Diabetic retinopathy detection and comparison of several CNN models

MSc Research Project
MSc. Data Analytics

Sanjeet Shekhar
Student ID: 23209470

School of Computing
National College of Ireland

Supervisor: Shubham Subhnil

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sanjeet Shekhar ……..………………………………………………………………………………………………
Student ID:	23209470 ……………………………………………………………………………………………..……
Programme:	MSc. Data Analytics …………………………………………………………… **Year:** 2024 …………………………..
Module:	MSc. Research Project ……………………………………………………………………………………….………
Supervisor:	Shubham Subhnil ……………………………………………………………………………………….………
Submission Due Date:	12/12/2024 ……………………………………………………………………………………….………
Project Title:	Diabetic retinopathy detection a comparison of several CNN models and optimizers ……………………………………………………………………………………….………
Word Count:	6156 ……………………………………… **Page Count** 15 …………………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………………………………………………………………………………………………………

Date: ………………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Diabetic retinopathy detection a comparison of several CNN models and optimizers

Sanjeet Shekhar
23209470

Abstract

Diabetic retinopathy is a type of medical condition that stems from diabetes and leads to partial or full loss of sight and is a leading contributor to blindness in most of the developed countries in the world. Early detection of the condition is not only effective by is crucial for the treatment of people suffering from diabetic retinopathy.

In this study, a convolutional neural network (CNN) – based framework for classification of several stages of diabetic retinopathy has been proposed. The study uses images from actual cases of several stages of the retinal disease and compares several models in order to find out the best-performing model in predicting the stage of the disease.

The study compares the ResNet model with EfficientNet and MobileNet to get a better understanding of the model suitable for the task study also compares how different optimizers compare with each other and concludes EfficientNet model with an accuracy of 82% trained using Adam optimizer to be the best choice for the use case.

1 Introduction

Diabetic retinopathy (DR) is an eye disease affecting 22.27 % of all people who have diabetes. Being such a chronic eye disease diabetic retinopathy is one of the major reasons for blindness among people suffering from diabetes. The disease is progressive and gradually leads to complete blindness. It starts with mild and if not treated moves to non-proliferative abnormality to severe proliferative diabetic retinopathy with a risk of irreversible vision loss, if untreated. The earlier the diagnosis the better the chances of treating the disease with no loss of vision. Nevertheless, lack of resources in many places high limits and constraints on proper screening and diagnosis are the main motivations behind making the process of diagnosis automated for efficient and reliable solutions, As addressed in the study, the challenges with automated diabetic retinopathy detection from retinal images are employing deep learning models based models.

The study proposes methods to take advantage of convolution neural networks (CNNs), in order to identify and classify different stages of diabetic retinopathy, by identifying the capability of CNNs in image classification and extraction of details that are necessary to identify various stages of DR. This gives advantage as compared to traditional methods in which identification of the disease and its stages mainly relies on manual examination of the examiner and is prone to human errors, an automated approach in this procedure can be more reliable fast and highly accurate while offloading some of the load on health care systems especially in remote and resource-constrained settings.

The main purpose of this work is to fill in the gap by evaluating and comparing three different well-known CNN models i.e. ResNet18, EfficientNetB0 and MobileNetV2 systematically on the task of Diabetic retinopathy classification based on different stages and the impact of two popular optimizers stochastic gradient descent (SGD) optimizer and Adaptive Moment Estimation (Adam) optimizer. The following study generates and evaluates a comprehensive analysis of model performance using retinal images from a publicly available dataset with rigorous pre-preprocessing and augmentation techniques.

The research question that is aiming to be achieved in this study is: For the purpose of detecting and classifying different stages of diabetic retinopathy how to choose the best combination of convolution neural network architecture and an optimizer? The study answers this question by assessing the accuracy and evaluating each model-optimizer pairing, to determine which pair performs the best for diabetic retinopathy detection and classification.

Advanced filtering to boost the performance of important features of the retina, as well as applying the augmentation in the training set including rotation and color normalization in order to boost model generalization are part of the methodology. The comparative analysis is expected to determine the best pairing of model and optimizer resulting in the most performed model, providing practical information on the design of automated diagnostic tools.

In the end, this study helps build the nascent field of AI powered medical diagnostics as an actionable recommendation for over DR detection frameworks. The findings could give rise to scalable solutions suitable for deployment in clinical and nonclinical settings in resource constrained environments. This research addresses critical gaps in model selection and optimization that will greatly improve the reliability and accessibility of automated DR diagnosis and will have a profound impact on patient outcomes worldwide.

2 Related Work

Diabetic retinopathy (DR) is a progressive ocular disease of people with diabetes and is a leading cause of preventable blindness worldwide. DR detection is revolutionized due to the advent of artificial intelligence, more specifically deep learning (DL) based technique to automate diagnosis using retinal image analysis. Various approaches to perform deep learning and a number of techniques to optimize and process the data have been studied to improve the detection accuracy, scalability and robustness of the models.

As (Slavom´ır Kajan *et al.,* 2020) have demonstrated through their studies, neural networks that can be pretrained and are extremely useful for DR classification. The paper explains how models such as VGG-16, ResNet-50 and Inception-v3 can be extremely effective if the models are pretrained. This can reduce the training time of the training process and increase the performance of the metrics. By extending this work, in another study, (Al-Kamachy, Hassanpour and Choupani, 2024) investigated the multi-class classification of DR stages. Then, they showed that such fine-tuned models, Inception V3 in this case, can provide higher accuracy in the classification of DR into different levels of severity. These findings affirm the use of pre-trained models as a starting point for the exploration of DR research, preserving moderate efficiency and accuracy as propounded by this study.

Generalized models are limited to their approach, where custom CNN architectures play a very important role in solving those problems. To detect microaneurysms and hemorrhages, Subramanian et al. (2023) designed a CNN architecture which could achieve high sensitivity, a critical metric for medical diagnostics. By doing this, Ghosh and Chatterjee (2023)(Ghosh and Chatterjee, 2023) further advanced this and integrated feature attention mechanisms in the VGG19 architecture. Besides improving discriminative abilities, this approach increased interpretability, which is crucial for clinical applications. These custom improvements ensure that models to optimize for exactly the tasks they are supposed to do such as concentrating on fine detail in the retina related to DR, since usually deep learning techniques are memory and computational bound. Including these methods in the current work can be used to improve the performance of CNN architectures for detecting subtle retinal abnormalities.

Retinal image segmentation improves classification accuracy by bringing out the relevant features through isolation. The hybrid methodology proposed in this research demonstrates how segmentation techniques can be combined with classification models. This method is demonstrated by (Yasashvini *et al.*, 2022)who studied hybrid architecture by combining CNNs with DenseNet and ResNet architectures, this improves the feature extraction and model performance by better utilizing the depth and the connectivity. By integrating these approaches with the research, it may help improve both the granularity and accuracy of DR detection.(Yasashvini *et al.*, 2022) One such another notable work is by (Mahilnan *et al.*, 2023) in which they developed Hybrid CNN SVM model. CNN and SVM were used together for feature extraction and classification respectively and the resulted system achieved better accuracy over standalone models.

CNN training depends on the ability to ensure quality of input data during data preprocessing. (Suedumrong *et al.*, 2024) showed that techniques (e.g., background removal and data augmentation) can significantly increase the robustness and the generalization of model. The authors found that their study achieved 90.6% validation accuracy, which highlights the need for high quality datasets. Rotation, scaling, and flipping are augmentation techniques that add diversity to dataset and drive model to learn invariant features as well as prevent overfitting. Similarly in this research, applying similar preprocessing pipelines will ensure we train the models on high quality data which would in turn lead to better generalization and performance.

For real time DR detection, deep learning models need to be efficient. A systematic review about computationally efficient models, especially vision transformers, that improved the accuracy at the cost of computations low was conducted by (Haq *et al.*, 2024). With these transformers, attention mechanisms fit seamlessly into CNNs so that we only process the most relevant features. Standard CNN architectures were reliable to use in binary classification tasks that led to high accuracy when trained on small datasets as mentioned by (Baba and Bala, 2022), in their work Detection of Diabetic Retinopathy with Retinal Images using CNN.

Furthermore, vision transformers propose a new mechanism to sacrifice some predictive power to improve computational efficiency, which lays the foundation for greater availability of AI in healthcare. Specifically, the developed elastic strain energy model gives roadmaps to the current research efforts that aim to balance the performance with the computational burden.

The training dynamics of CNNs are strongly affected by what optimizer is selected. Poojary and Pai (2024) studied fine tuned models with optimizers including SGD, Adam, and RMSProp and found SGD is the best optimizer minimizing learning curves in stabilizing training loss and improving convergence on ResNet50. In complement, Dogo et al. (2022) explored the effect of nine different optimization algorithms and concluded that Adam and Nadam performed the best for architectures that contain higher depth and width. In addition, these studies demonstrate that optimization choices can have a dramatic impact on training time and accuracy, and how understanding such performance dependencies can be critical to effective use of tailored approaches designed to meet particular dataset needs. Introducing insights from these studies into this research guarantees optimal training configurations leading to minimized convergence time and improved model accuracy.(Ramaprasad Poojary and Akul Pai, 2019; Dogo, Afolabi and Twala, 2022)

Comparative studies are very illuminating in the comparative strengths and weaknesses of different approaches. In a second work, (Haq *et al.,* 2024) provided a taxonomy for lightweight and accurate models using systematic comparisons of CNNs and vision transformers. Like the previous two, (Ramaprasad Poojary and Akul Pai, 2019) assessed the impacts of fine-tune and transfer learning, while outlining a framework to evaluate models by generalizing across diverse datasets. Comparison of these provides the need for adaptable methodologies, which are in direct support of the development of scalable solutions in this research. This research aims to create benchmarks for evaluating CNN and hybrid models in real world diagnostic tasks by synthesizing findings from these comparative frameworks.

3 Research Methodology

The methodology employed for the purpose of this project consists of five steps. The process starts with data gathering where images of eyes affected with diabetic retinopathy are collected and categorised as different stages of severity. The next steps include the preprocessing of the categorized images for better processing which leads to data transformation and then to data modelling where the three named "ResNet-18", "MobileNetV2", "EfficientNetB0" are trained and results for the training and modelling of the algorithms are evaluated and compared.

The first step mainly focuses on data collection and data loading. Diabetic retinopathy image dataset, a dataset containing retinal images of different severity used in the research is uploaded into Google drive. This dataset was imported from Google Drive into the Google Colaboratory for pre-processing and model training as shown in the fig 1. The data set is structured into directories according to severity levels of diabetic retinopathy (e.g., no DR named as 0, mild named as 1, moderate named as 2, severe named as 3, and proliferative DR named as 4). Each directory contains retinal images of respective severity.

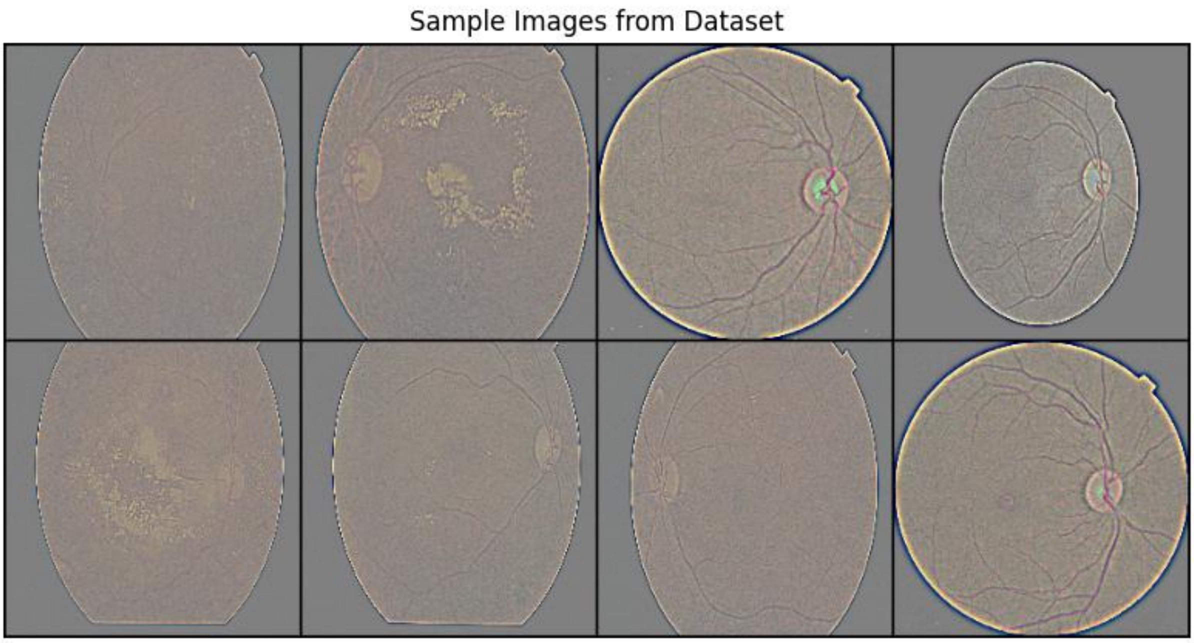

Fig 1. Sample dataset

The second step, data Pre-processing involves image reclassification to ensure a clean image dataset and prepare for transformation and model training. The images that were misclassified are reclassified and the information that is irrelevant and duplicated was removed to maintain dataset quality. This also process leads to an increase in the quality of the dataset. The result of data preprocessing for the diabetic retinopathy image dataset contains 1500 (approx.) images of the retina each as affected and non-affected retina and saved in JPG format ready for further processing and transformation.

The third step in the process is the perform data transformations this involves reducing noise in the picture and emphasize features that are responsible for accurate classification and applying techniques such as colour normalization and resizing of images to the standard dimensions in order to attain a robust training. Colour Normalization is applied on the image dataset in order to ensure that consistency across images is maintained, this helps in reducing variations caused by different lighting and reduces variety introduced by using different capturing devices. This step helps the algorithm to extract the actual differentiating factors and improves the feature extraction process during training. The resizing of images was performed next. The images used for training were of size 224X224 pixels, a standard input size for a deep learning model like Resnet, MobileNet and EfficientNet. Techniques such as image rotation, image flipping and cropping were also applied to introduce variability in the dataset. This step improved generalization by simulating different image perspectives.

The fourth steps involve model training in order to classify the retinal images. The paper compares three different model architectures for the purpose of classification. Each model was implemented using contains its own advantages and disadvantages, The dataset is divided into training, testing and validation sets. The optimizer chosen for the purpose is the stochastic gradient descent optimizer and Adam optimizer. The optimizers are also been compared to get a better understanding of under what model how the model performance changes. Hyperparameter tuning such as changing learning rates also highly affects the actual accuracy of the algorithm working.

The fifth step involves evaluating trained models and quantifying the results of each of the deep-learning classification models using accuracy and loss. The comparison of the models and two optimizers is visualized, and the optimal algorithm is selected from the research.

4 Design Specification

The deep learning framework architecture compares deep learning image classification models and optimizers with each other. The project aims to build a scalable and interpretable model that classifies retinal images into predefined categories, such as No DR, Mild, Moderate, Severe, and Proliferative DR.

The project relies on an end-to-end pipeline that takes retinal images as input, preprocesses them, and outputs classification results. Images are standardized through resizing and normalization to streamline processing. Noise is reduced using Gaussian filtering, and data augmentation techniques such as flipping, rotation, and brightness adjustments are employed to enhance model robustness and address dataset imbalance issues.

To this end, the solution utilizes advanced convolutional neural network (CNN) architectures such as ResNet and EfficientNet, in conjunction with transfer learning. They used pretrained models as feature extractors, fine tuned on target dataset for diabetic retinopathy classification. A fully connected layer followed by a softmax function is used to do the classification, and the output is a probabilistic result for each category. On multi class classification, cross entropy loss is to train the model and Adam optimizer and a learning rate scheduler for a stable and efficient convergence.

The primary programming language is Python and the framework for the development of model is PyTorch. Something like Torchvision is used to handle data processing tasks, and Matplotlib for visualization. Training on our models occurs on platforms such as Google Colab, with GPUs for quick computation and faster development necessary.

At the project level, a thorough evaluation is conducted, together with monitoring accuracy as the primary performance metric to maintain model reliability. Training and validation are done on different datasets; model generalization capability is evaluated on a separate test dataset with cross-validation. Data imbalance risks are addressed using strategies, such as weighted loss functions and data augmentation.

The emphasis is on devising a reliable, robust and efficient system for detecting diabetic retinopathy. The deliverables include a trained model, the documented codebase, and the deployment and core usage guidelines. The project addresses key challenges and integrates state of the art technologies to produce a clinically valuable tool for diagnosis of diabetic retinopathy.

5 Implementation

This project implemented (developed and compared) three advanced deep learning models, ResNet-18, MobileNet-V2, and EfficientNet-B0 to detect diabetic retinopathy in the retinal images. This task was structured from data preparation to the model training then evaluation and then finally comparing models for selecting the most effective.

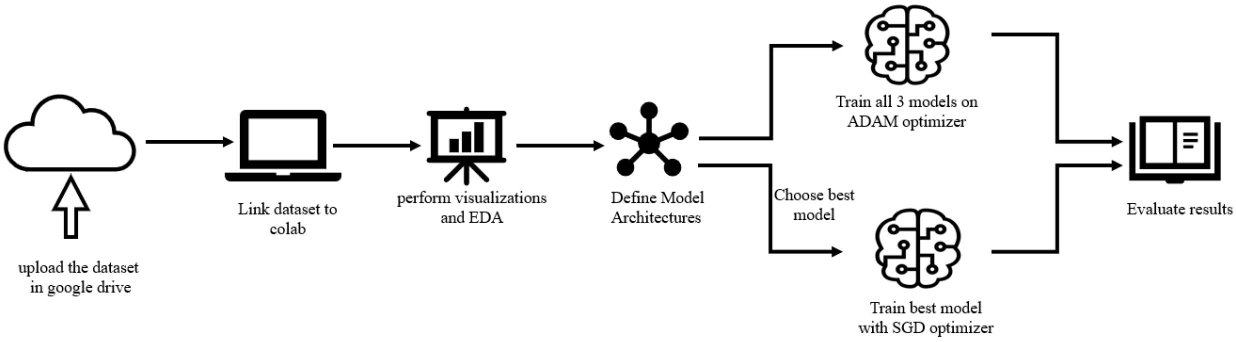

Fig 2. Implementation Workflow

As shown in the workflow diagram in Fig 1. The process starts with uploading the dataset into Google Drive so that they are easy to integrate with the Google Colab environment. The dataset was accessed in Colab and an extensive EDA was performed and various visualizations of the data to understand the variety of data in order to get some insights on how the dataset is structured, the distribution of images and potential challenges such as class imbalance and quality of images. After EDA, some augmentation techniques like random resizing, cropping, flipping and normalization were done on the data. The transformations performed with a goal to improve model robustness and increase in variability in training data in general lead to a decrease in overfitting and improved performance in real-world unseen data. The dataset is split into training, validation, and testing subsets in an 80:10:10 ratio for the models to be trained and evaluated on.

In this study, three model architectures with different strengths are used. A lightweight residual network, ResNet-18, with the ability to resolve the vanishing gradient problem of residual networks by utilizing residual connections, was chosen. MobileNet-V2 was included for its computational efficiency and the granted comfort of its running on resource-constrained environments. Finally, we select EfficientNet B0 as one of the state-of-the-art models with high performance at the same time is computationally efficient. Three models were pre-trained on ImageNet, followed by transfer learning to the diabetic retinopathy detection task. The final layers of these models were replaced by custom fully connected layers matching target number of classes so as to enable these models to make the distinction between different categories of severity of diabetic retinopathy.

The training in the project is divided as a two-stage training process. All three models were initially trained using Adam optimizer which is known for adaptive learning. This step allowed a baseline comparison of the models. Using the results, we selected the best performing model, EfficientNet, for use in the second stage of fine tuning. The model was trained here using the Stochastic Gradient Descent (SGD) with momentum optimizer, which just helps the convergence to avoid local minima. A learning rate scheduler was used in both stages of the training to dynamically adjust the learning rate and obtain optimal convergence. All the model architectures were trained over 200 epochs first with Adam and the best model is trained again with SGD with a batch size of 64 and their performances were monitored through metrics like accuracy and loss. These metrics helped to identify that how well the models are able to classify images and how the models perform under class imbalances scenarios. Finally, the final validation accuracies for ResNet, MobileNet as well as EfficientNet were calculated.

Python and PyTorch were used for the whole implementation, as they are flexible and scalable deep learning frameworks. We used the data transformation and loading from pre-trained models of PyTorch's torchvision library. Matplotlib was used to visualise the training trends and the drawing of evaluation results as clear and informative plots. The project was carried out in Google Colab, which is equipped with an NVIDIA-SMI 535.104.05 T4 GPU with 15 GB RAM for training purposes, leading to a substantial speedup of the training time. The data was saved in the Google Drive, which was perfectly integrated to the Colab environment.

6 Results and Evaluation

This section provides a comprehensive analysis of the experimental outcomes obtained by evaluating three prominent convolutional neural network (CNN) architectures: EfficientNet, MobileNet and ResNet. The models were trained on a dataset with five output classes for 200 epochs, and their performance was assessed using key metrics: Train loss, validation loss and accuracy. We report results that demonstrate the comparative strengths and weaknesses of these architectures in terms of accuracy, convergence pattern, and generalization capability.

6.1 Performance

A comparison of the three models was based on the best validation accuracy during training. Find Model Best Validation Accuracy compare the three models was the best validation accuracy achieved during training. The results, summarized in the table below, indicate that EfficientNet outperformed the other models in terms of accuracy, while maintaining a comparable level of loss:

Model	Best Validation Accuracy	Stepper function
ResNet	71.31%	Not present
MobileNet	77.22%	Not present
EfficientNet	82.08%	Present

Table 1. Model performance

MobileNet (74.49%), ResNet (71.31%), and EfficientNet (82.08%) performances were measured on the validation accuracy. We observed that ResNet showed the lowest accuracy among the three models and MobileNet had a strong balance between accuracy and computational efficiency. Interestingly, ResNet's training stability was remarkable, with sharp optimization behaviour across the 200 epochs, with no large swings.

All these results demonstrate the architectural benefits of EfficientNet, that is compound scaling for depth, width, and resolution enables better performance. Using MobileNet’s lightweight architecture, we were able to achieve competitive accuracy and deploy it on edge devices. ResNet had lower accuracy than others, but due to its stability and robustness, it is a good choice for applications that require a stable optimization.

6.2 Training and Validation Loss

For each of the three models, I trained on the dataset, and I saw the training loss decrease steadily over the epochs for each of the models. This implies that we are optimizing properly and getting some good learning of features in our dataset. Notably, the validation loss differed significantly among the models: The validation loss of MobileNet attain was 0.67. The validation loss was higher over ResNet with 0.76. The validation loss of 0.47 was achieved by EfficientNet. (These are the absolute lowest loss attained throughout the training for all models)

Validation loss indicates the amount of generalization that a model can achieve. EfficientNet demonstrated better generalization than the other networks as it has validated lower validation loss, therefore managed to minimize overfitting and output prediction for the validation set. Although MobileNet's validation loss was fairly low, its performance should still be optimized, especially in hyperparameter fine tuning. On the other hand, the model trained on ResNet Architecture fails to achieve higher generalization on the validation set.

Further, the consistent reduction in training loss in future epochs for all three models admits the data and training setup as being useful for optimization. However, it becomes apparent that EfficientNet’s validation loss is the lowest of all, at the cost of a relatively low accuracy, but it does show potential for application in settings where the generalization capabilities are to be strong.

6.3 Convergence pattern

The efficiency of a model is judged by the rate of convergence. We found that EfficientNet and MobileNet converged faster than ResNet. At Epoch 100, EfficientNet and MobileNet had become mostly steady with lower loss values, whereas ResNet fluctuated but only in smaller margins. These results suggest that EfficientNet and MobileNet have architectures with faster convergence and are suitable for tasks requiring the need of effective learning with a reasonable number of epochs.

This can be attributed to EfficientNet's superior convergence pattern to its compound scaling which achieves a good trade-off between network depth, width, and resolution. The balance in this model enables it to extract the relevant features far more efficiently and therefore comes to be optimized much quicker. MobileNet also had efficient convergence on account of its lightweight blueprint and streamlined caricature showing pleasing to restricted surroundings.

On the other hand, deeper architecture of ResNet resulted in slower convergence. This design guarantees the stability and the robustness, but in order to properly fine tune the model parameters, it needs more epochs. Although ResNet converges slower than several other methods, its trend in training loss shows a consistent decline of optimization and hence can be a reliable method for specific situations.

6.4 Model-wise detailed results

A closer examination of the individual models provides deeper insights into their performance characteristics:

6.4.1 ResNet

The other models converged faster than ResNet. At epoch 100, this was still fluctuating, with a clear separation from training and validation accuracy shown in Fig 3. that shows that it was moderately overfitting. Although somewhat slower in its learning curve, ResNet proved to train with amazing stability, where training loss steadily and smoothly declined. This shows that its good optimization process is robust which enables it to learn reliably through epochs.

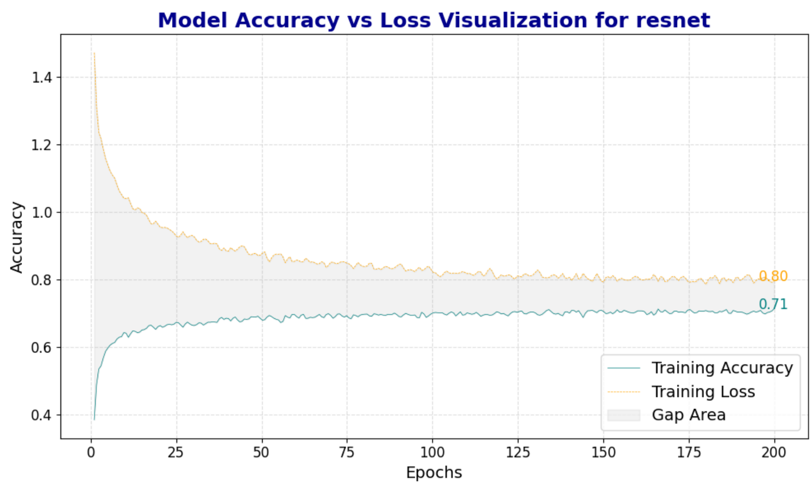

Fig 3. ResNet model accuracy vs loss

The accuracy is smaller and validation loss is higher in ResNet that means its not able to generalize on unseen data as good as Inception V3. Yet, it has stable and consistent performance so it may be a good choice for accurate tasks that need interpretability and robustness.

6.4.2 MobileNet

By epoch 50 MobileNet received a jump in validation accuracy showing really fast learning of the features. As a result, this is a good candidate for applications where training time or computational resources are limited. MobileNet converged faster than the ResNet can be seen in Fig 4., however, it showed a slightly higher validation loss compared to EfficientNet. These results indicate that even greater generalization performance could be achieved for MobileNet by fine tuning hyperparameters, e.g. learning rate and regularization techniques.

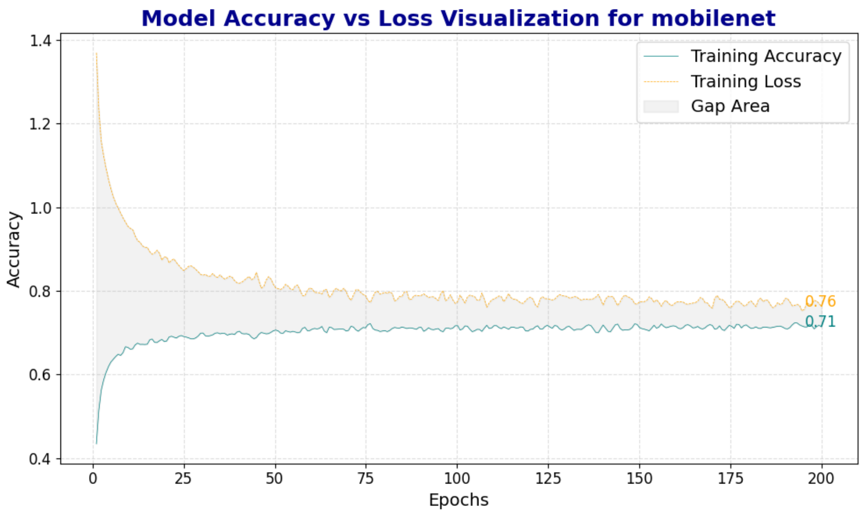

Fig 4. MobileNet model accuracy vs loss

MobileNet is lightweight in architecture and has competitive accuracy, making it a great choice for deployment in edge devices and constraint resource environments. The trade off between efficiency and performance is practical for real applications.

6.4.3 EfficientNet

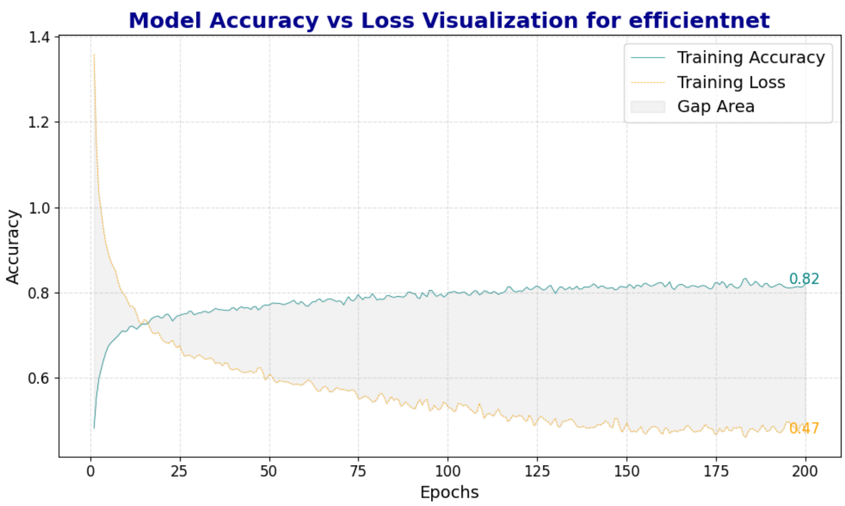

Fig 5. EfficientNet model accuracy vs loss

In fact, across all metrics, EfficientNet performed consistently better than ResNet and MobileNet. At epoch 70, the other models had surpassed it with respect to both loss and accuracy, having the lowest validation loss (0.47) and highest validation accuracy (82%).This is largely why EfficientNet is successful: its compound scaling technique enables a strong ability to capture features efficiently. Thanks to this architectural advantage, EfficientNet achieves both high accuracy and small computation. Its performance suggests that EfficientNet is a good option for high accuracy and generalization robust tasks. Overall, its versatility makes it the greatest model in this research as it can specialized to outperform MobileNet and ResNet on every key metric.

6.5 Adam Vs SGD

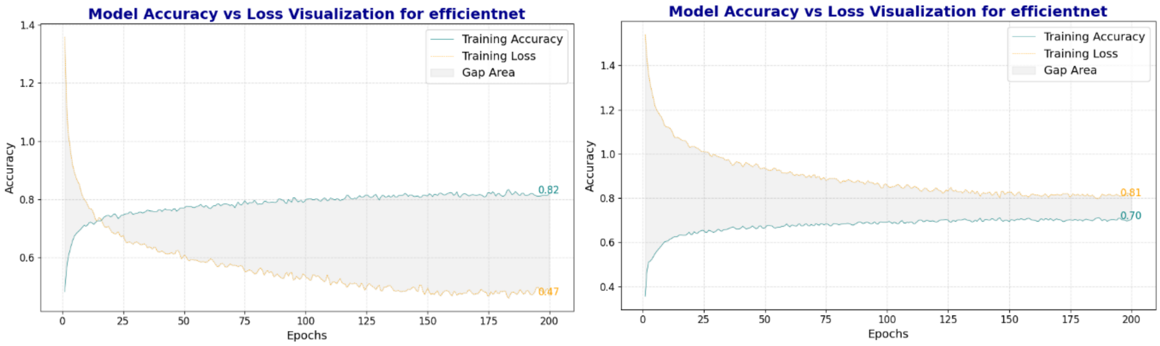

Fig 6 EfficientNet Model Accuracy vs Loss graph for Adam(left) and SGD(right)

As clearly observed in the graphs in Fig 6 the EfficientNet model performs much better than with Adam optimizer as compared to that with SGD this is mainly because ADAM optimizer integrates adaptive learning rates and momentum, it better updates each parameter dynamically according to the gradient's history, and in general performs better than SGD for image classification, with a faster learning rate and a lower L2 Norm of the intermediate weights. The reason why ADAM converges faster on such landscapes and is more effective generally comes from its adaptability. In contrast to SGD which needs to be carefully tuned by a fixed learning rate, it works out of the box and is more robust to hyperparameter settings. Moreover, ADAM's momentum makes updates smoother and aids greatly in convergence in deep neural networks. For the given dataset ADAM performs very well in terms of being faster and more stable training performance.

7 Discussion

The experimental findings provide valuable information about performance characteristics and architectural choices of ResNet, MobileNet, and EfficientNet. The results of this study highlight the central role that architecture plays in setting model performance and thus, the applicability across different machine learning tasks. The tailoring employed in each model showcases the trade offs in the design philosophies of each by showcasing the unique strengths of each model, as well as their abilities in accuracy, efficiency, and scalability. In this section we interpret the results in detail and discuss their implications for machine learning more generally.

7.1 Architectural Efficiency

The results confirm the influence of architectural efficiency in defining model performance. The compound scaling technique that optimizes depth, width and resolution produced the top-performing model — EfficientNet. In line with findings in existing literature, EfficientNet's capability to achieve the highest accuracy using relatively few computational resources is well suited for high-performance machine learning tasks.

The results for MobileNet, specifically for lightweight applications, showed their performance was competitive in terms of accuracy and efficiency, making them a good candidate for edge computing or to be used in resource-constrained environments. MobileNet achieves a great trade-off between computational cost and predictive power, preserving a level of performance that meets practical standards of performance but ensuring that its computational cost is minimized. Because of this, MobileNet is a great candidate for real-world applications that require compactness and adaptability.

Although the accuracy achieved by ResNet on this study is not the highest, being safe and robust to optimization makes it still a benchmark model. It possesses a guaranteed stable convergence and learning dynamics making it applicable in settings in which architectural simplicity and interpretability is critical. ResNet's performance is consistent because of its residual learning philosophy for deep networks where residual learning alleviates the vanishing gradient problem. These represent a robust tool to continue to be relevant in reliability demanding optimization tasks over raw accuracy.

7.2 Overfitting Analysis

Extremely capable of avoiding overfit, EfficientNet was able to keep this difference between training and validation loss to a minimum As the EfficientNet Architecture allows features such as dropout and batch normalization this boosts the generalization and prevents the model from overfitting. Also, the fact that EfficientNet cannot get overfit derives its ability of getting high accuracy and good generalization which is important when you want to deploy the model into some real world applications

On the other hand, ResNet had mild overfitting, as its training and validation accuracy gap was quite large. This implies that, though ResNet’s architecture is well understood for its stability, improvements on generalization through applying additional amount of regularization or data augmentation strategies should be further explored. In particular, MobileNet was lightweight enough so as not to underfit and to be able to avoid the problems of overfitting to the training data, but still competitive in validation performance.

7.3 Scalability

With EfficientNet scaling across varying computational constraints, it has been observed that it is a generalizable solution to large-scale image classification problems. Its architecture smoothly adapts itself to varying resource availability while providing same performance across different deployment scenarios.

MobileNet is very lightweight but retains a high accuracy, thus it is quite suitable for mobile or embedded applications. Our compact architecture makes it possible to deploy efficiently on edge devices, responding to the increasing need for AI solutions in IoT and other resource-constrained environments. MobileNet's achievable tradeoff between accuracy and computational cost makes it suitable for practical scenarios where accuracy is just as important as the cost of computation.

In applications with the need for interpretability and stability, simpler and more robust ResNet survives. With its lower complexity it is appropriate for high complexity datasets or situations where computational resources are constrained. Design philosophy of ResNet guarantees that it will be the reliable baseline performance for the various machine learning tasks.

7.4 Training Analysis

Performance characteristics of the three models are also understood through their training dynamics. It has been observed through the experiments that ResNet, with its deeper architecture, converged slower than our baseline model, reflecting the slower optimization process. However, this slower convergence can be a bottleneck in time-sensitive applications and for this, resnet has systematically and steadily learnt the problem, and that can be a plus point in resnet. The reason it's not slow is because it rarely returns bad results. That makes it good for tasks where stability is more important than speed.

In contrast, MobileNet had reached early stabilization before the training and had already made significant improvements in the validation accuracy by epoch 50. Such learning dynamics are efficient due to the streamlined design of this scheme, which is tailored for resource-efficient applications. The stabilization of MobileNet on fast and consistent performance proves its applicability for real-world conditions where rapid training is demanded.

It resulted in an EfficientNet with a balanced trade-off between complexity and easy convergence. Its architectural efficiency is based on its ability to achieve rapid convergence while maintaining high accuracy. EfficientNet outperformed the other models by epoch 70, proving its ability to learn and choose the important features and optimize performance within a reasonable number of epochs. Such balance makes EfficientNet applicable in a broad variety of machine learning problems.

8 Conclusion and Future Work

In this study we highlighted how architectural choice plays a critical role in defining the performance and applicability of machine learning models. Results show that EfficientNet is more accurate, more efficient, and more generalized than MobileNet and ResNet, and can be a solution for tasks that need high accuracy but whose resources scale. Finally, as the relative balance of efficiency and predictive power, MobileNet is revealed to be a prime contender for resource constrained environments and ResNet for applications that privilege stability, simplicity and interpretability.

Finally, this study has broader implications of increasing trend towards compact and yet powerful architectures in machine learning. It justifies the new trend towards performance optimized models from EfficientNet's perspective, and MobileNet casts both an accuracy and a computational efficiency perspective can be beneficial to real world applications. The significance of these baseline architectures is demonstrated by ResNet's continued relevance in shaping future innovations, and the necessity for robust baseline architectures to leapfrog their successors.

The results of this study will help guide the architectural decisions made in future research and deployments of machine learning. In this work, we examine the interplay of accuracy, efficiency, and scalability in order to provide useful guidelines to overcome the various challenges that face the field of today, staying relevant, powerful, and capable of being used in a large variety of applications.

These findings are the first step towards additional optimization and improvement of the considered models. Model architectures need to be fine-tuned. Tuning different hyperparameter configs such as learning rates, batch sizes and regularizations has a huge impact; to name a few. Further improving model robustness and adaptability can be achieved with customization of data augmentation strategies specific to the dataset’s specificities.

We examine model compression techniques with an aim of making these architectures more efficient and feasible for deployment in environments where resources are limited. Pruning, quantization, and knowledge distillation among others methods can reduce the computational demands without impeding prediction accuracy. EfficientNet and MobileNet are optimized for edge and IoT devices with special relevance to these approaches.

Another compelling direction for future work on hybrid architectures is proposed. We show that averaging over predictions from ResNet, MobileNet, or EfficientNet ensemble frameworks (i.e., ResNet50, MobileNetV3 large, and EfficientNet B4) or harnessing Neural Architecture Search (NAS) for innovative model designs based on these three CNN blocks can yield designs with excellent performance in a task-specific manner. By combining the two optimization strategies, the models can strike a better trade-off between computational expense and predictive power, enabling their use in a broader array of domains and computational scenarios. The capabilities of these advancements will facilitate significant boundaries and push for better model performance and adaptability, so that they may be applied to a broader set of real-world applications.

References

Al-Kamachy, I., Hassanpour, R. and Choupani, R. (2024) ‘Classification of Diabetic Retinopathy using Pre-Trained Deep Learning Models’. Available at: http://arxiv.org/abs/2403.19905.

Baba, S.M. and Bala, I. (2022) 'Detection of Diabetic Retinopathy with Retinal Images using CNN', in Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022. Institute of Electrical and Electronics Engineers Inc., pp. 1074–1080. Available at: https://doi.org/10.1109/ICICCS53718.2022.9788368.

Dogo, E.M., Afolabi, O.J. and Twala, B. (2022) 'On the Relative Impact of Optimizers on Convolutional Neural Networks with Varying Depth and Width for Image Classification', Applied Sciences (Switzerland), 12(23). Available at: https://doi.org/10.3390/app122311976.

Ghosh, S. and Chatterjee, A. (2023) 'Transfer-Ensemble Learning Based Deep Convolutional Neural Networks for Diabetic Retinopathy Classification', in 2023 3rd International Conference on Advancement in Electronics and Communication Engineering, AECE 2023. Institute of Electrical and Electronics Engineers Inc., pp. 489–493. Available at: https://doi.org/10.1109/AECE59614.2023.10428233.

Haq, N.U. et al. (2024) 'Computationally efficient deep learning models for diabetic retinopathy detection: a systematic literature review', Artificial Intelligence Review, 57(11). Available at: https://doi.org/10.1007/s10462-024-10942-9.

Mahilnan, V. et al. (2023) 'Diabetic Retinopathy Detection Using Hybrid CNN - SVM Model', in International Conference on Sustainable Communication Networks and Application, ICSCNA 2023 - Proceedings. Institute of Electrical and Electronics Engineers Inc., pp. 1575–1580. Available at: https://doi.org/10.1109/ICSCNA58489.2023.10370142.

Ramaprasad Poojary and Akul Pai (2019) 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA). IEEE.

Slavom´ır Kajan et al. (2020) Detection of Diabetic Retinopathy Using Pretrained Deep Neural Networks. IEEE.

Suedumrong, C. et al. (2024) 'Diabetic Retinopathy Detection Using Convolutional Neural Networks with Background Removal, and Data Augmentation', Applied Sciences (Switzerland), 14(19). Available at: https://doi.org/10.3390/app14198823.

Yasashvini, R. et al. (2022) 'Diabetic Retinopathy Classification Using CNN and Hybrid Deep Convolutional Neural Networks', Symmetry, 14(9). Available at: https://doi.org/10.3390/sym14091932.