Group #11 - ST1
BARRION, Mary Barbie
DE GUZMAN, Adrian Gil
MANABAT, Jervie

Final Project

December 07, 2024
CSCI 114 : Pattern Recognition

INTRODUCTION

Diabetic Retinopathy (DR) is a microvascular complication of diabetes and is the leading cause of preventable blindness among working-age adults globally. It is estimated to affect over 93 million people worldwide, with the International Diabetes Federation projecting an alarming increase in the prevalence of diabetes to over 600 million cases by 2045 (International Diabetes Federation, n.d.; Fong et al., 2004). DR is caused by prolonged periods of hyperglycemia, which leads to damage to the retina's blood vessels. It progresses through five distinct stages: no DR, mild, moderate, severe, and proliferative DR, each reflecting varying degrees of vascular abnormalities and retinal damage (EyePACS, n.d.). Early detection and timely treatment are crucial to mitigating vision loss, as studies have shown that blindness from DR can be reduced to as low as 5% with prompt intervention (Chappelow et al., 2012).

Despite the potential for intervention, the current diagnostic approach for DR heavily relies on manual evaluation by clinicians, requiring expertise and specialized equipment. This process involves examining digital color fundus photographs of the retina to identify specific pathological features, such as microaneurysms, hemorrhages, and exudates (Thanati, Chalakkal, & Abdulla, 2019). However, this manual process is both time-consuming and resource-intensive, creating significant barriers to care, particularly in resource-limited settings where the prevalence of diabetes is often highest. Moreover, delays in obtaining results from human reviewers can lead to missed follow-ups, miscommunication, and delayed treatment, exacerbating the burden of vision impairment (Deng et al., 2009).

To address these challenges, automated detection systems have emerged as a promising solution for DR classification. Recent advancements in machine learning, particularly deep learning, have demonstrated exceptional potential in medical image analysis. Convolutional Neural Networks (CNNs), for instance, have been extensively applied to DR detection, showing superior accuracy compared to traditional methods reliant on handcrafted feature extraction (Qummar et al., 2019; Jiang et al., 2019). However, the effectiveness of deep learning-based methods often depends on the availability of large, high-quality labeled datasets and computational resources, both of which may be limited in real-world scenarios (Marmanis et al., 2016).

Given the constraints of limited computational resources and time, the group recognizes the inherent challenges in achieving state-of-the-art performance in training a model for Diabetic Retinopathy (DR) classification. Without access to high-performance computing systems equipped with powerful GPUs, the team’s ability to train complex, deep learning models is restricted. However, the group is motivated by the growing prevalence of DR, a condition that is becoming a significant health issue, particularly in underserved communities where access to specialized care is often limited. The potential impact of DR on patients' vision, leading to blindness if not detected and managed early, serves as a strong driving force for the group to contribute in any way they can, even within these constraints.

Starting with this advocacy, the group is committed to developing a basic prototype that could lay the groundwork for future advancements. While the current project will focus

on achieving a modest accuracy target in the range of 30%-50%, the goal is to build a functional model that can at least demonstrate the potential of machine learning in assisting healthcare providers with the early detection of DR. Although this level of accuracy is far from the clinical standards needed for real-world implementation, the project serves as an important first step. It is an opportunity for the group to learn the foundational techniques in model development and data preprocessing, as well as gain hands-on experience in the challenges of working with medical image data.

The group acknowledges that their current model is very far from being suitable for clinical use, but they view this as part of a larger, ongoing effort. The work done in this initial phase will provide valuable insights and a base upon which more advanced models can be built in the future. Additionally, this project could inspire further research into cost-effective methods for detecting DR, particularly for regions where advanced technology and medical expertise are scarce. In this way, the project not only aims to contribute to the immediate goals of model development but also aligns with broader healthcare objectives of improving accessibility to preventive care and reducing the impact of diabetes-related blindness.

METRICS AND OBJECTIVES

Project Objectives

The following are the key objectives the group aims to accomplish by the conclusion of this project. These objectives serve as a roadmap, ensuring that the project aligns with its primary goals and delivers meaningful outcomes:

1. To build an initial model capable of classifying diabetic retinopathy into different stages using deep learning, with a focus on simplicity and feasibility given the available computational resources
2. To focus on gaining a fundamental understanding of essential preprocessing steps, such as image resizing and normalization, to work within the limitations of available resources
3. To achieve more than 50% in accuracy, precision, recall, and f1-score to provides a starting point for further development
4. To use basic evaluation metrics like accuracy and confusion matrices to assess the model’s performance and identify areas for improvement, without aiming for complex validation techniques
5. To provide the group with practical experience in building, training, and evaluating a deep learning model, while learning the challenges and limitations of working with medical image data

Metrics and Goals

This section presents the evaluation metrics the group aims to achieve. These metrics focus on measuring the model's basic functionality and its ability to classify diabetic retinopathy cases accurately within the project's limitations.

a. **Goal: Correctly classify at least 50% of the cases**
   Rationale: We want to prioritize the correctness of the models' predictions in more cases as possible to ensure its reliability as a tool in helping doctors in classifying retinopathy levels. For this reason, the group is targeting at least 50% level of correct predictions to make the results acceptable and assistive in diagnoses. However, we considered the difficulty of achieving excellent results given the limited amount of time and system limitations so the main target is to at least accurately detect more than half of the cases.
b. **Goal: Detect at least 50% of true cases at each severity level**
   Rationale: Our group aims to detect as many true cases as possible, regardless of severity, to ensure accurate diagnosis and ensure that patients receive the medications and care in accordance to their level.
c. **Goal: Identify at least 50% of patients without the condition**
   Rationale: With regards to detection of true cases, we also want to detect cases without retinopathy (retinopathy scale: 0) to avoid healthy patients from receiving unnecessary treatments or tests to avoid financial and medical complications.
d. **Goal: Achieve 50% reliability in balancing correct identification of cases and avoidance of false alarms**
   Rationale: We prioritize finding a balance between correct detection of patients with retinopathy and healthy patients to ensure that the model outputs results that are reliable and applicable to the diagnosis of any kind of patients.

DATASET

Dataset Selection

The dataset selected for this project is *Diabetic Retinopathy Arranged Dataset*, accessible on Kaggle at this link. This dataset consists of high-resolution retinal images captured using specialized fundus cameras designed for imaging the interior of the eye. These images have been anonymized to protect patient privacy and curated to remove any identifying information.

The dataset contains 35,126 retinal images categorized into five classes representing the progression stages of diabetic retinopathy (DR):

- 0: No DR
- 1: Mild DR
- 2: Moderate DR
- 3: Severe DR
- 4: Proliferative DR

This dataset provides a valuable resource for training and testing AI and machine learning models to diagnose diabetic retinopathy accurately. By facilitating automated analysis, the dataset contributes to advancing early detection and monitoring of DR, which are critical to preventing vision loss in diabetic patients.

Data Preprocessing Steps

Preprocessing is a crucial step to ensure the dataset is suitable for training a machine learning model and addresses potential issues such as imbalances, variability in image quality, and input inconsistencies. The following preprocessing steps will be applied:

Dataset Preparation

1. Dataset Balancing

   To ensure balanced representation across all classes, we will use 260 images per stage, with 234 images allocated for training and 26 for testing. We will use these numbers because stage 4 has 708 images but after further preprocessing images, only 260 are fit to be included—the least number among the stages. This approach aligns with the methodology described by (Li et al. 2014), where "in each testing session, samples from one group are chosen as the testing data, and all other samples from the remaining nine groups are used for training." By adopting this strategy, we address class imbalance, which can otherwise bias the model towards overrepresented categories and reduce its accuracy in predicting underrepresented stages.

2. Image Resizing

   All images will be resized to a standard dimension, such as 1024 x 768 pixels, using interpolation before using them for the models. This resizing standardizes the input size, simplifies the processing pipeline for neural networks, and reduces computational complexity.

3. Data Splitting

   The dataset will be divided into training and testing subsets, with 90% of the data allocated for training and 10% for testing. This approach follows the method described by (Soomro et al. 2019), where "the generated patches area combined with the original full images and 90% of the images is used for training and the remaining is used for validation." Stratified sampling will be used to maintain consistent class distributions across both subsets.

4. Image Quality Check

   Images with significant noise, artifacts, or incomplete retinal views will be removed from the dataset. This step eliminates poor-quality samples that could introduce noise and hinder the model's ability to learn meaningful patterns.

5. Label Encoding

   Class labels, initially represented as categorical descriptions (e.g., "No DR"), will be converted into numerical values (0-4). This transformation is necessary for supervised learning algorithms to process the labels effectively.

Dataset Preprocessing in the Models

1. Cropping to Square Dimensions

   When loading the images, each will be cropped into a square with size 256x256 pixels, to highlight the retina and guarantee a uniform input size. This stage guarantees that the subject of interest (the retina) is centrally positioned while minimizing irrelevant background information.

2. Color Conversion (BGR to RGB and Grayscale)

   The color space of the photographs will be changed from BGR to RGB to improve the readability of retinal veins and structures. This is especially useful for medical picture analysis, where minor color variations can have diagnostic implications. Moreover, for some models grayscale transformation will be applied to training subsets of images to simulate lighting conditions, ensuring that the model is robust to changing image properties.

3. Weighted Gaussian Blur

   A weighted Gaussian blur will be used to cut down on noise while maintaining important features. In order to improve the clarity of veins and lesions for classification, this procedure smoothens the image without obscuring important features.

4. Random Augmentation

   Horizontal flipping, random rotation (up to 10 degrees), random resized cropping (scaled between 80% and 100%), and minor color jitter (brightness, contrast, saturation alterations) are some of the augmentations that will be used. These augmentations increase the variety of dataset and improve the model's ability to generalize.

5. Normalization

   Images will be normalized using mean and standard deviation values ([0.485, 0.456, 0.406] and [0.229, 0.224, 0.225], respectively), aligning with pre-trained models' expectations. This standardization helps in faster convergence during training.

Observations and Dataset Limitations

The Diabetic Retinopathy Arranged Dataset is a robust foundation for building a classification model due to its size and diversity. However, several limitations should be considered:

I. Quality Variability - Some images may have low resolution or noise, which could negatively impact model performance.
II. Dataset Dependency - The dataset might not fully capture the variability found in real-world clinical settings, potentially limiting the model's generalizability.

III. Class Representation - In real-world scenarios, the distribution of DR cases may not be as balanced as the dataset, posing challenges when applying the model to clinical populations.
IV. Bias - The dataset may lack sufficient demographic diversity, such as variations in age or ethnicity, which could affect its applicability across broader populations.
V. Label Accuracy - There is no explicit guarantee that all labels were assigned by medical experts, potentially introducing inaccuracies in ground truth annotations.

Overall, Diabetic Retinopathy Arranged Dataset is well-suited for developing a machine learning model to classify DR stages. The proposed preprocessing steps aim to enhance the dataset's quality, address imbalances, and optimize its suitability for model training and evaluation. While the dataset has certain limitations, recognizing these challenges provides a foundation for improving the model's performance and ensuring its reliability for practical applications.

MODELS OVERVIEW

For this project, the group decided to train three models: one classical machine learning model, a Decision Tree, and two neural networks, U-Net CNN and ResNet-18 CNN. Each model was chosen for its distinct functionality and approach to solving the problem. The Decision Tree provides a straightforward, interpretable method for identifying patterns in data, while U-Net specializes in segmenting specific regions in medical images. In contrast, ResNet-18 is designed for extracting complex features and performing classification tasks on retinal images. The following sections outline the definition and rationale for selecting each model.

(1) DecisionTree

A Decision Tree is a machine learning model that makes decisions by dividing data into multiple branches based on its features. It functions similarly to a flowchart, in which each node represents a decision based on a feature and each leaf node represents the final outcome.

Rationale for Model Choice:

A Decision Tree is suitable for this project because of its simplicity in implementing and flexibility in working with numerical and categorical data. It has the ability to quickly identify key features affecting the severity of diabetic retinopathy without requiring a great amount of data. Most importantly, the main reason this is chosen on top of other classical models is due to the fact that it yielded the best results from the parameter and classical model testing.

(2) U-Net CNN

U-Net is a deep learning model commonly used for image segmentation and it consists of two main parts, namely a contracting path and

an expansive path. The contracting path collects features while decreasing the size, while the expansive path rebuilds the image to its original size using the features and skip connections that pass important information between the two parts (GeeksforGeeks, 2023).

Rationale for Model Choice:

U-Net is an effective image segmentation technique used to identify specific regions in medical images and its ability to deliver accurate results with limited data makes it popular for detecting anomalies or structures (Siddique et al., 2020). In the case of diabetic retinopathy, U-Net allows excellent isolation of small features like blood vessels, lesions, or microaneurysm, which are factors in determining the severity. This is done through its encoder-decoder structure which enables precise pixel-level classification, making it effective for analyzing retinal images.

(3) ResNet-18 CNN

Resnet-18 is a Convolutional Neural Network (CNN) architecture introduced in the Residual Networks (ResNet) family (Resnet18, n.d.). It is a convolutional neural network that is 18 layers deep (Resnet18, n.d.). The model is known for its use of residual learning, which addresses the vanishing gradient problem and enables the training of very deep networks by introducing skip connections or residual blocks. The architecture consists of 18 layers (hence the name), with a structure that includes convolutional layers, batch normalization, ReLU activations, and shortcut connections.

Rationale for Model Choice:

ResNet-18 is widely used for image classification tasks, known for its efficient performance and relatively low computational demands. For the Diabetic Retinopathy (DR) classification task, ResNet18 is a suitable choice due to its ability to extract complex features from retinal images, even with limited computational resources. Given the dataset's size and preprocessing enhancements like normalization and augmentation, ResNet18's robust feature extraction and generalization capabilities make it a practical model to achieve the project's modest accuracy goals while providing a foundation for future improvements. Additionally, its pre-trained weights on ImageNet allow transfer learning, accelerating convergence and enhancing performance with limited data.

MODEL DEVELOPMENT

I. Trained Models

A. Model Files

Python notebooks and trained model files folder can be accessed here: CSCI114_FinalProject_Group11

B. Short Description on Models Training

The Decision Tree model utilizes preprocessing techniques like image resizing to consistent dimensions and color-space transformations (e.g., GRAY, LAB, HSV) to optimize input features. Additionally, parameter testing is performed by varying image sizes and color spaces to identify the best-performing configuration, ensuring the classical model is fine-tuned for accuracy and efficiency.

ResNet takes advantage of residual connections to address vanishing gradients, enabling effective learning in deep layers. More techniques are used in learning including transfer learning with pretrained weights (e.g., ImageNet1K), data augmentation (random rotation, resizing, color jittering), and regularization through dropout. Fine-tuning with the AdamW optimizer and learning rate scheduling ensures stable and high-performance training.

The U-Net model uses a symmetrical encoder-decoder structure, where the encoder extracts features with convolution and max-pooling, and the decoder reconstructs segmented images via upsampling and skip connections. Skip connections are included to enhance precision by transferring spatial details from the encoder to the decoder. Training is also optimized with ReLU activations, batch normalization, and the Adam optimizer for efficiency.

II. Results of Evaluation Metrics on Training and Testing Dataset

A. Performance Metrics

1. Decision Tree

Decision Tree Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2
2	1.00	0.67	0.80	3
1	0.50	1.00	0.67	2
3	1.00	0.67	0.80	3
4	1.00	1.00	1.00	3
accuracy			0.85	13
macro avg	0.90	0.87	0.85	13
weighted avg	0.92	0.85	0.86	13

Evaluation Metrics:

Accuracy	0.85
Precision	0.92
Recall	0.85
F1 Score	0.86

2. ResNet-18 CNN

ResNet CNN Classification Report:

	precision	recall	f1-score	support
0	0.38	0.46	0.41	26
1	0.48	0.38	0.43	26
2	0.36	0.31	0.33	26
3	0.68	0.81	0.74	26
4	0.79	0.73	0.76	26
accuracy			0.54	130
macro avg	0.54	0.54	0.53	130
weighted avg	0.54	0.54	0.53	130

Evaluation Metrics:

Accuracy	0.54
Precision	0.54
Recall	0.54
F1 Score	0.53

3. U-Net CNN

U-Net Classification Report:

	precision	recall	f1-score	support
0	0.56	0.38	0.45	26
1	0.39	0.46	0.42	26
2	0.33	0.15	0.21	26
3	0.59	0.77	0.67	26
4	0.69	0.92	0.79	26
accuracy			0.54	130
macro avg	0.51	0.54	0.51	130
weighted avg	0.51	0.54	0.51	130

Evaluation Metrics:

Accuracy	0.54
Precision	0.51
Recall	0.54
F1 Score	0.51

III. Discussion of Model Results

A. Did the Model perform good or not? Explain why you think so.

The Decision Tree model achieved an accuracy of 85%, which significantly exceeded the 50% target set for correct classification. This suggests that the model is good at correctly identifying cases in a significant portion of the data. In addition, the model performed exceptionally well for class 0 and class 4, with 100% precision and recall, indicating that it can accurately identify patients without the condition and those at the highest severity level. However, performance for class 1 and class 2 was weaker, particularly for class 1, where precision dropped to 50%, showing some

difficulty in distinguishing these classes. Overall, the decision tree has a strong balance, but may struggle with certain severity levels.

The ResNet-18 CNN model achieved an accuracy of 54%, which aligns with the goal of correctly classifying at least 50% of the cases. However, this is lower than the Decision Tree, indicating that the CNN struggles more with the dataset, potentially due to its complexity and overfitting or lack of sufficient data. Precision, Recall, and F1-Score: The performance varies across classes, with particularly low performance in class 2, showing that it has difficulty with intermediate severity levels. For class 3 and 4, the recall is higher, indicating that the model does well in identifying more severe cases, but the overall precision is still low, indicating a higher rate of false positives.

The UNet CNN model also achieved an accuracy of 54%, which meets the 50% target but does not perform as well as the Decision Tree. This suggests that while it can identify some key patterns, its ability to generalize to the full range of diabetic retinopathy cases is still limited. Similar to the ResNet-18, the UNet CNN model struggles with intermediate severity classes (class 2 and class 1), but performs better with class 4, showing its ability to identify more severe cases. Its performance across the board in terms of precision, recall, and F1-score is lower compared to the Decision Tree, reflecting some weaknesses in detecting cases accurately across all severity levels.

In general, all of the models met the target goals for the evaluation metrics with the Decision Tree outperforming the ResNet and U-Net by a huge margin. The classical model also displayed great balance in terms of classifying across multiple levels. Furthermore, the results also highlighted the last two models' difficulty in classifying intermediate cases but displaying more accuracy in detecting more severe cases.

B. How did the models compare to one another? Explain how they differ

The Decision Tree model clearly outperformed the neural networks, achieving the highest overall accuracy (85%) and demonstrating balanced performance across multiple severity levels. Its simplicity and ability to work effectively with limited data gave it an advantage over the neural networks, particularly in classifying both the absence of diabetic retinopathy (class 0) and the most severe cases (class 4).

In contrast, the ResNet-18 CNN and U-Net CNN, both achieving 54% accuracy, struggled to match the Decision Tree's performance. The ResNet-18, designed for image classification, showed strength in detecting severe cases (class 3 and class 4) but struggled with intermediate classes, likely due to the dataset's complexity and potential overfitting. Similarly, the U-Net CNN, typically used for image segmentation, was effective in identifying severe cases but showed weaknesses in its ability to classify

intermediate severity levels and to generalize across the dataset.

In summary, the Decision Tree's classical approach provided a better balance and reliability, while the neural networks highlighted their potential for handling more severe cases but were limited by data and resource constraints.

C. Did any of the models reach your goal? What model would you recommend to use for this use case based on your goal?

All three models met the minimum goal of 50% accuracy, with the Decision Tree significantly exceeding expectations by achieving 85% accuracy. This model demonstrated the best balance and reliability, making it the most suitable choice for this use case.

Given the project's goal of building a functional prototype to assist in diabetic retinopathy classification, the Decision Tree is recommended. It did not only surpass the accuracy target but also displayed strong performance in classifying both mild and severe cases. While the ResNet-18 and U-Net CNNs met the minimum goals and showed potential for further development, their current limitations in handling intermediate severity levels make them less practical for immediate use.

However, despite the Decision Tree's superior performance in accuracy, we recommend using the ResNet-18 CNN for model deployment. Neural network-based models, particularly CNNs, are generally better suited for medical image classification due to their ability to capture intricate spatial features and patterns in images, which classical models like Decision Trees cannot effectively achieve. ResNet-18, as a convolutional neural network, can handle the complexity of medical imaging data and has the flexibility to scale for improvements with additional data or fine-tuning, making it more adaptable for real-world deployment.

Neural networks are particularly advantageous in medical imaging because of their capacity for hierarchical feature extraction. For example, while a Decision Tree relies on preselected features, a CNN like ResNet-18 can automatically learn relevant features, such as the texture and structure of retinal blood vessels, lesions, and microaneurysms, which are critical for diagnosing diabetic retinopathy (Litjens et al., 2017). Additionally, ResNet-18 includes residual learning, which addresses the vanishing gradient problem and allows for efficient training even with relatively deep architectures, making it more robust for handling complex visual data (He et al., 2016).

In terms of deployment, ResNet-18 also benefits from advancements in transfer learning. Using pre-trained weights (e.g., on ImageNet), the model can be fine-tuned for diabetic retinopathy classification, enabling faster convergence and improved performance even with limited datasets. CNNs

also integrate better with modern medical imaging workflows, as they can be optimized for GPU acceleration, allowing for faster predictions and scalability in clinical settings (Esteva et al., 2017).

While the Decision Tree provides simplicity and interpretability, its limitations in capturing the high-dimensional and complex patterns within retinal images make it less ideal for deployment in medical image classification tasks. In contrast, ResNet-18's flexibility, adaptability to larger datasets, and superior performance potential in image-based tasks make it the preferred choice for deployment in diabetic retinopathy classification systems.

CONCLUSION AND RECOMMENDATIONS

I. Conclusion

The evaluation of three distinct machine learning models—Decision Tree, ResNet-18, and U-Net—revealed notable differences in their ability to classify diabetic retinopathy severity levels. While the Decision Tree model achieved the highest accuracy, it is clear that neural networks, specifically ResNet-18, hold potential for better performance as they are further optimized and trained on larger, more diverse datasets. The Decision Tree's excellent performance across a broad range of severity levels makes it a valuable tool for immediate deployment in clinical settings where quick and reliable results are needed. However, the inherent complexity and capacity of deep learning models like ResNet-18 and U-Net suggest that, with additional training and data, they may offer significant improvements in handling more complex medical imaging tasks.

This study underscores the importance of selecting the right model based on specific healthcare needs. While classical models offer practical and interpretable solutions, neural networks offer the adaptability and potential for greater accuracy as datasets evolve. This distinction is crucial for the future of medical image analysis, particularly in the context of diabetic retinopathy screening.

Overall, the project successfully accomplished its objectives, including the development of an initial model capable of classifying diabetic retinopathy into various stages using deep learning techniques. Despite the constraints posed by available computational resources, the model achieved an accuracy exceeding 50%, which met the project's primary goal. Additionally, the group gained practical experience in essential preprocessing techniques, such as image resizing and normalization, which are critical for adapting medical image data to deep learning models. The evaluation process, conducted through basic metrics like accuracy and confusion matrices, provided a solid foundation for future improvements by identifying key performance gaps.

By completing this project, the team not only met the predefined objectives but also gained valuable insights into the challenges of working with medical image data, particularly within the constraints of computational resources. The results demonstrate that even with limited resources, deep learning models can be successfully applied to the classification of diabetic retinopathy

II. Recommendations

Given the current limitations in computational resources, the models tested in this project—specifically the Decision Tree, ResNet-18, and U-Net—were able to meet the basic objectives. However, to fully maximize the potential of these models and achieve high accuracy results, further studies and model developments are recommended, particularly if the project has access to more generous computational resources.

For instance, fine-tuning the ResNet-18 model could significantly improve its performance. By leveraging transfer learning and adjusting the model's hyperparameters, the model’s ability to extract complex features from diabetic retinopathy images could be enhanced. Additionally, more advanced versions of convolutional neural networks, such as ResNet-50 or DenseNet, could be explored for their ability to capture even deeper and more nuanced features. With sufficient computational power, models like these could provide superior classification accuracy, especially for challenging intermediate stages of diabetic retinopathy that were more difficult for the current models to identify.

Another avenue for future research includes developing hybrid models that combine classical machine learning approaches with deep learning techniques. Such models could potentially balance the strengths of both approaches, offering flexibility and interpretability while retaining the power of deep learning in feature extraction. Moreover, exploring more specialized architectures for medical image classification, such as Vision Transformers (ViTs) or Attention-based models, may offer better scalability and flexibility when working with complex and diverse medical imaging datasets.

Additionally, the expansion of the dataset used for training could greatly benefit model development. By incorporating a larger and more diverse range of retinal images—possibly sourced from multiple healthcare institutions or publicly available datasets—the model's generalization capabilities could be improved, leading to better performance across different populations and severity levels.

In summary, while the current models provide a solid foundation, the potential for improved accuracy, efficiency, and flexibility in medical image classification is vast. If computational resources allow, fine-tuning existing models and exploring advanced neural network architectures would be key steps toward maximizing the impact of deep learning in diabetic retinopathy diagnosis and broader medical image analysis applications.

MODEL DEPLOYMENT

1. Application Video Demo short description
    a. Input and Output

        The application is a website called *Magis Optics*. This prototype is developed with Python Django for the backend, HTML and CSS for the frontend, and the ResNet model's state dictionary for prediction. It has four sections: *Home*, *Overview*, *About Model* and *Try Model*. *Home* contains the welcome message and a tagline. *Overview* contains a short description of the topic problem, classification system used, and objectives. *About Model* contains the information about the chosen model, rationale of model choice, and how the model works. *Try Model* contains the instruction on how to test or use the model and buttons needed. Input from users will be a retina image in the format of .jpg or .png while the output will be displayed in a separate screen containing the predicted class of the uploaded picture by the model.

2. Video Demo
   Video Demonstration Link: Group11_Video_Demo

REFERENCES

Chen, Wanghu & Yang, Bo & Li, Jing & Wang, Jianwu. (2020). An Approach to Detecting Diabetic Retinopathy Based on Integrated Shallow Convolutional Neural Networks. *IEEE Access*, 8, 178552-178562. https://doi.org/10.1109/ACCESS.2020.3027794

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, Jun. 2009, pp. 248–255.

Dugas, E., Jared, Jorge, & Cukierski, W. (2015). *Diabetic Retinopathy Detection*. https://kaggle.com/competitions/diabetic-retinopathy-detection

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. https://doi.org/10.1038/nature21056

EyePACS. (n.d.). *Diabetic Retinopathy Arranged Dataset*. https://kaggle.com

EyePACS. (n.d.). *Diabetic Retinopathy Detection*. https://www.eyepacs.com

Fong, D. S., et al. (2004). Retinopathy in diabetes. *Diabetes Care*, 27(1), 7-10.

GeeksforGeeks. (2023, June 8). *UNet Architecture Explained*. https://www.geeksforgeeks.org/u-net-architecture-explained/

GeeksforGeeks. (2024, June 17). *Comprehensive Guide to Classification models in SciKitLearn*. https://www.geeksforgeeks.org/comprehensive-guide-to-classification-models-in-scikit-learn/#6-naive-bayes

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

International Diabetes Federation. (n.d.). *Diabetes Atlas*. https://diabetesatlas.org/

Jiang, H., Li, Y., Zhang, X., & Liu, Y. (2019). An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In *Proceedings of the 41st Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, Jul. 2019, pp. 2045–2048.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van der Laak, J. A. W. M. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88. https://doi.org/10.1016/j.media.2017.07.005

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014). Medical image classification with convolutional neural network. In *Proceedings of the 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, Singapore, 2014, pp. 844-848. https://doi.org/10.1109/ICARCV.2014.7064414

Marmanis, D., Gikas, V., & Soille, P. (2016). Deep learning Earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1), 105–109. https://doi.org/10.1109/LGRS.2015.2496422

Resnet18. (n.d.). *MathWorks*. https://www.mathworks.com/help/deeplearning/ref/resnet18.html

Siddique, N., Sidike, P., Elkin, C., & Devabhaktuni, V. (2020). U-Net and its variants for medical image segmentation: theory and applications. https://doi.org/10.48550/arXiv.2011.01118

Soomro, T. A., Qamar, A., Ali, M., & Malik, A. S. (2019). Impact of image enhancement technique on CNN model for retinal blood vessels segmentation. *IEEE Access*, 7, 158183-158197. https://doi.org/10.1109/ACCESS.2019.2950228

Thanati, H., Chalakkal, R. J., & Abdulla, W. H. (2019). On deep learning-based algorithms for detection of diabetic retinopathy. In *Proceedings of the International Conference on Electronics, Information, and Communications* (pp. 1–7), Auckland, New Zealand.

Qummar, S., Niazi, M. A., & Farooq, M. (2019). A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access*, 7, 150530–150539.