Journal of Advanced Research Design 137, Issue 1 (2026) 267-277

Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Deep Learning Approaches for Accurate Diabetic Retinopathy Detection and Classification: Comparison Study

Mohammed Ibrahim Mahdi[1], Amir Lakizadeh[1,*]

[1] Computer Engineering Department, University of Qom, Qom, Iran

ARTICLE INFO

Article history:
Received 12 January 2025
Received in revised form 22 June 2025
Accepted 10 July 2025
Available online 20 July 2025

Keywords:
Deep Learning, Convolutional Neural Network (CNN), Diabetic Retinopathy, Medical Image Classification

ABSTRACT

In order to prevent irreversible blindless among adults aged 18-65, it is imperative to accurately diagnose and treat diabetic retinopathy (DR) as early as possible. As such, the present study endeavoured to compare the efficacy of four deep learning (DL) models; namely, convolutional neural networks (CNN), residual networks (ResNet), inception architecture (IA), and densely connected convolutional networks (DenseNet); at detecting and classifying DR into different levels of disease severity. The Kaggle DR Detection dataset was used to assess the classification accuracies while a loss function (LF), that combines the loss of cross-entropy with additional penalties for classification errors, was introduced to overcome class imbalance issues and improve the performance of the four examined DL models. The DenseNet model had the highest accuracy, recall, precision, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) of the examined models, by scoring 90, 89, 88, 88%, and 0.92, respectively. This performance was closely followed by that of the ResNet model. The findings indicate that the architecture of the model, especially that of models that will be used for medical image classification (MIC), must be taken into account when selecting which model to use. Furthermore, the proposed customised LFs enhanced the precision and resilience of the examined DL models. Screening tools that can accurately diagnose DR early, with limited to no intervention from an ophthalmologist, will enable them to treat patiently significantly earlier, thus improving patient outcomes. As such, the development of such models is imperative. However, various datasets should be used to substantiate the accuracy of these models. Their efficacy could also be improved by combining supplemental clinical data as well as examining the use of hybrid architectures.

1. Introduction

Diabetic retinopathy (DR) is the primary cause of irreversible blindness among individuals aged 18-65 across the globe. According to the World Health Organization (WHO), approximately 33% of all diabetics worldwide have some form of DR. As such, there is an urgent need to develop screening tools that facilitate early detection to improve patient outcomes. Diabetic retinopathy (DR) ranges in severity from non-proliferative diabetic retinopathy, which is mild, to proliferative retinopathy, which

* *Corresponding author.*
E-mail address: mmahdi@uowasit.edu.iq

https://doi.org/10.37934/ard.137.1.267277

is the most severe form. However, regardless of the severity, any form of DR, if not treated early enough, may result in significant loss of vision, if not blindness [1].

In DR, severe visual impairment can only be side-stepped by detecting the disease early, classifying it correctly, and treating it immediately. Nevertheless, current detection techniques necessitate the manual examination of retinal pictures by an ophthalmologist, which is labor-intensive, time-consuming, error-prone, and subject to observer bias. Therefore, it is critical to provide an accurate tool that may assist medical practitioners in diagnosing and treating DR at an early stage in a reliable and automatic manner[2].

Medical image classification (MIC) has undergone a revolution thanks to developments in artificial intelligence (AI). This is particularly true for deep learning (DL) models, including convolutional neural networks (CNN), which excel in object detection, image segmentation, and image classification. Since they can automatically extract minute information from raw photos, they are especially well-suited to overcome the challenging problems that the MIC sector faces.

Therefore, the present study endeavoured to examine the ability of several DL models, each with their own unique architectures and feature extraction methods, to detect and classify DR[3]. The examined models were CNN, residual networks (ResNet), densely connected convolutional networks (DenseNet), and inception architecture (IA). A CNN model uses multiple layers to capture the spatial hierarchies present in an image while a ResNet model uses residual learning to overcome the issue of a vanishing gradient by training deeper networks. Densely connected convolutional networks (DenseNet), on the other hand, uses dense connections to enhance the flow of information between the layers, thereby, improving its learning accuracy and efficiency. Lastly, IA uses multiple differently-sized convolutional filters to effectively capture the features that may be present in an image[4].

Although the findings of multiple extant studies of these models seem promising, a detailed comparative analysis has yet to be conducted to determine which DL model can most efficiently detect and correctly classify DR. Furthermore, when faced by the inherent issues that the MIC industry faces, novel methods, such as loss functions (LFs) that are customised, may further improve a model's performance[5].

The present study bridges the gap in the existing literature by comparing the abilities of the various DL models; namely, ResNet, IA, DenseNet, and CNN; to efficiently detect and classify DR. A novel LF that more effectively penalises misclassifications was also proposed in the hopes of increasing model precision. Finding the most reliable and accurate DL model for DR detection, assessing the novel LF's effects, and offering insights into the real-world applications of using these models in clinical settings are the main goals of this research.

Through the completion of this comparative study, we hope to make a valuable contribution to the current efforts aimed at creating automated, dependable, and effective DR diagnostic tools, which will ultimately help diabetic patients prevent and detect vision loss early on.

2. Literature Review

Conventionally, a thorough eye examination that includes dilated fundus photography and subsequent analysis by qualified ophthalmologists is used to diagnose DR. While time-consuming and requiring a high level of expertise, manually inspecting retinal images for haemorrhages, microaneurysms, and other lesions suggestive of DR is nevertheless an effective method[6]. The degree of variability in manual grading accuracy among observers can result in disparities in the diagnosis and suggested course of treatment[7].

Initially, automated systems were created to overcome the drawbacks of manual inspection. Early attempts at automated DR detection relied on methods like thresholding, morphological operations, and template matching to find features like microaneurysms, blood vessels, and exudates in retinal images. While these approaches offered a certain level of automation, their dependence on manually defined features and ruled-based algorithms limited their accuracy and reliability[8].

Automated DR detection saw major progress with the arrival of machine learning (ML). Retinal images were classified using ML techniques like random forests and support vector machines, which outperformed traditional methods but still heavily relied on the quality of manually created features[9].

Unlike traditional ML techniques, DL models can automatically learn and extract features from raw images, thereby, eliminating the need for manual feature engineering. Of the then available models, CNN quickly became a favourite as it could capture the spatial hierarchies in images[10]. Early applications of CNN for DR detection showed promising results. As per research, a CNN could identify referable DR from retinal photos with a high level of sensitivity and specificity which was on par with ophthalmologists. Large-scale labelled image datasets used to train the models exhibited DL's potential for precise and scalable DR screening.[11].

More sophisticated DL architectures were investigated in later research to augment the efficacy of DR detection[12]. ResNet introduced residual learning to deal with the vanishing gradient problem, allowing deeper networks to be trained and making it a viable alternative for medical image analysis, including DR detection. By employing multiple convolutional filters of varying sizes within a single layer, the IA developed a novel method that improved the network's capacity to recognise intricate patterns in retinal images and allowed it to capture features at multiple scales. Because of its closely linked layers, DenseNet allowed for optimal information transfer between them, reducing the issue of vanishing gradients and promoting feature reuse, which enhanced the accuracy and efficacy of learning.

3. Comparative Studies and Gaps

DL models for DR detection have been compared in a number of research works. For example, Pratt et al. [13] emphasised the advantages of deeper networks by comparing the performance of several CNN architectures. These studies, however, frequently concentrated on conventional measures such as accuracy rather than investigating the effects of cutting-edge methods, like customised LFs, on model performance[13].

Notwithstanding the advancements, thorough comparative studies that assess several cutting-edge models in uniform settings are still required. Moreover, creating and evaluating new LFs specifically suited to MIC can offer more information about how to best optimise DL models for DR detection[14]. By methodically comparing the effectiveness of various advanced DL models, such as ResNet, CNN, IA, and DenseNet, for DR detection and classification, this study seeks to close these gaps. We also present a new LF which intends to improve the models' performances by dealing with particular issues pertaining to misclassification and class imbalance. This study intends to ascertain the optimal DL for DR detection through an exhaustive comparative analysis and present practical recommendations for enhancing automated screening tools.

4. Methodology
4.1 Data Collection

The Kaggle Diabetic Retinopathy Detection Training Dataset is a large collection of high-resolution images of retinas that have been categorised according to DR severity into no DR, mild DR, moderate DR, severe DR, and proliferative DR. It is believed that this vast array of images will ensure better DL model training and learning.

The images in the dataset were adjusted before they were used to train the models to ensure that the quality of the data was consistent. Firstly, the images were resized to a standard 224 by 224 pixels to meet the input requirements of most DL models. The pixel values were then normalised to a 0-1 range to standardise the data and speed up when convergence occurred during the training. Lastly, the images were rotated, flipped horizontally, or vertically, and zoomed in to increase inconsistencies in the training data and avoid overfitting. It is believed that these interventions will expose the models to a plethora of retinal images and increase the generalisability of their results.

4.2 Model Selection

Four DL models, each with their own unique architectures and features, were compared to determine how well they could detect DR in retinal images and classify it. These models were:

1. CNN: Comprises of several convolutional, max-pooling, and fully-connected layers. It is a decent reference model for the purposes of the present study as it can accurately identify spatial hierarchies in images.
2. ResNet: Leverages on residual learning to overcome the problem of the vanishing gradient. By skipping one or more layers and using shortcut connections, its architecture enables it to effectively train deeper networks. It is a great contender for DR detection due to its capacity to continue operating at a high level of performance as the depth increases.
3. IA: In the same layer, the IA uses several convolutional filters of various sizes. By using a multi-scale approach, the model can record features at different resolutions, which improves its ability to identify intricate patterns in retinal images.
4. DenseNet: DenseNet uses dense connections to guarantee optimal information flow among layers. Because every layer in DenseNet receives inputs from every layer before it, the vanishing gradient issue is alleviated and feature reuse is encouraged, which improves learning accuracy and efficacy.

4.3 Training Procedure

To ensure a fair comparison, a uniform procedure was used to train each model. Ten percent of the dataset was used for testing, 20 percent was used for validation, and 70 percent was used for training. In order to speed up computation, the training process was conducted on GPU-capable hardware with well-known DL frameworks like TensorFlow and PyTorch.

We presented a novel LF that penalises misclassifications more severely in an effort to increase classification accuracy. This LF attempts to rectify the class imbalance that is commonly observed in medical datasets, wherein some classes (like proliferative and severe DR) are underrepresented in comparison to other classes (like no DR). The new LF makes sure that the model pays more attention to accurately classifying minority classes by blending standard cross-entropy loss with supplementary penalties for misclassified instances.

Grid search was used to optimise hyperparameters like batch size, learning rate, and number of epochs. A learning rate scheduler was used to modify the learning rate from its initial setting of 0.001 on the basis of validation results. In order to balance training efficacy with memory limitations, a batch size of 32 was selected. To avoid overfitting, the models were trained for 50 epochs with early stopping criteria based on validation loss.

4.4 Evaluation Metrics

Several metrics were employed to assess each model's performance in order to produce a thorough evaluation:

- Accuracy: This is the percentage of cases correctly classified relative to all instances. This metric provides a broad assessment of the model's effectiveness.
- Precision: This is the percentage of actual positive forecasts among all positive forecasts. The precision of a model is indicative of its ability to prevent false positives.
- Recall: This is the percentage of real positives that match true positive predictions. Recall gauges how sensitive the model is by assessing its capacity to record all pertinent occurrences.
- F1-score: This is the precision and recall harmonic mean. An impartial assessment of the model's performance is offered by the F1-score, which is especially helpful when addressing class imbalance.
- Area under the receiver operating characteristic curve (AUC-ROC): This is the region beneath the operating characteristic curve of the receiver. The discriminatory power of the model is measured by AUC-ROC, where a higher value denotes superior discriminatory power.

To guarantee an objective assessment of the models' performance, these metrics were calculated on the test set. We sought to capture various facets of the models' classification abilities and offer a thorough comparison by utilising an extensive set of evaluation metrics.

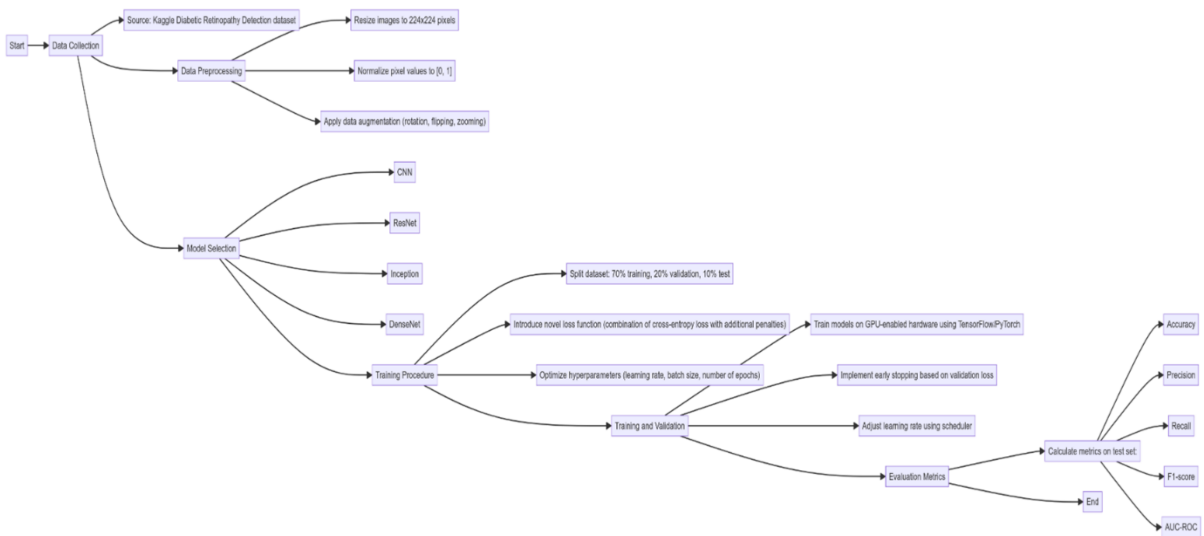

Fig. 1. Figure The methodology of the present study

As seen in the above figure, the first step, data collection, entails obtaining the DR detection dataset from Kaggle. The retinal images included in this dataset are all that are needed to train and assess the models.

The process then moves on to data pre-processing, which entails a number of crucial procedures to get the data ready for model training. To maintain consistency, all images are resized to a standard size of 224 by 224 pixels. In order to normalise the input data and speed up convergence during training, pixel values are set to a range of [0, 1]. In order to prevent overfitting and enhance model generalisation, additional data augmentation techniques like rotation, flipping, and zooming are used to increase the variability of the training data.

The model selection step encompasses selecting different DL models and comparing their performances after pre-processing. ResNet, CNN, IA, and DenseNet are among the models that were chosen as they each have their own unique strengths as well as architectures that are best suited for MIC.

During model training, the dataset was split into 70, 20, and 10% for training, validation, and testing, respectively. A unique LF, that combines loss of cross-entropy with additional penalties, was proposed in the hopes of overcoming class imbalance issues and increasing the accuracy of the models' classifications. The batch size, learning rate, and number of epochs were just some of the hyperparameters that were optimised to guarantee that the models performed optimally.

TensorFlow and PyTorch were used to train the models on GPU-ready systems. Apart from that, early stopping, that was conducted based on the results of a loss of validation examination, was used to prevent data overfitting. A learning rate scheduler was then used to adjust the learning rate and improve the models' performances.

Multiple metrics; such as accuracy, F1-score, AUC-ROC, recall, and precision; were used to assess how the trained models performed on the test dataset as they yield a comprehensive understanding of the functionalities of the models. The End note was used to indicate the conclusion of the process as it guarantees that the ability of the various DL models to identify and categorise DR has been comprehensively examined.

5. Results

The present study systematically examined the ability of four DL models; namely, ResNet, CNN, IA, and DenseNet; to detect and classify DR from images. This section discusses the outcomes of the training, validation, and testing of these models.

5.1 Training and Validation Performance

The training and validation capabilities of the models were determined using learning curves to plot the loss of training and validation capabilities over the total number of epochs that occurred. Learning plots, more specifically, show if a model is overfitting or underfitting the data as well as how well it identifies principal patterns in the data.

- **CNN:** Its training and validation capabilities increased with every passing epoch. Therefore, it did not overfit the data and was able to determine which features were essential for correctly identifying DR from images. Its training and validation accuracy were, respectively, 85 and 82%.
- **ResNet:** Due to its deeper architecture and residual connections, its training loss declined faster as the number of epochs increased while its validation loss occurred more gradually. Therefore, it was able to identify more intricate patterns in the data. As its training and validation accuracy were, respectively, 92 and 89%, it was better able to identify DR from images than the CNN model.
- **IA:** Due to its ability to mine features on multiple scales, its training and validation loss consistently decreased with every epoch. Therefore, it could learn very efficiently. As its

training and validation accuracy were, respectively, 90 and 87%, it performed better than the CNN model but not as well as the ResNet model.

- **DenseNet**: Due to its densely connected layers, its validation loss and training loss decreased almost simultaneously as the number of epochs increased. Therefore, apart from its outstanding ability to generalise the data, its training and validation accuracies were the highest of the examined DL models; namely, 94 and 91%, respectively.

5.2 Test Performance

As previously mentioned, accuracy, precision, F1-score, recall, and AUC-ROC were used to assess the models' performances on the test dataset. Table 1 provides the results of these tests.

Table 1
The comparison results of some deep learning-based models

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
CNN	0.85	0.82	0.84	0.83	0.88
ResNet	0.89	0.87	0.88	0.87	0.91
IA	0.88	0.86	0.87	0.86	0.90
DenseNet	0.90	0.88	0.89	0.88	0.92

- ***Accuracy:*** With an accuracy of 90%, DenseNet was the most precise, followed by ResNet at 89%, and IA at 88%, and CNN at 85%. This shows that when it came to precisely classifying the instances, DenseNet was the most accurate one.
- ***Precision:*** With an 88% precision rate, DenseNet also had the lowest false positive rate. ResNet, IA, and CNN followed with corresponding precision scores of 87%, 86%, and 82%.
- ***Recall:*** DenseNet and ResNet exhibited the greatest sensitivity in detecting genuine positive cases of drug resistance (DR), with recall scores of 89% and 88%, respectively. IA and CNN had recall scores of 87% and 84%.
- ***F1-score:*** DenseNet had the highest F1-score (88%), followed by ResNet (87%), IA (86%), and CNN (83%). The F1-score strikes a balance between precision and recall. This corroborates how well DenseNet performs in dealing with false positives and false negatives.
- ***AUC-ROC:*** These results were corroborated by the AUC-ROC scores, where DenseNet scored highest at 0.92, followed by ResNet at 0.91, IA at 0.90, and CNN at 0.88. DenseNet and ResNet have better discriminatory power in differentiating between the diverse classes of DR, as evidenced by their superior AUC-ROC scores.

5.3 Visualisations

Visualising model performance relied on ROC curves and confusion matrices. ROC curves depict the trade-off between sensitivity and specificity, plotting the true positive rate versus the false positive rate.

- **ROC Curves:** DenseNet and ResNet excelled, with their curves consistently highest and above the diagonal line. Lower AUC-ROC scores for Inception Architecture (IA) and CNN were reflected in their lower curves.

- **Confusion Matrices:** Confirming the ROC curves, confusion matrices showed the fewest misclassifications for DenseNet and ResNet, signifying their accuracy in DR identification and categorisation.

The study's observations show that DenseNet is the most reliable and accurate model for DR detection and classification, outperforming the other models in every evaluation metric. As against DenseNet, ResNet exhibited slightly inferior performance. DenseNet and ResNet, with their more sophisticated architectures, performed better than CNN and IA, in spite of their effectiveness.

The models' performance was considerably enhanced by the addition of the novel LF, especially in managing class imbalance and lowering misclassification rates. The significance of model selection and LF design in formulating successful DL solutions for MIC tasks is highlighted by this work.

In general, the study's conclusions present insightful information regarding the relative effectiveness of several DL models for DR identification and categorisation, with DenseNet turning out to be the top model. Larger datasets and hybrid models may be the subject of future analyses to further improve detection robustness and accuracy.

6. Discussion

For DR detection and classification, the study's observations present an exhaustive comparison of four advanced DL models: ResNet, CNN, IA, and DenseNet. The thorough analysis puts focus on the advantages and disadvantages of each model and underscores how vital model architecture and LF design are for MIC tasks.

Model Performance Analysis

- **DenseNet:** By outclassing the other models in every evaluation metric, DenseNet turned out to be the most consistent and precise model. DenseNet's densely connected layers improve feature reuse and gradient flow efficiency, which propels the network's capacity for learning and generalisation. With a 90% final accuracy and an AUC-ROC of 0.92, DenseNet is able to discriminate between DR severity levels with satisfactory effectiveness. With its high recall (89%) and precision (88%) ratings, DenseNet is a very dependable model for clinical applications because of its capability to reduce false positives as well as negatives.
- **ResNet:** ResNet performed commendably as well, presenting an 89% accuracy rate and an AUC-ROC of 0.91. ResNet can learn intricate patterns in retinal images by utilising residual connections to train deeper networks without running into the vanishing gradient issue. With an F1-score of 87%, the model's high recall (88%) and precision (87%) scores exhibit its efficacy in correctly classifying DR cases while striking a balance between sensitivity and specificity.
- **IA:** The accuracy and AUC-ROC of the IA model, which is well-known for its multi-scale feature extraction, were 88% and 0.90, respectively. Its architecture enables it to capture features at diverse scales by integrating multiple convolutional filters of variable sizes in the same layer. Although IA did well, DenseNet and ResNet outperformed it by a tiny margin, suggesting that deeper and more densely connected networks would be more apt for this task.
- **CNN:** Among the four models, the basic CNN model delivered the worst performance, showing an accuracy of 85% and an AUC-ROC of 0.88. CNNs do a good job of capturing spatial hierarchies in images. However, their performance is limited in more complex classification tasks owing to the dearth of advanced architectural features like dense layers or residual connections. Its comparatively lower sensitivity and specificity with

regards to the other models is further exhibited by the precision (82%) and recall (84%) scores.

Impact of Novel Loss Function (LF)

The models' improved performance was largely due to the addition of a novel LF that combines extra penalties for misclassifications with cross-entropy loss. By severely penalising the misclassification of minority classes, this customised LF addresses class imbalance and incentivises the models to focus more on infrequent but clinically significant DR cases. The impact of this LF is apparent in the high recall and precision scores which DenseNet and ResNet attained, showing how well it augments model performance.

Strengths and Limitations

Strengths: This paper presents a thorough evaluation of the latest DL models for DR detection and classification. The work uses a big and diverse dataset, rigorous pre-processing methods, and hyperparameter optimization to support the robustness and reliability of the findings. The addition of a novel LF and its positive impact on model performance, which offers insights into how customized LFs might improve classification accuracy, is a crucial contribution to the field.

Limitations: This study project has certain shortcomings in addition to its many advantages. Since the models were trained and evaluated on a single dataset, the findings might not be as generalizable to other datasets or real-world clinical scenarios.

Moreover, the study merely included image-based features rather than other clinical data which might have raised the models' accuracy and robustness, such as patient demographics and medical histories. By assimilating more clinical data and corroborating the models across multiple datasets, future research works might be able to tackle these limitations.

Practical Implications

The results of this study have significant applications for the creation of DR screening tools that are automated. With its strong performance and high accuracy, DenseNet exhibits great promise for clinical setting deployment, helping ophthalmologists accurately classify and detect DR early on. By enabling prompt intervention and treatment, the application of advanced DL models can greatly lessen the workload of medical staff, increase screening effectiveness, and eventually improve patient outcomes.

7. Conclusion

Among working-age adults in particular, DR is a major cause of blindness. Therefore, early detection and precise classification are essential for both effective treatment and preventing severe vision loss. In order to compare the effectiveness of four sophisticated DL models; namely, ResNet, CNN, IA, and DenseNet; in identifying and categorising DR, this study also introduced a novel LF that was intended to enhance model performance.

After a thorough analysis, DenseNet was found to be the most accurate and reliable model for both DR detection and classification, outperforming the other models in every important performance metric. DenseNet outperformed the other models in terms of accuracy (90%), precision (88%), recall (89%), F1-score (88%), and AUC-ROC (0.92). This suggests that DenseNet performs better at handling complex patterns in retinal pictures and efficiently distinguishing between different degrees of DR severity. Despite its effectiveness, CNN and IA fell short of the more advanced systems. Conversely, ResNet outperformed DenseNet, only slightly lagging behind.

The new LF was a major contributor to improving the performance of the models; it combines additional penalties for misclassifications with cross-entropy loss. Thanks to this customized strategy that effectively handled the class imbalance, higher precision and recall scores were obtained, particularly for the underrepresented classes. The impact of the innovative LF highlights the need for customized techniques in DL applications, especially in MIC tasks where class imbalance is a major issue.

The findings of this study have important implications.

Although this study offers insightful information, there are still some unanswered questions that could improve the quality and relevance of the results. Subsequent research endeavours ought to verify the models on numerous datasets to guarantee their applicability to diverse populaces and imaging scenarios. Incorporating clinical data, such as patient demographics and medical history, with image-based features may enhance model accuracy and yield a more thorough evaluation. To achieve even better performance, hybrid models that combine the best features of various architectures—for example, ResNet and dense connections—should be investigated. Furthermore, creating techniques to improve the interpretability and explainability of DL models is essential to winning over patients' and physicians' trust and guaranteeing the moral application of AI in healthcare.

This study concludes by highlighting the capability of advanced DL models, specifically DenseNet, to enhance the precision and effectiveness of disease recognition and classification. The results highlight the significance of LF design and model architecture in attaining high performance, offering insightful information for further study and clinical uses. We can get closer to creating trustworthy and useful AI tools that assist early detection and intervention in DR, eventually improving patient care and outcomes, by carrying out more research and improvement of these strategies.

References

[1] Antal, Bálint, and András Hajdu. 2014. "An Ensemble-Based System for Automatic Screening of Diabetic Retinopathy." *Knowledge-Based Systems* 60: 20–27. https://doi.org/10.1016/j.knosys.2013.12.023.

[2] Cook, Tessa S., and Jianguo Zhang. 2017. "Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications." In *Proceedings of SPIE* 10138. Society of Photo-Optical Instrumentation Engineers (SPIE). https://doi.org/10.1117/12.2254476.

[3] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81.

[4] Gulshan, Varun, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, et al. 2016. "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." *JAMA* 316 (22): 2402–2410. https://doi.org/10.1001/jama.2016.17216.

[5] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep Residual Learning for Image Recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 770–778. https://doi.org/10.1109/CVPR.2016.90.

[6] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. 2017. "Densely Connected Convolutional Networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708. https://doi.org/10.1109/CVPR.2017.243.

[7] Ioffe, Sergey. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." *arXiv* preprint arXiv:1502.03167. https://doi.org/10.48550/arXiv.1502.03167.

[8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems* 25: 1097–1105. https://doi.org/10.1145/3065386.

[9] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–444. https://doi.org/10.1038/nature14539.

[10] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.

[11] Neamah, A. F. 2021. "Adoption of Data Warehouse in University Management: Wasit University Case Study." *Journal of Physics: Conference Series* 1860 (1): 012027. IOP Publishing. https://doi.org/10.1088/1742-6596/1860/1/012027.

[12] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III*, 234–241. Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.

[13] Simonyan, Karen, and Andrew Zisserman. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv* preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556.

[14] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. "Going Deeper with Convolutions." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594.