A Comparative Study On Multi Stage Fine Tuning Approaches for Diabetic Retinopathy Detection

***Abstract*—Diabetic retinopathy (DR) is closely linked to the rising global prevalence of diabetes. It remains the leading cause of blindness among the working-age population worldwide, posing significant challenges to both quality of life and socio-economic stability. Early detection of retinal abnormalities associated with DR is critical in preventing irreversible vision loss. However, current screening methods face notable limitations, including high costs, scalability issues, and dependence on expert ophthalmologists for accurate diagnosis. With ongoing technological advancements, the use of Deep learning to automate and enhance the accuracy of DR detection has gained increasing attention. In the proposed study we compare the fine tuning of the network using three pre-trained models modifying the Fully Connected layer introducing Dropout layers for regularization and another linear layer mapping to the two classes of DR and No DR. Additional techniques such as data augmentation, transfer learning, fine-tuning and Early stopping criteria are included. Thus, with this multi-stage fine tuning approach the models were able to extract relevant features suited to the current domain. The usage of Transfer learning allows us to leverage and fine tune pretrained models without the need of training a deep neural network from scratch reducing the need for high computational power. EyePACS dataset provided by Kaggle is used within this study containing 35,126 retinal images taken with different lighting conditions and spanning across 5 severity levels of No DR (0), Mild DR (1), Moderate DR (2), Severe DR (3) and Proliferative DR (4). Due to the high prevalence of No DR and high-class imbalance, classes were combined into No DR and DR. The DR class consisted of severity levels from 1-4 and hence the study was converted into Binary Classification. The comparative analysis of fine tuning was done on three deep learning architectures known as ResNet50, DenseNet121 and Inception V3 with DenseNet121 surpassing its peers yielding the highest accuracy of 89.20% and an Area under the curve (AUC) of 86.94%. This study highlights the potential of automated screening systems to facilitate early detection. Future efforts will aim to incorporate these models into clinical practice and expand their use to identify different severity levels of DR.**

***Keywords*—Bio-informatics, fundus images, CNN, Artificial Intelligence, and Diabetic Retinopathy**

I. Introduction

Diabetic Retinopathy is the leading cause of vision impairment worldwide, particularly affecting diabetic patients. According to the International Diabetes Federation, approximately 463 million adults were living with diabetes in 2019, and this number is expected to rise to 700 million by 2045. The global burden of DR has far-reaching implications. It is not only the leading cause of preventable blindness among working-age adults but also significantly impacts quality of life, economic productivity, and mental well-being, placing additional strain on individuals, families, and healthcare systems. Recent comprehensive evidence conducted by Zhen et al stated the global burden of diabetic retinopathy in a meta-analysis. This systematic review analyzed 59 population-based studies and found that among people with diabetes, the global prevalence of DR is 22.27%, with vision-threatening DR affecting 6.17% and clinically significant macular edema present in 4.07% of diabetic patients. The researchers estimated that in 2020, approximately 103.12 million adults worldwide had DR, with this number projected to increase dramatically to 160.50 million by 2045 due to the rising diabetes prevalence. Notable geographical variations exist, with Africa (35.90%) and North America/Caribbean (33.30%) showing the highest prevalence rates, while South and Central America had the lowest (13.37%). After adjusting for various factors, the study found that Hispanic and Middle Eastern populations with diabetes face significantly higher risks of developing DR compared to Asians. These findings underscore the disproportionate burden of DR across different regions and populations, highlighting the urgent need for targeted screening and intervention programs, particularly in high-risk areas such as the Middle East, North Africa, and the Western Pacific.

DR occurs as a result of prolonged hyperglycemia, a condition characterized by consistently high blood sugar levels which damages the blood vessels of the retina which is the lightsensitive tissue at the back of the eye responsible for vision. Over time, the increase in high blood sugar levels cause microvascular complications that weaken the walls of retinal blood vessels, leading to microaneurysms, hemorrhages, and fluid leakage into the retina. Figure 1 displays the global prevalence of DR categorized by geographic regions. Each region is annotated with its corresponding DR prevalence percentage and the estimated number of individuals affected. Color intensity represents the number of individuals with DR [1].

As the disease progresses, a reduction in blood flow occurs, stimulating the growth of abnormal blood vessels neovascularization, which can result in retinal detachment and permanent vision impairment. DR isclassified into five stages based on the severity of retinal damage and progresses through each stage with increasing severity, these levels include- No DR (0) constitutes to no visible damage to the retina. The first signs of damage appear in Mild DR (1), where microaneurysms, small bulges in the retinal blood vessels, are present. As the condition progresses to Moderate DR (2), hemorrhages and other microvascular changes are present indicating further retinal damage. In Severe DR (3), blood vessel blockages result in

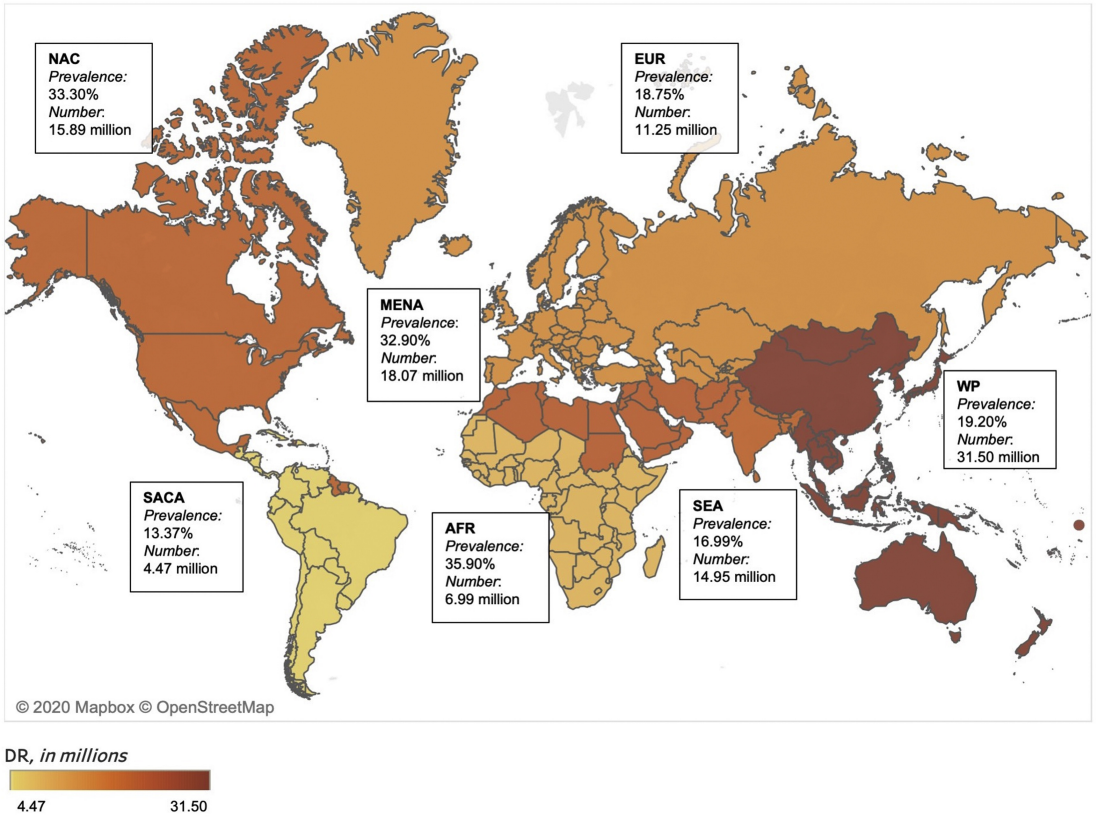

Fig. 1. Global prevalence of Diabetic Retinopathy (DR) by region, indicating prevalence percentages and affected population size. Darker regions represent a higher burden of DR. Adapted from *Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045*, Ophthalmology Journal, via Mapbox and OpenStreetMap.

widespread ischemia, which disrupts the retinal blood supply. The most advanced stage,Proliferative Diabetic Retinopathy (4), is characterized by the growth of new blood retinal vessels detachment (neovascularization), which can lead to a significant vitreous of hemorrhage vision or loss.Traditional methods for detecting diabetic retinopathy have served as a challenge to many people due to its expense and low availability of doctors, thus leveraging the state of art technologies allows for advancements in the field of healthcare and provides a solution for the people. Addressing this challenge is essential not only for reducing rates of preventable blindness but also for improving overall patient outcomes and reducing long-term healthcare burdens.

The existing approaches to DR screening are based on the manual analysis of the retinal fundus images by expert ophthalmologists. Such approaches also have several drawbacks: They are expensive; they are not easily extendable; they are dependent on trained medical specialists and so it cannot be used in widespread screening in resource poor settings. Additionally, identifying such subtle retinal features with different severity levels of DR remains difficult based with approach of manual diagnosis, particularly with the naked eye. Deep Convolutional Neural networks (DCNN) have been been the leading architecuture used in medical image analysis, DCNN's have also played a huge role in the detection of diabetic retinopathy [2]. DCNN is a multilayer neural network with deep learning capabilities, the architecture is designed in such a way that it progressively extract higher and complex features from the data at each layer to provide the desired output. Earlier layers within DCNN extract generic features whilst layer layers assemble these intermediate representations into highly specialized, task-specific features that directly contribute to final classification decisions. These deeper layers learn to recognize complex patterns unique to the specific training data with the main focus of DCNN being feature learning.

Recent studies have show the benefit of Deep learning techniques, specifically CNN's have been applied to the task of DR stage classification from retinal fundus images in recent years. Z. Khan recently proposed a deep learning architecture based on VGG-NIN for the purpose of DR detection [3]. Khan. achieved very good efficiency of the DR detection using the VGG-NIN model, their model has been designed to minimize training parameters number in order to improve the convergence speed and training time for classification tasks. The model was designed such that it has a small number of training parameters which have been made enough to learn the retinal features that are useful for multi-stage classification. With an area under the curve of 0.95, Khan's work established

a high benchmark for accuracy in the automated grading of diabetic retinopathy severity.

This study entails a comparative study comparing the usage of multi-stage fine tuning approaches on different Pretrained Models. Within this project we train several deep neural networks using transfer learning to detect the presence of diabetic retinopathy. The usage of Transfer learning is employed within this study, Transfer learning is a machine learning technique where knowledge gained on one task or domain is transferred to another task. Training a neural network from scratch requires substantial amount of data, high computational power, and long training runs. The usage of transfer learning within this study becomes logical as the parameters of the pretrained model have already been optimized such that their earlier layers have learnt how to detect these generic features. The pretrained models have been trained on the ImageNet dataset consisting of 14,197,122 natural images [4].

The remainder of this paper is organized as follows. Section II discusses the Literature review and existing approaches for diabetic retinopathy detection. Section III explains the preprocessing steps applied to the fundus images to prepare them for training. Section IV provides an overview of the methodology and the deep learning approach used in this study.Section V presents the results and discussion, including comparisons with state-of-the-art models. Finally, Section VI provides the conclusion and suggests directions for future work.

II. Related works

Traditional screening methods for DR rely heavily on manual interpretation of retinal fundus images by trained ophthalmologists, which is both time-intensive and resource-dependent. These limitations have motivated the adoption of automated solutions, leveraging the advancements in machine learning and deep learning techniques, to improve diagnostic efficiency and accessibility.tudies have demonstrated the value of deep learning in the classification and detection of DR stages. Gargeya and Leng [4] proposed a data-driven approach utilizing a CNN to analyze fundus images, achieving an Area Under the Curve of 0.97 on a dataset comprising over 75,000 retinal images. This study emphasized the importance of preprocessing techniques, such as brightness adjustment, and the application of heatmaps to enhance model interpretability. Datasets play a critical role in the development and validation of DR detection models. The APTOS 2019 dataset, which consists of 3,662 retinal fundus images labeled across five severity levels, has been widely used to train and evaluate CNN-based architectures. Its diversity and detailed annotations make it a valuable resource for deep learning applications. Additionally, the MESSIDOR dataset serves as a benchmark for external validation, with studies reporting F1-score scores exceeding expectation, affirming the dataset's importance in establishing the robustness of models.

Surpriya Mishra employed DenseNet121 architecture for DR classification using the APTOS 2019 dataset, achieving a Quadratic Weighted Kappa score of 0.8981. Their findings highlight DenseNet's superior performance over VGG16 in handling high-resolution images, underscoring the potential of more advanced architectures for accurate DR detection [5].

Ensemble methods are considered the state-of-the art solution for many Machine Learning challenges, they essentially combine multiple machine learning algorithms together to achieve a much better predictive performance as to opposed to using a single Machine Learning algorithm advanced the field. A study conducted by Ghosh et al presented an ensemble model combining VGG16 and Inception V3 networks [6]. By freezing certain layers and concatenating feature vectors from both models, the ensemble achieves a classification accuracy of 96.4% on the APTOS dataset, demonstrating the efficacy of combining multiple architectures.

Sikder et al [7] utilized a Randomized Tree classifier to develop an ensemble learning-based model for early DR detection, achieving a classification accuracy of 91% on the APTOS 2019 dataset.The approach used by Sikder demonstrated the importance of robust preprocessing and the ensemble's ability to mitigate the effects of noisy data. Whilst using such approaches like Ensemble methods may perform well, the computational complexity it provides means the training time is significantly increased.The advancements in machine learning, particularly deep learning, have significantly enhanced the capability of automated systems to detect and classify DR stages. The application of CNNs, ensemble learning, and preprocessing techniques has improved accuracy and robustness, while explainability tools have paved the way for clinical integration. Future work must focus on overcoming existing challenges, ensuring that these systems can be effectively deployed to improve global healthcare outcomes.

Several significant studies have explored different approaches to diabetic retinopathy classification. Research by [8] evaluated the impact of data augmentation strategies on model performance through two experimental configurations. Their investigation utilized DenseNet121 architecture trained on both APTOS and DDR datasets, with results demonstrating exceptional classification capabilities across all five DR severity stages. Their augmentation-enhanced approach achieved remarkable performance metrics on the APTOS dataset: 98.36% test accuracy with perfect top-2 and top-3 accuracy scores (100%). When applied to the more challenging DDR dataset, the model maintained strong performance with 79.67% test accuracy, 92.76% top-2 accuracy, and 98.94% top-3 accuracy. Additional evaluation metrics confirmed that higher-quality image inputs substantially improved model efficiency compared to both conventional techniques and non-augmented approaches.

III. Transfer Learning

Transfer learning is a machine learning technique where knowledge gained on one task or domain is transferred to another task. Training a neural network from scratch requires

substantial amount of data, high computational power, and long training runs. The usage of transfer learning within this study becomes logical as the parameters of the pre-trained model have already been optimized such that their earlier layers have learned how to detect these generic features. The pretrained models have been trained on the ImageNet dataset consisting of 14,197,122 natural images . To formalize our approach to transfer learning for diabetic retinopathy detection, we adopt the mathematical framework established by Pan and Yang [9], which has become the standard notation in the transfer learning literature.

A. Pretrained Models

CNN's pretrained on large scale datasets offer a compelling foundation for transfer learning applications in specialized domains. In this study, we leverage models pretrained on the ImageNet dataset, which contains over 14 million natural images across 1,000 object categories. The extensive feature representations learned from the diverse images provide a rich starting point for diabetic retinopathy detection. Our selection of specific architectures ResNet50 , InceptionV3 , and DenseNet121 was guided by both their proven performance in the general computer vision domain and their successful adaptation to medical imaging tasks in previous literature. These architectures represent distinct approaches to deep feature extraction, with varying network depths, connectivity patterns, and computational requirements.

The utilization of transfer learning in this context offers several advantages beyond mere computational efficiency. While the reduction in training time and hardware requirements is significant eliminating the need for weeks of training on high performance GPUs the primary benefit lies in addressing the limited availability of labeled medical imaging data. The features learned from ImageNet, though derived from natural images, contain generalizable representations of visual primitives such as edges, textures, shapes that can be effectively repurposed for detecting features in retinal fundus images.

Through systematic fine-tuning experiments, we investigate whether full fine-tuning or shallow tuning tuning yields optimal performance for diabetic retinopathy detection. This comparison allows us to understand whether the early convolutional layers of pretrained models that captured generic visual features require adaptation to the medical imaging domain, or if modification of the later, more deeper layers is sufficient. This analysis contributes to the broader understanding of cross domain feature transferability in medical image analysis.

B. Fine-tuning and shallow tuning

Full fine-tuning involves adjusting all network parameters of a pre-trained model during the training process. In this approach, weights from models pre-trained on natural images serve as initialization points, and every layer of the network is subsequently optimized using the target medical imaging dataset. This comprehensive adaptation is particularly valuable when the target domain (retinal fundus images) differs significantly from the source domain (natural images). While computationally intensive, full fine-tuning allows the entire network to adapt to the unique characteristics of medical images, potentially capturing domain-specific features at all levels of abstraction.Shallow fine-tuning, in contrast, operates on the principle that early convolutional layers in deep neural networks learn generic visual features (edges, textures, and basic shapes) that transfer well across domains, while later layers capture more domain-specific representations.

In our implementation, approximately 70% of each network's parameters remain frozen, with only the final 30% of layers and classification head undergoing optimization. Fig. 1 presents a schematic representation of the transfer learning approach employed in our study. The diagram illustrates how knowledge acquired from pre-trained CNN models is systematically transferred to our diabetic retinopathy detection task. This process involves adapting models initially trained on natural image datasets to the specialized domain of retinal fundus images. During transfer, we maintain the convolutional architecture while replacing the classification head, modifying the output dimension from the original 1000 ImageNet classes to our binary classification requirement (DR/No DR).

C. Fixed feature extraction

CNN'S comprise a hierarchical architecture where representations become increasingly abstract with network depth. The convolutional layers serve as feature extractors that learn progressively more complex patterns, while fully connected layers function as classifiers that map these high-dimensional features to output classes. This hierarchical structure creates a natural separation between feature extraction and classification components. The initial training iterations typically produce large gradient magnitudes due to the random initialization of the classification layers. By isolating training to only these layers, we prevent these large gradients from propagating through and potentially distorting the carefully pretrained convolutional weights in the early iterations.In our approach, we build upon these theoretical foundations by implementing a staged transfer learning methodology. We begin with a fixed feature extraction phase where convolutional layers pretrained on ImageNet remain frozen while only the custom classification head undergoes optimization. This initial phase serves multiple purposes beyond preventing gradient distortion. It establishes a performance baseline that quantifies how well ImageNet features transfer to retinal images without adaptation. Additionally, it allows the optimizer to adjust to the data distribution characteristics before introducing additional trainable parameters.The exploration of fixed feature extraction transitioning to selective fine-tuning represents a promising direction for balancing the competing demands of performance, computational efficiency, and generalization in medical imaging applications. Understanding precisely how much adaptation different network components require for optimal performance in diabetic retinopathy detection will provide valuable insights for developing efficient and accurate screening tools, potentially extending to other medical imaging domains.

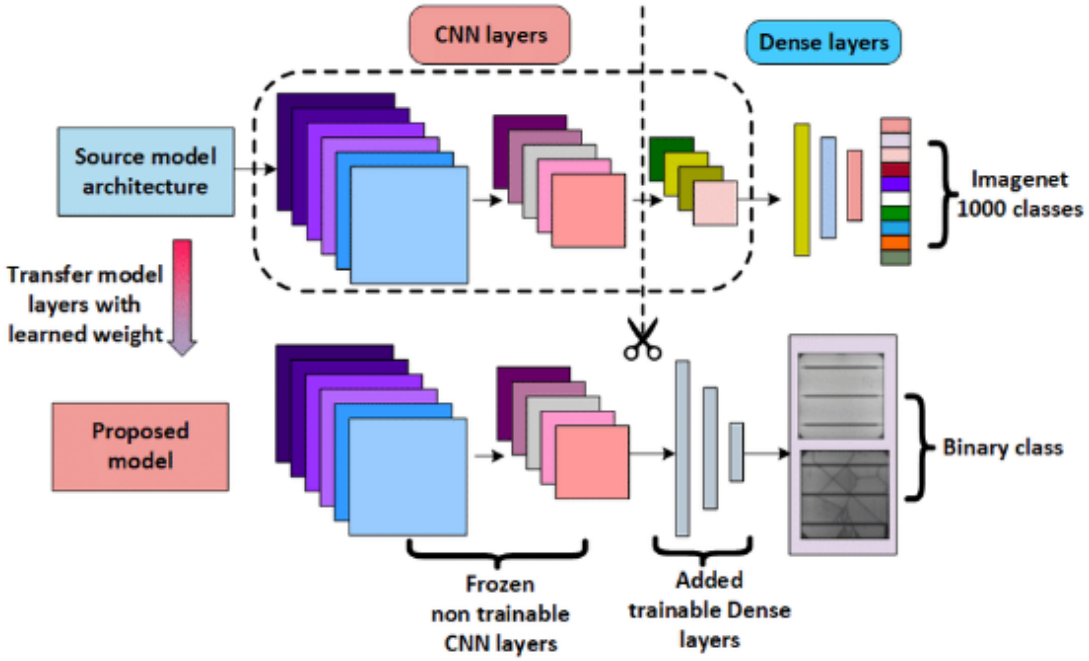

Fig. 2. Knowledge distribution between parameters in Transfer Learning [10]

IV. Dataset

The dataset used in this study is the EyePacs dataset, a widely recognized collection of retinal fundus images annotated for DR severity. The dataset includes 35,126 images categorized into five severity levels: No DR (0), Mild DR , Moderate DR (2), Severe DR (3), and Proliferative DR (4). The class distribution is highly imbalanced, with the prevalence of the No DR class being significantly higher compared to the other categories. Specifically, the distribution is as follows: No DR: 25810 images, Mild DR: 2443 images, Moderate DR: 5292 images, Severe DR: 873 images, and Proliferative DR: 708 images. This class imbalance poses a challenge as neural networks tend to overfit to the majority class, leading to biased predictions.The dataset is publicly available through Kaggle's data science competition [11].

Models trained on imbalanced data are likely to favor the majority class, reducing recall to minority classes, as seen in the study of Gulshan where they used sampling techniques to mitigate this [12]. This is particularly critical in DR detection, where underrepresenting severe cases (e.g., Proliferative DR) can lead to misdiagnosis and delayed treatment. To address these challenges, a series of preprocessing and data augmentation techniques were applied to balance the dataset and improve generalization, on top of the previous techniques we also converted the problem into binary classification where all images belonging to severity levels 1,2,3 and 4 were combined into a single class called DR(class 1), whilst class 0 was considered to be all healthy patients which is a severity level 0.

A. Data Preprocessing techniques

To prepare the dataset for training, several preprocessing steps were employed. All images were resized to a uniform dimension of 512x512 pixels, which is an input size suitable for many CNN architectures. Resizing ensures consistency across the dataset and compatibility with the model architecture while preserving important features of the retinal images. Pixel intensity values were standardized to a mean of 0 and unit variance of 1. Standardization helps to stabilize the training process and accelerates convergence by ensuring that the input values have similar magnitudes [13]. Table 1 displays the conversion of multi-classifaction to binary classification.

We used a scale radius normalization approach to standardize image sizes while maintaining proportions. Unlike simple resizing, which can distort important retinal structures, our scaling method identifies the approximate radius of the retinal fundus and uses this as a reference point for consistent scaling. This resizing ensures that retinal features maintain proportional relationships across all images, using a target scale of 500 pixels for the radius. we also removed non informative black borders from the images. Rather than applying arbitrary cropping dimensions, our technique analyzes pixel intensity distributions and preserves all regions containing potentially relevant information (defined as pixel values exceeding a tolerance threshold of 7).On top of this a high-boost filtering technique was used to improved the visibility of subtle pathological features. This approach combines the original image with its Gaussian-blurred negative to enhance fine details and edge information.

Data Augmentation on the fly was used which means the

TABLE I
TRANSFORMATION OF CLASS DISTRIBUTION FROM MULTI-CLASS TO BINARY CLASSIFICATION

Multi-class Distribution			Binary Distribution		
Class	**Images**	**%**	**Class**	**Images**	**%**
No DR	25,810	73.48%	No DR	25,810	73.48
Mild DR	2,443	6.95%	DR	9,316	26.52
Moderate DR	5,292	15.07			
Severe DR	873	2.49			
Proliferative DR	708	2.01			
Total	**35,126**	**100%**	**Total**	**35,126**	**100**

model sees different variations of the same image on each training iteration. This was used to improve the models generalizability as it helps the model learn invariant features rather than memorizing specific training patterns. The augmentation pipeline were applied exclusively to the training dataset and included random horizontal flipping (50% probability), random rotation (0-180 degrees), random vertical flipping (30% probability).

V. METHODOLOGY

Deep CNN's are a powerful class of neural networks designed to process spatially structured data, such as images. They function by learning features through successive layers, where lower layers capture simple patterns like edges and textures, and deeper layers identify more complex features like shapes and high-level structures.In this study we evaluated three pretrained models with different architectures. The three architectures considered are ResNet50, InceptionV3, and DenseNet121. All models were pretrained on the ImageNet dataset, providing a robust foundation of visual features that could be adapted to retinal fundus image analysis.

Initially, all three pre-trained models were loaded with weights obtained from training on the ImageNet dataset. During the first three epochs, we employed a fixed feature extraction approach, in which the convolutional base of each network was kept frozen to preserve pretrained representations. Only the modified classifier heads were trained at this stage, serving as an initial adaptation to the features present in the EyePACS dataset. The reasoning for freezing all layers except the modified head is avoid substantial distortion to pretrained weights from initially high gradients. Each pre-trained model architecture inherently included an average pooling layer, effectively reducing dimensionality and flattening convolutional outputs into one-dimensional vectors. Consistent modifications were applied across all models, with slight adjustments made according to each model's existing architecture. Specifically, ResNet50 and DenseNet121 architectures had their classifier heads modified by introducing a dropout layer with a probability of 0.5 for regularization, followed by a fully connected layer with 2048 neurons activated by a Rectified Linear Unit activation function (ReLU). This layer was succeeded by another dropout layer with a probability of 0.3 before the final classification layer, mapping features to two classes (DR and No DR). However, InceptionV3 already incorporates a dropout layer of probability 0.5 prior to the classification layer. Therefore, to prevent redundancy and potential over-regularization, the initial dropout layer of 0.5 was omitted in our modification of the InceptionV3 architecture.

Our training methodology followed a two-stage approach designed to optimize knowledge transfer. During the first three epochs, all convolutional layers remained frozen with only the custom classification head undergoing optimization. Beginning with epoch four, we unfroze all network parameters to enable comprehensive adaptation throughout each model architecture, allowing refinement of both low-level and high-level features for optimal performance on retinal images.

All models were trained for a maximum of 40 epochs using the Adam optimizer with an initial learning rate of 10^{-4}. This conservative learning rate facilitated incremental adaptation of pretrained weights rather than dramatic reconfiguration. To address the class imbalance in our dataset (73.48% No DR vs. 26.52% DR), we implemented a weighted cross-entropy loss function that applied class-specific penalties proportional to class distribution, ensuring appropriate emphasis on minority class detection. To optimize convergence, we implemented the ReduceLROnPlateau scheduling strategy with a reduction factor of 0.5 and patience value of 3. This approach systematically reduced the learning rate by half when validation loss failed to improve for three consecutive epochs, facilitating more refined weight adjustments as training progressed toward convergence. Figure 2 illustrates the learning rate evolution throughout training for InceptionV3.

Early stopping was implemented as a regularization technique to avoid overfitting. The training process was halted early if the validation loss did not improve for ten consecutive epochs, thereby identifying the optimal training point at which the model generalized best to unseen data. This approach prevented the models from learning redundant, non-generalizable features.

The dataset was divided into training, validation, and test subsets following a 70/15/15 split strategy. The training subset was utilized for parameter learning, while validation data monitored model performance during training. The test set, entirely held out during training, assessed the final performance metrics. An external clincal dataset provided by Biruni University Hospital was used as a test set on the best performing model. The dataset consisted of 111 diabetic retinopathy images spanning across 5 severity levels. These images underwent identical preprocessing steps as the EyePACS dataset to ensure consistency. This methodology enabled systematic evaluation of transfer learning effectiveness for diabetic retinopathy detection, with performance metrics including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) used to assess each model's diagnostic capability.

VI. EXPERIMENTAL RESULTS

We evaluated three pre-trained models—ResNet50, InceptionV3, and DenseNet121—and a custom CNN architecture

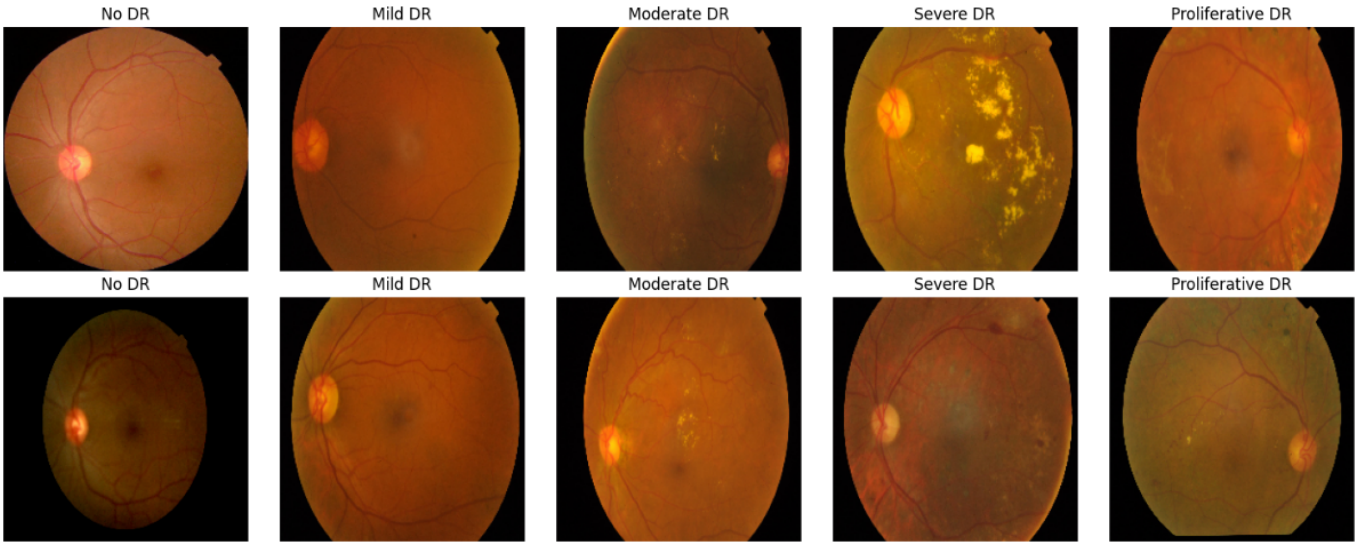

Fig. 3. Fundus Image of DR with severity gradings

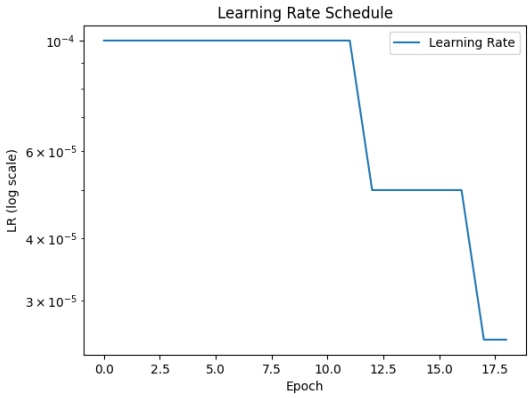

Fig. 4. Learning rate evolution throughout training for InceptionV3, showing the effects of the ReduceLROnPlateau scheduler with patience=3 and factor=0.5. The decreases in learning rate correspond to plateaus in validation loss.

using multiple performance metrics to comprehensively assess their diabetic retinopathy detection capabilities. All models were trained using the methodology described in Section **??**, with early stopping triggered at different epochs: epoch 18 for ResNet50, epoch 19 for InceptionV3, epoch 24 for DenseNet121, and epoch 15 for the custom CNN.

Table II presents the performance metrics for all four models evaluated on the EyePACS test dataset. DenseNet121 demonstrated the strongest overall performance with an accuracy of 89% and area under the ROC curve (AUC) of 0.8694, followed by ResNet50 (88% accuracy, 0.8591 AUC), InceptionV3 (88% accuracy, 0.8550 AUC), and the custom CNN (82% accuracy, 0.8123 AUC).

The recall values for DR detection ranged from 0.73 to 0.82, with DenseNet121 achieving the highest sensitivity. This metric is particularly important in a screening context, as it

TABLE II
PERFORMANCE COMPARISON OF MODELS ON EYEPACS TEST DATASET

Model	**Accuracy (%)**	**AUC**	**Precision (DR)**	**Recall (DR)**
ResNet50	88	0.8591	0.77	0.80
InceptionV3	88	0.8550	0.77	0.80
DenseNet121	89	0.8694	0.78	0.82

represents the model's ability to correctly identify patients with diabetic retinopathy. The superior recall of transfer learning approaches compared to the custom CNN demonstrates the value of leveraging pre-trained weights, even when the source and target domains differ substantially.

Figure 5 presents the training and validation loss curves for the three evaluated models: DenseNet-121, InceptionV3, and ResNet-50. Each plot illustrates how the model's performance evolved over training epochs. A consistent decrease in training loss is observed across all models, while the validation loss provides insight into generalization. DenseNet-121 shows smoother convergence with minimal overfitting, while InceptionV3 and ResNet-50 exhibit slightly more fluctuation in validation loss, indicating variability in learning stability. These visualizations help compare model robustness and convergence behavior.

VII. CONCLUSION

This study presented a comprehensive evaluation of transfer learning strategies using pretrained convolutional neural networks (CNNs) for the automated detection of diabetic retinopathy (DR). By applying both full-network fine-tuning and shallow-tuning techniques across DenseNet-121, InceptionV3, and ResNet-50, the study aimed to uncover performance trends under consistent experimental conditions. The findings revealed that DenseNet-121 achieved the most robust

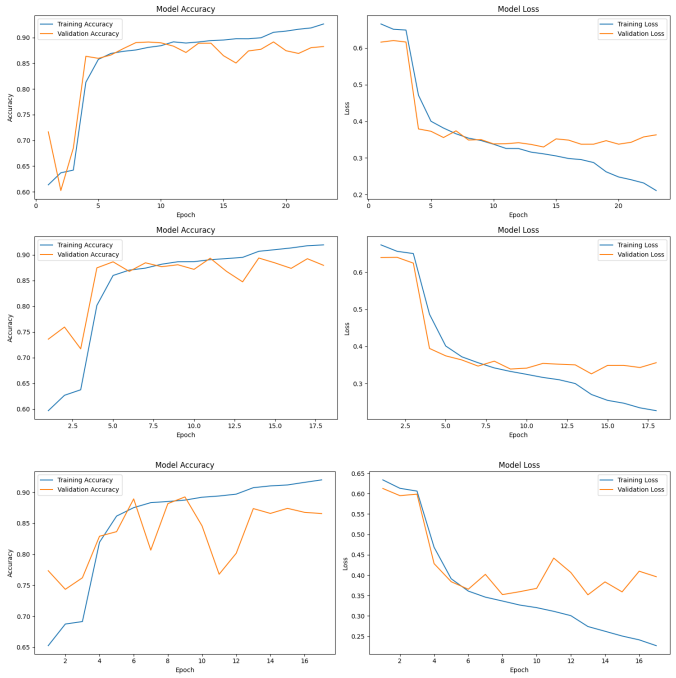

Fig. 5. Training and validation loss curves for (top) DenseNet-121, (middle) InceptionV3, and (bottom) ResNet-50.

performance, particularly in binary classification, demonstrating its capacity to balance recall and precision in a clinical context.

Furthermore, the research emphasized the importance of data preprocessing, class balancing, and careful selection of training strategies to mitigate overfitting and address class imbalance—challenges commonly found in medical imaging datasets. The results contribute to the growing body of evidence supporting the use of deep learning for DR screening and lay a foundation for future work in optimizing deployment-ready diagnostic systems, especially in under-resourced healthcare environments.

REFERENCES

[1] Z. L. Teo, Y.-C. Tham, M. Yu, M. L. Chee, T. H. Rim, N. Cheung, M. M. Bikbov, Y. X. Wang, Y. Tang, Y. Lu *et al.*, "Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis," *Ophthalmology*, vol. 128, no. 11, pp. 1580–1591, 2021.

[2] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, p. 100377, 2020.

[3] Z. Khan, F. G. Khan, A. Khan, Z. U. Rehman, S. Shah, S. Qummar, F. Ali, and S. Pack, "Diabetic retinopathy detection using vgg-nin a deep learning architecture," *IEEE Access*, vol. 9, pp. 61 408–61 416, 2021.

[4] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017.

[5] S. Mishra, S. Hanchate, and Z. Saquib, "Diabetic retinopathy detection using deep learning," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*. IEEE, 2020, pp. 515–520.

[6] S. Ghosh and A. Chatterjee, "Transfer-ensemble learning based deep convolutional neural networks for diabetic retinopathy classification," in *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)*. IEEE, 2023, pp. 489–493.

[7] N. Sikder, M. S. Chowdhury, A. S. M. Arif, and A.-A. Nahid, "Early blindness detection based on retinal images using ensemble learning," in *2019 22nd International conference on computer and information technology (ICCIT)*. IEEE, 2019, pp. 1–6.

[8] G. Alwakid, W. Gouda, M. Humayun, and N. Z. Jhanjhi, "Deep learning-enhanced diabetic retinopathy image classification," *Digital health*, vol. 9, p. 20552076231194942, 2023.

[9] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[10] M. R. Rahman, S. Tabassum, E. Haque, M. M. Nishat, F. Faisal, and E. Hossain, "Cnn-based deep learning approach for micro-crack detection of solar panels," in *2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE, 2021, pp. 1–6.

[11] Kaggle, "Diabetic retinopathy detection," Kaggle Competition, July 2015, accessed: May 09, 2020. [Online]. Available: http://www.kaggle.com/c/diabetic-retinopathy-detection

[12] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[13] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.