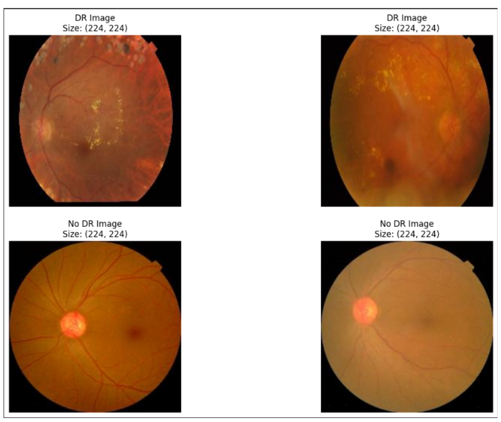
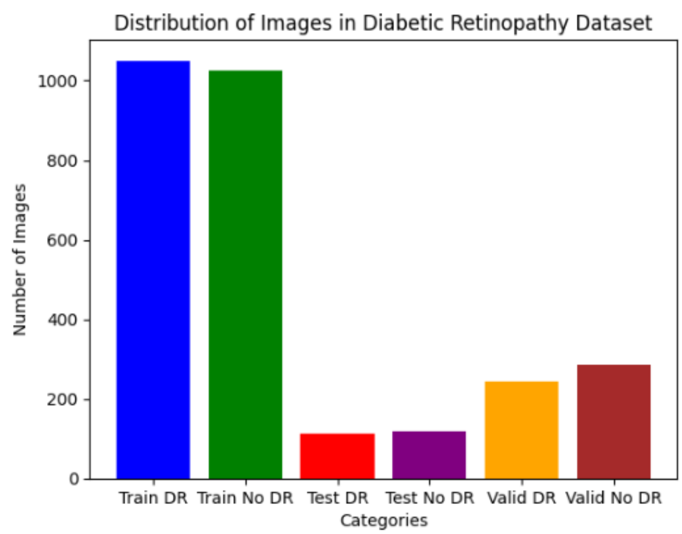
4980 Machine Learning
Muhtasim Fuad Jeet
Unit 6: Final Project

Section 1: About the dataset
The dataset comprises high-resolution retinal images suitable for training and evaluating automated systems for Diabetic Retinopathy (DR) detection and grading. The images capture various imaging conditions, reflecting real-world clinical scenarios. Each image has been meticulously assessed by a medical professional for the presence of DR and assigned a corresponding grade on a scale between 0 and 1, with the following interpretation: Grade 0: Diabetic Retinopathy present Grade 1: No Diabetic Retinopathy This rich dataset holds immense potential for advancing the development of automated DR detection and grading algorithms. Its large size, encompassing diverse image conditions, allows for robust training and reliable performance evaluation. Moreover, including expert-annotated grades provides valuable ground truth for training and assessing model performance. The dataset's key characteristics include: This comprehensive dataset will be invaluable for researchers and developers working on automated DR detection and grading systems, ultimately contributing to improved early detection, timely intervention, and personalized treatment for individuals with diabetes.

Figure 1: Example images and distribution of images in the dataset





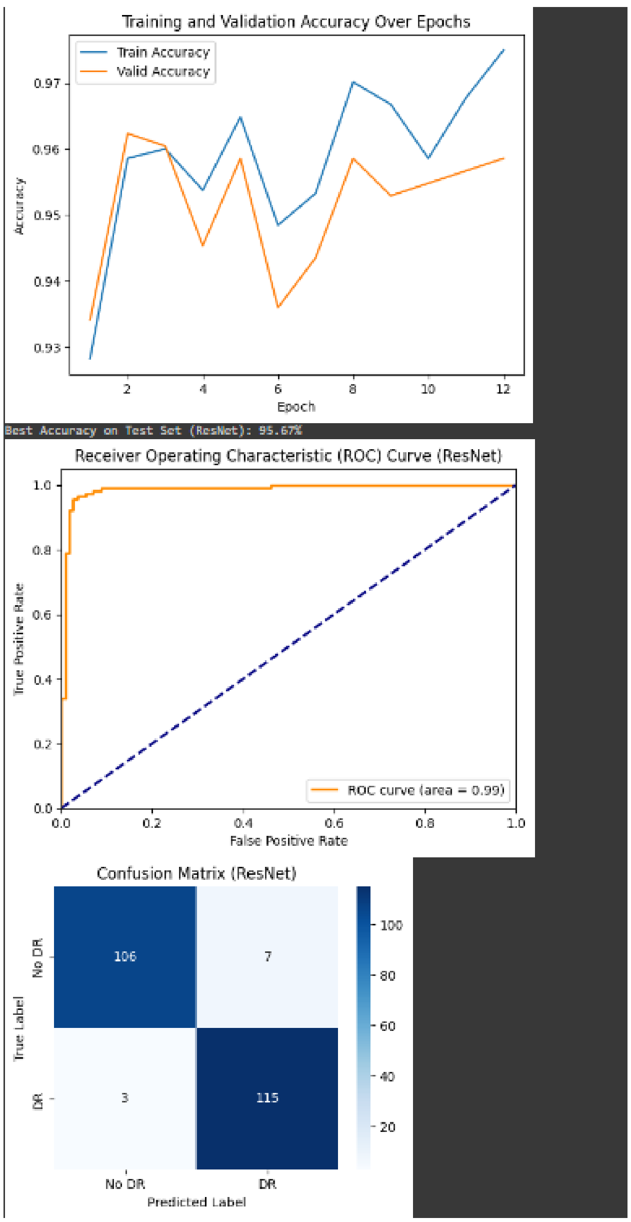
Section 2:
Models used for this dataset
For this project, we used the following models:

ResNet50: ResNet, short for Residual Networks, is a popular model in the field of deep learning for image recognition. The "50" in ResNet50 refers to the number of layers it has - in this case, 50 layers. This model is known for its use of residual connections which help in avoiding the problem of vanishing gradients, a common issue in deep neural networks as they get deeper.
EfficientNet: EfficientNet is a series of models that are designed to provide a scalable and efficient architecture for deep learning. These models use a compound coefficient to uniformly scale all dimensions of the network—depth, width, and resolution. EfficientNets are known for achieving higher accuracy with fewer parameters compared to other models, making them efficient in terms of computational resources.
MobileNet V3: MobileNet V3 is a version of the MobileNet model that is specifically designed for mobile and edge devices. It's known for its efficiency and small size, making it suitable for applications where resources are limited. MobileNet V3 uses techniques such as network architecture search and a new activation function, H-Swish, to improve performance and efficiency.

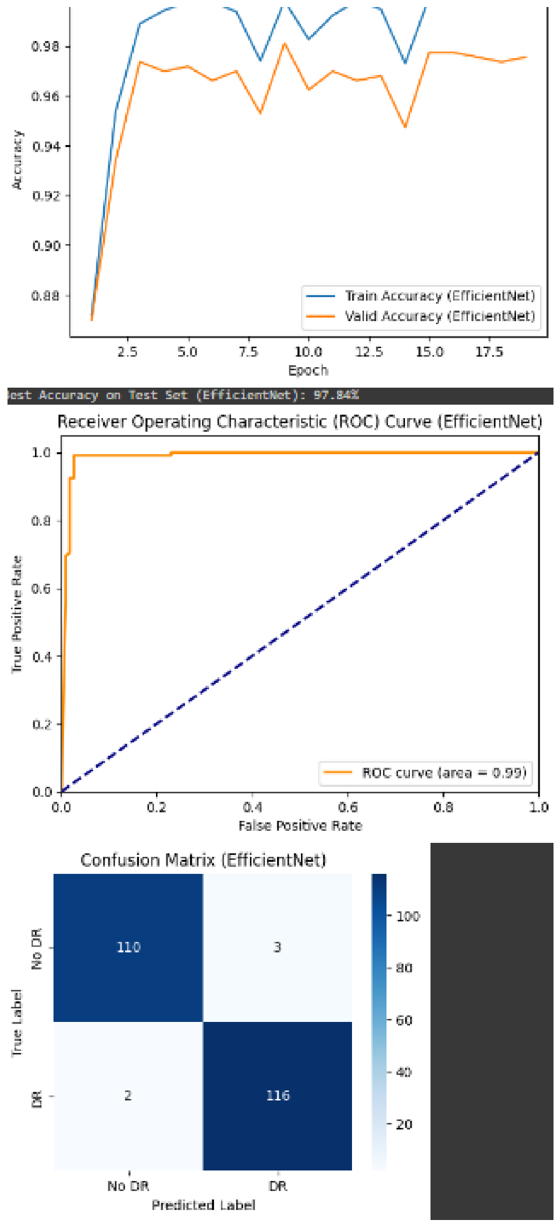
ViT (Vision Transformer): The Vision Transformer (ViT) model is a bit different from the other models as it applies the transformer architecture, commonly used in natural language processing, to image recognition tasks. Unlike conventional CNNs, ViT divides an image into patches and processes these patches as a sequence, similar to how words in a sentence are processed in NLP. This approach allows ViT to capture global information about the image, which can lead to improved performance on image recognition tasks.

Section 3: Workflow
For each of these models:
(i) We used a Data loader for training the models
(ii) We plot the ROC and Confusion Matrix
(iii) We set early stoppage at 100 epochs for the models and plot the accuracy change over epochs

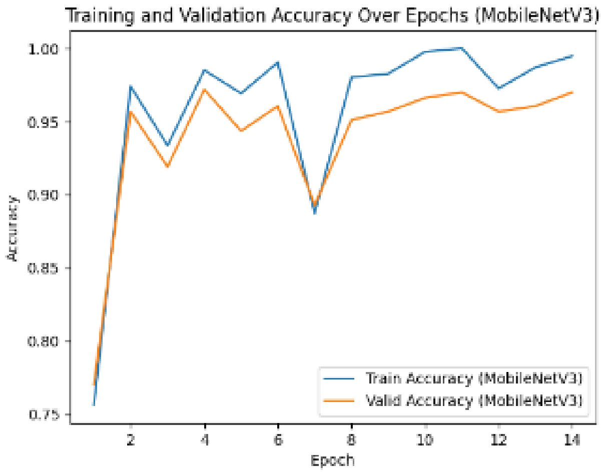
Renet50 Results

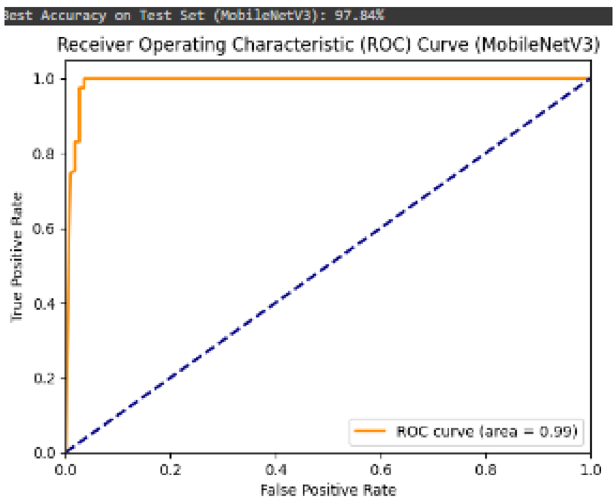


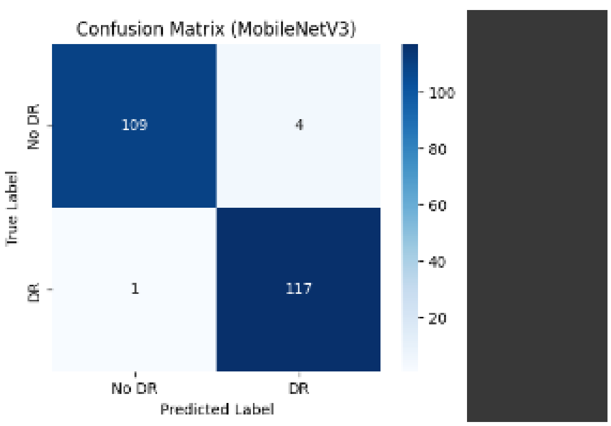
EfficientNet Results

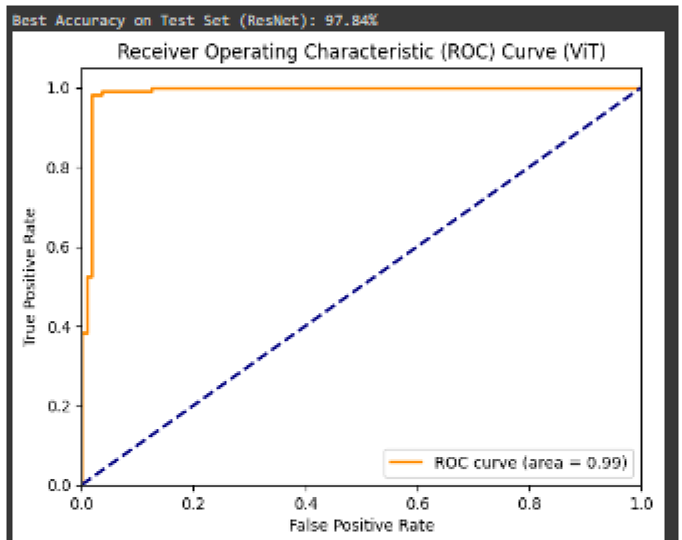


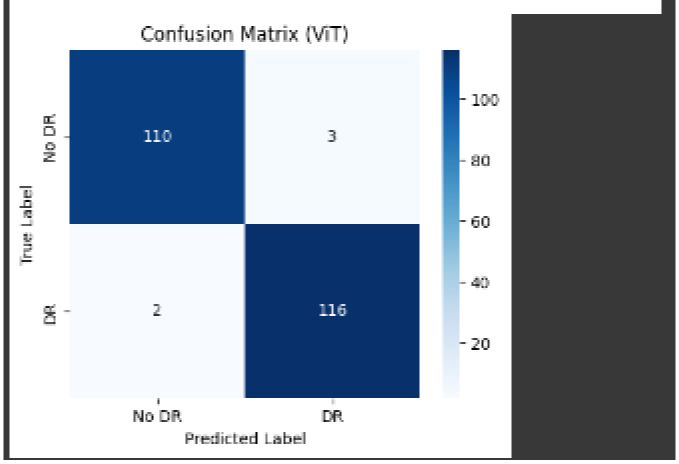
Results for MobileNetV3 **Results for ViT**





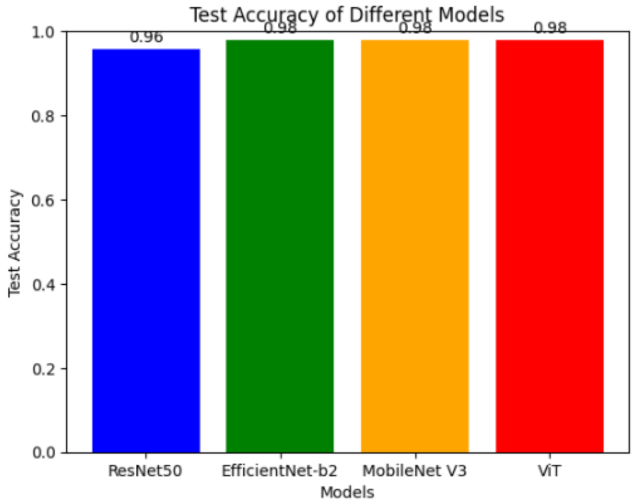






Section 4: Workflow Results

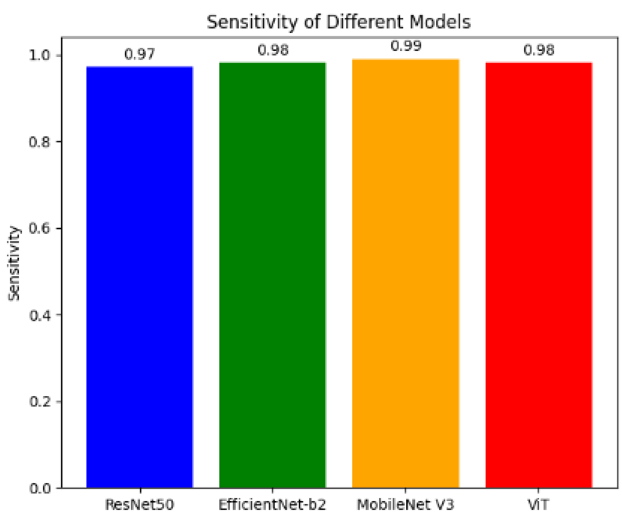
(i) After that, we compare the test accuracy of Different models.

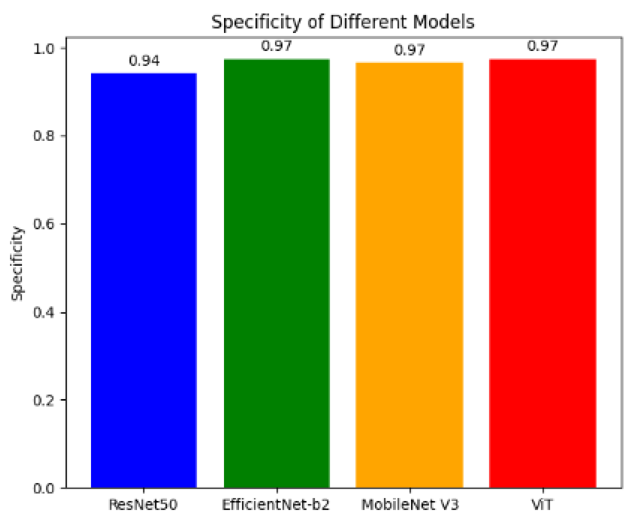


Based on the test accuracy results for different pre-trained models:

Comparing the Performance of Various Pre-trained Models for DR Diagnosis:

The test accuracies for the different models are as follows:ResNet50: 95.67% EfficientNet-b2: 97.83% MobileNet V3: 97.83% ViT (Vision Transformer): 97.83% These accuracy values suggest that all models perform very well in diagnosing diabetic retinopathy (DR). The test accuracies are quite high, with EfficientNet-b2, MobileNet V3, and ViT achieving the highest accuracy of 97.83%.





(ii) After that, we compare the sensitivity and specificity of different models.

Given that all models have a high ROC-AUC value (0.99), it indicates that they have near-perfect discriminatory power. Here's a brief interpretation:

A value of 0.99 is very close to the maximum value of 1.0, indicating that the models have an excellent ability to distinguish between positive and negative cases. The models' predictions result in high true positive rates and low false positive rates across different classification thresholds.

Consistency: The consistency of the high ROC-AUC across all models suggests that each model is performing very well in terms of discrimination. High Confidence: The models are making predictions with high confidence, which is a positive sign in a diagnostic or classification task.

Sensitivity: EfficientNet-b2, MobileNet V3, and ViT have similar sensitivity values, indicating that they perform similarly in correctly identifying positive cases. ResNet50 has a slightly lower sensitivity compared to the other models.

Specificity: All models have relatively high specificity values, and there is less variation between them. ViT has the highest specificity, followed closely by EfficientNet-b2 and MobileNet V3. ResNet50 has a slightly lower specificity compared to the other models.
Conclusion: In terms of sensitivity and specificity:

EfficientNet-b2, MobileNet V3, and ViT demonstrate similar and strong performance in correctly identifying both positive and negative cases. ResNet50 shows slightly lower sensitivity and specificity compared to the other models. For a more comprehensive evaluation, it's beneficial to consider both sensitivity and specificity, as they provide insights into different aspects of model performance. In this case, ViT appears to have a balanced and strong performance in terms of sensitivity and specificity.

Section 5: Ethical Implications:
Analyzing the ethical implications, potential harms, or biases related to the described dataset and models involves considering several key aspects:

Ethical Implications and Potential Harms Data Privacy and Consent: High-resolution retinal images are sensitive personal health data. It's crucial to ensure that the data was collected with informed consent and is being used in compliance with data protection laws (like GDPR or HIPAA). Misuse or unauthorized access to such data could lead to privacy breaches.
Misdiagnosis Risks: Automated systems for disease detection and grading, while beneficial, carry the risk of misdiagnosis. False negatives might lead to delayed treatment, while false positives could cause unnecessary anxiety or treatment. The reliability and accuracy of these systems must be continuously monitored and improved.
Over-reliance on Technology: There's a risk that healthcare providers might over-rely on automated systems, potentially undermining their clinical judgment. It's important to use these systems as assistive tools rather than definitive decision-makers.
Biases in the Dataset and Models Representation Bias: Does the dataset adequately represent the diverse population affected by Diabetic Retinopathy? Factors like ethnicity, age, and underlying health conditions can affect retinal appearance. A lack of diversity can lead to biased algorithms that perform well for some groups but not for others.
Annotation Bias: The grades assigned by medical professionals could be subjective. Inter-rater variability might introduce inconsistencies in the dataset, affecting the learning process of models.
Model Generalization: EfficientNet-b2, MobileNet V3, ViT, and ResNet50 have their own architectural strengths and weaknesses. Their performance in real-world, diverse clinical settings needs to be validated. A model that performs well on a specific dataset might not generalize well to other datasets or real-world scenarios.

Addressing These Issues Data Privacy:
Implementing strict data governance policies, ensuring all data is anonymized and stored securely.

Diverse and Inclusive Dataset: Ensuring the dataset covers a wide range of demographics and imaging conditions to reduce representation bias.
Continuous Evaluation: Regularly updating and testing the models against new data to ensure they adapt to changing real-world conditions.
Human Oversight: Maintaining human oversight in the decision-making process, especially for critical healthcare decisions.
Transparent Reporting: Reporting the performance metrics of each model, including their limitations, to ensure informed usage by healthcare professionals.
Cross-Validation with Multiple Models: Using a combination of models (like EfficientNet-b2, MobileNet V3, ViT, and ResNet50) for validation can reduce the risk of biases inherent in a single model.

In summary, while the dataset and models used hold immense potential for advancing DR detection, it's vital to address ethical concerns, potential harms, and biases to ensure they are used responsibly and effectively in clinical settings.

Link to project:
https://colab.research.google.com/drive/1DkQy_ILRffv_Ymb2UtIxqHp3S_miUGOg?usp=sharing