Diabetic Retinopathy Detection Using Deep Learning Models

Kulsoom Farhan*, Muneeba Ahmed*, Syed Sajjad Ali*
*Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan

***Abstract*—Abstract: This project explores the potential of artificial intelligence (AI) to address the challenge of eye disease diagnosis in low-resource environments, particularly in developing countries like Pakistan. Early detection of eye diseases is critical,especially for individuals at high risk due to family history, age (over 60), diabetes, or prior eye injuries/surgeries. While timely intervention can prevent permanent vision loss, many eye diseases lack early symptoms. Regular eye examinations are crucial, yet access to such services can be limited in resource-constrained settings.**

This project aims to develop a deep learning model for automated eye disease detection using retinal fundus images. We will evaluate the performance of a range of pre-trained architectures, including established CNNs like VGG16, ResNet50, DenseNet (121 & 169), and XceptionNet, alongside more recent advancements like InceptionV3, InceptionResNetV2, Swin Transformer, MobileNet, EfficientNet, and ViT models. This model will be designed to accurately classify various eye diseases, potentially improving diagnostic accuracy and efficiency compared to traditional methods. By eliminating geographical and economic barriers to essential eye health services, the project aspires to promote a healthier future for underserved populations.

***Index Terms*—Artificial Intelligence (AI) , Eye Disease Detection , Low-Resource Settings ,Retinal Fundus Images , Deep Learning , Convolutional Neural Networks (CNNs) , Deep Neural Networks (DNNs) , Transformers , VGG16 , InceptionV3 , InceptionResNetV2 ,ResNet50 , Swin Transformer , DenseNet-169 , XceptionNet , DenseNet-121 , MobileNet ,EfficientNet , ViT Model , Diabetic Retinopathy (DR)**

I. INTRODUCTION

BACKGROUND

Diabetic retinopathy (DR) is a microvascular complication of diabetes that affects the retina, the light-sensitive layer at the back of the eye. It is a leading cause of vision loss in working-age adults, particularly among diabetic individuals [8].Early detection and intervention are crucial for preventing vision impairment and blindness associated with DR.Traditionally, eye examinations by trained ophthalmologists are the gold standard for DR diagnosis. However, these examinations can be resource-intensive and limited by accessibility, especially in remote areas or for patients with limited mobility.

MOTIVATION

Developments in deep learning and artificial intelligence (AI) offer promising avenues for automated DR diagnosis using retinal images. This technology has the potential to overcome limitations in traditional screening methods by providing a cost-effective, scalable, and potentially more accessible solution for early DR detection.

PROBLEM STATEMENT

The accurate diagnosis of DR from retinal images requires robust models that can effectively differentiate between healthy and diseased states, as well as classify the severity of the disease across various stages. While several deep learning models have shown promise for DR diagnosis, a systematic evaluation is needed to identify the most suitable architecture for this specific medical image classification task.

OBJECTIVES

This study aims to:

1) Investigate the performance of various pre-trained deep learning models for diagnosing DR using a labelled retinal image dataset.
2) Evaluate the models based on their accuracy, precision, recall, F1-score, and confusion matrix analysis.
3) Identify the model that achieves the most effective and robust performance for DR classification.

By achieving these objectives, this study can contribute valuable insights into the suitability of different deep learning architectures for automated DR diagnosis. This knowledge can pave the way for the development of more reliable and efficient AI-powered tools to support early detection and management of DR, ultimately improving patient outcomes and reducing the burden of vision loss associated with diabetes.

II. LITERATURE REVIEW

F M Javed et al[1] in their research employed an image noise removal technique on the dataset using a median filter. This technique was successful in producing noise-free dataset by removing outliers and atypical intensity values caused by noise. The authors then applied gamma correction for image enhancement and claimed it leads to the increase in the result accuracy. Undersampling and oversampling augmentation techniques were used to balance the and increase the dataset. The authors then compared their proposed model DRNet13 with different deep learning models such as ZFNet, GoogLeNet, AlexNet, InceptionV3, InceptionResV2, SqueezeNet, ResNet50, ResNet101, VGG16, VGG19, ShuffleNet, Xception, DenseNet201, DarkNet19 and DarkNet53. Their model showed 97Abini M.A et al[2] employed pre-trained Convolutional Neural Network (CNN) VGG-16 and MobileNet-V2 to classify the five stages of diabetic retinopathy. Their model achieved 90Rajasekhar et al [3] used Contrast Limited Adaptive Histogram Equalization technique (CLAHE) and adaptive mean filtering for data augmentation. These techniques were successful in reducing noise amplification. The authors employed a

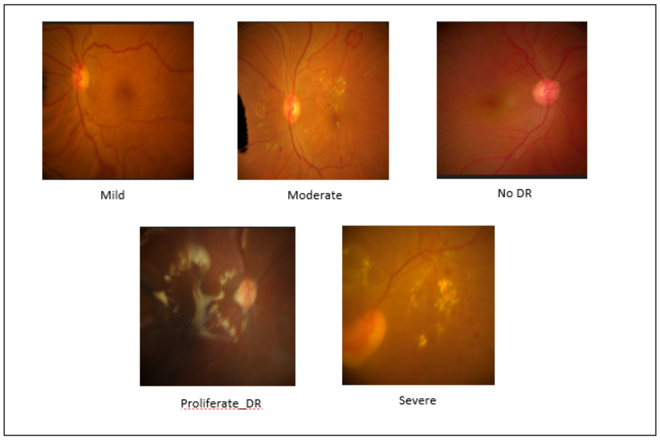

Fig. 1. Image Dataset

vision transformer called a Swin Transformer for the multi-class classification problem and compared the results with DenseNet169 , ResNet50 , XceptionNet ,InceptionV3, VGG16 and MobileNet. Mamoona et al[4] claimed that a model will learn better and show high efficiency when trained with high quality images. They used APTOS and DDR dataset to train DenseNet-121 model. The authors used two filtering algorithms, the CLAHE and the enhanced super resolution generative adversarial network (ESRGAN) to produce high quality datasets for APTOS and DDR. Various augmentation techniques were to deal with the imbalance datasets. The model achieved a 79.67Ritesh et al [5] employed CNN model and AlexNet model on APTOS dataset using Up-sampling technique to deal with imbalanced dataset. Kazi et al[6] employed transfer learning models such as DenseNet121, Resnet50, VGG16, VGG19, and InceptionV3, Xception, and machine learning models such as SVM, and neural network models like (RNN) for binary and multi-class classification problem. The dataset used is APTOS 2019 Blindness Detection dataset. Among these Xception model showed the highest accuracy. Abdul[7] proposed a light weight hyperparameter-optimized MobileNet V3 model for the classification task using the APTOS and EyePacs datasets. This model required fewer parameters to train, had fewer FLOPs , and less training time. The auther proposed that this model can be implemented as a mobile application and treat patients in remote areas. However, improvement is needed in the model's efficiency to detect DR from low-quality fundus images.

III. DATASET

A. *The Sight Insight Diabetic Retinopathy Dataset*

This research utilizes a labelled image dataset curated by Sight Insight to train a model for diagnosing diabetic retinopathy (DR). The dataset offers several advantages for this purpose.

*1) **Data Description:**:* The dataset consists of retinal images captured using a specially designed lens system. Each image is assigned a label corresponding to one of five DR severity classes:

- No Diabetic Retinopathy (No_DR)
- Mild DR
- Moderate DR
- Severe DR
- Proliferative Diabetic Retinopathy (Proliferate_DR)

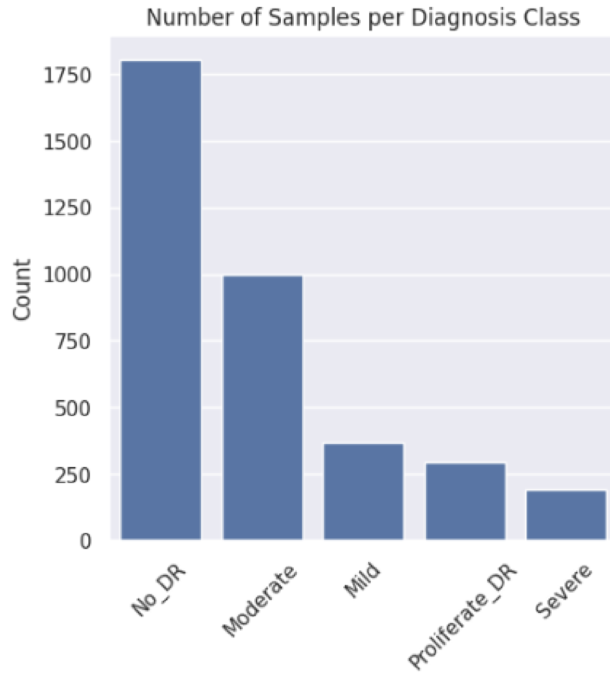

Fig. 2. Count of Diagnosis

S. No	Diagnosis	Count
1	No_DR	1805
2	Moderate	999
3	Mild	370
4	Proliferate_DR	295
5	Severe	193

TABLE I
COUNT OF DIAGNOSIS

This categorization scheme encompasses the entire spectrum of DR progression, allowing the model to be trained for multi-class classification across all stages of the disease.

*2) **Justification for Dataset Choice:**:* The Sight Insight dataset offers several key advantages for training a DR diagnosis model:

- **Comprehensiveness:** The inclusion of all five DR severity classes ensures the model can learn to identify each stage effectively.
- **Label Accuracy:** The pre-existing labels, presumably assigned by medical experts, provide a reliable ground truth for model training, fostering accurate classification in unseen images.
- **Clinical Relevance:** The five categories directly map to clinically relevant stages of DR, each requiring specific treatment approaches. By accurately classifying

these stages, the model has the potential to aid in recommending appropriate patient care.

*3) **Future Considerations:**:* Further investigation into the following aspects of the dataset would be beneficial:

- **Image Quality:** Assessing the image resolution, focus, and overall quality is crucial, as these factors directly impact the model's ability to extract meaningful features for classification.
- **Data Size:** The number of images within the dataset is an important factor. A larger dataset provides a richer pool of information for the model to learn from, potentially leading to improved generalizability and robustness.
- **Data Bias:** It is important to evaluate whether the dataset represents a diverse population in terms of factors such as ethnicity, age, and gender. Biases in the data can lead to the model performing poorly on specific demographics.

By carefully considering these additional aspects, we can ensure the Sight Insight dataset is optimally utilized for training a robust and accurate DR diagnosis model.

IV. METHODOLOGY

This study investigates the performance of various deep learning models for diagnosing diabetic retinopathy (DR) using a labelled retinal image dataset. The methodology can be segregated into three key stages: data preprocessing, model training, and performance evaluation.

*A. **Data Preprocessing:***

1) **Data Acquisition:** The retinal image dataset curated by Sight Insight will be employed for this study.
2) **Preprocessing:** The images will undergo preprocessing steps to improve model generalizability and training efficiency. This may include:
    - **Resizing:** All images will be resized to a standard resolution to ensure uniformity as input for the models.
    - **Normalization:** Pixel intensities will likely be normalized to a specific range (e.g., 0-1 or mean-standard deviation normalization) to facilitate training convergence.
    - **Data Augmentation (Optional):** Techniques like random cropping, flipping, rotation, and color jittering can be implemented to artificially expand the dataset and improve model robustness to variations in image appearance.

*B. **Model Training:***

1) **Model Selection:** This study will evaluate the performance of several pre-trained convolutional neural network (CNN) architectures, including:
    - **VGG16:** A classic CNN architecture with multiple convolutional layers arranged in a sequential manner. VGG16 is known for its depth and ability to learn complex image features .
    - **InceptionV3:** An architecture from Google that utilizes inception modules, which combine convolutional and pooling layers within a single unit, promoting efficient feature extraction .
    - **DenseNet:** A CNN architecture that connects each layer to all subsequent layers, fostering feature reuse and potentially alleviating the vanishing gradient problem.
    - **InceptionResNetV2:** A hybrid model that combines elements of Inception modules and residual connections (introduced in ResNets) for improved gradient flow and performance .
    - **ResNet50:** A deep CNN architecture that incorporates residual connections, allowing the network to learn from the cumulative knowledge of all preceding layers, addressing the degradation problem in deeper networks .
    - **Swin Transformer:** A relatively new architecture that utilizes a transformer-based approach, achieving state-of-the-art results on various image classification tasks. Swin Transformer breaks down an image into patches, processes them with transformers, and reconstructs the final output .
    - **DenseNet-169:** A denser version of the DenseNet architecture with 169 layers, potentially offering increased capacity for feature learning compared to DenseNet-121.
    - **XceptionNet:** A CNN architecture from Google that utilizes depthwise separable convolutions for efficient processing and achieves strong performance on image classification tasks.
    - **DenseNet-121:** A variant of DenseNet with 121 layers, offering a balance between complexity and performance.
    - **MobileNet:** A lightweight CNN architecture designed for mobile and embedded devices, achieving good accuracy with reduced computational cost.
    - **EfficientNet:** A family of CNN architectures that scales model complexity based on a compound scaling method, offering a good balance between accuracy and efficiency.
    - **ViT Model:** Similar to Swin Transformer, a vision transformer (ViT) model utilizes a transformer-based approach for image classification. ViTs break down an image into patches, process them through a transformer encoder, and use a classification head for prediction .
2) **Transfer Learning:** Since the pre-trained models were trained on extensive image datasets, their weights will be transferred as a starting point for fine-tuning on the DR classification task. This leverages the models' learned features while adapting them to the specific problem of DR diagnosis.
3) **Training Pipeline:** A standardized training pipeline will be established for all models. This pipeline will encompass:

- Data augmentation (if applicable)
- Data normalization
- Splitting the dataset into training and validation sets: A common approach is to use an 80/20 split, where 80% of the data is allocated for training and 20% for validation. The validation set is used to monitor model performance during training and prevent overfitting.
- Training with an optimizer (e.g., Adam) and a loss function (e.g., categorical cross-entropy) suitable for multi-class classification.

V. RESULT

The table II presents a performance evaluation of 11 deep learning models on a diabetic retinopathy (DR) classification task. The models are VGG 16, CNN, Vit, EfficientNET, InceptionV3, Mobilenet, DenseNet-121, XceptionNet, DenseNet-169, Swin Transformer, and ResNet50. The classification task involves identifying five different classes of DR: Mild, Moderate, No DR (no diabetic retinopathy), Proliferate DR, and Severe.

The table II shows the precision, recall, and F1 score of each model for each class. Precision measures the proportion of true positives among all positive predictions made by the model, while recall measures the proportion of true positives among all actual positive instances in the dataset. The F1 score is a measure of the model's accuracy on a specific class, and it's the harmonic mean of precision and recall.

Upon examining the table II, it becomes clear that some models perform well on certain classes but poorly on others. For instance, the VGG 16 model has a high precision and recall on the No DR class but performs poorly on the Proliferate DR class. On the other hand, some models are more consistent across classes, such as the DenseNet-121 model, which has a relatively high precision and recall across all classes. Additionally, some models are better suited for certain classes, like the InceptionV3 model, which has a high precision and recall on the Moderate class.

Overall, the table II provides a comprehensive evaluation of the performance of different deep learning models on a diabetic retinopathy classification task. By analyzing the table, researchers and practitioners can identify the strengths and weaknesses of each model and choose the best model for a specific task. This can ultimately lead to the development of more accurate and effective DR classification systems, which can improve patient outcomes and healthcare services.

TableIII shows the accuracy of 12 different models after training for 5 epochs. The models are listed in the first column, and their corresponding accuracy is listed in the fourth column. The second and third columns are not relevant to the accuracy of the models. The table shows that the DenseNet-169 and Swin Transformer models have the highest accuracy at 74%, while the ResNet50 model has the lowest accuracy at 10%. The other models have varying degrees of accuracy, with the MobileNet, DenseNet-121, and InceptionResV2 models achieving accuracy above 70%. The VGG 16, Vision

Models	**Classes**	**Precision**	**Recall**	**f1-score**
VGG 16	Mild	0	0	0
	Moderate	0	0	0
	No DR	0.49	1	0.66
	Proliferate DR	0	0	0
	Severe	0	0	0
CNN	Mild	0.4	0.31	0.35
	Moderate	0.53	0.72	0.61
	No DR	0.89	0.93	0.91
	Proliferate DR	0.17	0.03	0.06
	Severe	0.18	0.05	0.08
Vit	Mild	0	0	0
	Moderate	1	0.01	0.01
	No DR	0.52	0.96	0.67
	Proliferate DR	0.18	0.19	0.18
	Severe	0	0	0
EfficientNET	Mild	0	0	0
	Moderate	0	0	0
	No DR	0	0	0
	Proliferate DR	0	0	0
	Severe	0.05	1	0.1
InceptionV3	Mild	0.2	0.36	0.26
	Moderate	0.46	0.64	0.53
	No DR	0.96	0.84	0.9
	Proliferate DR	0	0	0
	Severe	0	0	0
Mobilenet	Mild	0.82	0.12	0.21
	Moderate	0.56	0.88	0.68
	No DR	0.96	0.94	0.95
	Proliferate DR	0.22	0.03	0.06
	Severe	0.22	0.24	0.23
DenseNet-121	Mild	0.53	0.26	0.35
	Moderate	0.55	0.83	0.66
	No DR	0.88	0.96	0.92
	Proliferate DR	0	0	0
	Severe	0	0	0
XceptionNet	Mild	0.45	0.43	0.44
	Moderate	0.57	0.88	0.69
	No DR	0.96	0.87	0.91
	Proliferate DR	0.21	0.05	0.08
	Severe	0.33	0.13	0.19
DenseNet-169	Mild	0.48	0.19	0.27
	Moderate	0.53	0.92	0.68
	No DR	0.97	0.94	0.95
	Proliferate DR	0.56	0.08	0.15
	Severe	0	0	0
Swin Transformer	Mild	0.67	0.08	0.14
	Moderate	0.55	0.88	0.68
	No DR	0.94	0.95	0.95
	Proliferate DR	0.52	0.25	0.34
	Severe	0.5	0.11	0.17
ResNet50	Mild	0.1	1	0.18
	Moderate	0	0	0
	No DR	0	0	0
	Proliferate DR	0	0	0
	Severe	0	0	0
InceptionResV2	Mild	0.61	0.19	0.29
	Moderate	0.58	0.73	0.65
	No DR	0.81	0.99	0.89
	Proliferate DR	0	0	0
	Severe	0.3	0.08	0.12

TABLE II
PERFORMANCE OF DIFFERENT MODELS ON DIFFERENT CLASSES

S. No	Models	Epochs	Accuracy
1	VGG 16	5	49%
2	CNN	5	69%
3	Vision Transformer	5	49%
4	EfficientNet	5	50%
5	InceptionV3	5	63%
6	MobileNet	5	73%
7	DenseNet-121	5	73%
8	XceptionNet	5	72%
9	DenseNet-169	5	74%
10	Swin Transformer	5	74%
11	ResNet50	5	10%
12	InceptionResV2	5	71%

TABLE III
PERFORMANCE OF DIFFERENT MODELS

Transformer, and EfficientNet models have accuracy below 50%.

VI. CONCLUSION

The Sight Insight Diabetic Retinopathy Dataset has proven to be valuable for training and evaluating deep learning models for diabetic retinopathy (DR) classification. Our analysis shows that different models excel in different areas of DR detection. For instance, DenseNet-121 demonstrated consistent high performance across all classes, while VGG 16 showed high precision and recall for the No DR class but struggled with the Proliferate DR class. InceptionV3 was particularly effective in identifying the Moderate class. Overall, the dataset's diversity and detailed labeling have allowed for the development of models that can accurately classify various stages of DR. This capability is crucial for improving diagnostic accuracy and, ultimately, patient care in the field of ophthalmology. Additionally, our analysis indicates a need to increase the number of training epochs. Extending the training period could further improve the models' performance and enhance their ability to accurately classify all DR stages.

VII. FUTURE RESEARCH DIRECTION

This study investigates the potential of various deep learning models for diagnosing diabetic retinopathy (DR) using a retinal image dataset. While the findings will provide valuable insights, several avenues exist for future research to further advance this field:

1. Incorporation of Clinical Data:

Integrate clinical data, such as a patient's medical history, blood sugar levels, and other relevant factors, alongside retinal images. This combined approach could potentially improve model performance and enable risk prediction beyond what retinal images alone can reveal.

2. Explainable AI (XAI):

Implement XAI techniques to understand the rationale behind the model's decisions. This would not only enhance trust in the model's predictions but also provide valuable insights into the image features most relevant for DR diagnosis.

3. Generalizability and External Validation:

Evaluate the performance of the best performing models on additional, externally sourced datasets. This external validation is crucial to ensure the model's generalizability to real-world clinical scenarios beyond the initial training data.

4. Multi-modal Learning:

Explore the potential of multi-modal learning by incorporating other imaging modalities, such as optical coherence tomography (OCT), alongside retinal fundus images. Combining information from different modalities could potentially lead to more robust and informative diagnoses.

5. Addressing Data Bias:

Investigate potential biases within the dataset, such as underrepresentation of certain demographics or disease severities. Develop strategies to mitigate these biases and ensure the model performs fairly across all patient populations.

6. Uncertainty Quantification:

Implement techniques for uncertainty quantification within the model. This would allow the model to not only provide a classification but also an associated confidence score, indicating how certain it is about its prediction.This information can be valuable for clinicians in determining the need for further investigations or specialist referrals.

7. Real-world Deployment:

Explore strategies for deploying the best performing model in real-world clinical settings. This could involve integration with existing healthcare infrastructure and development of user-friendly interfaces for clinicians to interact with the model.

VIII. REFERENCES

[1] Shamrat, F M & Shakil, Rashiduzzaman & Sharmin, Sharmin & ovy, Nazmul & Akter, Bonna & Ahmed, Md & Ahmed, Kawsar & Bui, Francis & Moni, Mohammad Ali. (2024). An advanced deep neural network for fundus image analysis and enhancing diabetic retinopathy detection. Healthcare Analytics. 5. 10.1016/j.health.2024.100303.

[2] S. Kajan, J. Goga, K. Lacko and J. Pavlovičová, "Detection of Diabetic Retinopathy Using Pretrained Deep Neural Networks," *2020 Cybernetics & Informatics (K&I)*, Velke Karlovice, Czech Republic, 2020, pp. 1-5, doi: 10.1109/KI48306.2020.9039793.

[3] Kommaraju, Rajasekhar & Anbarasi, M.. (2023). Detection of Diabetic Retinopathy (DR) Severity from Fundus Photographs using Swin Transformer. 1-6. 10.1109/ICRAIE59459.2023.10468500.

[4] G. Alwakid, W. Gouda, M. Humayun, and N. Z. Jhanjhi, "Deep learning enhanced diabetic retinopathy image classification," Digit. Health, vol. 9, Aug. 2023, Art. no. 20552076231194942.

[5] Chandra, Ritesh & Tiwari, Sadhana & Kumar, Shashi & Agarwal, Sonali. (2024). Diabetic Retinopathy Prediction Based on CNN and AlexNet Model. 382-387. 10.1109/Confluence60223.2024.10463351.

[6] Alavee, Kazi & Hasan, Mehedi & Zillanee, Abu & Mostakim, Moin & Uddin, Jia & Alvarado, Eduardo

& De la Torre Díez, Isabel & Ashraf, Imran & Samad, Md Abdus. (2024). Enhancing Early Detection of Diabetic Retinopathy Through the Integration of Deep Learning Models and Explainable Artificial Intelligence. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3405570.

[7] A. R. Wahab Sait, "A lightweight diabetic retinopathy detection model using a deep-learning technique," Diagnostics, vol. 13, no. 19, p. 3120, Oct. 2023

[8] Yau, J. W., Rogers, S. L., Kawasaki, R., Klein, R., Matsuura, T., Pacheco, D., .. & Wong, T. Y. (2012). Global prevalence and major risk factors for diabetic retinopathy. Diabetes care, 35(10), 2062-2068.