INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 08-08-2024

Accepted: 08-05-2025

Published: 30-05-2025

Editor: Special Issue Editors: Dr. N. Pothanna, Dr.T. Jayashree, Dr. V. Ganesh Kumar

Citation: Bai BM, Pulletikurthi G, Gannamaneni R, Janagama A, Gaddam SR (2025) Explainable Diabetic Retinopathy Detection For Reduced Image Size. Indian Journal of Science and Technology 18(SP1): 9-18. https://doi.org/10.17485/IJST/v18si1.icamada6

* **Corresponding author**.

mathurabai_b@vnrvjiet.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Explainable Diabetic Retinopathy Detection For Reduced Image Size

B Mathura Bai[1]*, Gnanesh Pulletikurthi[1], Rushichandra Gannamaneni[1], Abhiram Janagama[1], Shashank Reddy Gaddam[1]

1 Department of Information Technology, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

Abstract

Objectives: This study aims to enhance the accuracy of classifying different stages of diabetic retinopathy with low computational and space complexity. **Methods:** This study identifies optimal retinal image size using Fourier transform, sinusoidal waveform and Nyquist frequency. Additionally, developed modified versions of ResNet and DenseNet with fewer parameters so that they perform effectively on reduced retinal images. This study has utilized initially the Kaggle dataset, which was imbalanced and contains 35,108 images. Then balanced the dataset by applying data augmentation to it. After applying the data augmentation obtained 1,25,756 images, which is a balanced dataset on which the proposed model was applied. On the image size of 28*28 pixels, we applied a modified version of DenseNet and ResNet and achieved accuracies of 78.33% and 80.34%, respectively. On a 14*14 image size for the modified DenseNet and ResNet, we achieved accuracies of 76.71% and 74.91%, respectively. The modified version of DenseNet has 2,43,685 parameters, and the modified version of ResNet has 2,28,053 parameters. **Findings:** Every study previously has considered the entire retinal image. The optimal image size is 28*28 pixels are identified by the proposed method. Using those images, a model is built and an algorithm is designed for the worst-case scenario, and it also performs well in the best-case scenario. For the retinal images, the proposed model achieves excellent results 80.34% even for larger images with sizes 28*28 pixels and above. **Novelty:** This study introduces a new technique for enhancing the classification process of low-quality retina images. To obtain the phase and amplitude of the image, we use the 2D Fast Fourier Transform (FFT) and take a sinusoidal waveform plot. The proposed method then resizes the above sinusoidal waveform, repeats the previous steps with the resized sinusoidal waveform, and compares it with the original waveform. The proposed method also applied the Nyquist frequency criterion, which states that resizing is possible if the pixel pitch of the resized image is at least twice that of the original. It also uses Gaussian blur to overcome jagged edges during resizing. The proposed method identified 28*28 pixels as the optimal image size, which maintains waveform consistency with the original image. A modified version of ResNet and DenseNet is built using the proposed method, which has fewer parameters and gives good accuracy.

Keywords: Convolutional Neural Network; Reduced Image Size; Modified ResNet; Modified DenseNet; 2D Fast Fourier Transform; Nyquist Frequency; Sinusoidal Waveform; Gaussian Blur

1 Introduction

Diabetic Retinopathy (DR) is a severe eye disease that occurs in diabetic people. If the people who have diabetes have uncontrolled blood sugar levels for a long period of time and are uncontrolled, then there is a chance of getting diabetic retinopathy, which leads to vision loss. Initially, the condition of the eye starts with Non-Proliferative Diabetic Retinopathy (NPDR), then it increases to moderate and severe levels, and finally it leads to proliferative diabetic retinopathy, which is the last stage and which leads to vision loss. Regular eye examinations are required for the people who have diabetes to prevent vision loss; blood sugar level is also important to maintain for the people who have diabetes. Early detection of diabetic retinopathy is important in the medical industry. Figure 1, show a normal retinal image and a DR-affected retinal image[1].

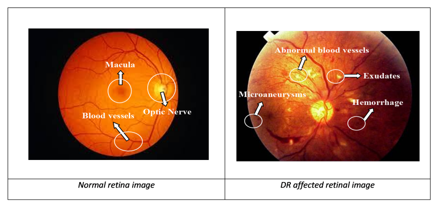


Fig 1. Normal retinal image and diabetic retinopathy affected retinal image

An ensemble-based approach, proposed by Sehrish Qummar et al[2], used a stacking classifier of Convolution Neural Network (CNN) architectures like ResNet50, Xception, DenseNet 121, DenseNet 169, and InceptionV3 to classify various stages of diabetic retinopathy. Their stacking model achieved an accuracy of 80.8% on the imbalanced Kaggle dataset. Eman Abdelmaksoud et al[3] have proposed a deep learning-based multi-label computer-aided diagnosis model to classify different stages of DR. They have resized all images to 512 x 512 pixels to build the model. They have used CNN architecture Unet to extract the important features in the image, and after extracting the features, they have used a multi-SVM classifier. Their model achieved 95.1% accuracy and 91.9% AUC on a publicly available Kaggle dataset. Ayesha Jabbar et al[4] have developed a transfer learning-based model for classifying different stages of diabetic retinopathy. They have used deep transfer learning algorithms, mainly VGGNet, for DR classification. Their model has achieved 97.6% accuracy on the publicly available Kaggle dataset. Fiaz Gul Khan et al[5] proposed a VGG-NiN neural network. Their model mainly aims to reduce the computation complexity and reduce parameters. The VGG16, spatial pyramid pooling layers, and Network in Network (NiN) method are used to classify various stages of diabetic retinopathy. Using Kaggle data for public use, the VGG-NiN method achieved a 56% recall. W. K. Wong et al[6] proposed a transfer learning method. The authors used ShuffleNet and ResNet-18 to classify various stages of diabetic retinopathy and ADE to extract features. The transfer learning model achieved 82% accuracy on the APTOS 5-class DR grading image. Transfer learning performed less with APTOS 2-class and EyePac Messidor-2 3-class images. Nasser Kehtarnavaz et al[7] introduced a deep learning-based technique to classify various stages of DR. Their model uses the approaches of both regression and classification and squeeze excitation in a densely connected deep neural network. Their model also included the instance learning process for lesion identification in the fundus image, as well as a full convolutional network for lesion detection. Their model achieved a weighted Kappa of 0.90 and 0.88 on APTOS and EyePACS, respectively. W. Nazih et al[8] have introduced a ViT model using transformers that performed better than other CNN architectures in classifying the different stages of diabetic retinopathy. Their model achieved an FGADR accuracy of 82.5%. Palaniswamy et al[9] proposed a model that uses both the Internet of Things (IoT) and deep learning (DL). They collected the data using an IoT device and processed and stored it on cloud servers. They have used image processing techniques for noise removal. They have used CNN architecture like DenseNet to extract features and used LSTM for classification. They had installed the above-summarized image in the LSTM classification. They have used the Honey Bee Optimization algorithm to update weights.

Everyone has considered the entire image, which is a high-resolution image, to build the model. When low-resolution images are provided, those models that build on high-resolution images may not perform accurately. To overcome that, 2D Fast Fourier Transform, Nyquist frequency, and Gaussian blur are used to identify the optimal image size, and found optimal image size is 28*28 pixels, which is a low-resolution image. So, the models are built on the reduced image size. Modified versions of ResNet and DenseNet are used to overcome the vanishing gradient problem with fewer parameters. The modified version of DenseNet has 243,685 parameters, and the modified version of ResNet has 228,053 parameters.

2 Methodology

2.1 Dataset Description

The diabetic retinopathy dataset under consideration was sourced from Kaggle, encompassing a total of 35,108 images categorized into five classes: no diabetic retinopathy, mild, moderate, severe, and proliferative. No diabetic retinopathy contains 25,802 images; mild diabetic retinopathy contains 2,438 images; moderate diabetic retinopathy contains 5,288 images; severe diabetic retinopathy contains 872 images; and proliferate diabetic retinopathy contains 708 images. The images of all stages of diabetic retinopathy are presented in Figure 2.

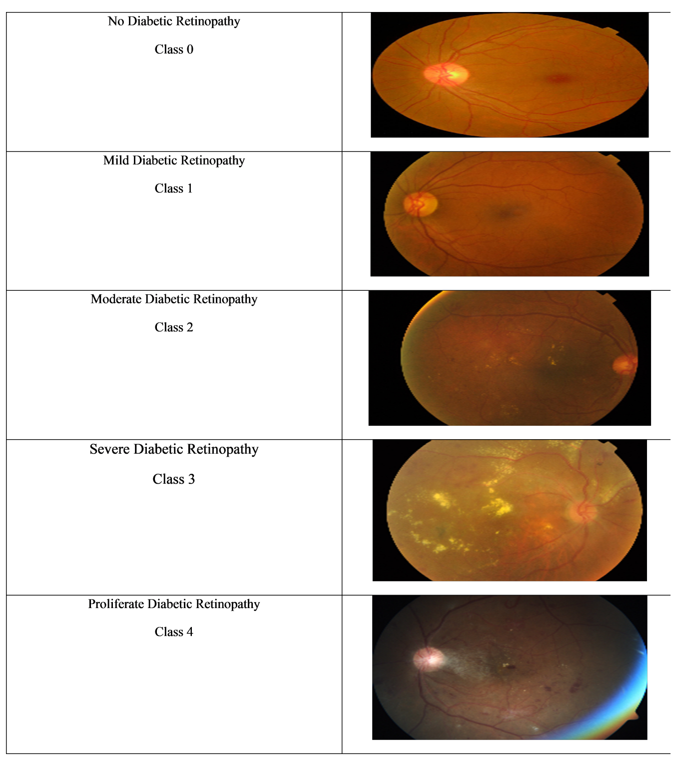


Fig 2. Different stages of diabetic retinopathy retinal images

An imbalanced dataset was present, so a balanced dataset is required to increase the classification accuracy. Data augmentation techniques like rotation, shear angles, and flipping both vertically and horizontally were applied to balance the data.

- Rotation: Image rotated by a specific angle, such as 10 degrees clockwise.
- Vertical flipping: Image was flipped along the vertical axis.
- Horizontal Flipping: Image was flipped along the horizontal axis.
- Shear Angle: Shifted the particular points of the image along the vertical or horizontal axis.

Table 1 shows the comparison of the number of retinal images before and after applying data augmentation.

Table 1. Comparison of retinal images before and after applying data augmentation

Class label	Number of retinal images before data augmentation	Number of retinal images after data augmentation
NO DIABETIC RETINOPATHY	25802	25802
MILD	2438	25893
MODERATE	5288	25875
SEVERE	872	25118
PROLIFERATE	708	23068

2.2 Image Resizing

To reduce the image size, applying 2D FFT to an image gives the amplitude and phase of the image. Using the amplitude and phase, plot a sinusoidal wave to verify the signal the image generates. The Nyquist frequency states that if the resized image pixel pitch is greater than or equal to twice the original image pixel pitch, then only resizing the image is possible without losing the important structural details of the image. When resizing the image, Gaussian blur is applied to avoid jagged edges, which also preserves the image quality.

When applying a 2D FFT to an image, a matrix in the form of complex numbers, where each element is represented as a+bi, with 'i' being the imaginary unit. After obtaining the complex numbers of the matrix, the amplitude and phase of the image can be extracted. The amplitude represents the intensity of different frequencies present in the image, while the phase indicates the spatial position of these frequencies within the image.

The mathematical representation for a two-dimensional Fast Fourier Transform (FFT) is expressed as follows:

Z(i, j) = $\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} z\left(a,b\right)\, e^{-2\pi i\left(\frac{ux}{N}+\frac{uy}{M}\right)}$

where:

- Z(i,j) represents the 2D Fourier transform of the input image in complex form.
- The pixel intensity at coordinates (a, b) within the image is denoted by z(a, b).
- The dimensions of the image are given by N and M, which represent the width and height, respectively.
- u and v correspond to the spatial frequencies along the horizontal and vertical axes, respectively.
- The symbol i represents the imaginary unit.

Amplitude is calculated by

$$A = \sqrt{Real\ Part^2 + Imaginary\ Part^2}$$

Phase is calculated by

$$\frac{Imaginary\ Part}{Real\ Part}$$

In Figure 3 , clearly shown how the amplitude and phase of an image is represented when applying the 2D Fast Fourier transform to it.

A concept known as nyquist frequency establishes the sampling rate of a signal or image, ensuring correct representation without aliasing. According to the Nyquist-Shannon sampling theorem, the subsampled image's frequency must be twice or more than the original image's maximum frequency. The Nyquist frequency for two-dimensional images is calculated differently for each dimension, considering the horizontal (x) and vertical (y) directions separately.

The Nyquist frequency in the x-direction is calculated as:

$f_{Nx} = \frac{1}{2} * \frac{1}{\triangle x}$

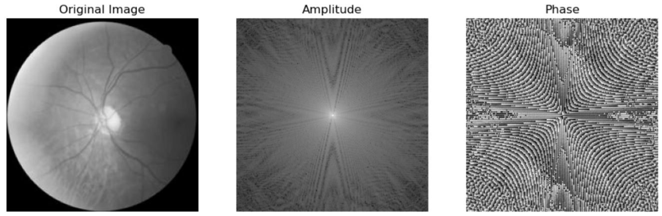


Fig 3. Original, Amplitude and Phase image

The distance between adjacent pixels in the x-direction, known as pixel pitch, is represented by Δx

The Nyquist frequency in the y-direction is calculated as:

$f_{Ny} = \frac{1}{2} * \frac{1}{\triangle y}$

The distance between adjacent pixels in the y-direction, known as pixel pitch, is represented by Δy.

The Nyquist frequency calculation in digital image processing is based on the pixel pitch in each direction, taking into account the minimum frequencies in the x- and y-directions. Tasks like resizing require proper sampling and reconstruction techniques to prevent aliasing and image quality loss, making this understanding crucial. Table 4 clearly displays the Nyquist frequencies in the x- and y-direction for various image sizes.

Table 2. Nyquist Frequencies for different sizes of images

Image size in pixels	Nyquist Frequency in pixel pitch
224*224	X axis = 0.002232 Y axis = 0.002232
112*112	X axis = 0.004464 Y axis = 0.004464
56*56	X axis = 0.008928 Y axis = 0.008928
28*28	X axis = 0.017857 Y axis = 0.017857
14*14	X axis = 0.035714 Y axis = 0.035714

Sinusoidal waveforms play a crucial role in image processing, particularly in resizing or resampling tasks, where dimensions are adjusted while preserving visual quality through interpolation. We have plotted the sinusoidal waveform based on the amplitude and phase of the image obtained when we applied 2D FFT to the image.

The general formula for a sinusoidal waveform is:

$f(x) = A.sin(\omega x / width + \phi)$

where:

- f(x) is the value of the waveform at position x.
- A is the amplitude of the waveform, which determines its peak value.
- The symbol ω denotes the angular frequency and is commonly expressed as ω=2πf, with f denoting the frequency.
- ϕ is the phase shift, which indicates how much the waveform is shifted horizontally.

In Figure 4 , the comparison is made between two different-sized images (224*224 pixels and 28*28 pixels) and observed that, upon analysis, they generate the same sinusoidal waveform. Further, compare this with an image size of 14*14 pixels in Table 5. It shows that further resizing the image to 14*14 pixels, generates a sinusoidal waveform with more noise.

Anti-aliasing is a technique in computer graphics to reduce visual artifacts like aliasing when displaying high-resolution images on lower-resolution screens. Gaussian blur is a popular method used to achieve this by convolving the image with a Gaussian function, preserving its overall structure while smoothing out jagged edges. The process involves identifying edges, creating a mask, applying Gaussian blur, and blending the blurred image with the original image. This reduces aliasing appearance and produces a smoother, more visually pleasing image. However, it comes with a computational cost, making optimization techniques essential for balancing visual quality with performance.

$G(x) = \frac{1}{2\pi\sigma^2} * e^{\frac{-(x^2+y^2)}{2\sigma^2}}$

where:

- G(x, y) represents the Gaussian function located at the specified coordinates (x, y).
- σ denotes the standard deviation of the Gaussian distribution, which regulates the degree of blurring.
- e stands for the base of the natural logarithm, being roughly equivalent to 2.71828.
- π symbolizes the mathematical constant pi, approximately equal to 3.14159.

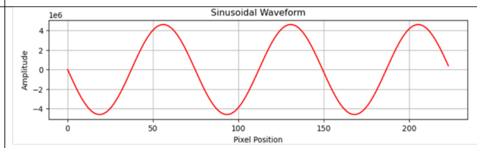
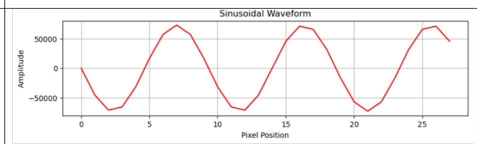
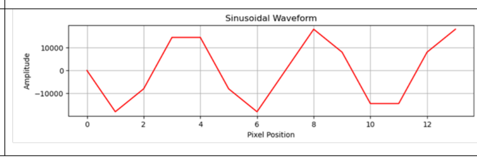
Image Size	Sinusoidal Waveform	Analysis
224*224		This is the signal for the original image 224*224 size.
28*28		This is the signal for the resized image 28*28 size signal, which is similar to the original image signal.
14*14		This is the resized image 14*14 size signal, completely distorted from the original image signal.

Fig 4. Sinusoidal waveform for different sizes of images

2.3 Modified ResNet

ResNet is a CNN architecture that uses skip connections to reduce the vanishing gradient problem. It consists of two types of blocks: identity blocks and convolutional blocks. A modification to the ResNet architecture reduces model complexity by employing a modified version of the residual block, as shown in Figure 5 . This modified version involves reducing convolutional layers and filter sizes and introducing different operations. Figure 6 illustrates the modified ResNet version, which employs modified residual blocks, max pooling, and average pooling. Our objective is to increase performance and decrease computing complexity. The model's final layer uses SoftMax as its activation function to create probabilities for different classes of DR, and categorical cross-entropy used as the loss function. Update the weights using the Adam optimizer.

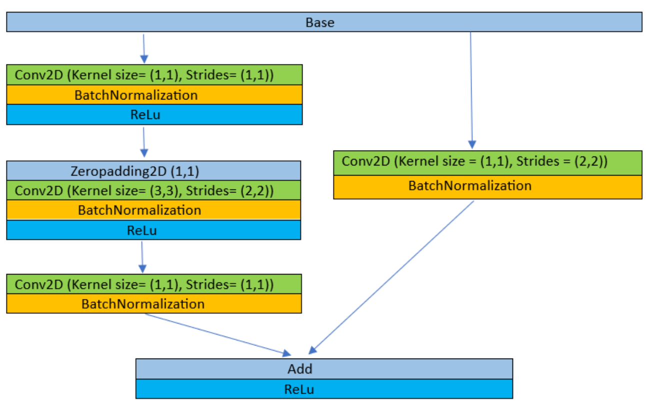


Fig 5. Modified Residual Block

2.4 Modified DenseNet

The DenseNet architecture consists of both dense blocks and transition blocks. The dense block, illustrated in Figure 7 , exemplifies this structure. The transition block manages feature map growth and data dimensionality by combining convolutional, batch normalisation, and pooling layers, reducing channels and spatial dimensions and reducing computational complexity.

The issue with concatenation arises when two images are combined, potentially merging their information. However, significant differences in dimensions can lead to a loss of information, particularly when the details of a smaller image are

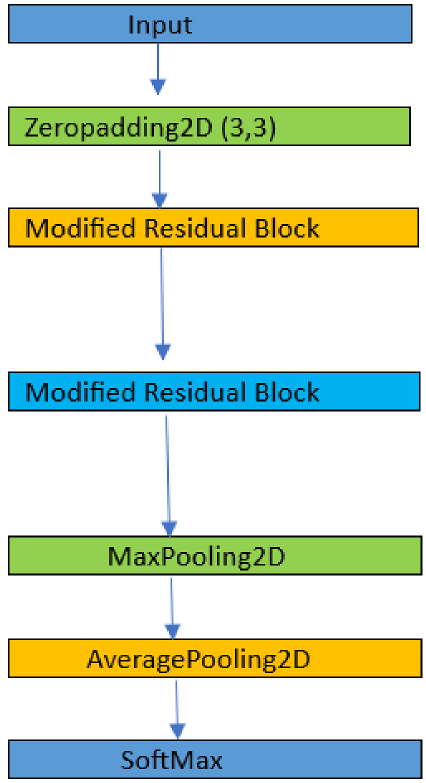


Fig 6. Modified ResNet

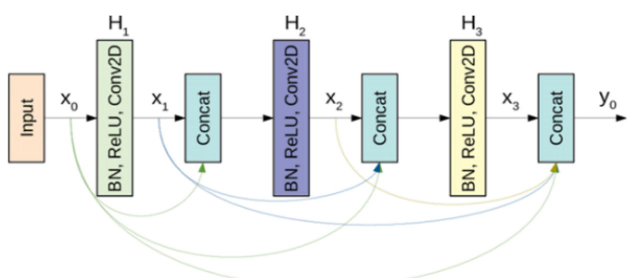


Fig 7. D ense Block

overshadowed by those of a larger one. The concatenation operation, illustrated in Figure 8 , exemplifies this structure.

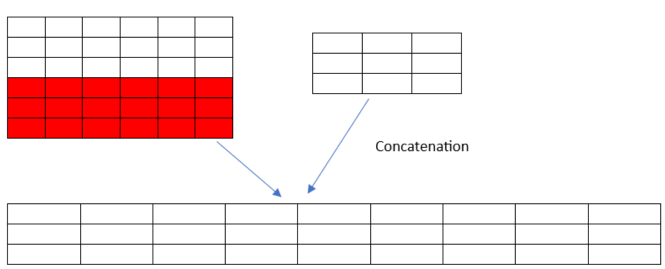


Fig 8. Concatenation Operation

The modified DenseNet avoids using the concatenation operation, instead utilising the add operation to consider the shaded information. The modified DenseNet uses a modified version of the residual block with different stride values. Figure 9 and Figure 6 illustrate the modified DenseNet and the modified version of the residual block, respectively. The modified version of DenseNet, which uses modified residual blocks and average pooling. The model's final layer uses SoftMax as its activation function to create probabilities for different classes of DR, and categorical cross-entropy used as the loss function. Update the weights using the Adam optimizer.

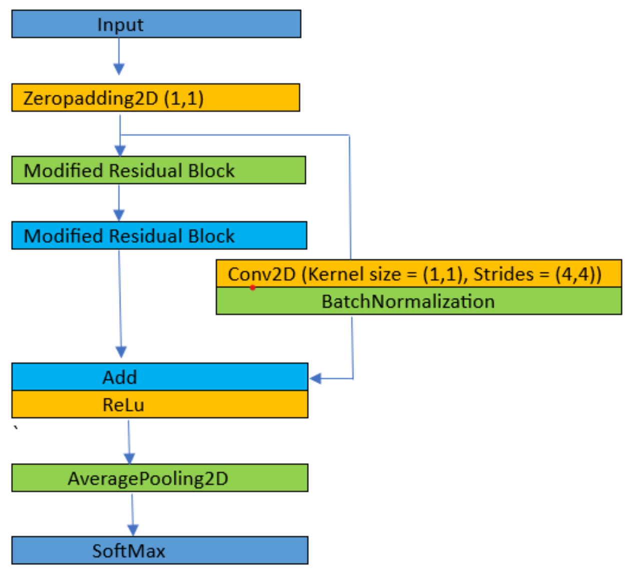


Fig 9. Modified Dense Block

3 Results and Discussion

The outcomes of the proposed model are compared with cutting-edge methodologies. The Central Processing Unit (CPU) plays a significant role in training a proposed model. There are various metrics that are to be considered to evaluate the proposed model, we have utilized accuracy, sensitivity, specificity, precision, and F1-score measures as performance metrics.

Compared to the previous studies, the current work proposes a novel approach to find optimal image size using 2D FFT and sinusoidal waveforms. And modified versions of ResNet and DenseNet to reduce the time and space complexity of the model. Hossein Shakibania et al[10] proposed a deep learning method for classifying various stages of diabetic retinopathy using retinal images. Their model uses two pre-trained models, ResNet50 and EfficientNetB0, to extract features and classify. It performs well in binary classification, with an accuracy of 98.50% on the APTOS 2019 dataset. To classify various stages of DR, Satya Prakash Singh et al[11] proposed a pre-trained EfficientNetB0 convolutional neural network (CNN) model. They trained the model by combining three datasets: IDRiD, Messidor, and APTOS 2019 Blindness Detection. The model achieved a maximum accuracy of 72%, but grayscale conversion improved it to 80%. The model also achieved accuracy of 72%, 76%, and 73% for green, red, and blue channels. The pre-trained models are effective for general image classification tasks because they use the weights of the model which was trained on the ImageNet Dataset. Which is not efficient for classifying retinal images because the feature representation changes. The current work overcomes the problem of pre-trained models. A. Jain et al[12] showed that VGG16 achieved 71.7% accuracy, VGG19 achieved 76.9% accuracy, and Inception v3 achieved 70.2% accuracy. Despite their lower accuracy, Inception v3 proved to be the best classifier for DR cases as it effectively reduces false negatives. S. Thorat et al[13] proposed a CNN model trained on a GPU using 35,126 retinal images publicly available by EyePACS on the Kaggle website, achieving an accuracy of around 81%. Xiaoling Luo et al[14] proposed a model that also learns features with local and long-distance dependence in images to improve feature representation to classify various stages of DR. They build the model using inception modules and long-range units, where long range units are used for learning the long-range dependence features in images. Their model achieved an accuracy of 83.6% on the EyePACS dataset. Considering the entire retinal image increases the space complexity of the model. To overcome a reduced retinal image is required. Our current work proposed a 2D FFT and sinusoidal waveform to find the optimal image. Table 6 illustrates the comparison of the proposed model with another's model.

3.1 Image size 28*28 pixels

Modified version of ResNet with 28*28 pixels image size achieves the accuracy, recall, specificity, precision, and f1-score of 78.33%, 78.33%, 94.58%, 78%, and 78% respectively. The modified ResNet model, evaluated on 28*28-pixel images, shows the following performance across individual classes. For classes 0, 1, 2, 3, and 4, the true positive rate is 87.97%, 71.50%, 57.61%, 85.91%, and 89.64%, respectively; similarly, the true negative rate is 94.85%, 95.01%, 92.35%, 94.23%, and 96.43%, respectively.

Modified version of DenseNet with 28*28 pixels image size achieves the accuracy, recall, specificity, precision, and f1-score of 80.34%, 80.35%, 95.08%, 80%, and 80% respectively. The modified DenseNet model, evaluated on 28*28-pixel images, shows the following performance across individual classes. For classes 0, 1, 2, 3, and 4, the true positive rate is 93.64%, 73.96%, 57.05%, 89.77%, and 88.04%, respectively; similarly, the true negative rate is 94.41%, 95.13%, 94.25%, 93.75%, and 97.83%, respectively.

3.2 Image size 14*14 pixels

Modified version of ResNet with 14*14 pixels image size achieves the accuracy, recall, specificity, precision, and f1-score of 74.91%, 74.91%, 93.72%, 75%, and 75%, respectively. The modified ResNet model, evaluated on 14*14-pixel images, shows the following performance across individual classes. For classes 0, 1, 2, 3, and 4, the true positive rate is 92.70%, 67.62%, 60.40%, 75.30%, and 78.95%, respectively; similarly, the true negative rate is 93.51%, 93.26%, 89.16%, 95.32%, and 97.30%, respectively.

Modified version of DenseNet with 14*14 pixels image size achieves the accuracy, recall, specificity, precision, and f1-score of 76.71%, 76.71%, 94.17%, 77%, and 76%, respectively. The modified DenseNet model, evaluated on 14*14-pixel images, shows the following performance across individual classes. For classes 0, 1, 2, 3, and 4, the true positive rate is 93.38%, 68.86%, 59.94%, 76.34%, and 84.89%, respectively; similarly, the true negative rate is 93.45%, 94.59%, 90.82%, 95.81%, and 96.19%, respectively.

Table 3. Comparing our model with other's model

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score	Total number of images
Qummar et al. [2] Ensemble CNN for imbalanced dataset.	80.8%	51.5%	86.72%	63.85%	53.74%	35,126
Qummar et al. [2] Ensemble CNN for balanced dataset (up sample).	64.2%	64.2%	66.2%	66.2%	64	1,29,050
Qummar et al. [2] Ensemble CNN for balanced dataset (down sample).	58%	58%	85%	78%	53.2%	3,540
Z. Khan et al. [5] VGG NiN	85%	55.6%	91%	67%	59.6%	35,126
Proposed model, Modified ResNet for image size 28*28 pixels	78.33%	78.33%	94.58%	78%	78%	1,25,756
Proposed model, Modified DenseNet for image size 28*28 pixels	80.34%	80.35%	95.08%	80%	80%	1,25,756

4 Conclusion

This study establishes a crucial advancement in the diagnosis of diabetic retinopathy by determining that an optimal image size of 28*28 pixels is necessary. The optimal image size was determined through Nyquist frequency and sinusoidal waveform characteristics. These findings ensure accurate predictions even with lower-resolution images, thereby increasing the model's accessibility across a variety of imaging methods and settings with limited resources. We employed modified versions of ResNet and DenseNet with reduced parameter counts 228,053 and 243,685, respectively achieving classification accuracies of 78.33% and 80.34%. To address the class imbalance in the current dataset, data augmentation techniques were used to balance the dataset. However, it is acknowledged that augmented data may not fully capture the characteristics of real pathological cases. Future work should focus on creating more balanced and representative datasets to further improve classification performance. Overall, this approach not only makes the model more efficient but also gives the way for future research into image processing and deep learning techniques that work well even with low-resolution medical images.

Acknowledgements

This paper has been presented in "International Conference on Applied Mathematics and Advanced Data Analytics for Industry 5.0 (ICAMADA-2024)" held during 25th to 27th April 2024 at VNR VJIET, in collaboration with APTSMS and CCOE. The authors express their sincere gratitude to the guest editors for their tireless efforts and dedication by means of comments, editing and overseeing the manuscripts to bring in this final form.

The APC is deferred partially by Indian Society for Education and Environment

References

1) Valarmathi S, Vijayabhanu R. A Survey on Diabetic Retinopathy Disease Detection and Classification using Deep Learning Techniques. In: 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII). 2021;p. 1–4. Available from: https://ieeexplore.ieee.org/document/9445163.
2) Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, et al. A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection. *IEEE Access*. 2019;7:150530–150539. Available from: https://dx.doi.org/10.1109/access.2019.2947484.
3) Abdelmaksoud E, El-Sappagh S, Barakat S, Abuhmed T, Elmogy M. Automatic Diabetic Retinopathy Grading System Based on Detecting Multiple Retinal Lesions. *IEEE Access*. 2021;9:15939–15960. Available from: https://dx.doi.org/10.1109/access.2021.3052870.
4) Jabbar A, Naseem S, Li J, Mahmood T, Jabbar MK, Rehman A, et al. Deep Transfer Learning-Based Automated Diabetic Retinopathy Detection Using Retinal Fundus Images in Remote Areas. *International Journal of Computational Intelligence Systems*. 2024;17. Available from: https://link.springer.com/article/10.1007/s44196-024-00520-w.
5) Khan Z, Khan FG, Khan A, Rehman ZU, Shah S, Qummar S, et al. Diabetic Retinopathy Detection Using VGG-NIN a Deep Learning Architecture. *IEEE Access*. 2021;9:61408–61416. Available from: https://dx.doi.org/10.1109/access.2021.3074422.
6) Wong WK, Juwono FH, Apriono C. Diabetic Retinopathy Detection and Grading: A Transfer Learning Approach Using Simultaneous Parameter Optimization and Feature-Weighted ECOC Ensemble. *IEEE Access*. 2023;11:83004–83016. Available from: https://dx.doi.org/10.1109/access.2023.3301618.
7) Majumder S, Kehtarnavaz N. Multitasking Deep Learning Model for Detection of Five Stages of Diabetic Retinopathy. *IEEE Access*. 2021;9:123220–123230. Available from: https://dx.doi.org/10.1109/access.2021.3109240.
8) Nazih W, Aseeri AO, Atallah OY, El-Sappagh S. Vision Transformer Model for Predicting the Severity of Diabetic Retinopathy in Fundus Photography-Based Retina Images. *IEEE Access*. 2023;11:117546–117561. Available from: https://dx.doi.org/10.1109/access.2023.3326528.
9) Palaniswamy T, Vellingiri M. Internet of Things and Deep Learning Enabled Diabetic Retinopathy Diagnosis Using Retinal Fundus Images. *IEEE Access*. 2023;11:27590–27601. Available from: https://dx.doi.org/10.1109/access.2023.3257988.
10) Shakibania H, Raoufi S, Pourafkham B, Khotanlou H, Mansoorizadeh M. Dual branch deep learning network for detection and stage grading of diabetic retinopathy. *Biomedical Signal Processing and Control*. 2024;93. Available from: https://dx.doi.org/10.1016/j.bspc.2024.106168.
11) Singh SP, Gupta P, Dung R. Diabetic retinopathy detection by fundus images using fine-tuned deep learning model. *Multimedia Tools and Applications*. 2024;83:86657–86679. Available from: https://link.springer.com/article/10.1007/s11042-024-19687-7.
12) Jain A, Jalui A, Jasani J, Lahoti Y, Karani R. Deep Learning for Detection and Severity Classification of Diabetic Retinopathy. In: 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT). IEEE. 2019;p. 1–6. Available from: https://ieeexplore.ieee.org/document/8741456.
13) Thorat S, Chavan A, Sawant P, Kulkarni S, Sisodiya N, Kolapkar A. Diabetic Retinopathy Detection by means of Deep Learning. In: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE. 2021;p. 996–999. Available from: https://ieeexplore.ieee.org/document/9432075.
14) Luo X, Wang W, Xu Y, Lai Z, Jin X, Zhang B, et al. A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. *CAAI Transactions on Intelligence Technology*. 2024;9(1):153–166. Available from: https://dx.doi.org/10.1049/cit2.12155.