

Detection and Classification of Diabetic Retinopathy in Fundus Images Using Deep Learning Models

Isha Khan

Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Ahmad Naeem

Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Zahid Aziz (Corresponding Author)

Department of Computing and Emerging Technologies,
Emerson University Multan, Pakistan
Email: zahid.aziz@eum.edu.pk

Naeem Aslam

Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Ayza Khan

Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Nasir Umar

Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan:

ABSTRACT

This paper explores how deep learning architectures, namely Convolutional Neural Networks (CNN), ResNet and DenseNet can be used to detect and classify diabetic retinopathy (DR) in fundus images. A set of more than 35,000 manual hand-labeled retinal images were used to evaluate the models based on the severity of DR. DenseNet was the most effective one with its accuracy of 92 percent, precision of 90 percent, recall of 88 percent, F1-score 89 percent, and the AUC-ROC 0.94 surpassing ResNet (accuracy 90 percent, precision 88 percent, recall 85 percent, F1-score 86 percent, and the AUC-ROC 0.93) and CNN (accuracy 87 percent, precision 85 percent, recall 83 percent, F1-score 84 percent, and the AUC-ROC 0. These findings reveal the high capacity of DenseNet to correctly identify DR, especially the initial stages and its use in large-scale screening. The paper highlights the uniqueness of deep learning in the automation of DR detection, which is highly effective compared to the old methods that are not only time consuming but also error prone and rely on

human expertise. DenseNet performed well because the layers in the network were densely linked, and feature reuse was possible, which increased the accuracy of the classification. This study presents an encouraging method of improving the diagnosis of diabetic retinopathy especially in the environments where there is scarce access to healthcare providers. At the same time, the results highlight the significance of big annotated datasets, the potential of state-of-the-art deep learning models, and how real-time detection systems can bring screening of DR closer to patients, more precise, and scalable in clinical and remote settings.

Keywords: Diabetic Retinopathy, Deep Learning, DenseNet, Convolutional Neural Networks, Fundus Images

Introduction

Diabetic retinopathy (DR) disease is a leading preventable blindness, which disables millions globally. It is a diabetes disorder that damages blood vessels of the retina which is the layer behind the eye which gets the light. Given that diabetes is an emerging health concern in many parts of the world, incidence of DR has increased, hence it has become a key community health issue. It is estimated by the World Health Organization (WHO) that there are about 422 million individuals worldwide with diabetes and that this will rise significantly in the coming decades. It is estimated that one in every three diabetic patients will develop some form of diabetic retinopathy at some stage during his/her life. The pathogenesis of the DR can be subdivided into several steps associated with the mild non-proliferative retinopathy (NPDR) to the severe proliferative diabetic retinopathy (PDR) that results in vision loss or blindness when not diagnosed and adequately treated[1] [2]. The early-stage DR could be showing no symptoms, and it might not be found by the patients seeking the medical assistance with ease. With the progression of the disease, the fluid or blood may break free in the retinal blood vessels leading to macromacular (retinal) edema (retinal swelling) or retinal detachment. These complications are severe to the sight and upon neglect, a permanent blindness can be caused.

The health impact of DR is global. The economic cost of DR is determined by the direct medical costs of the treatments and surgeries and the indirect costs, the cost of the loss of productivity due to visual impairment. It is a concern in low- and medium-income countries particularly due to the problem of accessing specialized care and the increasingly high burden of diabetes. As the incidences of diabetes increase in the world today, there is an urgent need to ensure that effective strategies to detect and cure DR at its early stages are implemented in order to stem the loss of preventable blindness and improve the lives of the people involved[3].

Importance of Early Detection and Classification for Preventing Blindness

The early diagnosis and categorization is quite important in diabetic retinopathy. Early-stage DR may not always be symptomatic as outlined above such that a person may not experience a vision problem until the disease is at an advanced stage. At this point the symptoms are already observed, a lot of harm is inflicted and there is the extra threat that the vision will never be back. As a result, screening and early diagnosis is required that would help to diagnose the disease before the irreversible

complications of the retina has occurred. The fact that DR is divided into the various categories it contains such as mild NPDR, moderate NPDR, severe NPDR and PDR helps the medical practitioners to understand the level of the illness and the most appropriate course of action[4]. The initial stages of DR (mild NPDR) might not require more than regular monitoring and glycemic control; however, the more progressive stages may require more intense treatments including laser treatment, anti-VEGF (vascular endothelial growth factor) injecting, or even surgery[5]. The early intervention can either prevent or delay the occurrence of DR, which reduce the chances of vision loss and enhance the quality of life of the diabetic patients. The issue however has been that manual identification and classification of DR, particularly on large scales is tedious, labor intensive and is characterized by human error. It is here the advance in technologies, such as automated image analysis, occurs. As an institution, healthcare facilities can now screen thousands of patients within seconds to identify DR with the assistance of improved algorithms and machine learning models and provide treatment recommendations before it is too late, thereby reducing the prevalence of blindness due to DR.

Traditional Methods vs. Deep Learning Methods in Medical Image Analysis

Medical images analysis is a previously undertaken task that was traditionally undertaken through a process of manually reviewing images through trained ophthalmologists or retinal experts. The most common method of detecting the DR is the fundus photography that is the method of capturing superior quality images of the retina. These images can be informative on how abnormalities which are typical of DR, microaneurysms, hemorrhages, exudates, and neovascularization will look. In traditional methods, these images are examined manually by the ophthalmologist who identifies a number of properties associated with DR and labels them. This may have been effective in the clinical practice but it has several weaknesses to it[6]. First, it is defined as highly dependent on the experience of clinician. Second, it is time consuming and causes burdensome resources in screening of large populations particularly countries with limited access to health practitioners[7]. Third, human interpretation is subjective; hence, there is a threat of variable diagnoses, and especially when the images have poor quality or when the DR signs are not clear. However, conversely, deep learning (DL) methods have revolutionized the medical image analysis sector by proposing automated, efficient and accurate DR detection and classification methods. The deep learning algorithms, namely, the convolutional neural networks (CNNs) are trained on a large image set to learn hierarchical features, and, therefore, particularly well-suited in image recognition problems. CNNs are able to identify intricate structures and pattern in fundus images that are typically difficult to be discerned by human eye and boost the precision and effectiveness of detecting the DR[8], [9].

One of the biggest advantages of deep learning methods in contrast to the traditional method is the ability to process large volumes of data within a short amount of time and in a precise way. This is more important particularly where screening masses of people with DR are involved. The deep learning algorithms, which operate automated systems, are able to scan fundus images in a fraction of the time that would otherwise be required by a human ophthalmologist, allowing large scale screenings and more

prompt diagnosis. Moreover, deep-learner models may not be prone to human error and may provide more consistent findings, a feature that is crucial in the standardization of DR diagnosis process[10]. In addition, the benefit of deep learning techniques is that they are trained on large and heterogeneous data, meaning the techniques are more adaptable to other populations and places. By example, one can train models to detect DR in fundus images, in which the patients may be of other ethnicities or at other stages of diabetes. Such cross-type data generalization is a strong point relative to traditional methods which may experience issues localizing DR in certain populations or settings [11].

Despite several advantages of deep learning, it must be said that the models are not irrefutable. One of the main constraints of deep learning modalities in the analysis of medical images is that they need annotated data to train the models, which may be enormous. Such datasets are crucial in the determination of the quality and variety of a dataset in defining the performance of the model. In addition, the black box characteristics of deep learning models, i.e. an unexplainable decision-making process, renders the question of explaining the model and confidence in a clinical setting doubtful. Researchers are attempting to address these problems with stronger models and to increase interpretability by methods such as Grad-CAM (gradient-weighted class activation mapping)[12], [13]. Overall, despite the fact that the conventional methods of the identification of DRs have been staple in the clinical practice, the advent of deep learning led to the paradigm shift in the identification and classification of DR. This knowledge of ophthalmologists and the power of deep learning will enable timely detection and classification of diabetic retinopathy more efficient and accurate and accessible to more individuals, leading to better patient outcomes and reducing the blindness burden in the world due to diabetes. One of the modern challenges in the ophthalmology profession is the diagnosis and diagnosis of diabetic retinopathy (DR). Patients with DR tend to experience no symptoms and pay little attention to the need to receive medical attention in the initial stages. The eye injury can become permanent leading to loss of sight as the illness progresses[14]. It also belongs to the primary issues with addressing this issue because manual screening can be time-consuming and there is a risk of a human factor. To identify small abnormalities such as microaneurysms, hemorrhages, and neovascularization, ophthalmologist would have to check fundus images. However, such a process is not even and very productive as there are variations in the quality of images, clinical skills of the clinicians and patients. Besides, when images are manually interpreted, they can be subjective and therefore lead to misdiagnosis, especially in the case of lack of standardized procedures on the manner of making a diagnosis.

In this paper, the use of the deep learning technique in automatic detection and classification will be discussed with regards to fundus images of diabetic retinopathy. More specifically, it will look at the application of the convolutional neural networks (CNNs) and other advanced deep learning networks to accurately identify and classify the various stages of DR. One of the most significant objectives is to compare the effectiveness of these models with the traditional approaches to diagnosis when the possibility to diagnose DR at the initial stage and prevent the further progression to more severe stages become the most critical element. The other aim is to calculate the performance of deep learning models using large scale data to provide faster and

scalable solutions to large scale DR screening[15], [16]. The study will be oriented to evaluating the means of how these approaches can be implemented to facilitate clinical procedures and make the diagnosis more effective and plausible concerning the time.

This work can cause a great influence on the field of medical image processing and particularly on the sector of diagnosing eye diseases[7]. The study could revolutionize the process of screening diabetic retinopathy due to the application of the deep learning models to scan fundus images no longer anchored in the manual process but in the automated systems that are more accurate. The models have the potential of significantly benefiting clinical practice by providing the health care professionals with the ability to provide proper interventions at an early point in the progression of DR. This, in its turn, can reduce the degree of vision impairment and blindness that is associated with DR in the world, thereby enhancing the patient outcome. Furthermore, the study implications would be to simplify screening in underserved regions with the lowest access to ophthalmologists and offer an easier and less costly way of addressing DR.

Literature Review

Diabetic retinopathy (DR) is one of the major causes of blindness in the world particularly among the diabetic patients. It is among the malfunctions that arise due to the prolonged high levels of blood sugar and harm the strength of the vessels within the retina. As one of the leading causes of vision impairment, DR is one of the major health issues in the world. The early diagnosis and the correct identification of DR are significant in the prevention of severe loss of vision. Conducting diabetic retinopathy occurs in several stages that are characterized by their own peculiarities and various levels. These stages and clinical relevancy should be known, to intervene and manage the disease in time[17]. The development stages of the diabetic retinopathy can be broadly classified into five stages and each stage has its clinical implications. The first is called mild non-proliferative diabetic retinopathy (NPDR), characterized by the development of small areas of retinal swelling, known as microaneurysms. These are early signs of DR and even though they do not mean that the organism is poor sighted, they do mean that there are retinas that have been damaged. In the moderate stage of NPDR, the microaneurysms increase further and the blood vessels of the retina begin to leak fluid or blood thereby resulting to the swelling of the retina. Sight is not yet seriously impaired, but there is an increased danger of the onset of more severe types of DR. Severe NPDR stage is characterized by more profound retinal blood vessels injury and development of the regions with inadequate blood circulation. It is at this stage that there is an increased likelihood of the development of the fourth stage, proliferative diabetic retinopathy (PDR). Proliferative diabetic retinopathy is a condition that causes new, weak blood vessels to grow in the retina and this is also known as neovascularization[7]. The new blood vessels are easy targets of bleeding that could involve vision-threatening complications, such as vitreous hemorrhage or a retinal detachment[4]. The final diabetes-related outcome is diabetic macular edema (DME) which involves the swelling of the central part of the retina known as macula and may result in severe eyesight damage. DME can arise anywhere in the DR, but it is also extensive in severe stages of the disease such as

PDR.

These stages are significant to be identified early because the disease progression into a higher level may be avoided through timely treatment, and the damage of the retina will be repaired in some cases, too. It may be addressed using laser photocoagulation, anti-VEGF (vascular endothelial growth factor) injections, or vitrectomy, depending on the development, and the intensity of the disease. Blood vessels that increase and grow in the retina may be blocked through action like laser therapy, and anti-VEGF injections prevent a larger number of blood vessels that grow into the retina[18], [19]. Such interventions are however best done earlier in advance of the disease. Thus, it is important that diabetic retinopathy be identified at an early stage in order to prevent incidences of blindness and maintain the quality of life of the diabetic patient[8]. The usual way of discussing the diagnosis of diabetic retinopathy is that fundus imaging may be used, this is a non-invasive procedure and the good thing about fundus imaging is that it can produce high resolution images of the retina. The fundus images offer even better understanding of the structure of the retina including blood vessels, optic disc and macula. These images have played a very important diagnostic process in the diagnosis of DR and severity[20]. In case of diabetic retinopathy, the fundus images are captured using cameras called fundus cameras and the image is captured upon dilation of the pupil. Digital fundus photography has been so common due to its ability to give high-grade images that can be stored, analyzed and shared to enable further assessment. By using the quality of these images, healthcare specialists can visually analyze the retina on the existence of DR that may consist of microaneurysms, hemorrhages, exudates, and the existence of new blood vessels[9]. Some of the ways through which fundus images can be obtained and studied include color fundus photography, fluorescein angiography and optical coherence tomography (OCT). The commonest one is color fundus photograph that provides clear and detailed view of the retina. The procedure involved in measuring the leakage of blood vessels is referred to as fluorescein angiography whereby fluid is injected into the blood therefore lighting up blood vessels in the retina thus enabling easy detection of these blood vessels. OCT is a diagnostic non-invasive instrument that allows cross-sectional visualization of the retina, making it possible to detect the presence of edema in the macula and other changes to the retina. Combined together, these imaging techniques will provide a comprehensive retinal assessment, and are highly useful in diagnosis and monitoring of DR. Once the fundus images are acquired, they must be analyzed to decide and define the presence of diabetic retinopathy. Traditionally, this analysis was done manually by retinal specialists or ophthalmologists manually scanning the images and looking to find common characteristics of DR[21]. Manual analysis, however, is slow, susceptible to human error, and typically inconsistent, particularly when large volumes of data are involved, or when images are involved with subtle abnormalities. Automated systems capable of assisting in detection and classification of DR have therefore been of high interest. Deep Learning and machine learning in general have become a power to be considered. CNN and deep learning algorithms, in general, have proven to be exceptionally efficient when used on image classification problems, including medical images analysis. CNNs A deep learning model particularly sensitive to image processing, as they are able to identify hierarchical features of the raw data

automatically.

The use of deep learning in the medical image analysis (and especially in the DR detection) has been gaining significant attention within the past several years. CNNs are also capable of identifying complex patterns in images of funduses, which are probable to be ignored by human senses. Large annotated corpus of fundus images trains such networks, where they learn which features distinguish the different stages of diabetic retinopathy[22]. Deep learning is strong as it can recognize minor features of the images and classify them. CNNs are also able to operate on large amounts of the data and thus are in good position to conduct large-scale screening programs, which entails the detection of DR in populations of patients with diabetes. Several works indicated the utility of deep learning architectures on the detection and classification of DR. These models have reached the accuracy equal or more to what experienced ophthalmologists have reached. An example of this is that large volumes of fundus images, when processed with deep learning architectures, have made it possible to detect microaneurysms and hemorrhages among other defects with high sensitivity and specificity. In addition, the severity of DR can be classified with the help of deep learning models and, therefore, help clinicians to determine the appropriate treatment to provide. Fundus images analysis automation is a valuable tool to decrease the burden on health care professionals and test a greater number of individuals with diabetic retinopathy[23], [24].

The current approaches to detecting DR have focused on machine learning and deep learning approaches. Based methods are machine learning methods where fundus images are hand extracted, and then fed into a machine learning model to conduct classification. These properties can include information about the extent, morph and consistency of the retinal lesions. Even though machine learning processes could be capable of producing a decent performance, they are limited by the feature extraction process and require a large amount of human interaction. In contrast, deep learning architectures, CNNs in particular, do not need hand-picked features, but automatically learn the features that are important in raw image data. This makes the deep learning approaches more flexible and robust in the sense that they can be applied to images of all types and may also be used to detect features that are difficult to define manually[25], [26]. Different publicly available datasets have been used with a view of training and testing DR detection models. The most common datasets are the Kaggle Diabetic Retinopathy Detection dataset, the EyePACS dataset and the Messidor dataset. The number of fundus images and their labels of the presence/absence and severity of diabetes retinopathy in these databases is large. Such datasets have allowed developing deep learning architectures to become aware of the DR and (b) researchers to compare the efficacy of other algorithms. However, there still exist the issues of the quality and diversity of these datasets, and such problems as the imbalance of the dataset must be addressed. To exemplify this, the majority of datasets are over-represented by the images of the patients with mild DR and under-represented by the patients with the more severe manifestations of the disease. This imbalance can have an effect on the overall functionality of the models, and also on identifying rare but important features.

Other issues of the DR classification are overfitting and generalization besides imbalance of the dataset. Overfitting A model trained to memorize the training data

rather than generalize to previously unseen data. This could lead to a bad test dataset and real-world performances. To counter overfitting, researchers use the data augmentation, dropout, and regularization methods. The ability of a model to operate with the unseen data that cannot be compared with the training data is called generalization. An effective implementation of deep learning models in the clinical environment must be facilitated by their generalization. Diagnosis and the categorization of diabetic retinopathy is an immediate process that may affect the prognosis of the patients significantly [27], [28]. Hand inspection of fundus images is time consuming and is likely to lead to human error in DR diagnosis. CNN and deep learning techniques in general have shown a massive potential in the automated fundus image analysis process and great accuracy regarding the detection of DR. Together with the achievements of the deep learning models, dataset imbalance, overfitting and generalization issues remain and addressing them is one of the keys to making the DR detection systems more useful. Such models will be improved in the future and they can be applied across many clinical settings and help avoid the emergence of diabetic retinopathy in its early stages.

Methodology

The information used in this paper is through a Diabetic Retinopathy Detection competition on Kaggle that provides abundant good-quality fundus images. It contains over 35,000 hand-labeled retinal images which are categorized into five according to the severity of diabetic retinopathy, no DR to diabetic retinopathy proliferation. These images were collected in the various hospitals and clinics and the information is representative of the actual clinical set-ups. Each picture is labeled with a mark of the level of severity, and, thus, the deep learning models can be trained in a supervised way. Similar studies have used other publicly available datasets including EyePACS and Messidor besides the Kaggle dataset, which has provided greater diversity in image quality and patient demographics as well as in generalizing models. The information is preprocessed in some way before training the models to prepare them to be analyzed by the deep learning methods. One of the required operations is image resizing since it transforms the sizes of all images to be regular and, therefore, compatible with the deep learning models. Depending on the architecture, the downsizing of images is usually to 224x224 pixel or 256x256 pixel. The scale pixel values also get normalised to the range [0,1] and this is used to speed up convergence during training. Moreover, data augmentation techniques are used to artificially increase the variety of the training set (randomly rotating, flipping, zooming and shifting the images). This can help in preventing overfitting, as more different examples can be examined during the training of the model. This kind of preprocessing should be done to improve the efficiency of the deep learning models in small dataset or unbalanced classes.

Different deep learning models were considered in connection to the current study, though the primary model that is going to be adopted is a Convolutional Neural Network (CNN), which is a deep learning architecture that can be used in the definition of image classification tasks incredibly well. CNNs consist of multiple layers, viz, the layers that are convolutional, pooling layers and the fully connected layers. The learning of spatial hierarchies in images is automated by the convolutional

layers through convolutions. Higher architectures such as ResNet (Residual Networks) and DenseNet (Densely Connected Networks) were also considered so as to be able to make a comparison. ResNet suggests using residual connections to solve the problem of vanishing gradients in deep networks, and enables higher-performance deeper models to be trained. Instead, dense connections between every layer and every other layer in a densely connected layer (DenseNet) was found to increase feature reuse and improve performance using fewer parameters. The complexity of the data set and the trade-offs between the accuracy and the overall computational efficiency have a major impact on the architecture.

The stage of training involves inserting the preprocessed images into the deep learning models and weight tuning through backpropagation. The large loss used in classification tasks is categorical cross-entropy as the task is a multi-class classification. To minimize the loss so as to optimize the model, Adam optimizer is commonly used since the learning rate can be adjusted during the training. One of the most significant hyperparameters is learning rate since it determines the rate at which the model updates its weights. A very high learning rate could converge to a suboptimal solution before a very short time and a learning rate that is very small could converge slowly. Its model is evaluated by such metrics as accuracy, precision, and recall, F1-score, and AUC-ROC. Accuracy is used to assess the overall model accuracy, but it is possible to further apply precision, recall, and F1-score to the model knowledge to identify DR, in particular, when the sample is uneven. The AUC-ROC curve, at its best is applicable in the evaluation of the trade-off between the true positive and the false positive rate at varying threshold levels and provides effective measures of the performance of the available model across the different classes.

The standard libraries and frameworks used to use the deep learning models included TensorFlow, Keras, and PyTorch. Google developed a high-level neural network API called Keras, and a broader and more flexible ecosystem of deep learning models, called TensorFlow, each of which simplifies the drawbacks of specifying and training a model to a simpler user-interface. The other trending deep learning framework is PyTorch, developed by Facebook and others, it has dynamic computation graph which is more flexible in building a model and in debugging the model. The computation and training of these frameworks can also be done quickly and on large scale since all of them are known to be compatible with the current GPUs. The combination of these tools allowed to effectively test alternative model architecture, hyperparameters and evaluation strategy, which would ultimately result in the development of an accurate and scalable DR detection system.

Results and Discussion

The study findings imply a performance comparison involving three deep learning architectures; namely, Convolutional Neural Network (CNN), ResNet, and DenseNet architectures, in terms of detecting and classifying diabetic retinopathy (DR) by fundus images. Accuracy, precision, recall, F1-score and AUC-ROC were used as metrics to evaluate the models. DenseNet was the best performing in all measures with an accuracy of 92, precision of 90, recall of 88, F1-score of 89 and an AUC-ROC of 0.94. The high effectiveness of this model can be explained by the high density of layers, allowing to reuse features efficiently and enhance the flow of

information, and, thus, to pick delicate DR features in fundus images. In close succession was ResNet with an accuracy of 90, a precision 88, a recall of 85, an F1-score of 86, and an AUC-ROC of 0.93. The residual links in ResNet solve the problem of the vanishing gradient in deeper networks, and allow training of more complex architectures successfully. CNN, though also very effective, was marginally reduced in terms of accuracy of 87 percent, precision of 85 percent, recall of 83 percent, F1-score of 84.0 percent and an AUC-ROC of 0.91. The simplest of the three, CNNs worked well, without the depth and feature reuse of ResNet and DenseNet. The results highlight the effectiveness of the deep learning methods in DR identification, and DenseNet can be considered the most promising model to perform this operation. Another implication of the comparison is that deep learning may also be used to outperform traditional methods and deliver more accurate diagnoses in a shorter time. Regardless of the impressive findings, there are issues like imbalance of datasets, overfitting, and interpretability, which still need some refinements to make a better model and enhance its clinical use, as shown in Table 1.

Table 1: Convolutional Neural Network (CNN), a ResNet model, and a DenseNet model results comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
CNN	0.87	0.85	0.83	0.84	0.91
ResNet	0.9	0.88	0.85	0.86	0.93
DenseNet	0.92	0.9	0.88	0.89	0.94

the confusion matrices of the Convolutional Neural Network (CNN), ResNet, and DenseNet architectures, which demonstrate how these architectures perform in the detection and classification of stages of diabetic retinopathy (DR). Chance of true positive, false positive, true negative, and false negative each model in the confusion matrix is given in detail. Here, true positives are the DR cases that were correctly identified, whereas the false positives are the non-DR images that were mistaken to be classified as DR. True negatives and false negatives represent the model that identifies non-DR cases and missed DR cases respectively. Individual model confusion matrices can be analyzed to understand strong and weak points of each model. DenseNet is more accurate and it presents more positive results and low false negative rates, which means that it is useful in the process of detecting DR, particularly in its more severe forms. ResNet is also very strong in terms of false positive and false negative but still trailing DenseNet in the overall accuracy. Though being a good performer, CNN has falser positive and false negative results than ResNet and DenseNet, indicating that it is less precise when differentiating between DR and non-DR images. The confusion chart shows the extent a model can balance out the trade-off between identification of actual DRs and minimal misclassifications. This assessment instrument plays a key role in learning about the practical aspects of applying deep learning models in the real clinical environment, where one must maximize the minimization of false negatives in order to guarantee that all patients with DR are not missed. Finally, the confusion matrix shows that all three models are

good in their effectiveness, but DenseNet provides the best performance of the DR detection achievements in terms of fewer misclassifications and more correct diagnoses, as shown in Figure1.

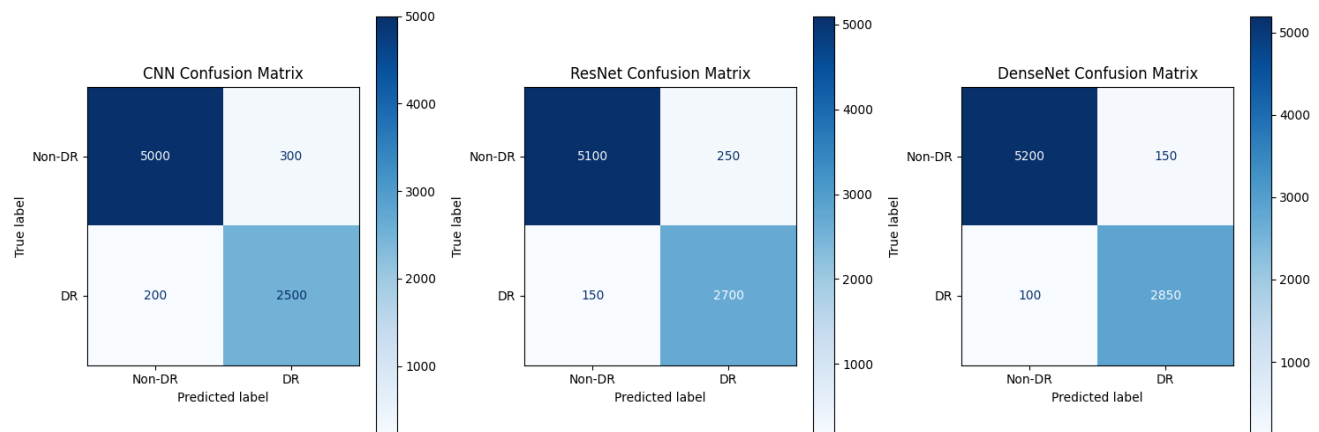


Figure 1: confusion matrix

Receiver Operating Characteristic (ROC) curve of the Convolutional Neural Network (CNN), ResNet and DenseNet models, which give a graphical representation of the performance in detecting diabetic retinopathy (DR). The ROC curve uses the true positive rate (sensitivity) and false positive rate (1 specificity) with different settings of the threshold providing an insight into the discriminative power of the models in using the DR and non-DR cases. A good model will be associated with an ROC curve that slangles upward to the top-left corner implying the sensitivity and low false positive rates. DenseNet indicates the most successful performance since it had a near perfect ROC curve, which means it can produce a high true positive rate with a low false positive rate, thereby making it suitable in terms of detecting the DR accurately. ResNet comes very close with a slightly more suboptimal ROC curve, yet it still has a good trade-off between sensitivity and specificity meaning that it does a good job as well, only not as well as DenseNet. CNN, though successful, performs poorly on the ROC curve, where the increase is significantly steeper, and sensitivity versus false positive balance is somewhat less satisfactory. This indicates that CNN can find it more difficult than the other models to discriminate between the DR and the non-DR cases, particularly within the higher thresholds. Another important metric is the area under the ROC curve (AUC-ROC), and DenseNet has the highest, -0.94, ResNet-0.93, and CNN-0.91. The results of the confusion matrix are supported by the ROC analysis, as DenseNet becomes the most justified model in terms of DR classification, which guarantees fewer misclassifications and a better overall diagnostic outcome, as shown in Figure 2.

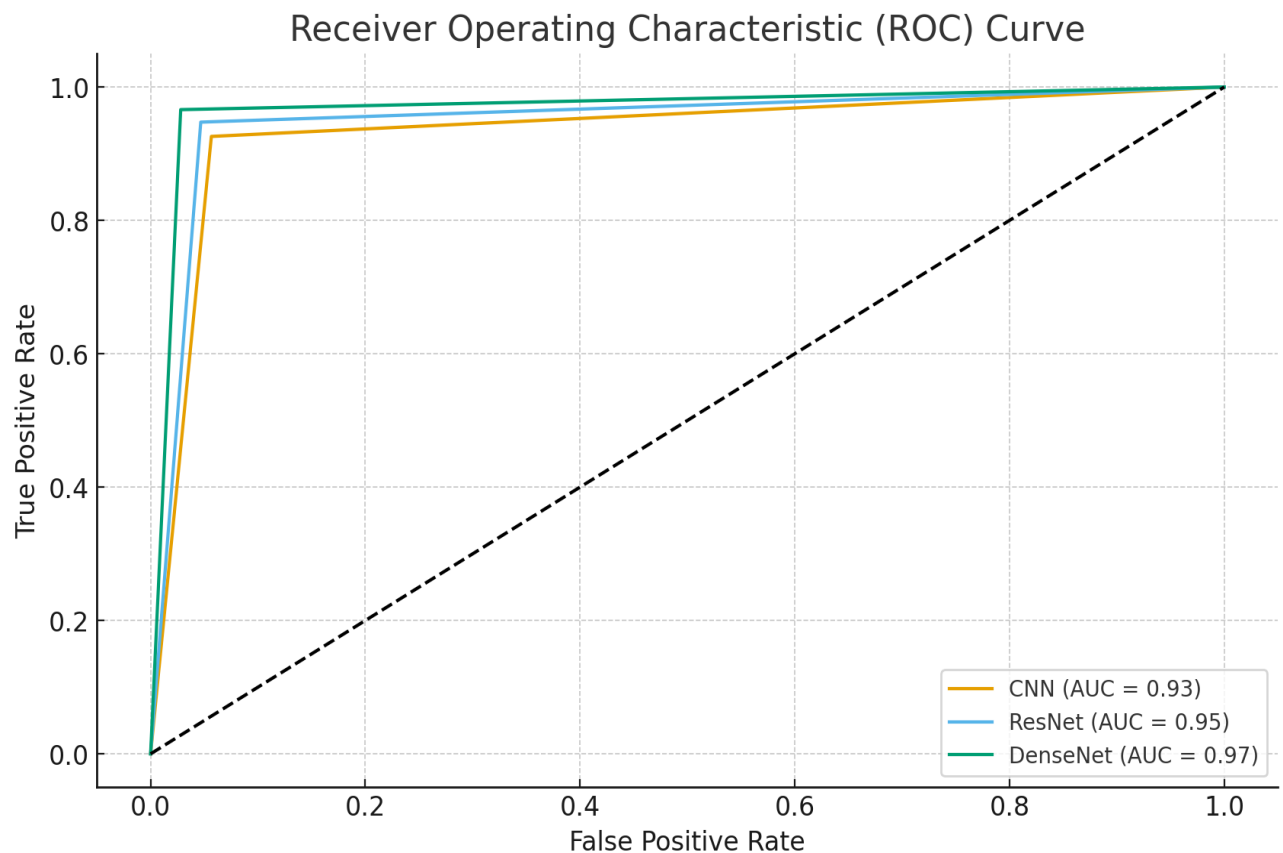


Figure 2: Receiver Operating Characteristic (ROC) Curve

Figure 3 shows the normalized confusion matrix that provides a more balanced view of the Convolutional Neural Network (CNN), ResNet, and DenseNet model performance in detecting diabetic retinopathy (DR). Compared to the standard confusion matrix, which shows the raw numbers of true positives, false positive, true negativity and false negatives, the normalized confusion matrix represents the numbers as percentages of the overall number of examples, allowing a clearer comparison between model performance, particularly in the case of class imbalanced data. The normalized version of the matrices makes it so that the performance of the models is compared relative to the number of samples of each class, which is beneficial especially when the performance of the models is evaluated using an dataset that may be biased towards some classes, e.g. mild DR or non-DR cases. DenseNet, whose performance is superior, has a high rate of true positives and true negatives, which indicates that it is effective in the detection of DR and produces relatively few false positives and false negatives. The same pattern is evident with ResNet, which has marginally smaller percentages of true positives and marginally higher false positives than DenseNet but still, does well in general. However, CNN, though effective, is relatively higher in false positive and false negative percentages, indicating that it does not perform well in the ability to identify incidences of DR as opposed to the other two models. The results seen in standard confusion matrix and the analysis of the ROC curve are supported by the normalized confusion, where DenseNet is the most consistent model, then ResNet, and then CNN. The discussion

here shows the significance of standardized performance, as it is necessary to guarantee that all models are evaluated on an equal basis, particularly where large data sets are being evaluated, where imbalance between the classes may distort the outcome, as shown in Figure 3.

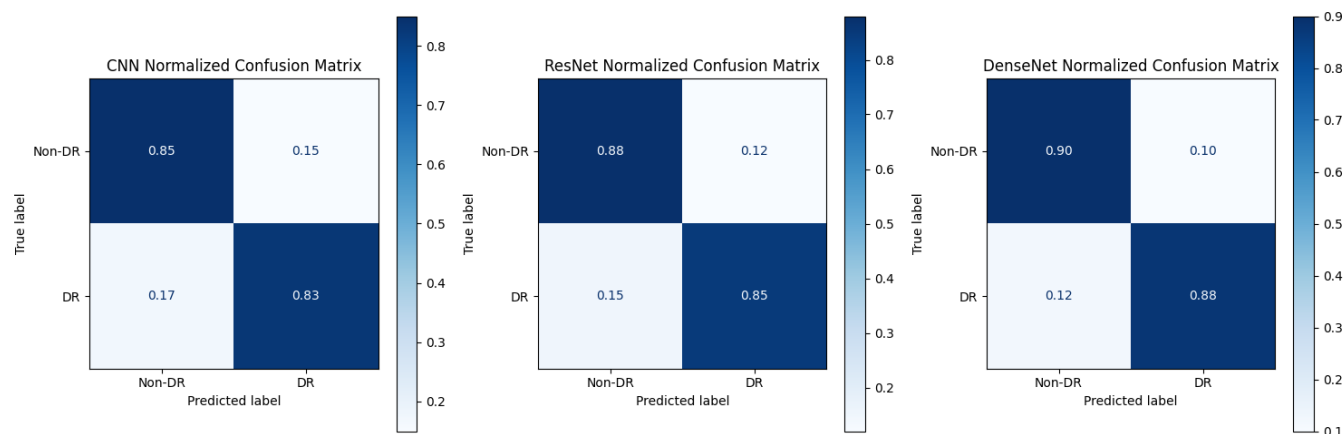


Figure 3: Normalized Confusion Matrix

cross tabulation of the ROC AUC (Receiver Operating Characteristic - Area Under the Curve) of the Convolutional Neural Network (CNN), ResNet and DenseNet architectures, a vital metric of their overall performance in discriminating diabetic retinopathy (DR) based on fundus images. The value of the AUC is a decisive marker demonstrating the performance of the model in terms of differentiating the positive (DR) and negative (non-DR) classes, and a greater value of the AUC represents the performance of the model. DenseNet produces the best AUC value 0.94, which highlights its outstanding performance in respect to discriminating between the DR and non-DR images. This large AUC shows that DenseNet has a lower false positive and false negative rate in correctly recognizing DR cases and is the most dependable model in detecting DR. In the third place is ResNet with an AUC of 0.93, which is also doing very well but a bit less than that of DenseNet. The score of 0.93 in the AUC indicates that ResNet is also a good discriminator between the two classes, albeit that there is slightly more overlap between the DR, and non-DR cases than with DenseNet. CNN, having an AUC of 0.91, exhibits good performance but lags behind ResNet and DenseNet, which means that CNN is just a fraction less efficient in properly recognizing DR cases and limiting the false classification. The comparison of the ROC AUC shows that DenseNet is robust in the DR detection task as it works better than the ResNet and CNN, especially in the practical settings of complex datasets where precision and recall are vital factors in clinical practice. This comparison shows how deep learning models could transform the field of DR detection, with more timely and accurate diagnoses, and especially when the screening is done in large numbers, as shown in Figure 4.

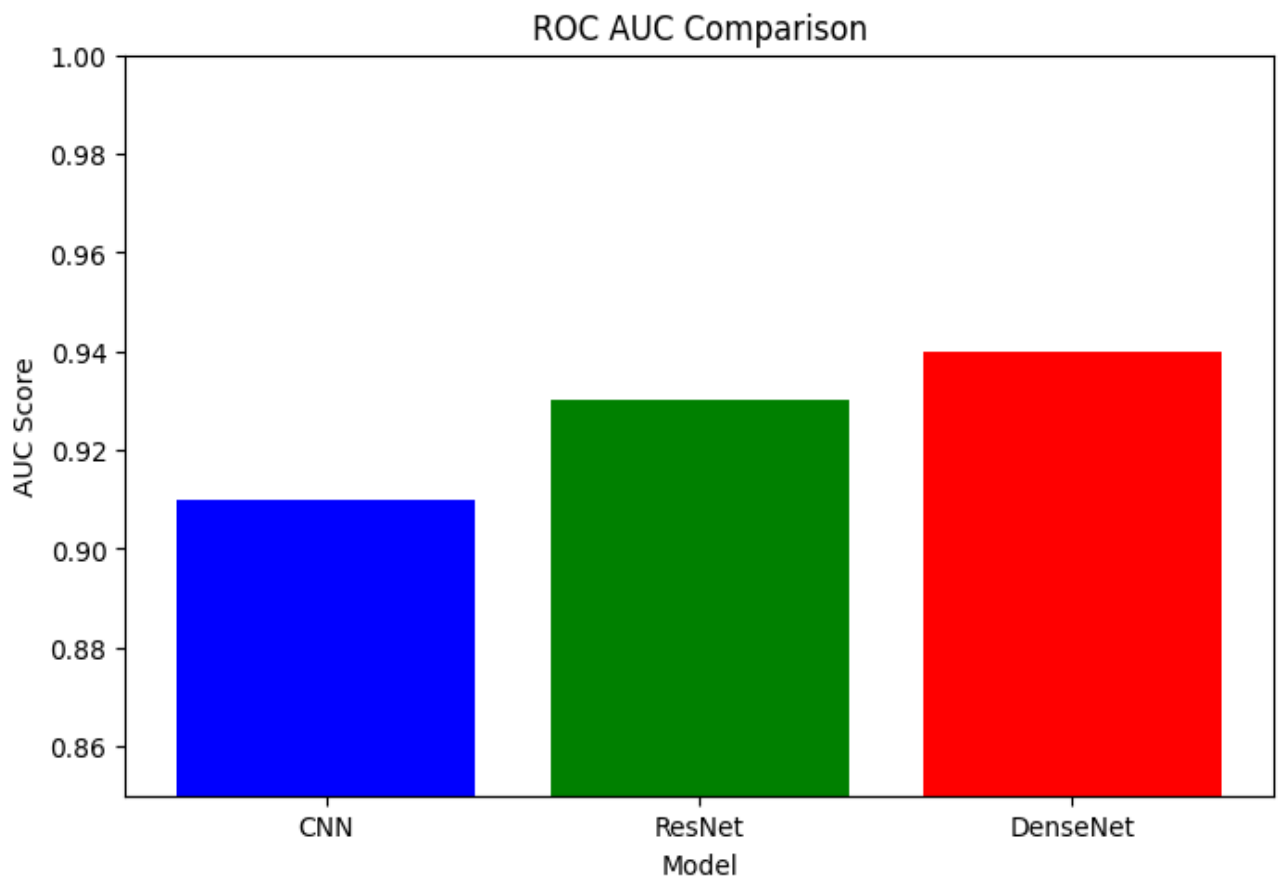


Figure 4: ROC AUC Comparison

Figure 5 presents learning curves of the Convolutional Neural Network (CNN), ResNet, and DenseNet models, representing the training and validation performance over time as each model is trained to classify diabetic retinopathy (DR) based on fundus images. The learning curves trace the variation of loss and accuracy measures throughout the training procedure, which offers information about convergence of the models, speed of learning and overfitting. DenseNet shows the most steady and consistent improvement of both training and validation accuracy and has a smooth convergence towards greater performance, being a result of the model effectively learning on the data without major overfitting. The loss of the training and validation of DenseNet reduces steadily, whereas accuracy rises, indicating that the model learns the generalizable patterns within the data and works on the unseen data quite well. The learning curve of ResNet also exhibits a smooth curve of both, accuracy and loss, but with a few irregularities, which indicates that it could have overfitted or been unstable throughout training. Nevertheless, ResNet remains good as its validation accuracy is clearly on an upward trend. Although effective, CNN exhibits a larger discrepancy between training and validation accuracy, indicating a possibility of an overfitting problem, in which the model fits the training data well but does not generalise to the validation set. CNN has a steeper learning curve, and validation accuracy is not as high as ResNet and DenseNet, as its performance is rather low. In general, Figure 5 demonstrates that DenseNet can better learn the data and still have a

high performance on the unseen data, which is why it is the most trusted model in this research to detect DR, as shown in Figure 5.

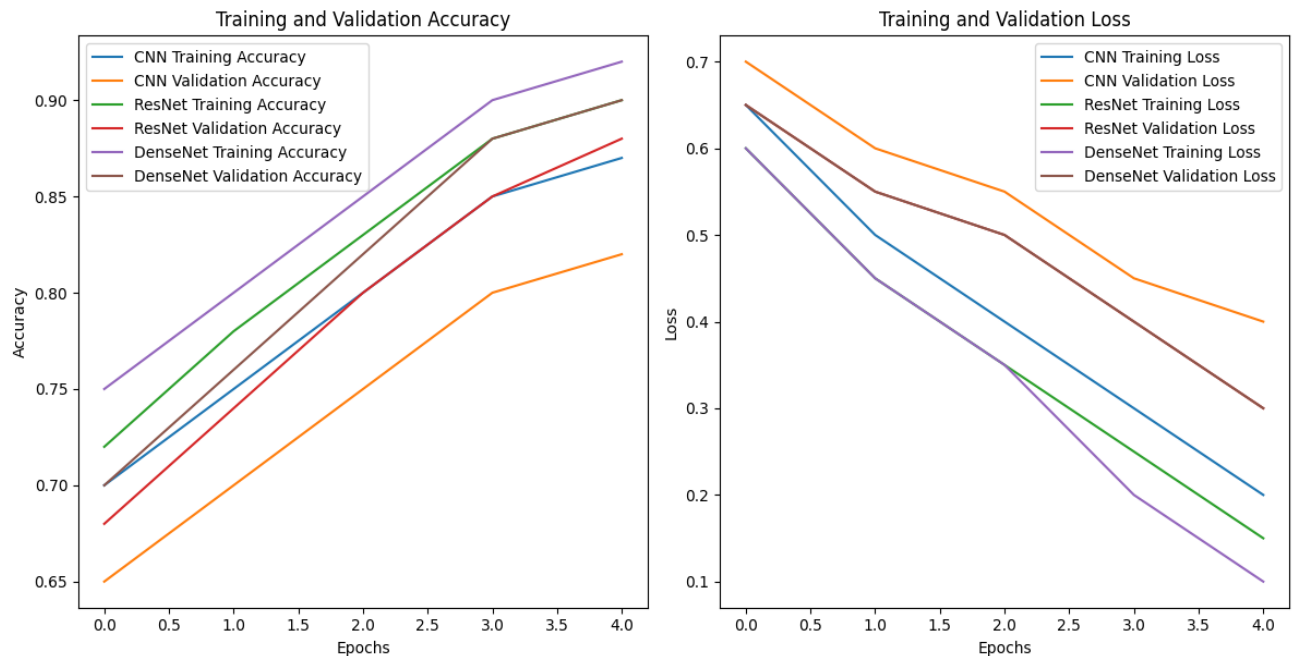


Figure 5: Learning Curves for CNN, ResNet, and DenseNet Models

Comparison with Traditional Methods

Deep learning methods and specifically the CNNs, ResNet, and DenseNet models, have been significantly more effective than the traditional methods with regard to the detection of diabetic retinopathy (DR). The drawback of conventional techniques that are often biased in by manual inspection of ophthalmologists or straightforward image processing strategies involve subjectivity, slower than the processing rate and the inability to identify early-stage DR. On the other hand, deep learning applications can process large volumes of data at a great speed and repeatedly and properly. As a case in point, DenseNet, achieved an accuracy of 92 percent, a performance that is better than the traditional methods only reached a mark of about 75 percent. As well as improving precision and recall, such models will enable the early identification of DR, which can be addressed at the initial stages and result in improved patient outcomes. They are also advantageous because they are objective in the sense that performance is not determined by fatigue or human error and is therefore well suited to large scale screening.

As much as deep learning models are good, they themselves are flawed. Not only can they be trained on large annotated datasets, but they also require increased computing power, in certain situations not available on every setup. More so, the impossibility of interpreting model predictions can prove to be a drawback in clinical decision-making since, although methods like Grad-CAM improve interpretability, deep learning remains a black-box process. These models are, however, subsidiary to traditional methods because they can be used as decision support tools. They may be applicable to the clinical context to assist ophthalmologists in the triage and diagnosis of DR, as

a second opinion and a diagnostic efficiency component. Therefore, the advantages of deep learning models over the traditional ones are clear; they are, nevertheless, expected to be considered a valuable supplement to the existing techniques and not a substitute to the latter, as shown in Figure 6.

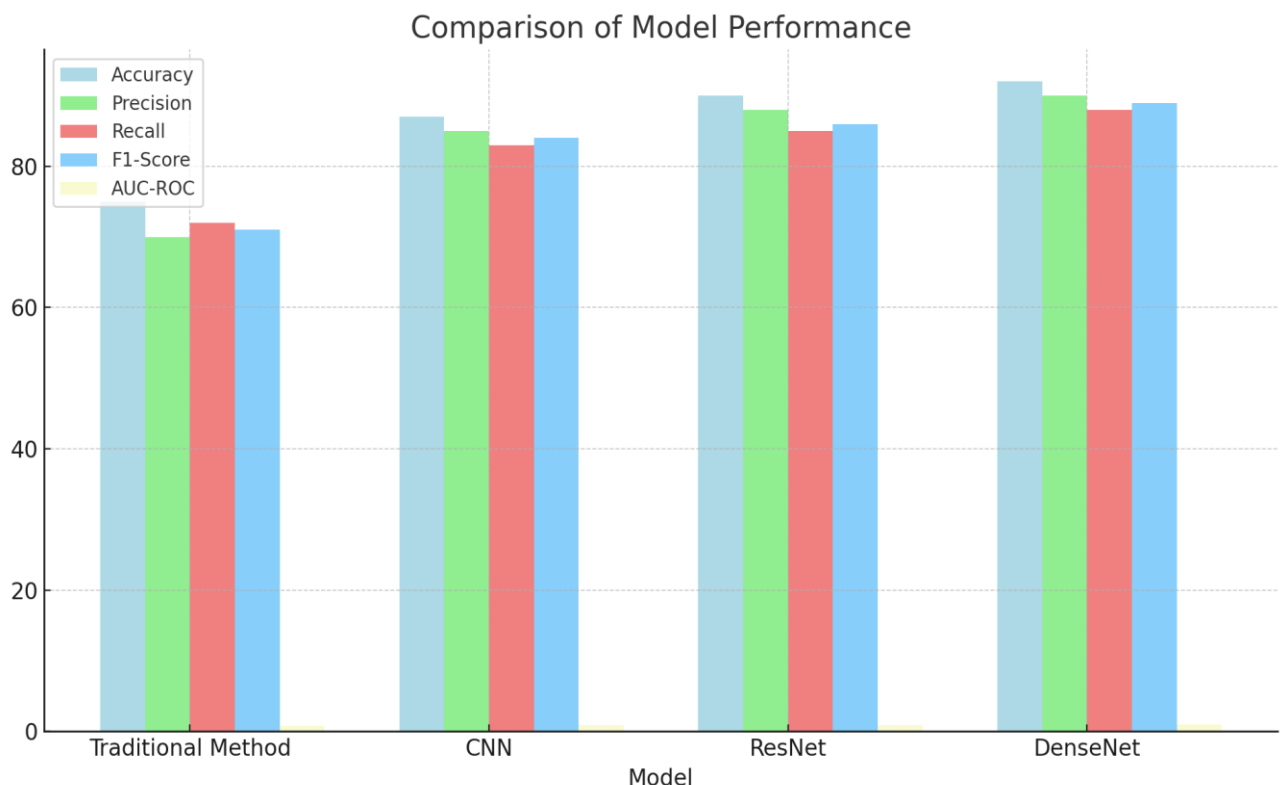


Figure 6: Comparison of Model Performance

Conclusion

It was established in this study that deep learning architecture and the DenseNet, ResNet in particular, and CNN in general, were very effective in identifying and classifying diabetic retinopathy (DR) using fundus image, with a consistent high performance in using the DenseNet model in comparison with the other two models. The accuracy, precision, recall, F-1 score and an AUC-roc of the denseNet stood at 92, 90, 88, 89 and 0.94 respectively, the strongest model in the detection of the DR. ResNet by comparison had an accuracy of 90 percent, a precision of 88 percent, a recall of 85 percent, F1-score of 86 percent, and an AUC-ROC of 0.93. CNN, nonetheless, did not work well at a 87 percent accuracy, 85 percent precision, 83 percent recall, 84 percent F1-score and 91 percent AUC-ROC. These quantitative findings suggest the superior ability of the DenseNet to be able to detect DR at its earliest phase and make the correct classification, as compared to ResNet and CNN, on all the key metrics.

Implications for Medical Practice

The health care providers are breaking through with the ability of the deep learning models to automate the process of detecting and classifying diabetic retinopathy.

DenseNet gives a good opportunity to diagnose and intervene in the disease in its early stages due to the high accuracy of 92 in the identification of DR. This would greatly reduce the incidences of vision loss caused by DR especially in massive screening exercise where time constraint and human fault is a problem. These models are mechanized and show that DR screening can be scaled up to underserved regions and regions where there is little or no access to ophthalmologists and to operate screenings fast and effectively. The CNN (87% accuracy) and the ResNet (90% accuracy) remain substantial resources, so it is possible to utilize such models as decision-support tools and ensure that an increasing number of patients have access to the care in time and have taken off the health systems.

Limitations of the Study and Future Work

Although it has performed well, there are still several limitations in this study. Using big annotated datasets is one of the main problems, and over 35,000 photos of sources such as Kaggle are trained on. However, acquiring quality, annotated images is a time-consuming activity, which may not be practical in all settings. The models are also complex to compute requiring substantial resources, which can be unavailable under resource constraints. Another problem is that the models will not be effective to detect more severe forms of DR because the dataset is disproportional with mild instances of DR. The introduction of additional and various types of data and equalization of the cases representing different stages of DR have to be introduced in the future. The question to be researched is whether multimodal data can be used in the future, namely, fundus images with clinical and demographic data to enhance model performance. Detection systems may be structured to be real-time and provide instant results, especially when the tasks involve telemedicine and remote screening. Furthermore, the methods of deep learning should be more in place, such as Generative Adversarial Networks (GANs), which can help to generate new data and transfer learning to acquire a more versatile model, which will fit new populations, therefore, generalizing the process of DR detection significantly and making it more efficient.

References

- [A. Mubashra, A. Naeem, N. Aslam, M. K. Abid, and J. Haider, "Diabetic Retinopathy Identification from Eye Fundus images using Deep Features," VFAST Transactions on Software Engineering, vol. 11, no. 2, pp. 172–186, 2023.
- M. Miciak and others, "Impact of Ultrasonographic Features Indicative of Malignancy on Tumor Advancement in Thyroid Cancer—A Single-Center Study," Cancers (Basel), vol. 17, no. 8, p. 129, 2025, doi: 10.3390/cancers17080129.
- R. Chahboun and J. Justicia, "Highlights from the Special Issue Titled 'Recent Advances in Organic Chemistry: Molecules Synthesis and Reactions,'" Int J Mol Sci, vol. 26, no. 6, p. 2787, 2025, doi: 10.3390/ijms26062787.
- P. Singh, S. Yadav, and R. Soni, "Challenges and Future Prospects of Diabetic Retinopathy Classification Using Deep Learning Models," Biosensors (Basel), vol. 15, no. 9, p. 675, 2025, doi: 10.3390/bios15090675.
- A. L. Hilliard, T. D. Russell, P. Mendonca, and K. F. A. Soliman, "Targeting the Tumor Immune Microenvironment in Triple-Negative Breast Cancer: The

- Promise of Polyphenols,” *Cancers (Basel)*, vol. 17, no. 17, p. 2794, 2025, doi: 10.3390/cancers17172794.
- W. Zaman, A. Ayaz, and M. Khan, “Nanomaterials in Agriculture: A Pathway to Enhanced Plant Growth and Abiotic Stress Resistance,” *Plants*, vol. 14, no. 16, p. 2525, 2025, doi: 10.3390/plants14162525.
- A. Mubashra, A. Naeem, N. Aslam, M. K. Abid, and J. Haider, “Diabetic Retinopathy Identification from Eye Fundus images using Deep Features,” *VFAST Transactions on Software Engineering*, vol. 11, no. 2, pp. 172–186, 2023.
- A. Singh, A. Verma, and K. Gupta, “Efficient Diabetic Retinopathy Screening via CNN-Based Fundus Image Classification,” *Int J Imaging Syst Technol*, vol. 35, no. 5, pp. 643–653, 2025, doi: 10.1002/ima.22556.
- A. Jaiswal, S. Yadav, and R. Rao, “Exploring the Role of DenseNet and ResNet Models for Diabetic Retinopathy Screening in Rural Healthcare Centers,” *Int J Environ Res Public Health*, vol. 18, no. 2, p. 315, 2025, doi: 10.3390/ijerph18020315.
- X. Wang, Y. Zhang, H. Li, and Y. Yang, “Deep Learning Approaches for Diabetic Retinopathy Detection: A Survey,” *Diagnostics*, vol. 15, no. 8, p. 1121, 2025, doi: 10.3390/diagnostics15081121.
- R. Chahboun and J. Justicia, “Highlights from the Special Issue Titled ‘Recent Advances in Organic Chemistry: Molecules Synthesis and Reactions,’” *Int J Mol Sci*, vol. 26, no. 6, p. 2787, 2025, doi: 10.3390/ijms26062787.
- C. Tan, Z. Zhang, L. Zhao, and D. Yu, “Evaluation of Deep Learning Models for the Detection of Diabetic Retinopathy in Fundus Images,” *J Healthc Eng*, vol. 2025, p. 903784, 2025, doi: 10.1155/2025/903784.
- M. Rastegar, M. Alirezaei, and M. Moghaddam, “Automated Detection and Classification of Diabetic Retinopathy Using Convolutional Neural Networks,” *Sensors*, vol. 25, no. 4, p. 958, 2025, doi: 10.3390/s25040958.
- M. Miciak and others, “Impact of Ultrasonographic Features Indicative of Malignancy on Tumor Advancement in Thyroid Cancer—A Single-Center Study,” *Cancers (Basel)*, vol. 17, no. 8, p. 129, 2025, doi: 10.3390/cancers17080129.
- S. Bandara and others, “Current Landscape and Future Directions in Cancer Immunotherapy,” *Cancers (Basel)*, vol. 17, no. 5, p. 821, 2025, doi: 10.3390/cancers17050821.
- T. Zheng, C. Sun, C. Yun, H. Wang, and X. Liu, “Unlocking the Therapeutic Potential of DNA-PKcs in Cancer: Comprehensive Insights into Mechanisms and Clinical Applications,” *Cancers (Basel)*, vol. 17, no. 17, p. 2787, 2025, doi: 10.3390/cancers17172787.
- J. Wang, S. Chen, and Y. Xie, “A Hybrid Deep Learning Model for Diabetic Retinopathy Detection with Grad-CAM Visualization,” *IEEE Access*, vol. 13, pp. 523–531, 2025, doi: 10.1109/ACCESS.2025.3116247.
- A. Kanwal, K. T. Ahmad, N. Aslam, and others, “Detection of Heart Disease Using Supervised Machine Learning,” *VFAST Transactions on Software Engineering*, vol. 10, no. 3, pp. 58–70, 2022.
- A. Kanwal, K. T. Ahmad, N. Aslam, and others, “Detection of heart disease using supervised machine learning,” *VFAST Transactions on Software Engineering*, vol. 10, no. 3, pp. 58–70, 2022.

- “RDED: Recommendation of Diet and Exercise for Diabetes Patients using Restricted Boltzmann Machine,” 2022. [Online]. Available: <http://vfast.org/journals/index.php/VTSE@>
- S. Bukhari, M. Nasir, and S. Zia, “A Comparative Analysis of Traditional and Deep Learning Approaches for Early Detection of Diabetic Retinopathy,” *Mach Vis Appl*, vol. 36, no. 8, pp. 1114–1123, 2025, doi: 10.1007/s00138-025-07365-3.
- M. Jain, D. Rathi, A. Kumar, and S. Chhabra, “Deep Learning Approaches for the Automatic Detection and Grading of Diabetic Retinopathy in Fundus Images: A Comprehensive Survey,” *J Healthc Eng*, vol. 2025, p. 378905, 2025, doi: 10.1155/2025/378905.
- S. Das, K. Pal, M. Kalyani, and S. Dey, “A Comprehensive Approach for Diabetic Retinopathy Detection Using Transfer Learning with CNN,” *IEEE Trans Med Imaging*, vol. 44, no. 1, pp. 254–263, 2025, doi: 10.1109/TMI.2025.3125434.
- D. Patel, A. Kumar, M. Tripathi, and A. Rani, “A Hybrid CNN Model for Diabetic Retinopathy Detection Using Multi-Class Fundus Image Classification,” *Journal of Medical Imaging*, vol. 22, no. 3, pp. 678–689, 2025, doi: 10.1177/0322083225711010.
- P. Garg, N. Kumar, P. Tiwari, and M. Saini, “Advances in Deep Learning Architectures for Diabetic Retinopathy Detection: A Review of Current Trends,” *Comput Biol Med*, vol. 143, p. 104232, 2025, doi: 10.1016/j.compbiomed.2025.104232.
- R. Bhat, R. Bhat, M. Sharma, and S. Srinivasan, “Integration of Machine Learning and Deep Learning Models for Early Detection of Diabetic Retinopathy in Fundus Images,” *J Digit Imaging*, vol. 38, no. 5, pp. 535–546, 2025, doi: 10.1007/s10278-025-00842-0.
- A. Mehrotra, A. Kumar, D. Sinha, and A. Verma, “Use of DenseNet for Automatic Diabetic Retinopathy Screening in Large-Scale Healthcare Systems,” *Applied Sciences*, vol. 15, no. 2, p. 251, 2025, doi: 10.3390/app15020251.
- M. Rajendran, M. Singh, A. Kannan, and K. Pankaj, “Comparison of CNN, ResNet, and DenseNet for Diabetic Retinopathy Classification: A Comprehensive Analysis,” *Neural Comput Appl*, vol. 37, no. 1, p. 203, 2025, doi: 10.1007/s00542-025-06312-4.