TPM Vol. 32, No. S5, 2025
ISSN: 1972-6325
https://www.tpmap.org/

DIABETIC RETINOPATHY CLASSIFICATION USING VARIOUS MACHINE LEARNING TECHNIQUES

PREETHI KOLLURU RAMANAIAH[1], NIKILA GS[2] ,VINOD HC[3], DEEPAK L[4]
[1]CLOUD ARCHITECT, EMAIL: preethiram4@gmail.com
[2]SOFTWARE ENGINEER, EMAIL: gs.nikila@gmail.com
[3]ASSOCIATE PROFESSOR, DAYANANDA SAGAR COLLEGE OF ENGINEERING, EMAIL: vinodhc-csd@dayanandasagar.edu
[4]ARTIFICIAL INTELLIGENCE AND DATA SCIENCE, R P SARATHY INSTITUTE OF TECHNOLOGY, SALEM-636305, EMAIL: deepakuniverse2004@gmail.com

Abstract
Diabetic retinopathy is the major visual impairment of persons with diabetes mellitus and early diagnosis is the key to effective management of the patient. The current work explores the state-of-the-art machine learning-based pipeline of diabetic retinopathy classification of fundus images. It is a model built upon VGG16, Inception V3 and DenseNet 201 to extract discriminative feature of retinal imaging. As clinical datasets are heterogeneous, this study uses the concept of transfer learning which allows a strong fine-tuning of a large image repository that contains the severity range of diabetic retinopathy. Model training is complemented by the data augmentation technique that serves as a two-in-one method of enhancing generalization and reducing overfitting. The evaluation of the performance is based on a complete set of measures, namely accuracy, precision, recall, and F1-score, which will be integrated to decide on the diagnostic performance of each model. The findings show that the proposed methodology provides better accuracy and thus it is appropriate to apply it in a clinical setting as a decision support tool to detect diabetic retinopathy and stratify it by risk to allow timely evidence-based interventions to be implemented before the loss of visual loss occurs.
Keywords: Diabetic retinopathy, VGG16, Inception V3, DenseNet 201, Machine learning

1. INTRODUCTION

The prevalence of diabetes has led to diabetic retinopathy (DR) becoming the main cause of vision deterioration between low- and middle-income countries according to Lin et al. [9]. The process of detecting diabetic retinopathy early depends on manual fundus image interpretation yet this method brings challenges of being time-consuming while being subjective according to Tufail et al. [16].

The present developments in deep learning (DL) provide a potential opportunity. It is important to remember that a weakly supervised heat-map mechanism, which was reported by Al Mukhtar et al. [1], showed strong results in diagnosing and localizing diabetic-retinopathy (DR) lesions. Moreover, a study by Dai et al. [5] known as DeepDR attained a better area under the receiver-operating-characteristic curve (AUC) when used in the detection of DR. Another study that used Inception V3 by Hakeem et al. [6] also provided better classification performance.

In this work, an integrative model that uses VGG16, Inception V3, and DenseNet 201 networks, which are pre-trained on the image net dataset, is proposed to automatically detect diabetic retinopathy. The data of annotated fundus images were subjected to ordinary transfer learning methodology, thus empowering the models to distinguish different classes of disease severity. Rotation, which was used alongside flipping and scaling, was used to augment data in order to increase generalization of the datasets.

The performance of the systems is usually assessed using precision and evaluation metrics in attempts to provide accurate and robust outcomes. The automated deep learning-based classification method used in this framework is aimed at reducing the workload of the diagnostic process and extending screening opportunities, which is likely to prevent interventions in the more progressive stages.

2. RELATED WORK

With the implementation of artificial intelligence (AI), the process of detecting diabetic retinopathy (DR) has advanced at a fast pace. The literature review conducted in the current paper thoroughly reviews the latest advances in the given area, comparing a range of AI models, deep learning systems, and predictive algorithms and outlining their potential to enable earlier detection and treatment. It is expected that such implements will positively affect patient care and eventually lead to better outcomes in terms of vision.

Some of the major studies include a weakly supervised heat map method that was reported by Al Mukhtar et al. [1] to achieve an accuracy of 98.65 percent on 1,200 fundus images. Alyoubi et al. [2] applied VGGNet using transfer-learning techniques and attained 96.6 percent accuracy in the detection of early DR. Da Rocha et al. [4] tested the performance of VGG16 in the detection of DR and Dai et al. [5] tested DeepDR with a cohort of 466,247 images, the AUC scores of which were between 0.901 and 0.972. Hakeem et al. [6] used Inception V3 and transfer learning to reach a 99.35 percent accuracy level in fundus analysis, and Jabbar et al. [8] reached a 96.6 percent accuracy using this method.

According to Hamza et al. [7], when CNN was used to classify retinal images in the diagnosis of diabetic retinopathy using a set of about 15,000 images, the overall accuracy was 97.2 %. Mostafa et al. [10] also analysed the performance of DenseNet201, ResNet50, VGG19 and MobileNetV2 on the APTOS 2019 dataset but with the addition of a similar investigation, finding that MobileNetV2 provided the best balance between accuracy (78.22 %) and computational efficiency. Saini and Susan [12] analyzed the algorithmic appropriateness of solving the problem of dataset imbalance in the case of pre-trained CNNs (VGG16, ResNet50, and DenseNet121). The results obtained showed that DenseNet121 was the best model in terms of classification, whereas EfficientDet D0 SSD (MobileNetV1) and PSPNet with focal loss were suggested in the case of object detection and segmentation, accordingly.

The mechanisms of diabetic retinopathy are explored comprehensively by Ansari et al. [3] in terms of their classification, epidemiology, risk factors, diagnostic and therapeutic difficulties and thus the multifactorial nature of Aetiology of DR as a medical condition. Lin et al. [9] also consider DR development and analyze biomarkers, imaging technology and interventions, and come to the conclusion that despite the higher prevalence in women, men experience worse outcomes. Tan and Wong [14] extrapolate into the year 2030, with the current developments in widefield imaging and OCT angiography, as well as with a forecast of the growth in the field of artificial intelligence applications. Tufail et al. [16] evaluate automated DR assessment systems, EyeArt and Retmarker, showing high sensitivity on referable DR and there is a possibility of reducing screening cost by using such technology on 20,258 patients.

The review of 40 studies of machine-learning in medical imaging by Rana and Bhushan [11] proved that deep-learning models tend to perform better than traditional methods of classification. Salehi et al. evaluated the means by which convolutional neural networks (CNNs) applied in transfer learning have improved diagnostic performance in medical imaging, but they also pointed out the need of large datasets and the fact that the models themselves lack interpretability [13]. Tiwari et al. introduced a six layered learnable CNN that could attain 99 percent accuracy in tumor type classification on MRI images [15].

Recent studies by Zhou et al. (2020) focused on the DenseNet architecture, which is a methodology that combines dense connections and sharing of information to optimize the assessment of medical images. At the same time, Vimal and Shirivastava (2020) developed a model that is based on the VGG 16 framework to simultaneously detect faces and face masks; this model demonstrated 93 % accuracy on 10,000 training images. It was also based on the findings of the study by Wang et al. (2020), to modify the Inception V3 model to recognize the style of the ancient architecture, where the dropout method was utilized to avoid overfitting, and the final success rate amounted to 98.64% on 5,513 instances. Finally, Yacouby and Axman (2020) proposed probabilistic variants of precision, recall and F1, which produced a substantial improvement in generalization on the SNLI dataset.

2.1 Research Gap Identified:

The review of the works available in the area of AI and their application diabetic DR synthesed the results of research that uses various model architectures and outlines ways in which deep learning methods may help in the earlier diagnosis and intervention. The information summarised in these attempts outlines independent systems that process videos, deep learning-based solutions designed to identify retinal lesions, and image processing routines with an AI algorithm verification programme. Predictive models of DR development are discussed in the analysis and the computer-aided diagnostic systems are also commented, but the limitations of the model applicability, lack of interpretability and nonconvergent assessment processes are also outlined.

3. Technologies

The study deals with the evaluation of the automated diagnostic device based on deep learning in the diagnosis of diabetic retinopathy. The method applied will be to optimize 3 popular convolutional neural network architectures, i.e., VGG16, Inception v3, and DenseNet-201 and use them to classify fundus images. Transfer learning helps to adapt the best model and data augmentation helps to enhance the robustness of the model. The accuracy measure, precision, recall and F1-score measures are the measures of performance evaluation. The comparative analyses that follow establish the most appropriate architecture that can be used in this task, therefore, leading to the proactive detection of diabetic retinopathy and early therapeutic intervention.

3.1 VGG16

VGG16 is a convolutional neural network model that is depicted in Figure 1; it has emerged as a common modality in the computer-vision literature where it is used in image classification (Al Mukhtar et al. [1]). It consists of 16 constituent layers of which 12 are convolutional and 3 fully-connected. The convolution with 3 x 3 filters and stride equal to 1 is used in each stage, and 2 x 2 max-pooling with stride equal to 2 is implemented after the last stage. This structure results in a small compact architecture of 138 million parameters and at the same time has effective feature extraction. VGG16 has become famous because of its good results in ImageNet dataset, and the portability of the model has led to its use in other, task-specific tasks. However, this significant number of parameters may also have computational restrictions when operating on resource-limited devices (Hakeem et al. [6]).

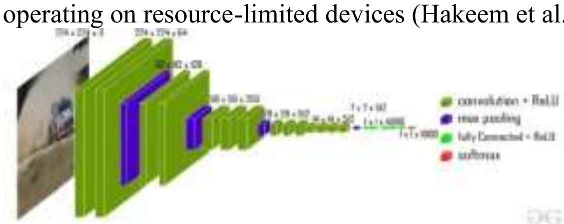


Fig. 1: Architecture of VGG16

3.2 Inception v3

Inception v3 The Inception v3 architecture was developed by Google with the aim of image classification and recognition (Hakeem et al. [6]). It is designed in such a way that it has many deep layers combined with Inception modules that maximize computational efficiency and enable extraction of complex visual features in images. The 23 million trainable parameters of the network allow high performance on benchmark tests, and also consume relatively few resources when participating in the ImageNet Large Scale Visual Recognition Competition. Thus, Inception v3 attains an excellent accuracy and low computation cost which makes it a famous model in visual computing.

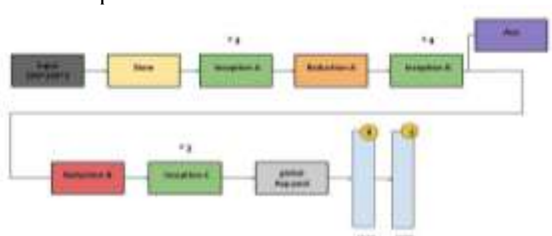


Fig. 2: Architecture of Inception v3

3.3 DenseNet 201

DenseNet-201 is an extension of DenseNet, by Zhou et al. [20], in which all the input layers are directly connected to the subsequent ones. This structure allows this 201-layered net to address the vanishing gradient issue by promoting gradient flow and can thereby lead to better performance and reduction of the Inception-related computational expense. Experimental studies show that DenseNet-201 shows a higher accuracy rate across different benchmarked datasets and its performance in terms of the computational cost justifies its use as a visual recognition model of choice by academic and commercial institutions alike.

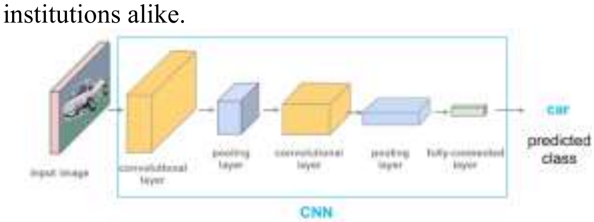


Fig. 3: Architecture of DenseNet -201

4. PROPOSED WORK

The Proposed system in this work describes a computational system based on the latest machine-learning techniques that would automate the diagnostic classification of diabetic retinopathy in fundus images. Three deep-learning structures are used in the system, i.e., VGG16, Inception V3, and DenseNet 201, in order to extract meaningful features in retinal scans. With the use of transfer learning, these pre-trained models are adapted to an annotated dataset which includes a different range of severity, and as a result, allows effective learning of the related visual patterns. During training, there is also generalization and avoidance of overfitting because of the data augmentation methods. The parameters of accuracy, precision, recall and F1-score critically analyse the performance of the system and the goal is to establish robustness and reliability. Finally, the system will be used to equip clinicians with a reliable diagnostic tool, which will help them diagnose DR early and stratify its risk, thus allowing them to intervene in time to avoid visual loss in diabetic patients.

Fig. 4 shows the system architecture that combines a large amount of fundus image datasets and uses transfer learning to conduct the systematic screening of the candidate models in a setting of diabetic retinopathy screening.

Image Dataset: The progression of a robust machine learning system is initiated through a properly curated richly labelled varied fundus image database. This type of data is essential to models that are to be used in reliable classification.

Data Analysis: The exploration of the data distribution, the detection of label imbalance and the isolation of the visual patterns. In order to reveal underlying structures, dimensionality-reduction methods are employed, the most common of them is principal component analysis.

Image Preprocessing: These stages include noise- and artifact-removal by means of resizing, cropping, and normalization. Out-of-order processing leads to inconsistency in analysis, and worse performance of the model.

Data Splitting: The obtained images are divided into training and testing databases. The first one can be used to develop the model, but the second one is only used to evaluate performance and generalizability.

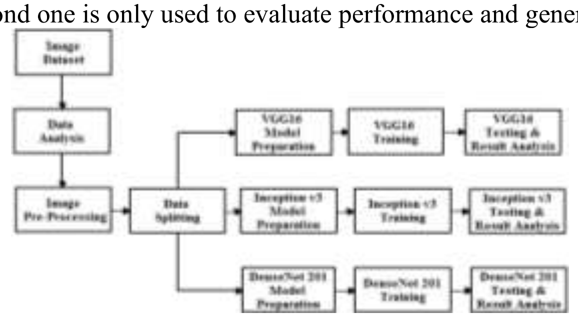


Fig. 4: System Architecture

Model Training: Optimization of three pre-trained models VGG16, Inception V3, and DenseNet 201, with the curated data. In this approach, the large feature set of ImageNet was used to increase the detection of DR.

Testing and Result Analysis: The models were tested and accuracy, precision, recall, and F1-score were calculated on a reserved set to determine the best model in DR screening after training.

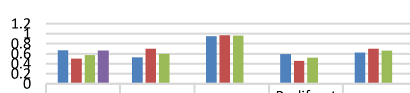
Dataset Description: The DR dataset was made of retinal images of size 224x224 which were classified into 5 categories, No_DR, Mild, Moderate, Severe, and Proliferate_DR, based on the APTOS 2019 Blindness Detection dataset and were engineered to acquire an accurate representation of the severity of DR.

The diagnostic system is will make the screening of diabetic retinopathy more efficient, thus ensuring that the patients at risk are identified early.

5. Experimental Results and Discussion

In this study, the classification of diabetic retinopathy is assessed with the help of evaluation metrics for comparison and analysis of VGG16, InceptionV3, and DenseNet 201 are carried out to choose the best model.

5.1 VGG16 Classification Report



	Mild	Moderate	No_DR	Proliferate_DR	Severe
VGG16 Precision	0.67	0.53	0.95	0.59	0.62
VGG16 Recall	0.5	0.7	0.97	0.46	0.7
VGG16 F1-Score	0.57	0.6	0.96	0.52	0.66
VGG16 Accuracy	0.67				

Fig. 6: Classification report for VGG16

Figure 6 demonstrates the performance of the five classes of the severity of DR. The model gives very good precision (95%) and recall (97%) on the class of No_DR, but gives only moderate F1-scores on the classes of 'Moderate (63%)

and Severe (60%)’. The classes of Mild (57%) and Proliferate_DR (52%) have much lower F1-scores. The macro and weighted averages are also close to the overall accuracy, which stands at 67%. These results show that non-DR cases were well identified, yet the need to enhance detection in defining the levels of DR severity was emphasized.

5.2 VGG16 Confusion Matrix

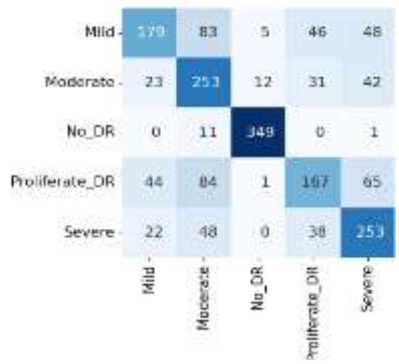


Fig. 7: Confusion Matrix for VGG16

The results in Fig. 7 are the confusion matrix of vgg16 that was used to conduct DR classification. The findings show that the model accurately detected the absence of DR (No DR) (349/361) and severe DR (253/361). The outcomes related to Moderate DR and Proliferate DR were much lower and were significant: 253/361 and 167/361, respectively. Moderate DR showed an even number of correct identifications (179/361) and incorrect identifications due to other classes (182), so it seems that the model tends to undergeneralize with regard to this category. Collectively, these observations emphasize the effectiveness of the model in distinguishing no evidence of DR and at the same time, the necessity to continue refining, specifically in Moderate DR and Proliferate DR.

5.3 VGG16 – Accuracy and Loss Graph

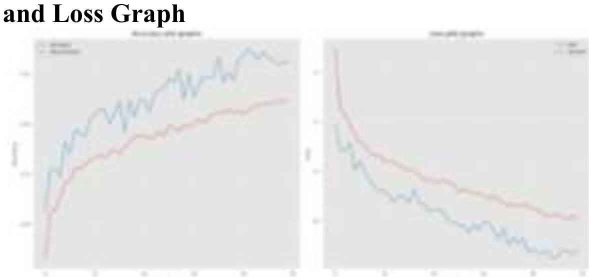
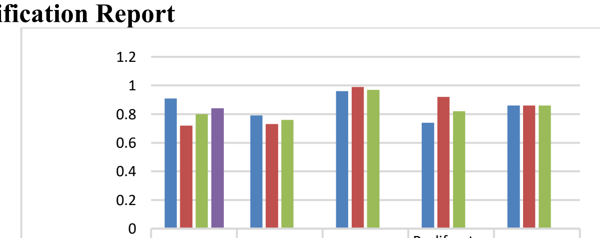


Fig 8: VGG16 Accuracy and loss graph

Fig 8 presents training and validation accuracy curves (blue and green, respectively) and training and validation loss curves (blue and green, respectively) over a few epochs. The accuracy of the training increases progressively by about 60 % on epoch 1 to nearly 95 % on epoch 50 and validation accuracy by about 55 % to 88 %. At the same time, both training and validation loss decrease: training loss is reduced by about 1.2-0.15, and validation loss is reduced by about 1.3-0.35. Though the gap between the training and validation curves increases as time goes by, which is a sign of possible overfitting, the consistent negative movement in both losses and positive movement in both accuracies ensures that learning is taking place.

5.4 InceptionV3 Classification Report



	Mild	Moderate	No_DR	Proliferate_DR	Severe
Precision	0.91	0.79	0.96	0.74	0.86
Recall	0.72	0.73	0.99	0.92	0.86
F1-Score	0.8	0.76	0.97	0.82	0.86
Accuracy	0.84				

Fig. 9: Classification report for InceptionV3

The classification report in Fig 9 illustrates strong performance. The model provides a precision of 96% and recalls of 99% in No DR and performs well in all the other levels with strong precision in Mild (91%), Moderate (79%), Proliferate DR (74%), and Severe (86 %). Also, there is 92 % recall of Proliferate DR. F1-scores have such trends with very good values of No DR (97 %), Mild (80 %), Proliferate DR (82%), and Severe (86 %). The overall accuracy (84%) is closely related to macro average (84%) and weighted average (84%) and this proves that there is no imbalance in its performance in all categories.

5.5 InceptionV3 Confusion Matrix

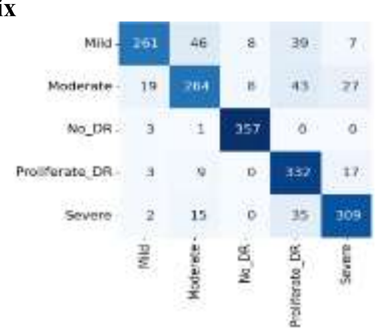


Fig. 10: Confusion Matrix for InceptionV3

The confusion matrix in fig. 10 shows that the No DR class is correctly classified 357 times (4 times misclassified), and Proliferate DR is also correctly classified 332 times (29 times misclassified). The severe DR is correctly classified in 309 cases (52 misclassifications), Moderate DR is correct in 264 cases (97 misclassifications) and Mild DR is correct in all 261 cases. In their turn, 100 other cases are classified as Mild DR erroneously.

5.6 InceptionV3– Accuracy and Loss Graph

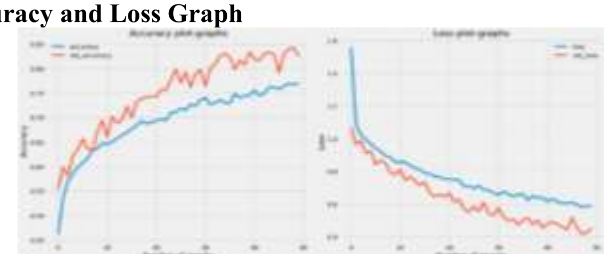


Fig 11: InceptionV3Accuracy and loss graph

The training curve of accuracy and loss can be seen in Fig 11. The training accuracy (red line) started at around 55 percent and steadily increased to around 85 percent after the epoch 40, but the validation accuracy (blue line) originally was at around 50 percent and increased to around 80 percent. Such an increasing gap suggests that there is some overfitting. In the case of loss, the training loss (blue line) decreases at a very high rate of about 1.6 to 0.4, but the validation loss (red line) decreases at a slow rate of about 1.5 to 0.8 and then the line shows a plateau after about 20 epochs. The overfitting as a reason is also supported by the validation loss that goes at a slower rate of decline and then levels.

5.7 DenseNet 201 Classification Report

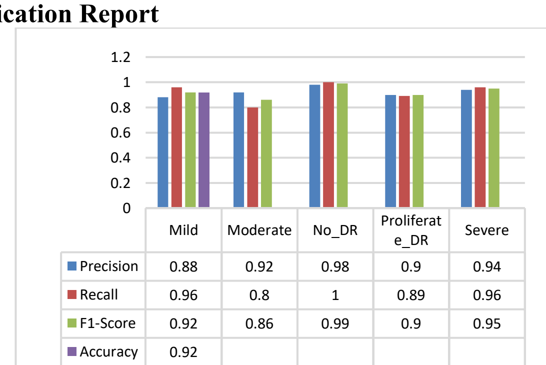


Fig. 12: Classification report for DenseNet 201

Fig.12 classification report for DR classification categorizes as follows. The resultant model is highly precise, between 88% and 98%, and its recall is between 80% and 100%. The F1-scores are similar, and range between 86% and 99%.

TPM Vol. 32, No. S5, 2025
ISSN: 1972-6325
https://www.tpmap.org/

The macro and weighted averages indicate that the accuracy is not biased in any of the levels of severity, as it is 92% and 92% respectively.

5.8 DenseNet 201Confusion Matrix

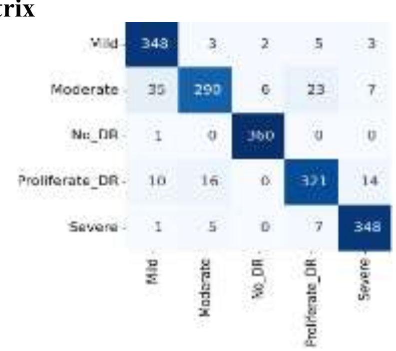


Fig 13: Confusion Matrix for DenseNet 201

The confusion matrix in Fig. 13 shows that the model is very accurate in terms of classifying the severity of the DR. In 360 out of 361 cases, no DR is properly identified, which is equal to 1 misclassification; in 348 out of 361 cases, Severe DR is properly identified, which is equal to 13 misidentifications. The accuracy of moderate DR falls into category 290 and the misclassifications into 71 and Proliferate DR falls into category 321 and the misclassifications into 40. The mild DR has a 100 percent accuracy and identifies 348 correctly and 13 incorrectly in total. However, such results call the need of complementary measures in order to have a full evaluation of performance.

5.9 DenseNet 201– Accuracy and Loss Graph

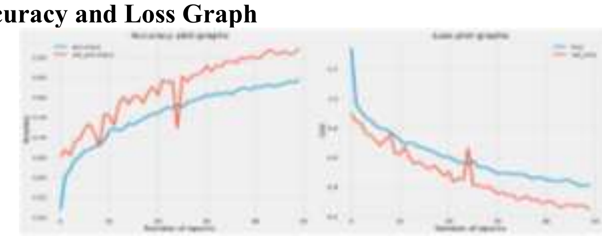


Fig. 14: DenseNet 201 Accuracy and loss graph

The learning dynamics shown in Fig. 14 present a number of patterns. First, the training accuracy (blue line) is seen to be in a steady increase with an initial value of about 60 % to about 95 % within 40 epochs and then converges. Second, validation accuracy (demonstrated by the red line) also grows in parallel, and it begins at approximately 55 % and converges at approximately 90 % after approximately 20 epochs. In the plot, the loss gradually decreases on both the training curve (about 1.2 to 0.3) and validation curve (about 1.1 to 0.4). The lack of a late divergence in such convergence is an indication of low probability of overfitting. The statistics therefore indicate that the model is not only learning well but it is also generalizing well.

5.10 Comparison

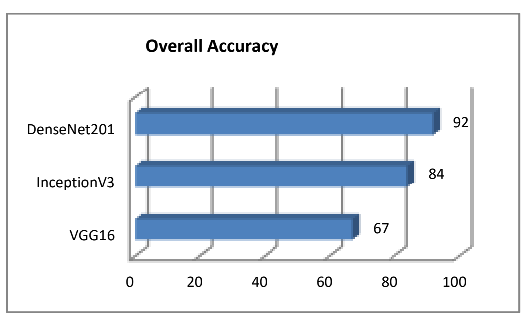


Fig 15: Accuracy Comparison Graph

In Fig. 15, a comparative evaluation of three machine-learning architectures of VGG16, InceptionV3, and DenseNet 201 as applied to classifying diabetic retinopathy (DR) is shown. The measure of evaluation is the overall accuracy that is the percentage of the correctly predicted images. Among the analyzed models, DenseNet 201 has the best accuracy of 92 %, which is better than InceptionV3 (84 %) and VGG16 (67 %). The advantage of DenseNet 201 can probably be explained by the fact that it performs full feature extraction and by the fact that it is less prone to overfitting compared to the other models. The outcomes reveal that DenseNet 201 has a great potential to become a powerful tool in the task of DR classification.

6. CONCLUSION & FUTURE WORK

This study tests the effectiveness of various machine learning paradigms, in this case, especially deep learning architectures, on the problem of automated diabetic retinopathy classification. The models VGG16, Inception V3, and DenseNet 201 were fine-tuned with a large amount of fundus images using transfer learning and data augmentation, and produced high amounts of accuracy and precision in the corresponding classification results. The developed model can be used as a decision support system by clinicians to identify diabetic retinopathy in a timely manner and stratify the risk. The study will facilitate patient care in addition to enhancing the screening service availability, particularly among underserved communities through the introduction of automation and advanced image-analysis techniques.
Future research may require an attempt to improve existing models, multi-modal data stream integration and performance evaluation in a wide range of clinical environments to increase applicability, scalability and overall effect on diabetic complications and blindness that are preventable.

REFERENCES

[1] Al-Mukhtar, M., Morad, A.H., Albadri, M. and Islam, M.S., 2021. Weakly Supervised Sensitive Heatmap framework to classify and localize diabetic retinopathy lesions. *Scientific Reports*, *11*(1), p.23631.

[2] Alyoubi, W.L., Abulkhair, M.F. and Shalash, W.M., 2021. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, *21*(11), p.3704.

[3] Ansari, P., Tabasumma, N., Snigdha, N.N., Siam, N.H., Panduru, R.V., Azam, S., Hannan, J.M.A. and Abdel-Wahab, Y.H., 2022. Diabetic retinopathy: an overview on mechanisms, pathophysiology and pharmacotherapy. *Diabetology*, *3*(1), pp.159-175.

[4] Da Rocha, D.A., Ferreira, F.M.F. and Peixoto, Z.M.A., 2022. Diabetic retinopathy classification using VGG16 neural network. *Research on Biomedical Engineering*, *38*(2), pp.761-772.

[5] Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., Liu, R., Wang, X., Hou, X., Liu, Y. and Long, X., 2021. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature communications*, *12*(1),

[6] Hakeem, A., Mehra, M. and Singh, R.P., 2022. Inception V3 and CNN Approach to Classify Diabetic Retinopathy Disease. *Int. J. Res. Appl. Sci. Eng. Technol.*, *10*(11), pp.660-671.

[7] Hamza, M., 2024. Optimizing early detection of diabetes through retinal imaging: A comparative analysis of deep learning and machine learning algorithms. *Journal of Computational Informatics & Business*, *1*(1).

[8] Jabbar, M.K., Yan, J., Xu, H., Ur Rehman, Z. and Jabbar, A., 2022. Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sciences*, *12*(5), p.535.

[9] Lin, K.Y., Hsih, W.H., Lin, Y.B., Wen, C.Y. and Chang, T.J., 2021. Update in the epidemiology, risk factors, screening, and treatment of diabetic retinopathy. *Journal of diabetes investigation*, *12*(8), pp.1322-1325.

[10] Mostafa, F., Khan, H., Farhana, F. and Miah, M.A.H., 2025. Application with Deep Learning Framework for Early Prediction of Diabetic Retinopathy. *AppliedMath*, *5*(1), p.11.

[11] Rana, M. and Bhushan, M., 2023. Machine learning and deep learning approach for medical image analysis: diagnosis to detection. *Multimedia Tools and Applications*, *82*(17), pp.26731-26769.

[12] Saini, M. and Susan, S., 2022. Diabetic retinopathy screening using deep learning for multi-class imbalanced datasets. *Computers in Biology and Medicine*, *149*, p.105989.

[13] Salehi, A.W., Khan, S., Gupta, G., Alabduallah, B.I., Almjally, A., Alsolai, H., Siddiqui, T. and Mellit, A., 2023. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, *15*(7), p.5930.

[14] Tan, T.E. and Wong, T.Y., 2023. Diabetic retinopathy: Looking forward to 2030. *Frontiers in Endocrinology*, *13*, p.1077669.

[15] Tiwari, P., Pant, B., Elarabawy, M.M., Abd-Elnaby, M., Mohd, N., Dhiman, G. and Sharma, S., 2022. Cnn based multiclass brain tumor detection using medical imaging. *Computational Intelligence and Neuroscience*, *2022*(1), p.1830010.

[16] Tufail, A., Rudisill, C., Egan, C., Kapetanakis, V.V., Salas-Vega, S., Owen, C.G., Lee, A., Louw, V., Anderson, J., Liew, G. and Bolter, L., 2017. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. Ophthalmology, 124(3), pp.343-351.

[17] Vimal, C. and Shirivastava, N., 2022. Face and face-mask detection system using vgg-16 architecture based on convolutional neural network. *International Journal of Computer Applications*, *183*(50), pp.16-21.

[18] Wang, X., Li, J., Tao, J., Wu, L., Mou, C., Bai, W., Zheng, X., Zhu, Z. and Deng, Z., 2022. A recognition method of ancient architectures based on the improved inception v3 model. *Symmetry*, *14*(12), p.2679.

[19] Yacouby, R. and Axman, D., 2020, November. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In *Proceedings of the first workshop on evaluation and comparison of NLP systems* (pp. 79-91).

[20] Zhou, T., Ye, X., Lu, H., Zheng, X., Qiu, S. and Liu, Y., 2022. Dense convolutional network and its application in medical image analysis. *BioMed Research International*, *2022*(1), p.2384830.