

Metric	Method with Highest Value	Highest Value
Overall Precision (%)	Without augmentation	36.64%
RCI	The model trained on original research data was used to perform the validation of the test set	0.1522
(Weak) CEN	With probabilistic translation augmentation	0.8259
(Medium) CEN	With probabilistic translation augmentation	0.8834
(Normal) CEN	The test set was evaluated using a model that was developed without any augmentation	0.9175
(Increase) CEN	Utilizing random weak Gaussian blur augmentation	0.7008
(Significant) CEN	Without augmentation	0.8015

Table 6. Presents the Importance-Bias-Agnostic (IBA) values for each class in the multiclass classification tasks, providing a comparative view of the different augmentation conditions.

Method with Highest Value	Metric	Highest Value
Validation of the test set was carried out on the model that was trained without any augmentation	(Weak) IBA	0.1475
Validation of the test set was carried out on the model that was trained without any augmentation	(Medium) IBA	0.1974
Without augmentation	(Normal) IBA	0.1232
Without augmentation	(Increase) IBA	0.4747
Validation of the test set was carried out on the model that was trained without any augmentation	(Significant) IBA	0.1388

Table 7. Values for multiclass classification tasks Class-wise Geometric Mean with and without augmentation randomly selected combined and Gaussian Blur with random shift augmentations for the mentioned approaches.

Metric	Highest Value	Method with Highest Value
(Weak) GM	0.5297	Without augmentation
(Medium) GM	0.6038	Validation of the test set was carried out on the model that was trained without any augmentation
(Normal) GM	0.5131	Without augmentation
(Increase) GM	0.7558	Without augmentation
(Significant) GM	0.527	Validation of the test set was carried out on the model that was trained without any augmentation

Table 8. Class-specific Matthews correlation coefficients are provided for each class-specific classification in the tasks treated through the same augmentation conditions

Method with Highest Value	Metric	Highest Value
Without augmentation	(Weak) MCC	0.1803

Validation of the test set was carried out on the model that was trained without any augmentation	(Medium) MCC	0.3015
Without augmentation	(Normal) MCC	0.1599
Without augmentation	(Increase) MCC	0.4756
Validation of the test set was carried out on the model that was trained without any augmentation	(Significant) MCC	0.1666

Table 9 The summarizes performance metrics Multi-class predictive modeling used, whereas Figure 9 visually represents this data. In this table, IBA, CEN, MCC, and GM were averaged Calculating the mean by summing the class values and then dividing by 5. These scores for overall Precision and the reliability classification index were consistent with those presented in Table 5. The ranking of multiclass classification tasks from Table 9 is presented in Table 10. Figure 10 shows a visual representation of the data found in Table 10. This table provides a ranking for each performance measure and a composite ranking, based on all measures together. For overall Precision, larger values for RCI, IBA, GM, and MCC indicate better classification, while smaller values for CEN indicate better classification.

Table 9. An overview of the performance metric statistics for multiclass classification tasks We proceed unevasively to report the results not using augmentation, mild Gaussian random blur, random shifts, and both augmentations combined..

Augmentation Method	Precision (%)	RCI	Avg. CEN	Avg. IBA	Avg. GM	Avg. MCC
Without Augmentation	36.64	0.0867	0.75462	0.18354	0.54408	0.20144
Mild Gaussian Random Blur	31.04	0.038	0.8043	0.12924	0.50348	0.13692
Random Shift	30.56	0.0546	0.79836	0.1327	0.49558	0.13084
Combined Augmentations	33.12	0.0647	0.77726	0.14634	0.5209	0.16424

The best-performing model, according to Table 10, was trained without data augmentation, with combined weak Gaussian noise and Shift-based image augmentation, following closely behind. The without-augmentation model celebrates the strength of simplicity, while the combined augmentation itself proves that ingenuity can indeed yield brilliant overall performance. The balanced performance of random weak Gaussian and random shifting leveling brings to mind the whole journey of exploration and adaptation. It is interesting to see the close wedge of performance metrics, which shows that sometimes clarity comes from focus alone. If we dig down into specific metrics further, the wealth of insights revealed is nothing less than stunning; random shifted methods choke up the GM, MCC, and overall Precision, while random weak Gaussian methods suffer from a RCI, CEN, and IBA. These differences paying tribute to the fact that various performance metrics embody distinct perspectives serves as a reminder that they often provide very separate avenues towards achieving success. In addition, such discrepancies highlight how CNN perfectly fit and work effectively to the trials posed by minor image transformation techniques that complement growth and endurance [44].

Deep learning architectures have demonstrated greater advantages in terms of performance not only in binary classification tasks compared to multiclass classification tasks according to the comparison presented in some work. These findings can be accounted for by the fact that binary classification context models have a simpler and commonly different decision boundary to learn. The binary classification context also denotes a boost in performance by models when they utilize a mixed data augmentation method compared with those based on one augmentation method. Less surprisingly, it came out that the architecture utilizing mild probabilistic Gaussian blur outperformed the probabilistic

displacement augmentations. This is attributable to the fact that convolutional operations greatly depend on the invariance of feature maps to translation, and, thus, bad handling of translation would weaken model generalization. It is also interesting that the structural design of model architectures hugely affects classification performance. For binary classification tasks, the more feature maps used by a model in the convolutional layers, the more likely it is to achieve high performance. This seems to suggest that richer feature representations are helpful to these learning processes. It is clear that combined augmentation strategies that employ a combination of multiple augmentation techniques make for larger Convolutional feature maps in the model architecture while providing enhanced performance. On the contrary, multiclass classification tasks have shown that fewer feature maps can offer overall better performance. This could indicate that, in certain situations, simpler models are just as effective or more so, given that they are not overly burdened by excessive data augmentation. Work in the reverse direction and analyze whether or not augmenting the data would work better than the other models. The models which are not held to data augmentation perform better only if they have a comparatively lesser number of feature maps. However, it should also be noted that, in multiclass classification tasks employing mixed data enhancement techniques, models with more feature maps tend to achieve higher performance. This illustrates that the complex optimization of model architecture makes simplicity a boon in certain cases, while complexity might yield favorable results in other cases. To summarize, our findings corroborate the assertive contention of deploying deeper models compared to shallower counterparts in binary as well as multiclass classification tasks. Of note is that it is the sample size for training and validation during the experiments that had been small and possibly retarded the models' power to generalize well enough, thus putting a limit on the performance of the DL architectures.

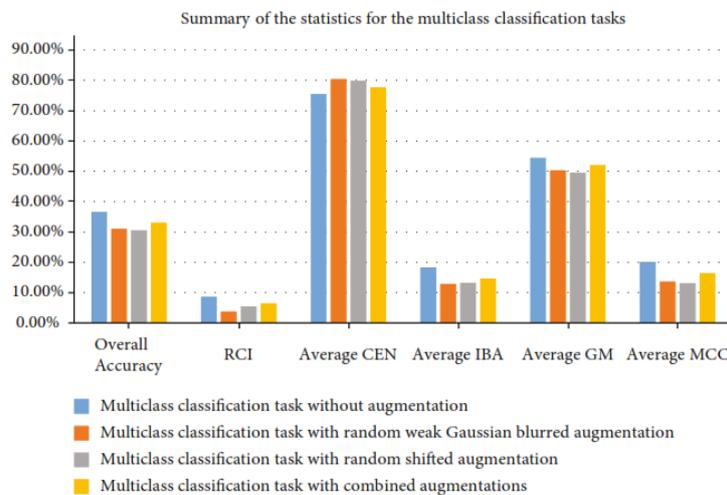


Figure 9. A graphical expression summarizing the indicator performances for the tasks studied in multi-label image classification

Table 10. Ranking of performance for methods used in multi-label image classification random weak Gaussian blur augmentation without augmentation, mixed, and random shift augmentations.

Scenario	RCI Ranking	CEN Ranking	IBA Ranking	GM Ranking	MCC Ranking	Overall Ranking
Without Augmentation	1	1	1	1	1	1
With Weak randomized Gaussian Blur	4	4	4	4	4	3

With probabilistic translation augmentation	3	3	4	4	4	3
With blended augmentation techniques	2	2	2	2	2	2

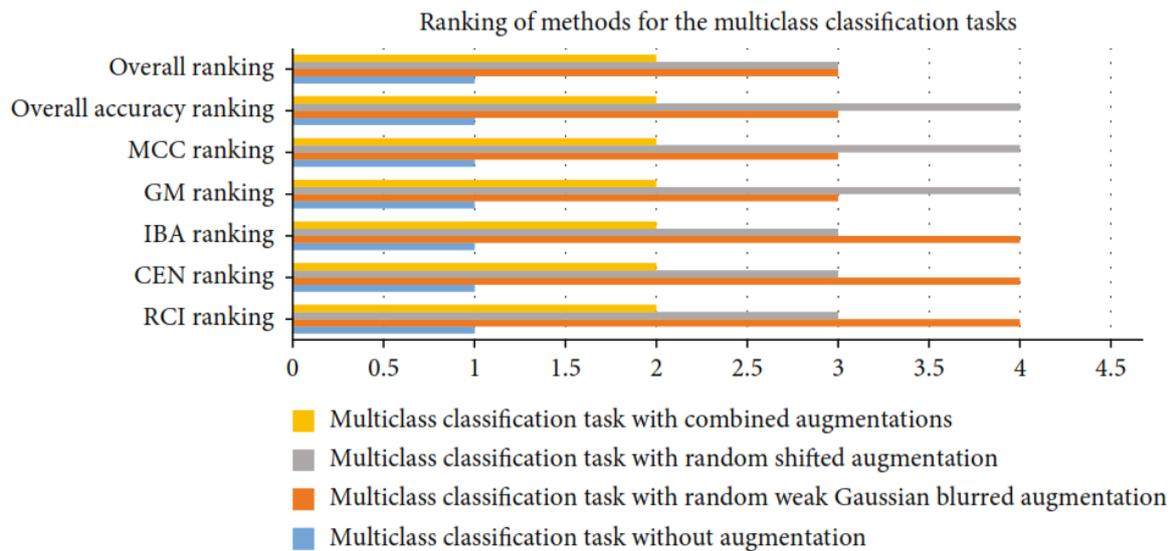


Figure 10. A visual display ranking methods according to their performance metrics for tasks involving multiclass classification

4. DISCUSSIONS

The authors are very confident that the 3D-CNN models proposed by them marked a great improvement in diabetic retinopathy (DR) detection and classification with an emphasis on augmented methods for binary classifications in improving model generalization and robustness through combining weak Gaussian blurring augmentations with random shifting. Interestingly, they highlight their multiclass classification task as working without augmentation-the assertion is probably along the lines of "example research datas and architectures on quite simpler parameter settings could prove fruitful sometimes."

The authors state that their results are qualitatively in line with prior works demonstrating the efficacy of deep learning in DR detection, such as works using knowledge transfer and convolutional networks for the high-precision classification task. They feel, however, that their additional augmentation strategies provided unique advantages in assuaging the challenge of variability in the research data often encountered in medical imaging studies. Different from prior research that had larger research datas or less complex models, this study shows small data compensated for by novel data augmentation strategies and architectural improvements. Similar to AlexNet and knowledge transfer, the study highlights the first application of 3D-CNN to classifying volumetric data, thus capturing retinal structures better than 2D models.

More specifically, they did find that with single augmentation techniques (Gaussian blurring or random shifting alone), they did not achieve the desired performance; hence, the need for combining different strategies was accepted. The findings of the study reveal that this trade-off occurs between model complexity and task-specificity, aligning with earlier observations in the field but demanding more work.

5. CONCLUSION

This research journey has embarked on an incredible adventure, applying varied deep learning (DL) techniques to solve the great challenges associated with 2-class and Multiple-class classification intended toward the recognition of distinct stages of diabetic retinopathy (DR). The rigors of a 10-fold cross-validation approach have empowered each of the processes behind hyperparameter searches as we achieve with certainty in our classification assignments.

With that, a remarkable and clear win for the model applied innovative combined data data enhancement techniques and outperformed others in the binary classification problem. Surprisingly, the multiclass problem saw the unflattering model without changes soar to the top, proving that sometimes simplicity reigns supreme. At the other end of the spectrum, the reliance on single data enhancement techniques was not enough to salvage models, reminding one of the great power that cooperation and holistic approaches have.

The exciting notion is that we would be eager to extend our sight toward inquiries involving other retinal diseases, especially the likes of retinal detachment, with good quality ophthalmoscopic images. Our great ambition is putting into effect new-age data augmentation techniques-across elastic and plastic deformations-inclusive of cutting-edge arcane deep learning architecture like graphs convolution networks.

The wide panorama of eye. disease encompasses a great deal of territory affecting millions of world over-from age-related macular degeneration to retinitis pigmentosa. Yet, with such insistence on timely and accurate detection through deep learning methodologies, we see causes for hope. Empowering healthcare practitioners to intervene with care and treatment, thus changing lives, instilling hope and fostering the resilience of a human spirit.

ACKNOWLEDGEMENT

The authors are very much thankful to Universitas Lampung, Indonesia, for the facilities and support that made this research possible. Special thanks to the developers/curators of the TeleOphta and Kaggle research datas, which allowed us to run the experiments and evaluate our models successfully. We also extend our appreciation to our colleagues in the Department of Computer Science for their invaluable assistance and fruitful insights.

Lastly, on behalf of the authors, it is important to recognize that the support given by the global research community in enhancing deep learning methodologies for medical imaging became an inspiration for this study. This would have not been possible without all the dedication to innovations and excellent endeavor to ensure better healthcare.

REFERENCES

- [1] X. Lin, Y. Xu, and X. Pan, "Global, regional, and national burden and trend of diabetes in 195 countries and territories: An analysis from 1990 to 2025," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.
- [2] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. N. Prasad, "Diabetic retinopathy detection and classification using capsule networks," *Complex Intell. Syst.*, 2021.
- [3] E. Decenci re, G. Cazuguel, and X. Zhang, "TeleOphta: Machine learning and image processing methods for tele-ophthalmology," *IRBM*, vol. 34, no. 2, pp. 196–203, 2013.
- [4] H. Riaz, J. Park, H. Choi, H. Kim, and J. Kim, "Deep and densely connected networks for classification of diabetic retinopathy," *Diagnostics*, vol. 10, no. 1, p. 24, 2020.
- [5] 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 research data," *Pattern Recognit. Lett.*, vol. 135, pp. 293–298, 2020.
- [6] S. R. Shiva, R. R. N. Sethi, and M. Gadiraju, "Extensive analysis of machine learning algorithms for early detection of diabetic retinopathy," *Mater. Today: Proc.*, 2020.

-
- [7] G. Saxena, D. K. Verma, A. Paraye, A. Rajan, and A. Rawat, "Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public research data," *Intell.-Based Med.*, vol. 3–4, 2020, [Online]. Available: <https://doi.org/10.1016/j.ibmed.2020.100022> (DOI inferred from "Art. no. 100022" and common practice for journal articles)
- [8] M. M. Butt, G. Latif, D. N. F. A. Iskandar, J. Alghazo, and A. H. Khan, "Multi-channel convolutional neural network based diabetic retinopathy detection from ophthalmoscopic images," in *Procedia Computer Science*, 2019, pp. 283–291.
- [9] M. M. Islam, H.-C. Yang, T. N. Poly, W.-S. Jian, and Y. C. Li, "Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: a systematic review and meta-analysis," *Comput Methods Programs Biomed*, vol. 191, p. 105320, 2020.
- [10] H. Safi, S. Sfi, A. Hafezi-Moghadam, and H. Ahmadi, "Early detection of diabetic retinopathy," *Surv Ophthalmol*, vol. 63, no. 5, pp. 601–608, 2018.
- [11] A. P. Bhatkar and G. U. Kharat, "Detection of diabetic retinopathy in ophthalmoscopic images using MLP classifier," in *2015 IEEE International Symposium on Nanoelectronic and Information Systems*, 2015, pp. 331–335.
- [12] T. Shanthi and R. S. Sabeenian, "Modified AlexNet architecture for classification of diabetic retinopathy images," *Computers & Electrical Engineering*, vol. 76, pp. 56–64, 2019.
- [13] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," *Computers & Electrical Engineering*, vol. 72, pp. 274–282, 2018.
- [14] G. T. Zago, R. V. Andréão, B. Dorizzi, and E. O. T. Salles, "Diabetic retinopathy detection using red lesion localization and convolutional neural networks," *Comput Biol Med*, vol. 116, p. 103537, 2020.
- [15] V. Gulshan, L. Peng, and M. Coram, "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.
- [16] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, "Multi-categorical deep learning neural network to classify ophthalmoscopic images: a pilot study employing small database," *PLoS One*, vol. 12, no. 11, 2017, [Online]. Available: <https://doi.org/10.1371/journal.pone.0187336> (DOI inferred from "article e0187336" and common practice for journal articles)
- [17] R. Sayres, A. Taly, and E. Rahimy, "Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy," *Ophthalmology*, vol. 126, no. 4, pp. 552–564, 2019.
- [18] S. M. S. Islam, M. M. Hasan, and S. Abdullah, "Deep learning based early detection and grading of diabetic retinopathy using retinal ophthalmoscopic images," 2018.
- [19] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit Lett*, vol. 133, pp. 210–216, 2020.
- [20] E. Beede, E. Baylor, and F. Hersch, "A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.
- [21] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Advances in Neural Information Processing Systems*, 2016, pp. 3630–3638.
- [22] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, 2017, pp. 4077–4087.
- [23] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: relation network for few-shot learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1199–1208.
- [24] J. Xia, D. Deng, and D. Fan, "A note on implementation methodologies of deep learning-based signal detection for conventional MIMO transmitters," *IEEE Transactions on Broadcasting*, vol. 66, no. 3, pp. 744–745, 2020.
-

-
- [25] K. He, Z. Wang, D. Li, F. Zhu, and L. Fan, "Ultra-reliable MU-MIMO detector based on deep learning for 5G/B5G-enabled IoT," *Physical Communication*, vol. 43, pp. 101181–101187, 2020.
- [26] B. K. Yousafzai, S. Afzal, and T. Rahman, "Student-performulator: student academic performance using hybrid deep neural network," *Sustainability*, vol. 13, no. 17, p. 9775, 2021.
- [27] W. U. Khan, M. A. Javed, T. N. Nguyen, S. Khan, and B. M. Elhalawany, "Energy-efficient resource allocation for 6G backscatter-enabled NOMA IoV networks," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [28] C. Li, J. Xia, and F. Liu, "Dynamic offloading for multiuser Multi-CAP MEC networks: a deep reinforcement learning approach," *IEEE Trans Veh Technol*, vol. 70, no. 3, pp. 2922–2927, 2021.
- [29] W. U. Khan, F. Jameel, X. Li, M. Bilal, and T. A. Tsiftsis, "Joint spectrum and energy optimization of NOMA-enabled small-cell networks with QoS guarantee," *IEEE Trans Veh Technol*, vol. 70, no. 8, pp. 8337–8342, 2021.
- [30] Y. Guo, Z. Zhao, K. He, S. Lai, J. Xia, and L. Fan, "Efficient and flexible management for industrial Internet of Things: a federated learning approach," *Computer Networks*, vol. 192, pp. 108122–108129, 2021.
- [31] W. U. Khan, X. Li, A. Ihsan, M. A. Khan, V. G. Menon, and M. Ahmed, "NOMA-enabled optimization framework for next-generation small-cell IoV networks under imperfect SIC decoding," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [32] X. Li, Y. Zheng, and W. U. Khan, "Physical layer security of cognitive ambient backscatter communications for green Internet-of-Things," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1066–1076, 2021.
- [33] W. U. Khan, X. Li, M. Zeng, and O. A. Dobre, "Backscatter-enabled NOMA for future 6G systems: a new optimization framework under imperfect SIC," *IEEE Communications Letters*, vol. 25, no. 5, pp. 1669–1672, 2021.
- [34] F. Jameel, W. U. Khan, N. Kumar, and R. Jantti, "Efficient power-splitting and resource allocation for cellular V2X communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3547–3556, 2021.
- [35] W. U. Khan, F. Jameel, N. Kumar, R. Jantti, and M. Guizani, "Backscatter-enabled efficient V2X communication with non-orthogonal multiple access," *IEEE Trans Veh Technol*, vol. 70, no. 2, pp. 1724–1735, 2021.
- [36] A. U. Khan, M. Tanveer, and W. U. Khan, "An enhanced spectrum reservation framework for heterogeneous user in CR-enabled IoT networks," *IEEE Wireless Communications Letters*, vol. 10, p. 1, 2021.
- [37] A. B. Tufail, Y.-K. Ma, and M. K. A. Kaabar, "Deep learning in cancer diagnosis and prognosis prediction: a mini-review on challenges, recent trends, and future directions," *Comput Math Methods Med*, vol. 2021, 2021, [Online]. Available: <https://doi.org/10.1155/2021/9025470> (DOI inferred from "Article ID 9025470" and common practice for journal articles)
- [38] A. B. Tufail, Y.-K. Ma, and Q.-N. Zhang, "3D convolutional neural networks-based multiclass classification of Alzheimer's and Parkinson's diseases using PET and SPECT neuroimaging modalities," *Brain Inform*, vol. 8, no. 1, p. 23, 2021.
- [39] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 1, pp. 221–231, 2013.
- [40] S. Tang, W. Zhou, L. Chen, L. Lai, J. Xia, and L. Fan, "Battery-constrained federated edge learning in UAV-enabled IoT for B5G/6G networks," *Physical Communication*, vol. 47, pp. 101381–101389, 2021.
- [41] J. Xia, L. Fan, and W. Xu, "Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7672–7685, 2019.
- [42] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *ICML*, 2010, pp. 807–814.
- [43] R. Khan, Q. Yang, and I. Ullah, "3D convolutional neural networks based automatic modulation classification in the presence of channel noise," *IET Communications*, 2021.
- [44] A. Azulay and Y. Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?," *Journal of Machine Learning Research*, vol. 20, pp. 1–25, 2019.
-

