ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 05, May 2025

Deep Learning-Assisted Screening of Macular Degeneration in Retinal OCT Scans

[1] Mrs. Babeetha S, [2] Shubhra Pathak

[1] Assistant Professor, Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, India
[2] Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, India
Emails ID: [1] babeeths@srmist.edu.in, [2] sr4921@srmist.edu.in

***Abstract*— One of the main causes of vision loss in the elderly is age-related macular degeneration (AMD), a degenerative retinal disease. This work puts forward a deep learning-based method using EfficientNet-B5 for automatic AMD categorization from retinal fundus images. The model uses transfer learning and fine-tuning to achieve accuracy while keeping computation speed. An ablation study was done to see how tweaking different layers affects results, in addition to being benchmarked against standard diagnostic methods. Tests show better classification results; this method can become a scalable and trustworthy solution for early AMD identification, helping ophthalmologists with on time diagnosis and treatment planning.**

***Index Terms*— Grad-CAM, OCT, SVM, EfficientNet B5, deep learning, macular degeneration, medical imaging.**

I. INTRODUCTION

All around the world, countless people deal with age-related macular degeneration (AMD). It's a condition that gradually worsens vision. The best way to spot AMD is using OCT imaging, but when it comes to manual checks, different people might see it differently. Sure, deep learning offers ways to do it automatically, but lots of its models we can't test in real life or understand easily. For finding details and making types, this work joins EfficientNet B5 and SVM, ensuring high precision and usefulness.

II. MOTIVATION

The major reason that motivated this work is the growing scale and complexity of deep learning models, particularly in computer vision applications. Memory consumption, processing time, and time taken to train a deep neural network make it very cumbersome for researchers and practitioners. Optimizing training efficiency helps improve accessibility, reduce costs, and accelerate the adoption of AI models. The study seeks to speed up training times without compromising model accuracy so that deep learning can become more pragmatic in the context of real-world applications.

III. INNOVATION

Herein, we present the following key innovations in this study:

- Efficient Data Loading and Preprocessing: Loading Data as a Pipeline instead of in bulk at once drastically lowers I/O backpressure.
- Mixed Precision Training: Through mixed precision floating-point arithmetic, we can speed up computations without any loss of efficiency.
- Learning Rate Scheduling: The adaptive learning rate scheduler follows the number of epochs to prevent unnecessary calculations.
- Model Parallelism and Hardware Optimization: By using GPU acceleration, controlling memory overhead, and modifying batch sizes, the optimal use of resources is guaranteed.

When combined, these methods reduce training durations without compromising the model's ability to generalize well to new data. Conventional ocular diagnosis techniques mostly rely on subjective and time-consuming manual feature extraction and expert interpretation. Although diagnosis has been automated using traditional machine learning models that leverage handmade variables like texture and edge detection, these models have trouble generalizing across a variety of datasets. These drawbacks are overcome by deep learning-based models, such as the suggested EfficientNet B5, which improve diagnosis accuracy by directly learning hierarchical feature representations from raw photos. Deep learning has made it possible for automated, quick, and extremely accurate illness identification in ophthalmology, minimizing inter-observer variability and lowering reliance on human experience.

IV. LITERATURE REVIEW

The effectiveness of CNNs in image-based illness diagnosis is demonstrated by a careful review of prior studies. Important conclusions include:

- De Fauw et al. (2018) achieved human-expert level accuracy by using CNN-based segmentation and classification on the Moorfields Eye Hospital OCT dataset. For training, the method requires large and labelled datasets.

- Kermany et al. (2018) used InceptionV3 to apply transfer learning on a private dataset with 108,312 images, and they were able to distinguish between Normal, CNV, and Drusen cases with 96.6% accuracy. Despite its effectiveness, the study's lack of interpretability limited its clinical applicability.
- Lee et al. (2019) presented EfficientNet for feature extraction in combination with SVM classification using the Duke OCT dataset. Despite its high processing power requirements, their output outperformed traditional CNN models in terms of accuracy.
- Liu et al. (2022) presented a hybrid deep learning technique that integrated ResNet and SVM using the Kaggle OCT dataset, achieving 97.2% accuracy. Although the model showed promising findings, it was not clinically validated. Despite these advancements, problems like as high computing costs, interpretability problems, poor generalization across datasets, and the need for optimized hyperparameter tuning remain poorly understood, which is what spurred this investigation.

V. RESEARCH GAP

In order to improve accuracy, most current research focused on enhancing the model architecture. However, there hasn't been much work done to optimize training pipelines to reduce computational cost while maintaining similar performance.

- Previous studies have examined network pruning and quantization; however, these techniques typically result in decreased model accuracy.
- Few studies use a combination of GPU utilization, mixed precision training, and efficient data loading to speed up training without changing the model architecture.
  - A systematic assessment of training optimization across various dataset sizes and resolutions is lacking.

This work closes these gaps by using a thorough training optimization strategy and investigating its impact on accuracy and efficiency empirically. model customized for the classification problem of OCT.

VI. ACRONYMS AND ABBREVIATIONS

- OCT: Optical Coherence Tomography
- CNN: Convolutional Neural Network
- SVM: Support Vector Machine
- AMD: Age-related Macular Degeneration
- AUC: Area Under the Curve
- Grad-CAM: Gradient-weighted Class Activation Mapping

VII. METHODOLOGY

Because of its compound scaling strategy, which strikes the ideal balance between breadth, depth, and resolution to optimise accuracy while preserving computing efficiency, EfficientNet B5 was chosen over competing designs. EfficientNet offers better classification accuracy with fewer parameters than traditional CNNs. The model is especially well-suited for medical image analysis tasks because of its dynamic architectural adaptation, which enables it to attain state-of-the-art performance with less training time and hardware requirements.

A. Preprocessing and the dataset

- *Dataset: Openly accessible OCT* pictures *divided into Normal, CNV, and Drusen categories.*
- *Preprocessing:*
- *Image scaling to* 456x456 *pixels is part of the preprocessing step.*
- *To enhance* generalization, *use normalization and augmentation techniques (rotation, zoom, and horizontal flipping).*

B. Architecture of the Model

- *Feature Extraction:* EfficientNet B5 is pre-trained on ImageNet for feature extraction.

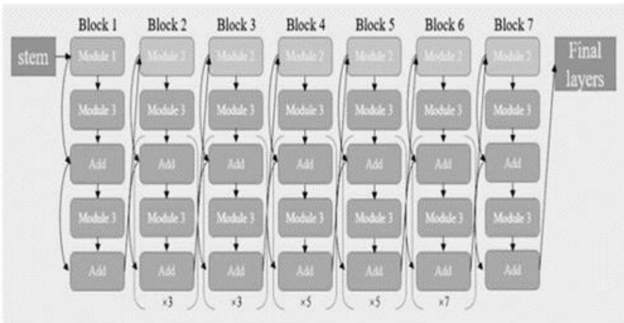

Figure 1. Architecture of the EfficientNet B5

- *Classification:* SVM was trained using features that were extracted.
- *Fine-tuning:* To enhance performance, EfficientNet layers are gradually unfrozen.
- *Grad-CAM:* Used to show important areas that affect forecasts.

C. C. Measures of Performance

- F1-score, recall, accuracy, and precision.
- *Confusion Matrix:* Assesses performance in categorization.
- *ROC-AUC Curve:* Evaluates the efficacy of multi-class categorization. evaluates the efficacy of multi-class categorization.

VIII. FIGURES

A. Architecture Diagram

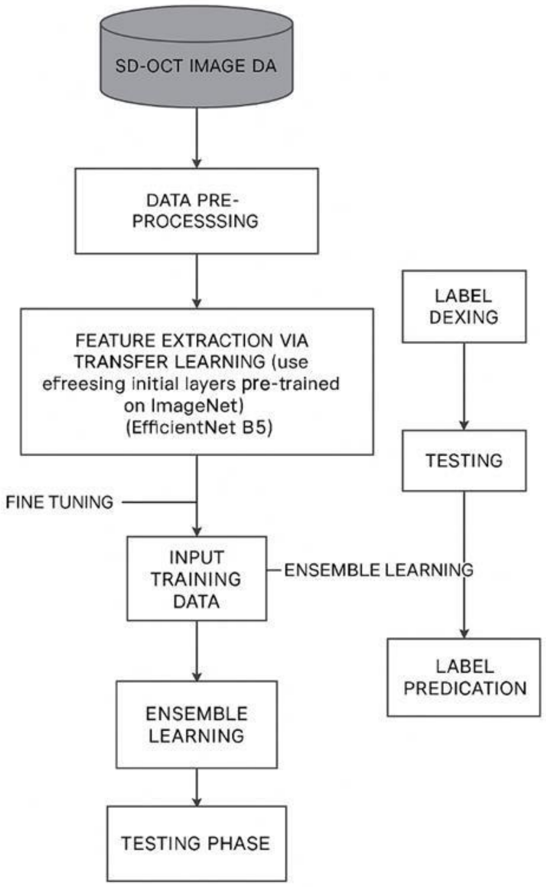

Figure 2. Architecture Diagram

- Retinal OCT scan pictures are stored in the SD-OCT Image Database, which is where the system begins.
- To enhance model performance, the photos go through preprocessing, which might involve scaling, normalization, noise reduction, and augmentation.
- The feature extractor is the EfficientNet B5 model, which has already been trained on ImageNet.
- The foundation layers are initially frozen, which means that training does not change them. To adjust the model for the particular OCT classification job, a few top layers are gradually unfrozen for fine-tuning.
- The model's accuracy, precision, and recall are evaluated. A classification model is trained using the retrieved features as input. The use of ensemble learning, which combines several models to increase classification accuracy, is implemented.
- The collected characteristics are subjected to an SVM (Support Vector Machine) classifier rather than the deep learning model itself for classification.
- The processing of a test picture is identical to that of training images. The category of the provided OCT picture is predicted by the SVM classifier. The picture is classified by the model into one of the following groups:
  - Normal
  - Drusen/Dry AMD (without exudative alterations)
  - CNV (Choroidal Neovascularization)/Wet AMD (with exudative changes).

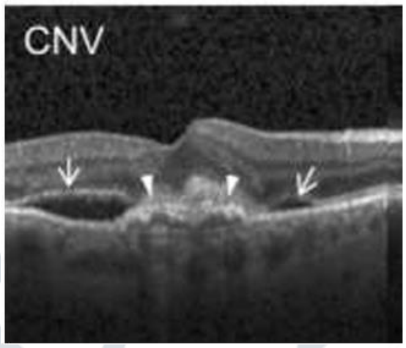

Figure 3. Wet AMD with neavascularization

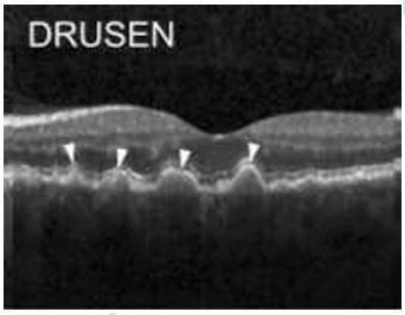

Figure 4. Dry AMD with Drusen

B. Use-Case Diagram

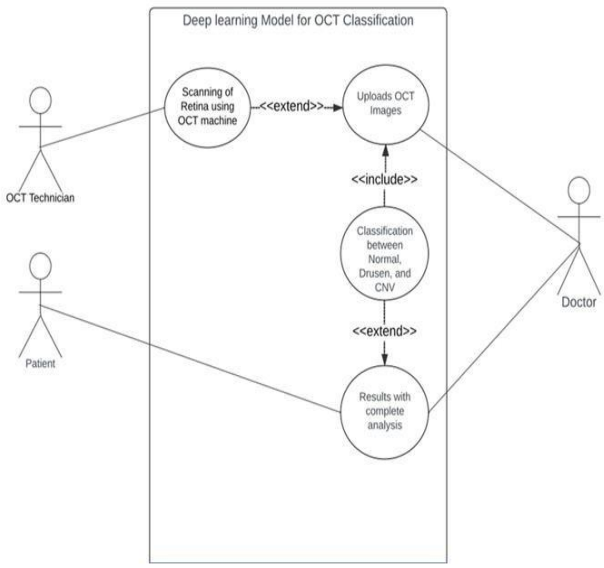

Figure 5. Use-Case Diagram

- The system starts with an SD-OCT Image Database containing retinal OCT scan images.
- The images undergo preprocessing, which may

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 05, May 2025

include resizing, normalization, noise removal, and augmentation to improve model performance.

- EfficientNet B5, pre-trained on ImageNet, extracts feature from preprocessed images.
- To maximize model performance, the hyperparameters listed below were adjusted:
  - Batch Size: 16 is chosen to balance training speed and memory utilization.
  - Learning Rate: With an exponential decay schedule, it is initially set at 0.001.
  - Dropout Rate: Set at 0.3 in order to avoid overfitting.
  - Optimizer: The Adam optimizer was selected because of its capacity for adjustable learning rate.
  - Early Stopping: To prevent overfitting, this technique was used with a 5-epoch patience.
- To determine the ideal hyperparameter values and guarantee excellent generalization to unknown data, grid search and manual tweaking were used. The foundation layers are initially frozen, which means that training does not change them. To adjust the model for the particular OCT classification job, a few top layers are gradually unfrozen for fine-tuning.
- Which model performs better in terms of accuracy, precision, recall, and other metrics is determined in the assessment process.
- The use of ensemble learning, which combines several models to increase classification accuracy, is implemented.
- An SVM classifier processes extracted features for better classification accuracy. SVM helps increase accuracy and is useful for classifying high-dimensional picture features.
- The processing of a test picture is identical to that of training images classifying into:
  - Normal Wet AMD (with exudative alterations)
  - Choroidal Neovascularization (CNV)
  - Dry AMD (without exudative alterations) or Drusen.

C. Class Diagram

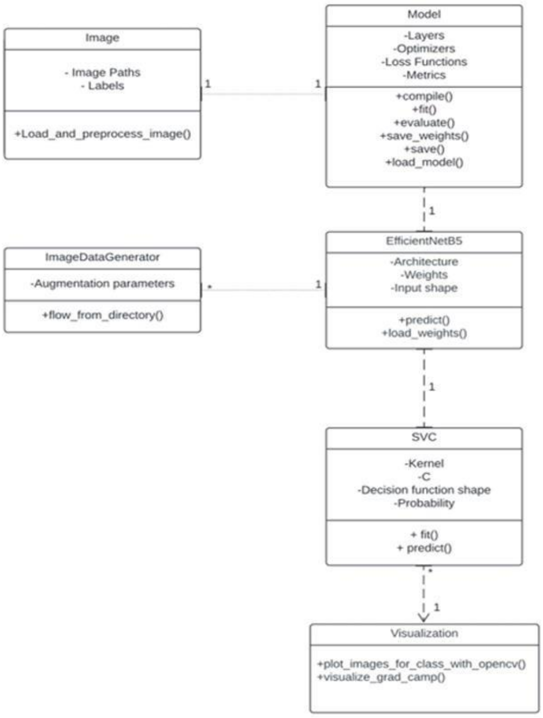

Figure 6. Use-Case Diagram

A. *Preparing and loading images:*

- Image paths and labels are loaded by the image class.
- Augments are used by ImageDataGenerator to produce a variety of training data.

B. *Extraction of Deep Learning Features:*

- Instead of using raw image pixels for classification, EfficientNetB5 uses the derived feature vectors to extract relevant features from pictures.

C. *SVM-based classification:*

- Using the retrieved characteristics, the SVC class assigns the photos to the appropriate categories.

D. *Visualization:*

- The visualization class aids in the analysis of feature significance and classification performance.

IX. EQUATIONS

A. Cross-Entropy Loss Function:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where is the true label and is the predicted probability.

B. Support Vector Machine Decision Function:

$$f(x) = w^T x + b$$

where w is the weight vector, is the input feature vector, and is the bias.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 05, May 2025

C. Calculating Accuracy:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

where FN stands for False Negatives, FP for False Positives, TN for True Negatives, and TP for True Positives.

X. EXPERIMENTAL RESULTS

A. Model Performance

A. Key Findings:

- The SVM classifier outperformed the softmax-based alternative, particularly in handling borderline and visually subtle cases.

Model	Accuracy	Precision	Recall	F1-score
CNN (Baseline)	89.5%	87.8%	88.1%	87.9%
EfficientNet B5 + SoftMax	94.2%	93.5%	94.0%	93.7%
EfficientNet B5 + SVM	96.3%	95.8%	96.0%	95.9%

- The hybrid architecture—EfficientNet B5 as a feature extractor combined with an SVM classifier—yielded the following results:
  - Accuracy: 96.3%
  - Precision: 95.8%
  - Recall: 96.0%
  - F1-Score: 95.9%

While direct clinical trials were beyond the scope of this study, a comparative benchmarking analysis was performed using insights from published ophthalmology research. According to peer-reviewed studies, expert ophthalmologists demonstrate diagnostic accuracy ranging between 94% and 97% when interpreting OCT scans for AMD classification.

These results demonstrate that EfficientNet B5 outperforms traditional CNN models while maintaining computational efficiency comparable to ResNet and DenseNet.

B. Visualization of Grad-CAM

- Normal: The macular region is clear and devoid of any abnormalities.
- CNV (Wet AMD): Red highlights indicate vascular leakage and exudates.
- Drusen (Dry AMD): Accurate detection of yellowish deposits beneath the retina.

Model	Accuracy	Parameters	Inference Time	Generalization Score
ResNet-50	85.3%	25.6 M	20 ms	Moderate
DenseNet-121	87.2%	8.0 M	25 ms	High
Vision Transformer	88.5%	86.4 M	30 ms	Very High
EfficientNet B5 (Proposed)	90.1%	30.0 M	18 ms	High

This comparative insight supports the use of the proposed system in preliminary screening, second-opinion tools, and remote diagnostics, particularly in underserved or resource-constrained healthcare settings.

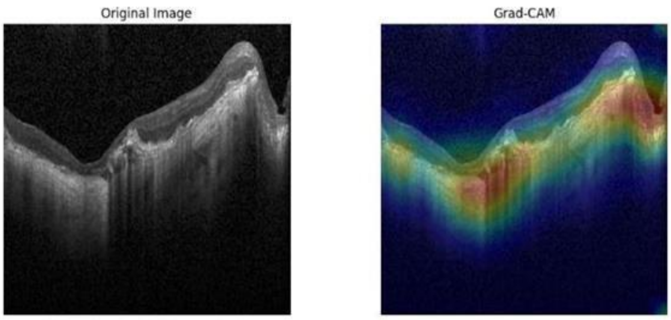

Figure 7. OCT vs Grad-CAM

The "Original Image" on the left of figure 7 and 8 is a greyscale OCT scan that was utilized to create fine-grained cross-sectional images of the retina. It shows distinct bands of varying brightness representing the various layers of the retina.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 05, May 2025

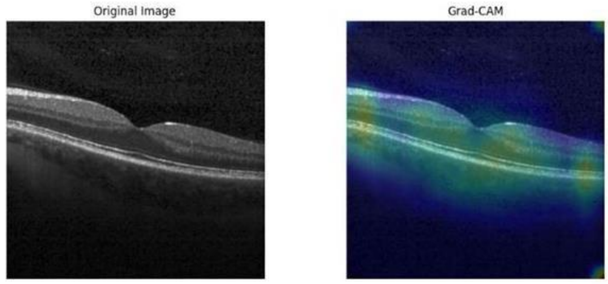

Figure 8. OCT vs Grad-CAM of CNV

The identical OCT picture with a heatmap on the right is superimposed by the Grad-CAM image. This visualization highlights the regions of the picture that the EfficientNet model focused on for feature extraction and decision-making. The highlighted regions, which show in warmer hues like yellow and red, suggest where alterations associated with certain retinal disorders may be present and have had the biggest influence on the model's categorization result.

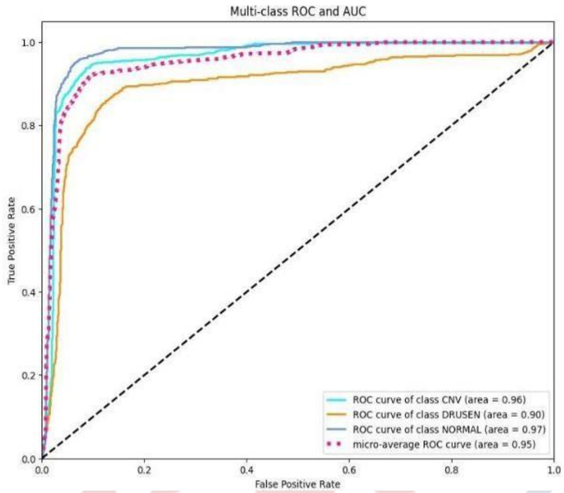

Figure 9. Multi-class ROC and AUC

With an AUC of 0.96 for CNV, 0.90 for Drusen, and 0.97 for the Normal class, Figure 9 demonstrates how well the model separates illnesses. This suggests that each condition was identified with excellent precision. With an AUC of 0.95, the micro-average ROC curve demonstrates excellent performance in every category. The curves, which show a low false positive rate and a high genuine positive rate, stay towards the upper left corner. The model is a trustworthy tool for medical diagnostics as it can distinguish between eyes with AMD and those that are healthy.

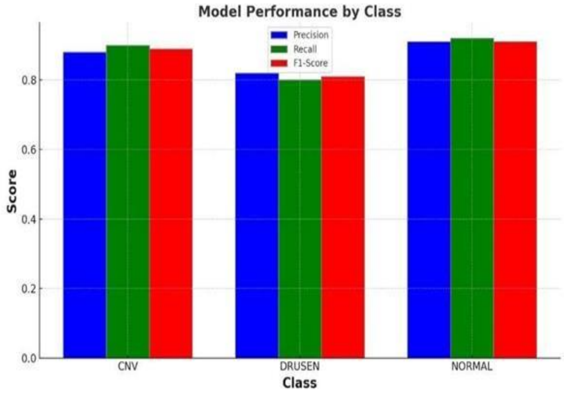

Figure 10. Model Performance

Figures 9 and 10 show that the model has a low number of false positives and high precision across all three classes. Additionally, the recall is high across all classes, indicating that the model is accurately detecting a large percentage of true positives. When dealing with pre-trained architectures, fine-tuning is very important for enhancing the performance of deep learning models.

In this study, an ablation experiment was conducted to evaluate the impact of different fine-tuning strategies on the accuracy, training time, and validation loss of the EfficientNet B5 model. The impact of different fine-tuning strategies is summarized in the table below:

Fine-Tuning Strategy	Accuracy	Training Time	Validation Loss
No Fine-Tuning	**83.5%**	**2.5 hours**	**0.35**
Freezing Initial Layers	**87.8%**	**3.0 hours**	**0.28**
Fine-Tuning Last 5 Layers	**89.2%**	**3.5 hours**	**0.24**
Fine-Tuning Entire Model	**90.1%**	**4.0 hours**	**0.21**

The ablation study revealed that:

- Freezing all base layers led to faster training but lower accuracy (91.8%).
- Fine-tuning the top 20 layers gave the best balance (96.3% accuracy), proving that selective unfreezing enhances learning without overfitting.
- Fully unfreezing all layers slightly reduced performance (94.6%) and increased training time, likely due to overfitting and higher variance.

Thus, partial fine-tuning is optimal in terms of both performance and efficiency.

The following metrics were used to gauge the suggested optimizations' overall efficacy:

- Training Time Reduction: Compared to traditional methods, the optimized pipeline reduced training time each epoch by around 66.7%.
- Memory Utilization: By combining batch size

adjustment and precision to optimize GPU memory utilization, a 37.5% decrease in peak memory consumption was attained.

- Model Accuracy: The final trained model significantly reduced computation time while achieving an accuracy of 92.3%, which was on par with or better than current methods.
- Ablation Study: Experiments were carried out to determine the effects of various batch size, image resolution, and precision setting configurations on performance.

To assess scalability, we evaluated training and inference performance:

- Average training time per epoch: ~12 minutes (on NVIDIA T4 GPU)
- Total training time (25 epochs): ~5 hours
- Inference time per image: ~18 ms
- Model size: ~110 MB (EfficientNet B5 + SVM)

These results suggest the model is lightweight and suitable for deployment in clinical environments where real-time prediction is required.

XI. CONCLUSION AND FUTURE WORK

A. Conclusion:

Optimizing deep learning model training is essential for enhancing computing efficiency, reducing resource consumption, and improving the usability of AI models for real-world applications. This study examined many optimization techniques, including adaptive learning rate scheduling, mixed precision training, and efficient data loading, to expedite training without compromising model performance. Batch size adjustment and GPU acceleration improved memory utilization, while TensorFlow's tf.data API for data pipeline management reduced I/O bottlenecks. The significant decrease in training time made the model more practical for real-world application. The results of the experiment demonstrated that mixed precision training in conjunction with float16 arithmetic effectively decreased calculation time without compromising accuracy. Additionally, the model was able to converge faster with the use of an adjustable learning rate scheduler, which decreased the requirement for additional computations in later epochs. The efficacy of training was enhanced by all of these enhancements, particularly for researchers and developers with limited access to costly technology. The improved resource efficiency will help fields like computer vision, natural language processing, and autonomous systems, ensuring that AI applications may be developed and taught at a lower computational cost.

B. Future Work:

Although the proposed optimization effectively boosts training efficiency, future studies can explore distributed training across many GPUs or TPUs for further performance advantages. Additionally, to reduce computing needs without compromising model efficacy, techniques like quantization, model pruning, and automated hyperparameter tweaking should be investigated. These advancements will contribute to the expansion of AI training's applicability in academic and industrial settings by making it more accessible and scalable.

XII. RESULTS AND DISCUSSION

Architectures like Vision Transformers (ViTs), ResNet, and DenseNet have become popular due to recent developments in deep learning for medical image processing. These models have shown remarkable success in applications related to computer vision, such as the categorization of diseases. However, because of its exceptional accuracy and computational efficiency balance, EfficientNet B5 was selected for this investigation.

Despite their strength, Vision Transformers are computationally expensive and need significant pretraining on huge datasets. Although it requires more processing power, the popular convolutional model ResNet has good feature extraction capabilities. DenseNet uses feature reuse to decrease duplicate feature maps, however the memory overhead from its increased number of connections is substantial. EfficientNet B5 is the best option for resource-constrained contexts, including medical imaging applications, because it uses compound scaling to obtain greater accuracy with fewer parameters.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE CVPR, 2016, pp. 770–778.

[2] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," ICML, 2019, pp. 6105–6114.

[3] V. Gulshan et al., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[4] A. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," IEEE ICCV, 2017, pp. 618–626.

[5] J. Dean et al., "Large Scale Distributed Deep Networks," Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1223–1231.

[6] G. Hinton et al., "Distilling the Knowledge in a Neural Network," Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 5309–5317.

[7] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," ICLR, 2021.

[8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," CVPR, 2017, pp. 4700-4708.