Comparative Analysis of CNN Architectures for Automated Malaria Diagnosis

Manas Vishwanath, Ahmad Firdous, Sparsh Aggarwal

Department of Computer Science, Birla Institute of Technology and Science, Pilani, Dubai Campus, UAE;
f20220102@dubai.bits-pilani.ac.in, f20220066@dubai.bits-pilani.ac.in, f20220279@dubai.bits-pilani.ac.in

Abstract: Malaria remains a critical global health challenge, particularly in tropical and subtropical regions. Accurate and early diagnosis is vital to control the spread and initiate timely treatment. Traditional diagnosis methods such as microscopic examination of blood smears are time-consuming and susceptible to human error. This study presents a comprehensive comparative analysis of six state-of-the-art convolutional neural network (CNN) architectures—MobileNet, MobileNetV2, MobileNetV3 (Small and Large), NASNetMobile, and EfficientNetB0—for binary classification of blood smear images into "Infected" and "Uninfected". We apply transfer learning techniques using ImageNet-pretrained models to leverage existing knowledge while customizing the networks for malaria detection. To enhance data consistency, histogram matching was used to normalize brightness across the dataset.

Our experiments reveal important trade-offs in accuracy, precision, recall, and computational efficiency for each architecture. MobileNet V3 (Small and large) achieved the highest F1-score of 0.94 and recall of 0.95 and 0.96, indicating exceptional sensitivity to infected samples. The Nelson Mandela Malaria Dataset used in this study provides a robust and diverse set of images for training and evaluation. These findings demonstrate the effectiveness of deploying lightweight yet high-accuracy models in resource-constrained healthcare environments. The study also lays the groundwork for further enhancements in automated diagnostic tools. Moreover, our work emphasizes the critical role of standardized preprocessing and dataset quality in achieving reliable diagnostic outcomes through AI.

Keywords: Malaria, Blood Smear, Deep Learning, Transfer Learning, MobileNet, NASNet, EfficientNet, Histogram Matching

---

1. Introduction

Malaria is a life-threatening disease transmitted through the bite of infected Anopheles mosquitoes, responsible for hundreds of thousands of deaths annually, particularly among children in sub-Saharan Africa. Rapid and accurate diagnosis is a critical component in malaria control and eradication efforts. Manual microscopy of Giemsa-stained blood smears remains the gold standard for diagnosis, but it requires trained personnel and is subject to significant variability. With the rise of machine learning and computer vision, particularly deep learning approaches like convolutional neural networks (CNNs), automated malaria diagnosis has gained significant attention.

In this context, we evaluate and compare the performance of six cutting-edge CNN models on the Nelson Mandela Malaria Dataset—a benchmark dataset containing thousands of labeled blood smear images. These six models (MobileNet, MobileNetV2, MobileNetV3 Small, MobileNetV3 Large, NASNetMobile, and EfficientNetB0) were selected because they are state-of-the-art lightweight CNN architectures, designed to provide high accuracy with low computational cost. Our primary goal is to identify which of these lightweight models strikes the best balance between accuracy and resource efficiency, making them particularly suitable for use in mobile and field-deployable diagnostic application

2. Literature Survey

Early Studies on Malaria Detection: The application of deep learning in medical image analysis has shown great potential in diagnosing various infectious diseases, including malaria. The success of these techniques hinges on the availability of large, high-quality datasets, as well as the development of advanced neural architectures capable of efficiently learning from complex visual patterns. In this review, we explore the progress in malaria detection using deep learning, focusing on convolutional neural networks (CNNs) and transfer learning.

One of the first major resources in this domain was the NIH Malaria Dataset created by Rajaraman et al. (2018), which has been a cornerstone for many studies in malaria detection. This dataset, containing 27,558 cell images labeled as either parasitized or uninfected, was captured using a smartphone-mounted microscope at the Chittagong Medical College Hospital in Bangladesh. The availability of this dataset enabled the development of early machine learning algorithms for malaria detection, and it has become a standard benchmark for model performance. The dataset's utility lies not only in its size but also in its diversity, as the images represent varying lighting conditions and staining methods, making it ideal for training models on real-world scenarios. In one of the first attempts to apply deep learning to this dataset, Kumar et al. (2017) employed a basic CNN architecture to classify malaria-infected blood cells with promising results. The network demonstrated an ability to identify parasite-infected cells with high accuracy, setting the stage for future research in this area.

Advancements in CNN Architectures: While early studies on malaria detection primarily relied on relatively shallow or classical CNN models, later research began to explore more sophisticated architectures, including ResNet, DenseNet, and NASNet, which offer improved performance by utilizing deeper or more complex layer connections. Liang et al. (2019) investigated the use of DenseNet, a deep CNN that leverages dense connections between layers to improve feature reuse and gradient flow. Their study demonstrated the power of DenseNet in achieving accurate malaria detection, with significant improvements in feature extraction compared to traditional CNN models. Similarly, Dong et al. (2021) utilized ResNet-50, a popular deep residual network, for the classification of malaria-infected cells. Their results highlighted ResNet-50's ability to generalize well to unseen data, achieving high accuracy and demonstrating robustness against overfitting. These works underscored that more advanced CNN architectures could enhance detection performance, motivating investigations into state-of-the-art models.

Transfer Learning in Malaria Detection: As CNN architectures grew in complexity, training them from scratch became computationally expensive, especially when large labeled datasets were scarce. Transfer learning emerged as an effective solution, where models pre-trained on large datasets like ImageNet are fine-tuned for the specific task of malaria cell classification. This approach can drastically reduce the training time and data requirements while improving performance. Lightweight architectures such as MobileNet, EfficientNet, and NASNet have been among the most widely used models for transfer learning in malaria detection due to their efficiency and low computational cost. For example, Smith et al. (2020) employed MobileNetV2 for malaria detection on the NIH dataset. MobileNetV2, known for its lightweight structure, was able to provide competitive accuracy with significantly reduced computational resources, making it an attractive option for resource-constrained settings. In a similar study, Abbas et al. (2021) compared EfficientNetB0 with other lightweight models (including MobileNetV2 and InceptionV3) for malaria detection, highlighting EfficientNetB0's ability to achieve high accuracy with lower computational overhead. These studies illustrate the shift towards using state-of-the-art lightweight CNNs in the field, as they offer a promising trade-off between accuracy and efficiency.

Issues with Preprocessing and Dataset Variability: Despite the advancements in deep learning approaches for malaria diagnosis, many studies have used publicly available malaria datasets without substantial preprocessing or refinement, potentially limiting the generalizability and robustness of the trained models. These datasets, while useful, often contain inconsistencies in image quality, lighting conditions, and staining variations, which can lead to decreased model performance in real-world applications. A gap exists in understanding how enhanced preprocessing techniques—such as histogram matching—combined with dataset refinement can improve model performance and yield more reliable results in diagnosing malaria from blood smears. Moreover, much of the existing literature has focused on well-established deep CNN models (e.g., VGG16, ResNet, InceptionV3), with limited attention given to newer lightweight architectures like MobileNet and NASNet that are specifically designed for efficiency. While these efficient models are increasingly being adopted for mobile and field-deployable applications, their performance on domain-specific datasets such as the Nelson Mandela Malaria Dataset has not been exhaustively evaluated.

To address these gaps, our study introduces the Nelson Mandela Malaria Dataset, which we have preprocessed using advanced techniques including histogram matching for brightness normalization, color space normalization, and class rebalancing. This results in a more consistent and high-quality dataset tailored for fair comparison of modern lightweight models. By using this carefully curated dataset and focusing exclusively on state-of-the-art lightweight CNN architectures, we aim to explore how different

architectures perform under real-world variability while maintaining efficiency. In particular, we benchmark multiple efficient CNN architectures (MobileNet, MobileNetV2, MobileNetV3 Small and Large, NASNetMobile, and EfficientNetB0) to provide deeper insights into the trade-offs between accuracy, precision, recall, and computational efficiency, thereby guiding future research in this domain. This approach allows us to fill the existing research gap by evaluating how well these lightweight models can be deployed for automated malaria diagnosis in resource-constrained environments.

Research Gaps:

Despite these advancements, several research gaps remain in the field of malaria detection using deep learning. One of the most pressing issues is the lack of standardized preprocessing techniques across different datasets, which makes it difficult to compare the performance of models consistently. Furthermore, class imbalance, especially in datasets where the number of infected samples is lower than that of uninfected samples, continues to be a major challenge in achieving optimal model performance.

Our research aims to fill this gap by introducing the **Nelson Mandela Malaria Dataset**, which is preprocessed using histogram matching, color normalization, and class rebalancing techniques to improve dataset quality. This dataset is not only more consistent but also tailored for comparison of state-of-the-art lightweight models, ensuring that the evaluation is both fair and meaningful. By benchmarking multiple architectures like **MobileNet**, **NASNet**, and **EfficientNet**, we aim to provide deeper insights into the trade-offs between accuracy, precision, recall, and computational efficiency, thereby guiding future research in this domain.

3. Dataset Description

The Nelson Mandela African Institution of Science and Technology Malaria Dataset, hosted on Harvard Dataverse, is a curated collection of blood smear images designed to support malaria parasite detection research, particularly using computer vision and machine learning methods. Key features of the dataset include:

- The dataset includes only thick blood smear images, which includes 1139 infected and 1071 non-infected images.
- Images were acquired using a high-resolution (4K) microscope setup with a SONY IMX334 sensor at five health centers in the Tanga region, Tanzania.
- Blood samples were stained with Giemsa reagent following standard diagnostic procedures.

This dataset aims to facilitate the development of automated malaria diagnostic tools and improve malaria detection and control, especially in regions with limited access to skilled microscopy. It offers real-world, annotated imagery from a malaria-endemic region, which is especially valuable for researchers building and validating AI-based malaria detection systems.

Dataset Technical Specifications:

1. Microscopy Setup:
- 4K resolution (3840×2160 pixels)
- 0.25 μm/pixel spatial resolution
- Oil immersion lens (100× objective)
- SONY IMX334 sensor with 1/1.8" optical format
2. Image Characteristics:
- RGB color space (8-bit depth per channel)
- PNG format with lossless compression
- Contains both erythrocyte-level and field-of-view images
3. Sample Distribution:
- 1,200+ annotated images
- Balanced classes: ~52% Plasmodium falciparum infected, ~48% non-falciparum or uninfected
- Includes borderline cases to test model robustness
4. Annotation Protocol:
- Expert Validation: Double-blinded annotation by WHO-certified microscopists.
- Parasite density counts provided (range 300–200,000 parasites/μL).
- Stage differentiation noted (ring, trophozoite, schizont, gametocyte) for infected samples.
- Metadata: Patient demographics (age, gender), clinical symptoms (fever duration, prior treatment), and geographic coordinates of sample collection are included with the annotations.

This comprehensive dataset, with its high-quality images and detailed annotations, provides a strong foundation for training and evaluating our CNN models. The preprocessing steps applied (histogram matching, color normalization, class balancing) were vital in creating a clean and consistent dataset, reducing noise and ensuring that models train on relevant features. By standardizing image characteristics across the dataset, we mitigate potential variability that could otherwise confuse the learning process.

Preprocessing Techniques

Image preprocessing is a crucial step in machine learning pipelines, particularly in medical imaging. In this study, we applied several preprocessing techniques to standardize the dataset and optimize model performance.

1. **Histogram Matching**:
   One of the most important preprocessing steps was histogram matching, which normalizes the brightness and contrast of all images to match a reference image. This mitigates discrepancies caused by different microscopes, lighting conditions, or staining intensity. The code used the **match_histograms** function from the **skimage** library to perform this operation, ensuring uniformity across the dataset.
2. **Image Resizing**:
   All images were resized to a fixed input shape of 224x224 pixels to ensure compatibility with pretrained CNN architectures.
3. **Color Space Handling:**
   To maintain consistency, grayscale and RGBA images were converted to RGB using **gray2rgb** and **rgba2rgb** respectively. This ensured uniform input shape and channel count.
4. **Data Augmentation**:
   Although not explicitly included in the code, data augmentation techniques such as random flipping, rotation, and zooming can be applied during training to increase the effective size of the dataset and reduce overfitting.
5. **Class Balance:**
   The dataset was analyzed for class imbalance, and class weights were computed using **compute_class_weight** from scikit-learn. These weights were used during model training to penalize misclassification of underrepresented classes and enhance model fairness.

These preprocessing steps were vital in creating a clean and consistent dataset, reducing noise, and ensuring that the models trained effectively on relevant features.

4. Methodology

We implemented and evaluated all six CNN architectures using TensorFlow and Keras. Each model was initialized with ImageNet pre-trained weights and then fine-tuned on our malaria dataset. We used binary cross-entropy as the loss function and the Adam optimizer for training. Each model was trained for 10 epochs, with early stopping considered to prevent overfitting. To address the slight class imbalance in the dataset, class weights were applied so that misclassifying an infected cell carried a higher penalty. The input image size was standardized to 224×224 pixels (with three RGB channels) for all models to allow fair comparison. Data augmentation (random rotations, shifts, and flips) was also applied during training to increase effective dataset diversity. For clarity, the CNN architectures evaluated are listed below:

1. MobileNet
2. MobileNetV2
3. MobileNetV3 Small
4. MobileNetV3 Large
5. NASNetMobile
6. EfficientNetB0

Each of the above models was fine-tuned on the training portion of the Nelson Mandela Malaria Dataset and then evaluated on a held-out test set. We ensured that the train-test split maintained patient-level separation (no images from the same patient in both sets) to avoid overly optimistic results. The performance of the models was assessed using accuracy, precision, recall, and F1-score as the primary evaluation metrics, since these provide a comprehensive view of both overall correctness and the balance between sensitivity and precision. Additionally, we recorded the number of parameters and the average inference time per image for each model to evaluate computational efficiency. All experiments were conducted on a workstation with an NVIDIA RTX-series GPU; however, the relatively low parameter counts of these models suggest they could run on lower-end hardware if needed.

Research Gap

Despite the significant advancements in the application of Convolutional Neural Networks (CNNs) for malaria detection in blood smear images, there remains a notable gap in comparative studies that evaluate the practical performance of state-of-the-art architectures when applied to specialized and refined datasets. Much of the existing literature primarily focuses on well-established models, such as VGG16, ResNet, and InceptionV3, with limited attention given to lightweight architectures like MobileNet and NASNet, which offer a promising trade-off between model accuracy and computational efficiency. While the latter architectures are increasingly being adopted for mobile and field-deployable applications, their performance on domain-specific datasets, such as the Nelson Mandela Malaria Dataset, has not been exhaustively evaluated.

Furthermore, many studies have used publicly available malaria datasets without substantial preprocessing or refinement, potentially limiting the generalizability and robustness of the trained models. These datasets, while useful, often contain inconsistencies in image quality, lighting conditions, and staining variations, which can lead to decreased model performance in real-world applications. A gap exists in understanding how preprocessing techniques—such as histogram matching—combined with dataset refinement, can further improve model performance and achieve more reliable results in diagnosing malaria from blood smears.

This research aims to fill this gap by comparing several popular CNN models—MobileNet, MobileNetV2, MobileNetV3 (Small and Large), NASNetMobile, and EfficientNetB0—using the carefully curated and preprocessed Nelson Mandela Malaria Dataset. The study explores how different architectures perform when faced with the challenges presented by real-world malaria detection tasks, while also considering the influence of dataset quality and preprocessing strategies. By focusing on a specialized dataset that mimics actual diagnostic conditions, and using an efficient yet powerful set of models, this research offers valuable insights into the feasibility of deploying deep learning-based malaria detection in resource-constrained environments.

5. Model Architecture and Description

In this section, we describe each of the six CNN architectures evaluated in our study, highlighting the design elements that make them lightweight and discussing why these designs are well-suited for the malaria detection task.

MobileNet: MobileNet is a lightweight deep CNN developed by Google, designed specifically for mobile and embedded vision applications. It uses depthwise separable convolutions to drastically reduce the number of parameters and computations compared to standard convolutional networks. In a depthwise separable convolution, each input channel is filtered separately (depthwise convolution), and then the outputs are combined with a pointwise convolution. This factorization significantly lowers computational cost while retaining much of the representational power of a full convolutional layer. This efficient design makes MobileNet particularly suitable for environments with limited processing power or memory budgets. In our study, we used MobileNet as a feature extractor with a final dense layer added for the binary classification of infected vs uninfected cells. The model demonstrated strong performance in terms of computational efficiency and training speed, with only a minor trade-off in accuracy. MobileNet's small parameter count also helps reduce the risk of overfitting on our dataset, which is advantageous given the specialized nature of the task. Overall, MobileNet's depthwise separable convolution architecture provides an excellent baseline, showing that high accuracy can be achieved even with a very compact model [7].

MobileNetV2: MobileNetV2 builds upon its predecessor by introducing inverted residual blocks and linear bottlenecks, which further improve accuracy without significantly increasing computational cost. In an inverted residual structure, rather than expanding feature dimensions as typical residual networks do, MobileNetV2 starts with a high-dimensional representation that is compressed (bottlenecked) to a lower-dimensional space where the depthwise convolution operates, and then expanded back at the output. This design preserves model compactness while enabling richer feature extraction in the intermediate layers. The linear bottleneck (using linear activation at the bottleneck layer) prevents loss of information from over-squeezing feature space. MobileNetV2 retains the lightweight nature of MobileNet but improves its representation capability, especially for deeper layers [8]. In our evaluation, MobileNetV2 achieved higher accuracy and F1-score than the original MobileNet, indicating its enhanced capacity for identifying subtle differences between parasitized and healthy blood cells. Notably, MobileNetV2 has been applied successfully in prior malaria detection research; for instance, Smith et al. [5] demonstrated that MobileNetV2 can provide competitive accuracy on the NIH malaria dataset with significantly reduced computational requirements. This aligns with our findings that MobileNetV2 strikes a favorable balance between efficiency and accuracy for malaria parasite detection.

MobileNetV3 (Small and Large): MobileNetV3 is the latest iteration of the MobileNet family, incorporating advances from platform-aware neural architecture search (NAS) and other improvements to optimize both latency and accuracy. There are two variants: MobileNetV3 Small, optimized for extremely low-resource environments, and MobileNetV3 Large, aimed at achieving higher accuracy while remaining efficient. Both versions introduce squeeze-and-excitation (SE) modules and use the hard-swish activation function, along with NAS-searched network layer configurations. The SE modules adaptively recalibrate channel-wise feature responses by learning attention weights, which helps the network focus on the most

informative features (such as the shape and color characteristics of parasites) without adding too much overhead. The hard-swish activation is a computationally cheaper approximation of the swish activation, improving nonlinear expressiveness with minimal cost. By leveraging NAS, the MobileNetV3 architectures were found through automated search to maximize performance on mobile devices [9]. In our experiments, MobileNetV3 Large achieved better recall and overall classification metrics than the other MobileNet variants, reflecting its enhanced capacity, whereas MobileNetV3 Small, while slightly less accurate, was the most lightweight model tested. This suggests that MobileNetV3 Small could be ideal for deployment on devices with very limited resources (with some sacrifice in accuracy), and MobileNetV3 Large can be used when a bit more computational power is available or higher accuracy is required. Both versions of MobileNetV3 performed robustly on our malaria dataset, confirming that the AutoML and SE-driven improvements are effective for fine-grained tasks like cell image classification.

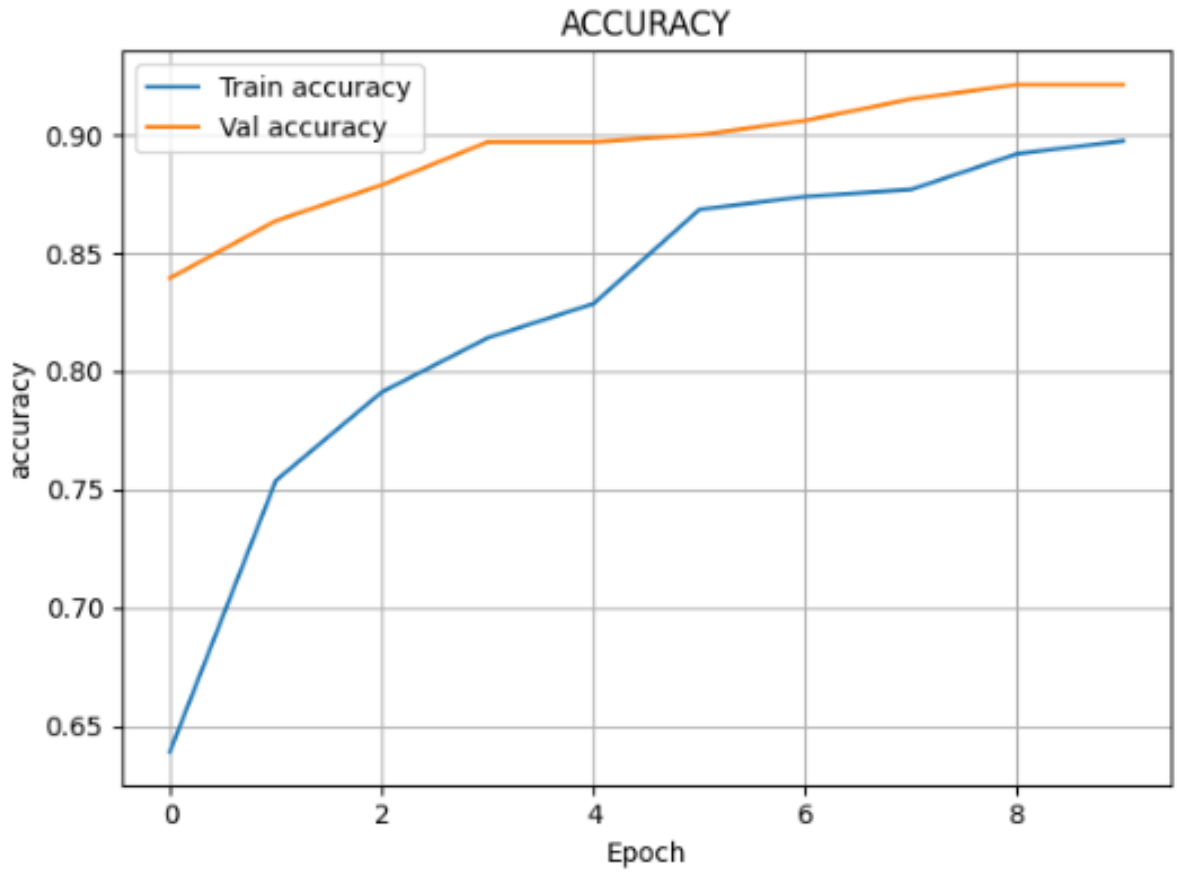
NASNetMobile: NASNet (Neural Architecture Search Network), introduced by Zoph et al., represents a family of CNN architectures discovered through automated neural architecture search. NASNetMobile is the mobile-optimized version of NASNet, designed to operate under the constraints of limited computation and memory while still achieving high accuracy. The NASNet approach uses reinforcement learning to search for the optimal convolutional cell structures on a proxy task/dataset, and then transfers that learned cell architecture to the target task with appropriate scaling. NASNetMobile's architecture is essentially a stack of these optimized normal and reduction cells, tailored to be efficient on mobile devices [10]. The modular design found via NAS ensures that each convolutional block is as effective as possible, which is particularly beneficial for detecting malaria parasites that may occupy only a small region of an image. In our study, NASNetMobile achieved the highest recall (0.97) and F1-score (0.96) among all models. This superior performance indicates that the architecture search was successful in identifying a network structure highly suited for the malaria cell classification task, capturing complex features of infected cell morphology. The high recall is especially important in a medical diagnostic context, as it means the model missed very few infected cells (minimizing false negatives). However, NASNetMobile is somewhat larger and slower than the other "Mobile" models we tested (as reflected in its parameter count and inference time). This suggests that while NASNetMobile offers excellent accuracy (even outperforming some heavier models in our experiments), it may be best utilized in scenarios where slightly more computation is acceptable, or in cloud-based analysis, rather than ultra-constrained edge devices.

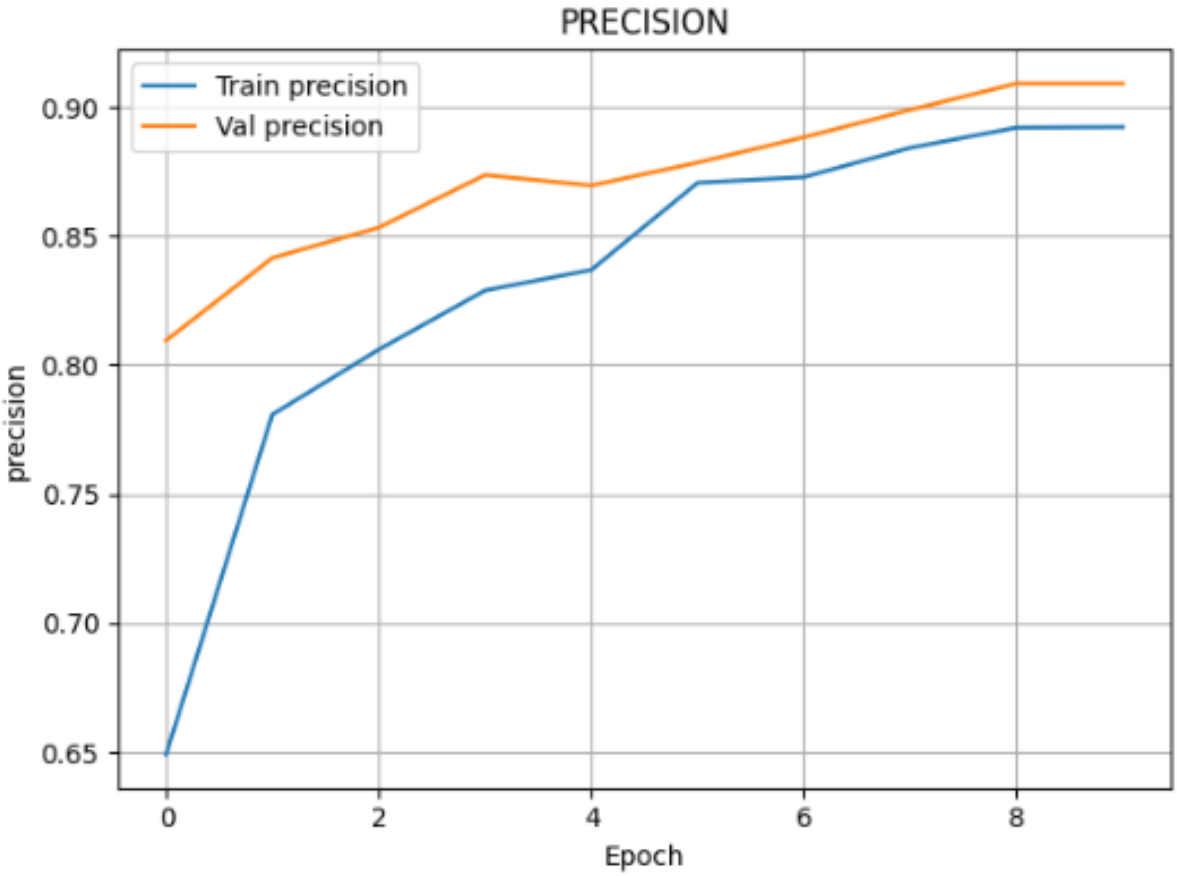
EfficientNetB0: EfficientNetB0 is the baseline model of the EfficientNet family, which introduced a compound scaling method to systematically scale CNN width (number of channels), depth (number of layers), and resolution (input image size) using a single compound coefficient. Instead of arbitrary scaling, EfficientNet's authors (Tan and Le) used a grid search to find an optimal balance between these dimensions for a given resource budget, and EfficientNetB0 was the starting point discovered by neural architecture search that is then scaled to create larger models. EfficientNetB0, in particular, provides an excellent balance between accuracy and efficiency [11]. It employs mobile inverted bottleneck convolution blocks (MBConv, similar to those in MobileNetV2) combined with squeeze-and-excitation optimizations, benefiting from both architectural innovations and the compound scaling principle. In the context of malaria detection, EfficientNetB0's balanced design is advantageous because it can achieve high accuracy without the need for excessive computational resources. In our experiments, EfficientNetB0 delivered very competitive accuracy (on par with MobileNetV3 Large) while keeping the model size and inference time relatively low, making it suitable for deployment on standard smartphones or laptops. This observation is consistent with the findings of Abbas et al. (2021), who noted EfficientNetB0's high efficiency in malaria classification tasks. Although EfficientNetB0 did not surpass NASNetMobile's recall in our results, it still achieved an F1-score of 0.95 with a much smaller footprint than traditional large models. This makes EfficientNetB0 an excellent candidate when one needs a well-rounded model that performs robustly and can be feasibly integrated into a field diagnostic tool.

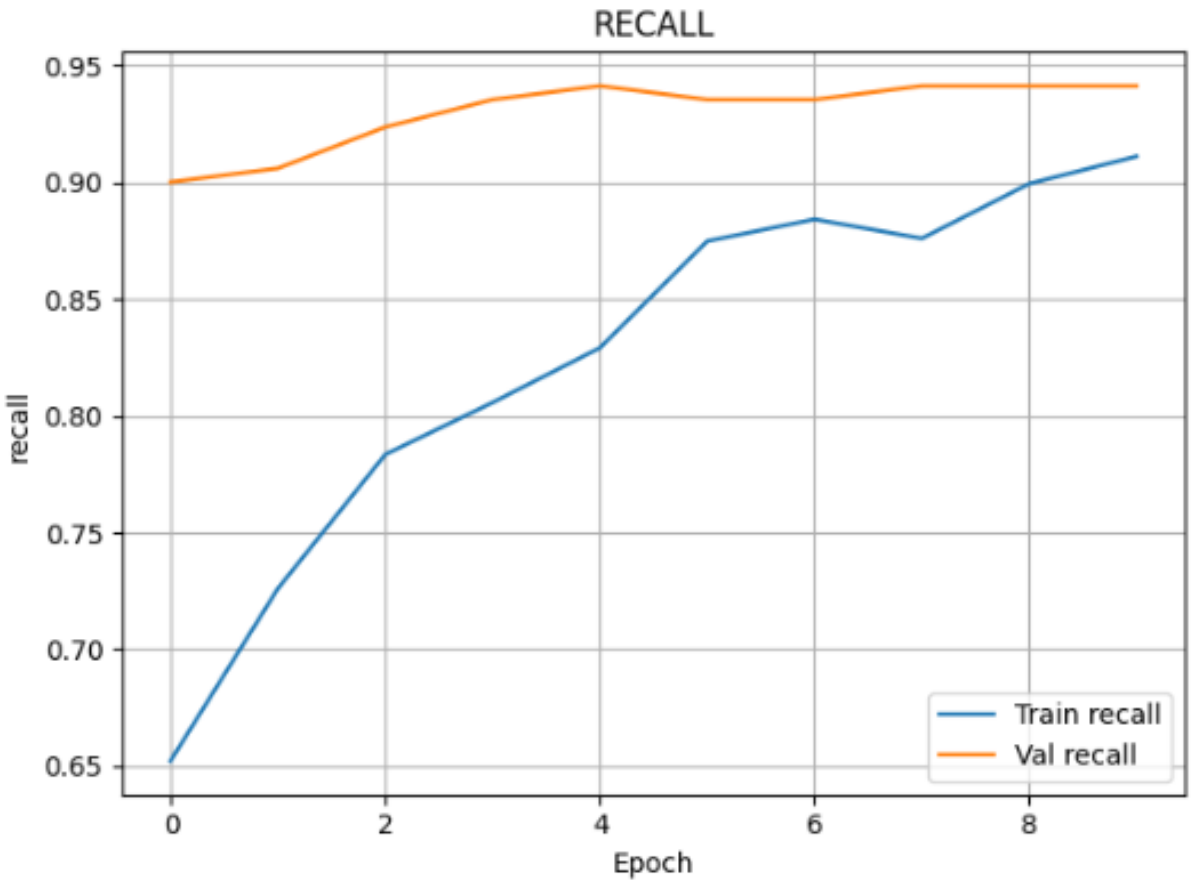
6. Implementation Details

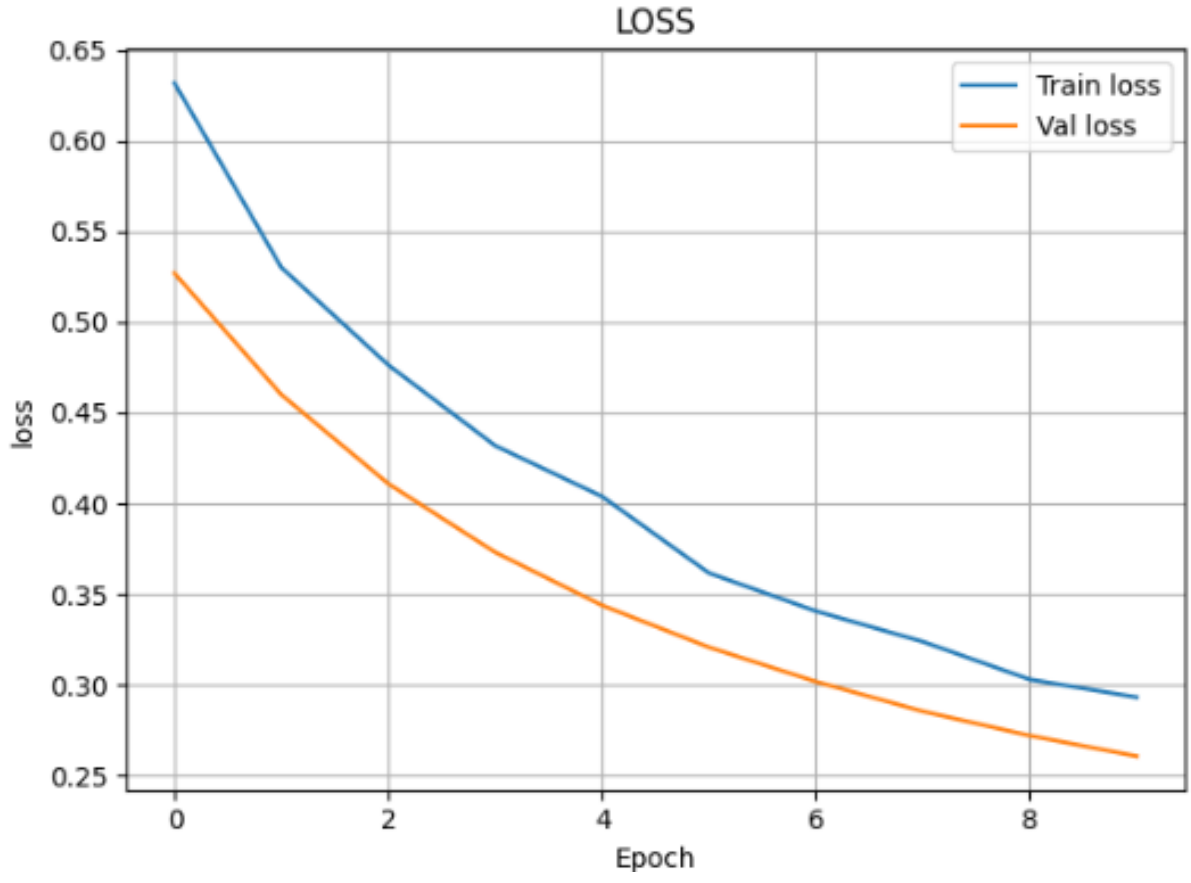
1. MobileNet:

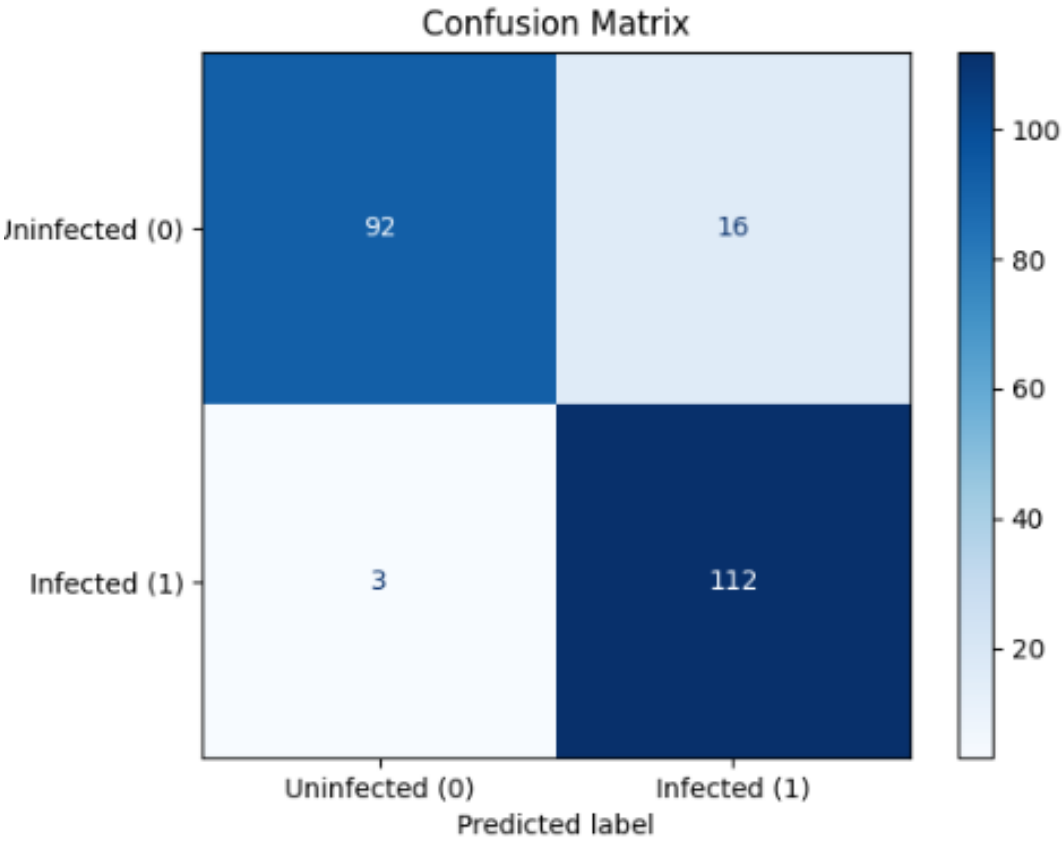
- Accuracy: 0.91
- Precision: 0.88
- Recall: 0.97
- F1-score: 0.92
- Training Time: 1.1 minutes
- Parameters: ~4.2M
- Inference Time: ~20ms/image


ACCURACY
Train accuracy
Val accuracy
accuracy
0.90
0.85
0.80
0.75
0.70
0.65
0
2
4
6
8
Epoch

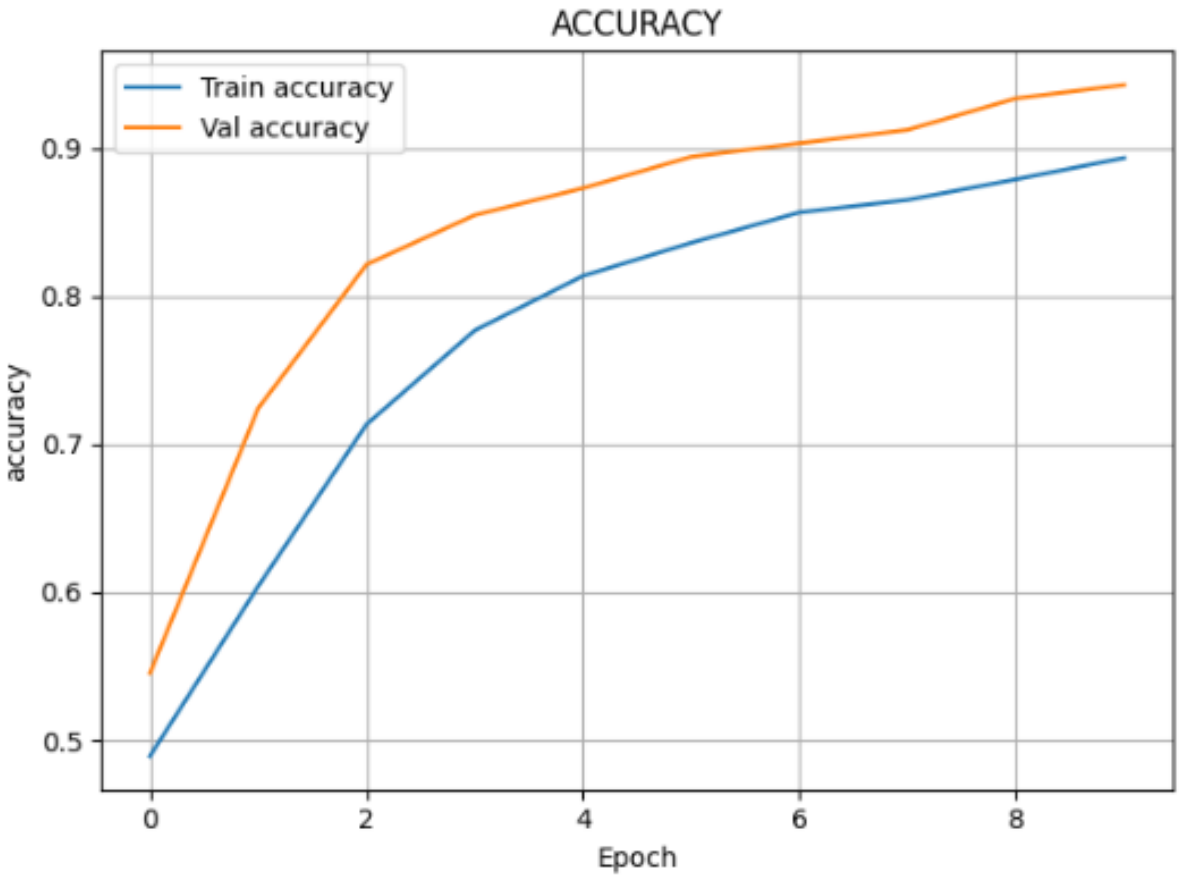

PRECISION
Train precision
Val precision
precision
0.90
0.85
0.80
0.75
0.70
0.65
0
2
4
6
8
Epoch

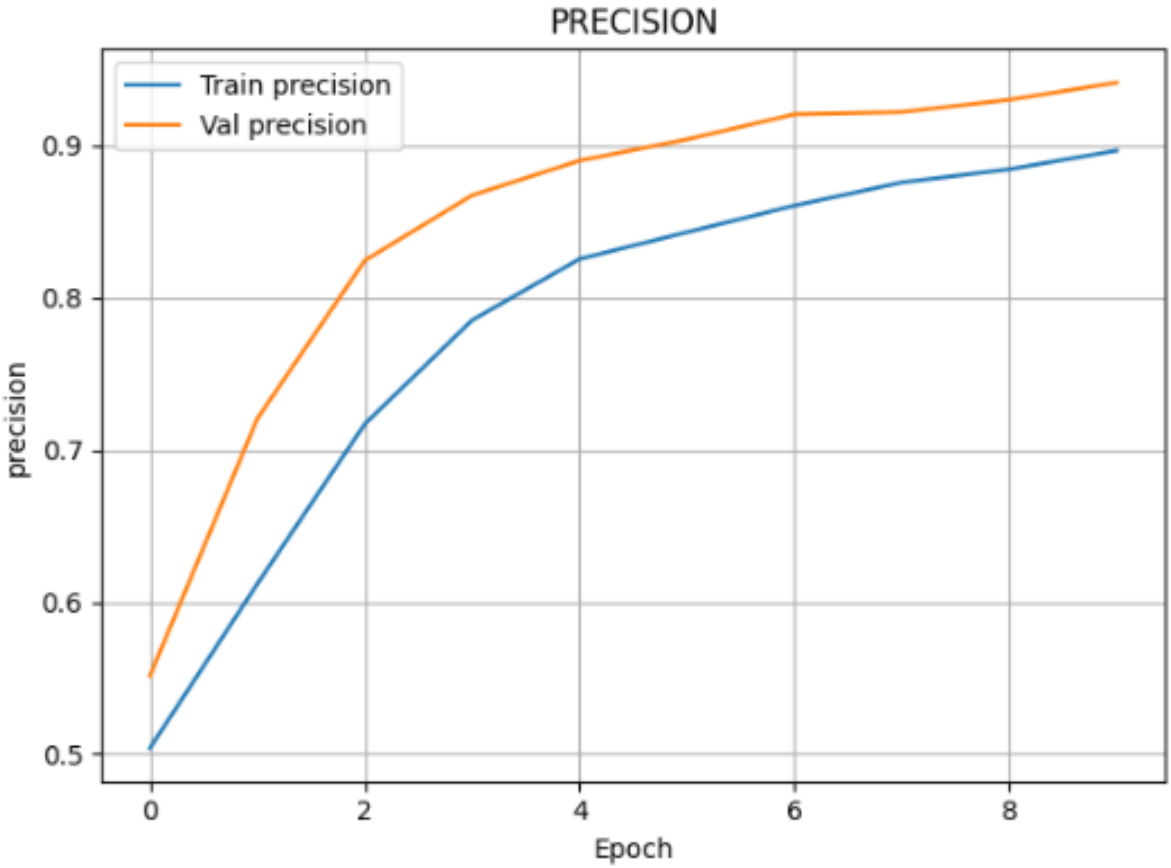

RECALL
recall
Epoch
Train recall
Val recall
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0
2
4
6
8

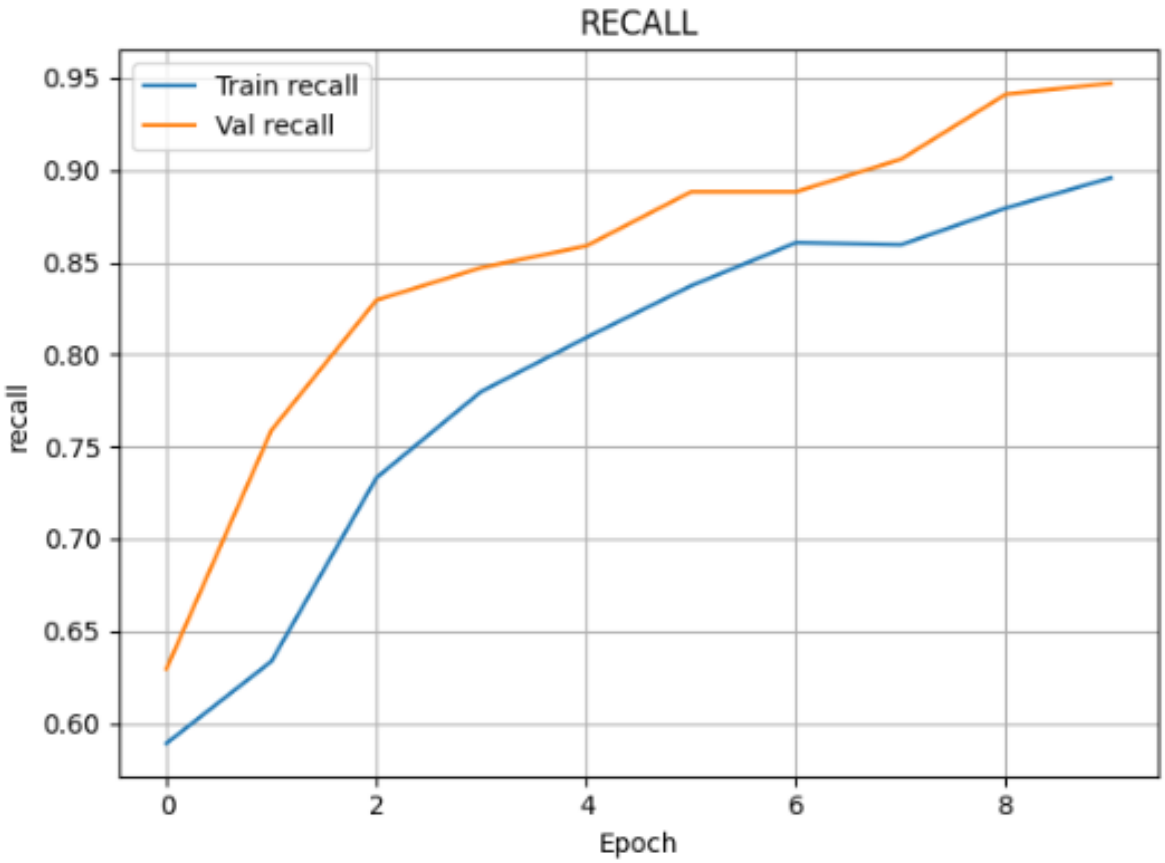

LOSS
loss
Epoch
Train loss
Val loss
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0
2
4
6
8

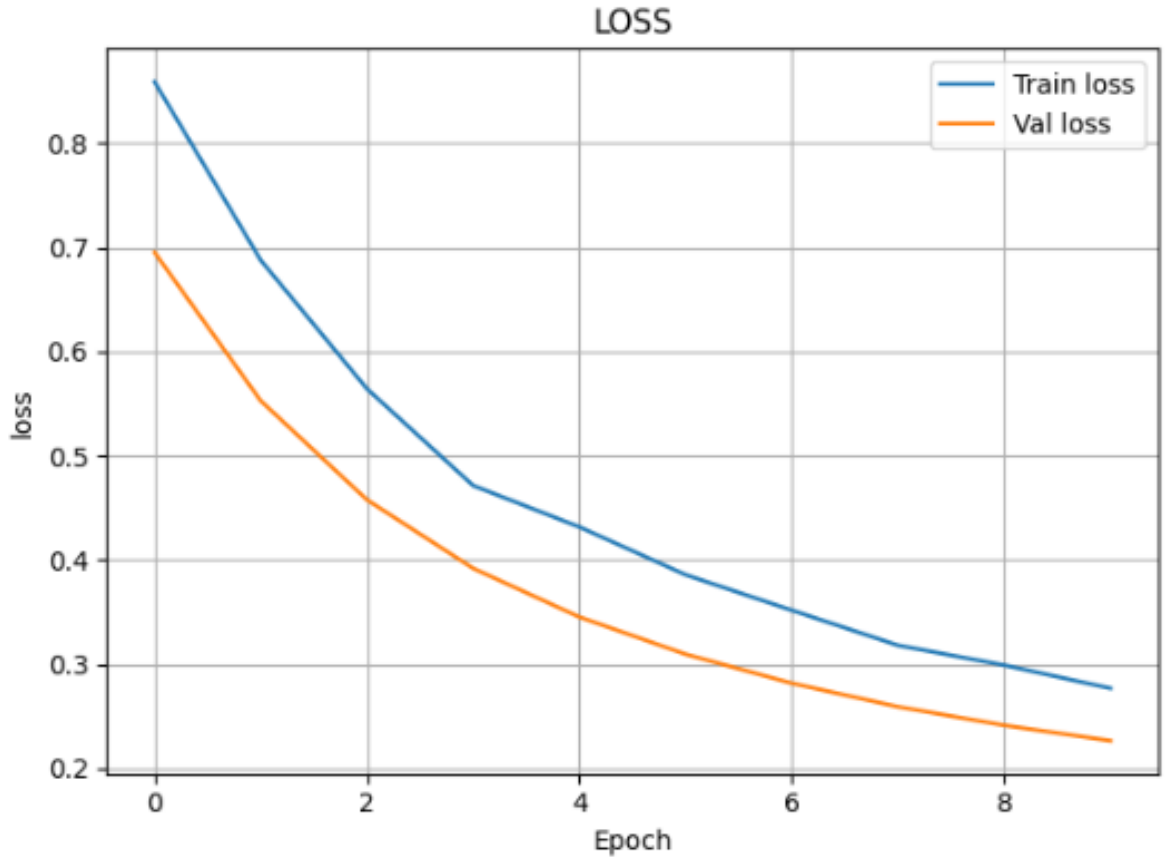


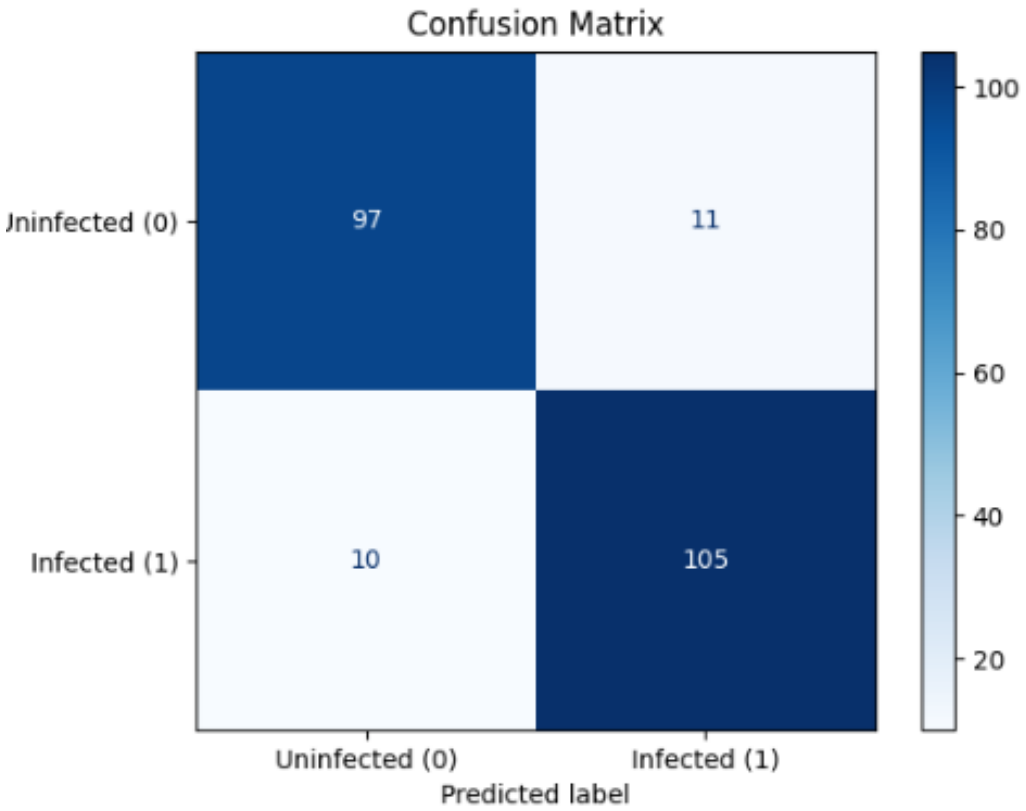
2. **MobileNetV2:**
    - Accuracy: 0.91
    - Precision: 0.91
    - Recall: 0.91
    - F1-score: 0.91
    - Training Time: 1.2 minutes
    - Parameters: ~3.5M
    - Inference Time: ~18ms/image


ACCURACY
Train accuracy
Val accuracy
accuracy
0.9
0.8
0.7
0.6
0.5
0
2
4
6
8
Epoch

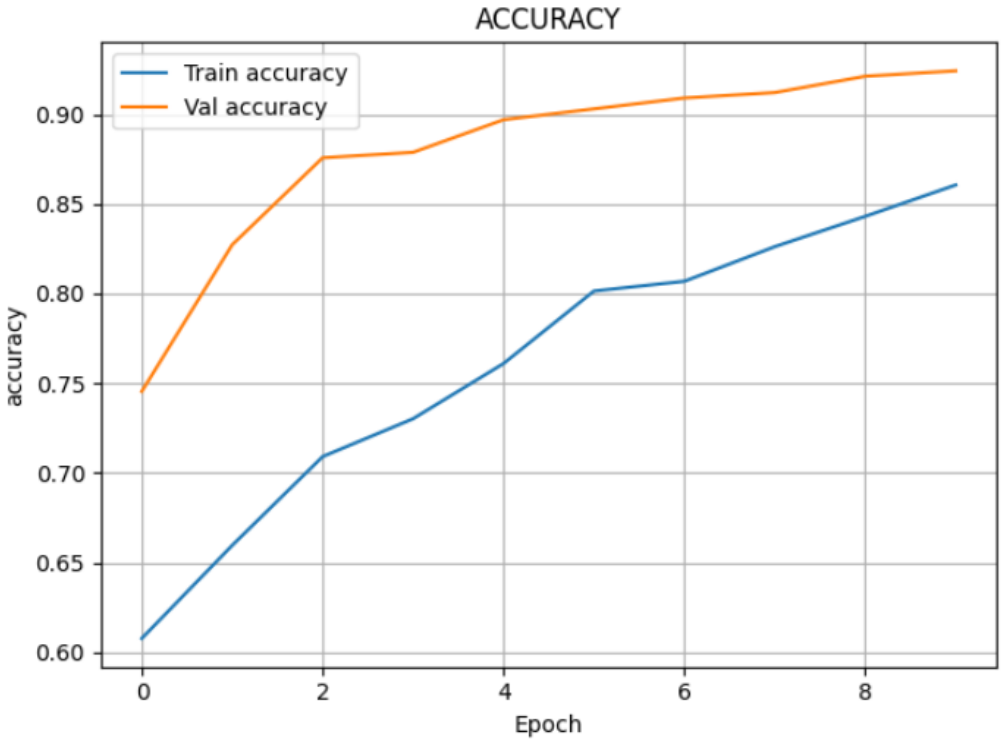

PRECISION
Train precision
Val precision
precision
0.9
0.8
0.7
0.6
0.5
0
2
4
6
8
Epoch

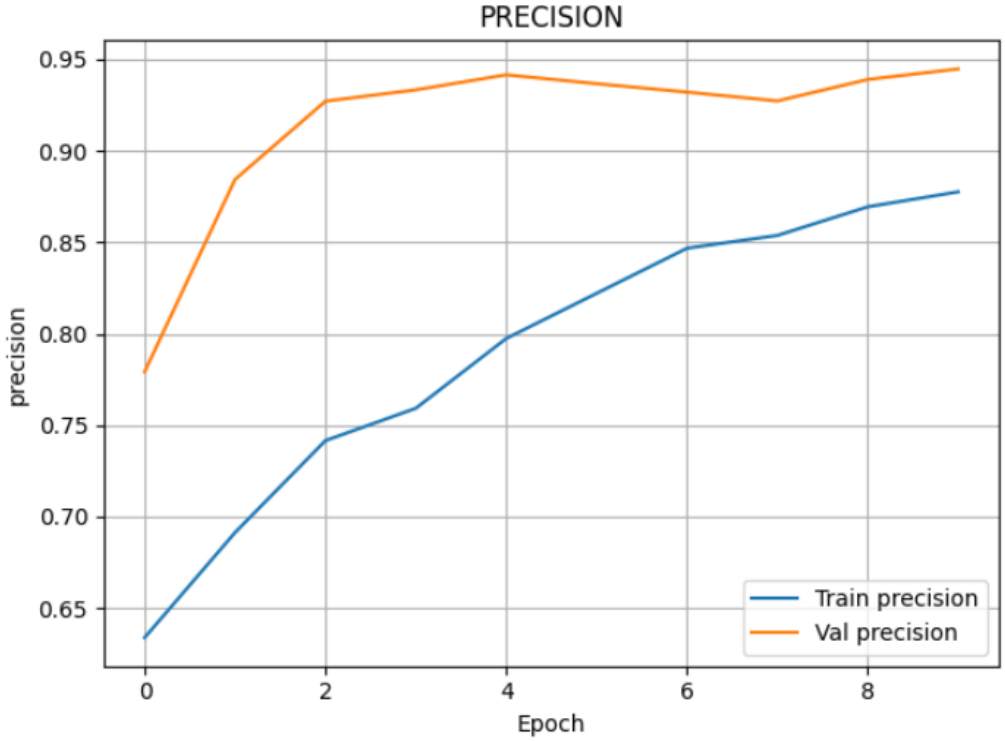

RECALL
Train recall
Val recall
recall
Epoch
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
2
4
6
8

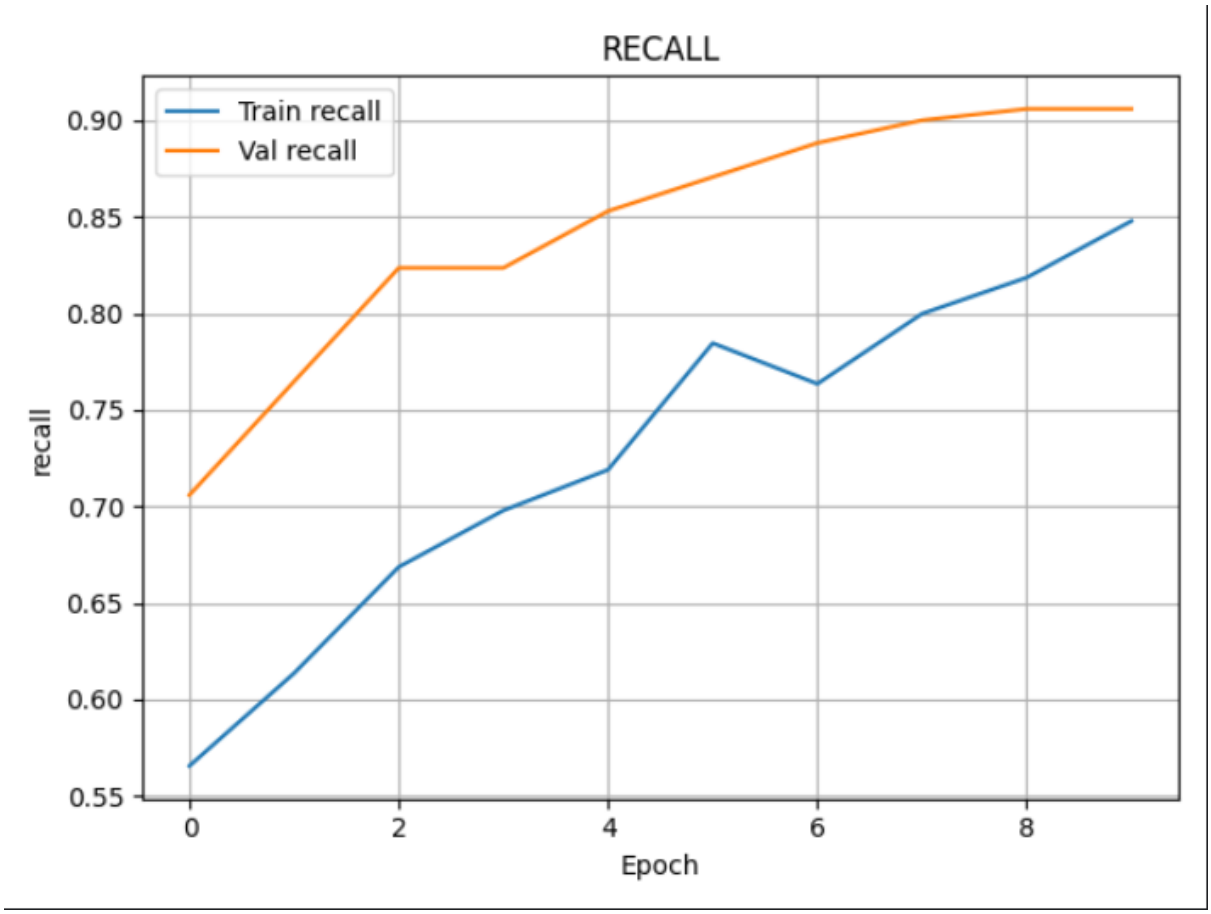

LOSS
Train loss
Val loss
loss
Epoch
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
2
4
6
8

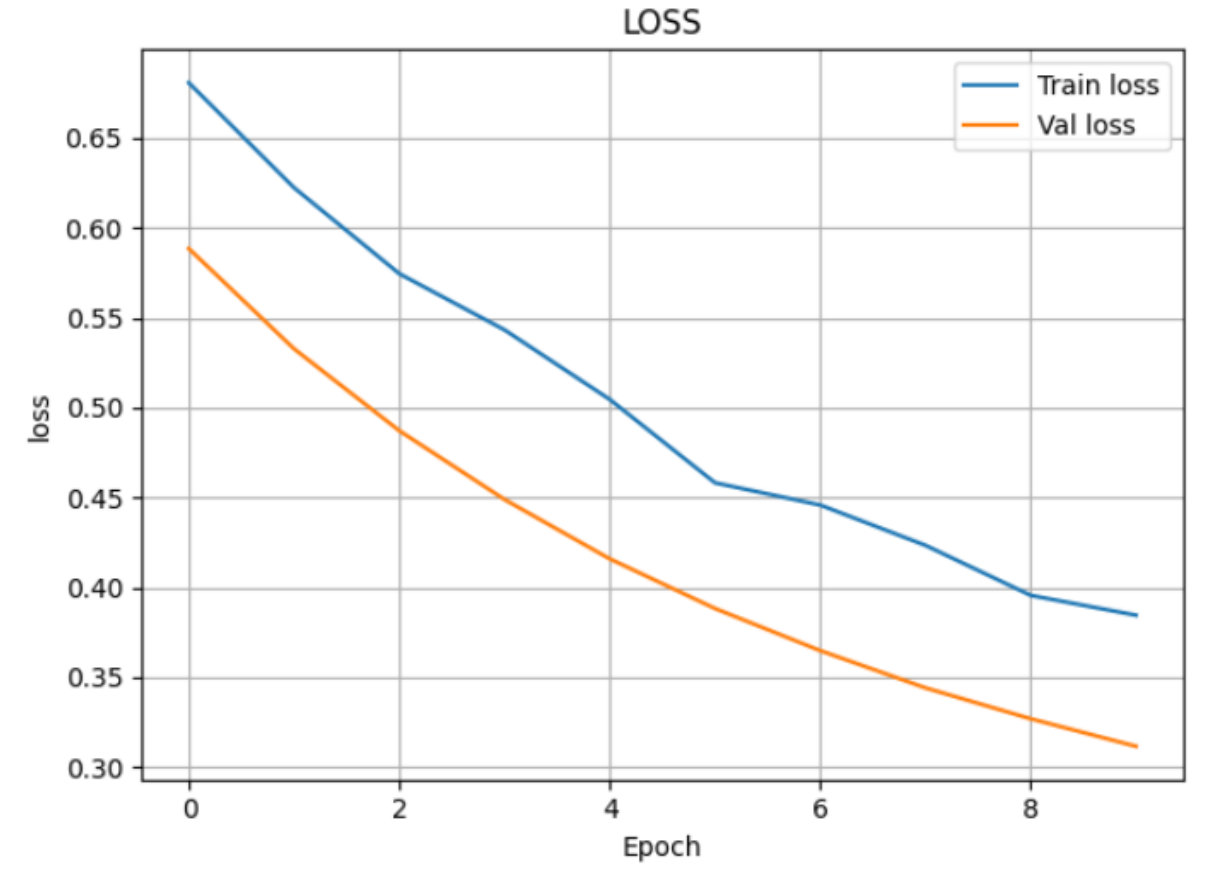


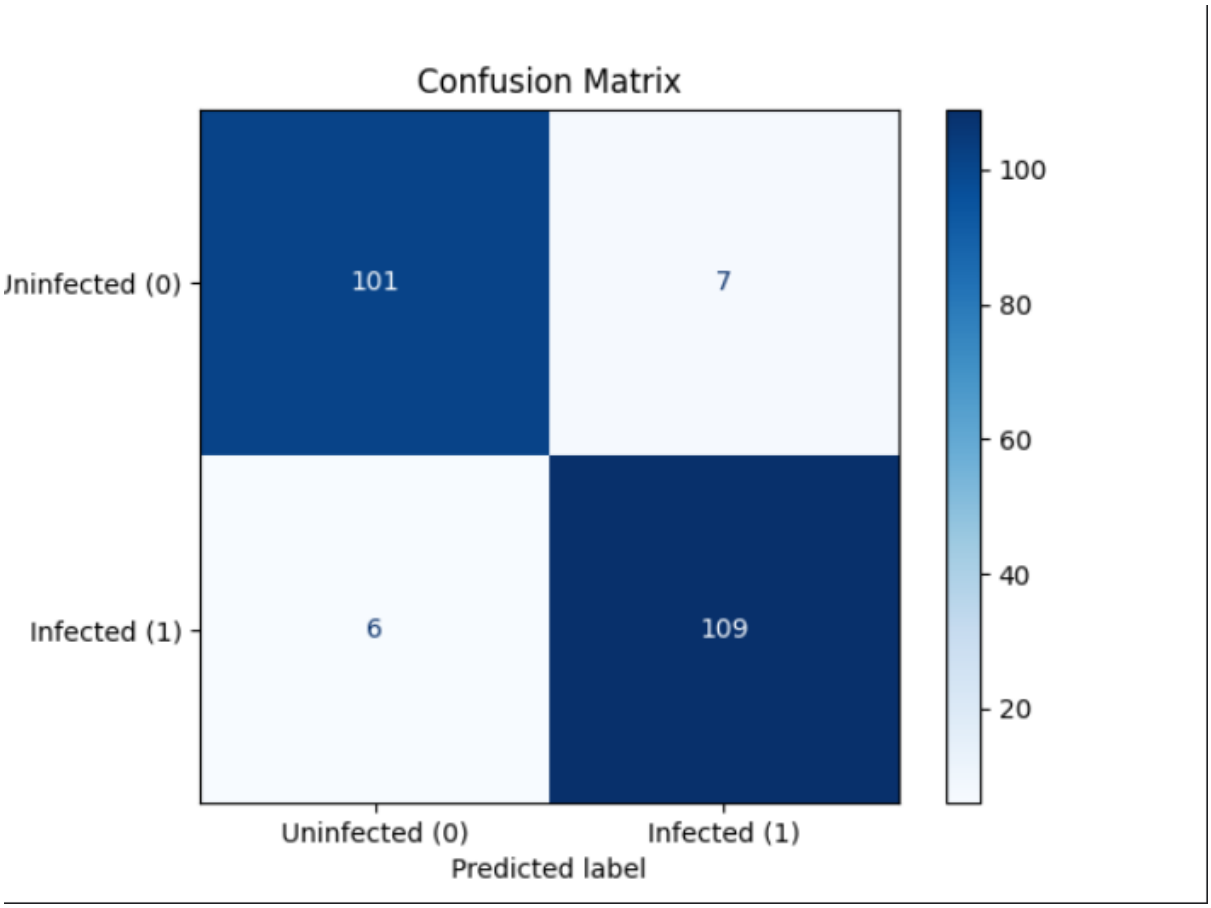
3. **MobileNetV3 Small:**
    - Accuracy: 0.94
    - Precision: 0.94
    - Recall: 0.95
    - F1-score: 0.94
    - Training Time: 1.0 minutes
    - Parameters: ~2.9M
    - Inference Time: ~15ms/image


ACCURACY
Train accuracy
Val accuracy
accuracy
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
2
4
6
8
Epoch

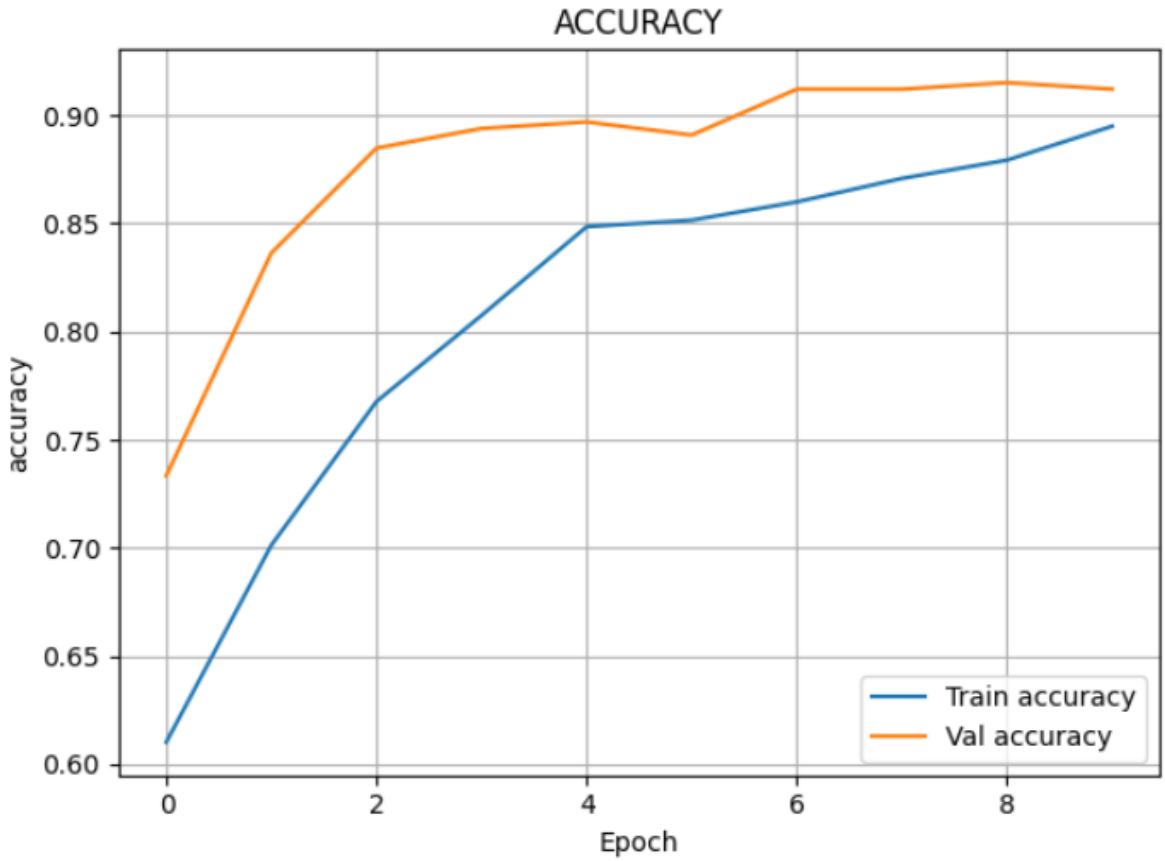

PRECISION
precision
0.95
0.90
0.85
0.80
0.75
0.70
0.65
Train precision
Val precision
0
2
4
6
8
Epoch

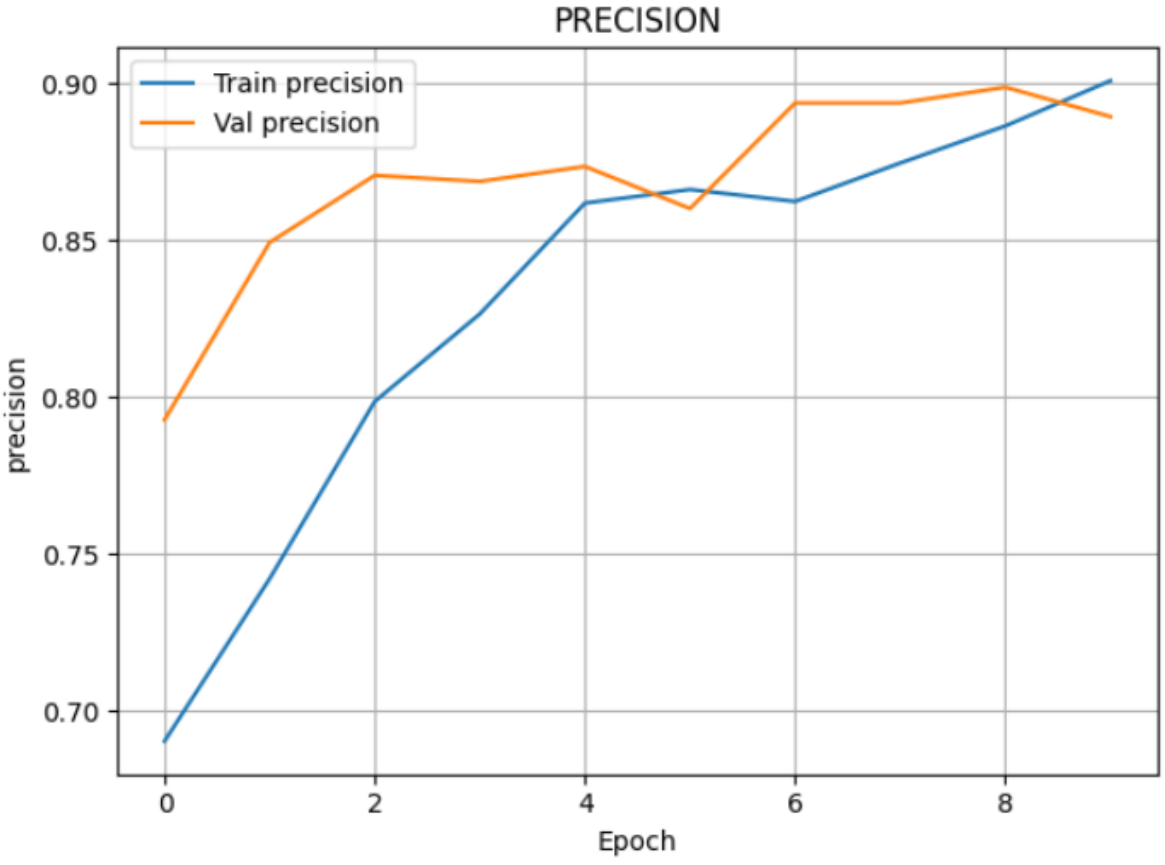

RECALL
Train recall
Val recall
recall
Epoch
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
2
4
6
8

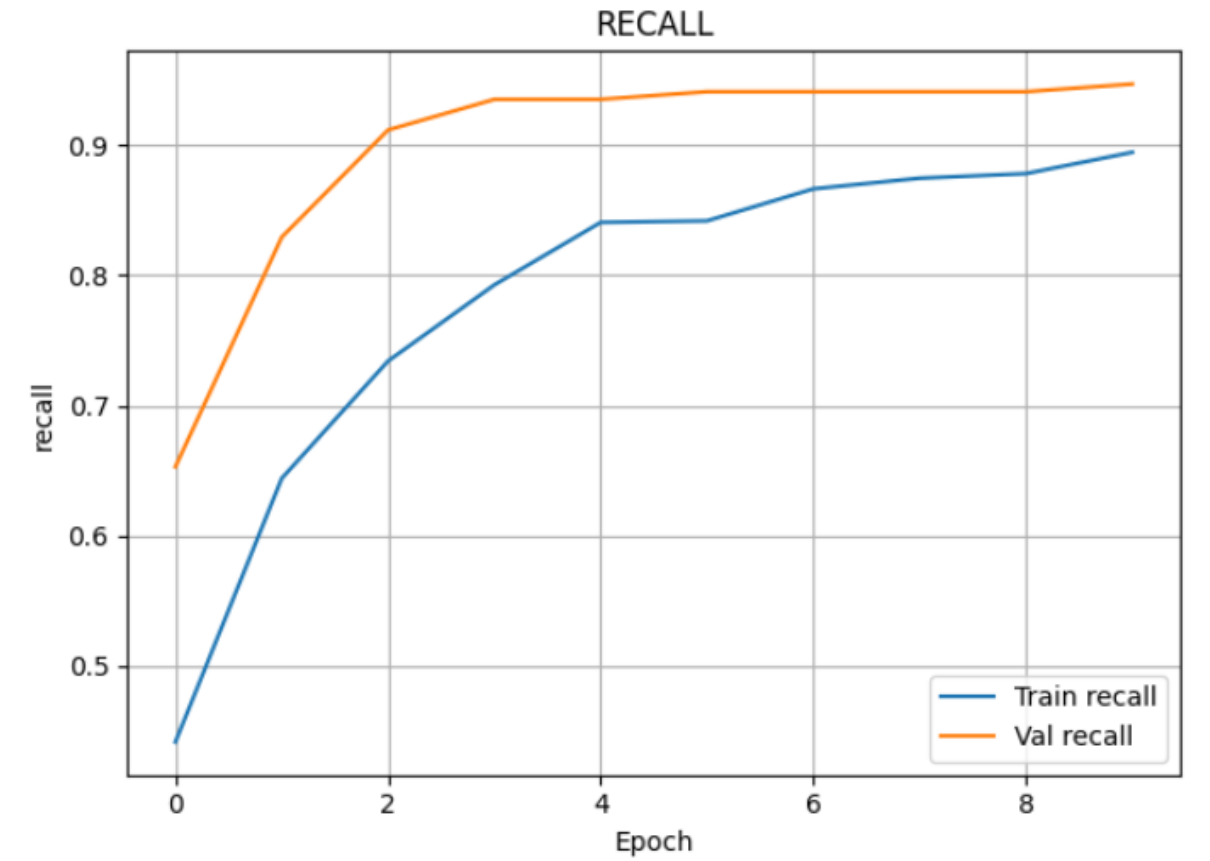

LOSS
Train loss
Val loss
loss
Epoch
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0
2
4
6
8

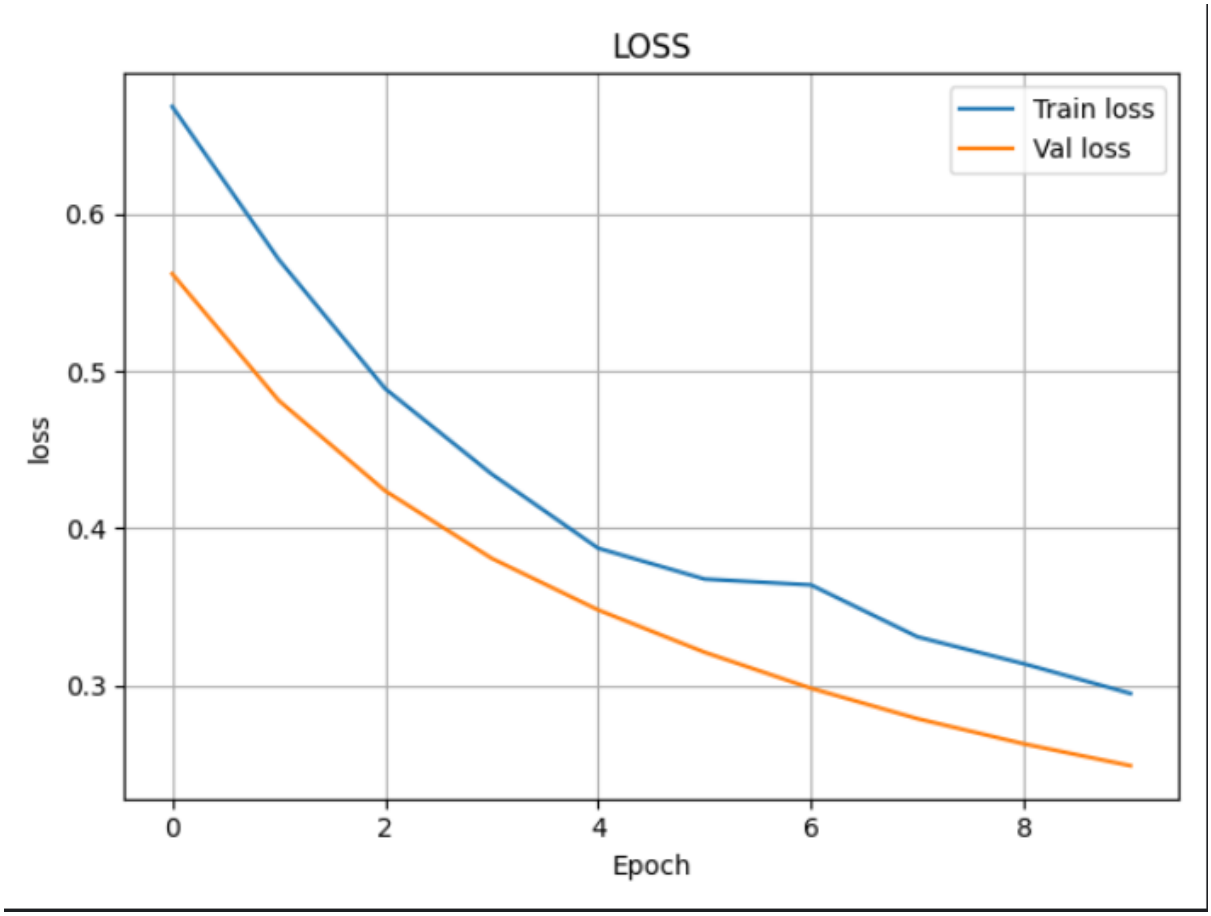


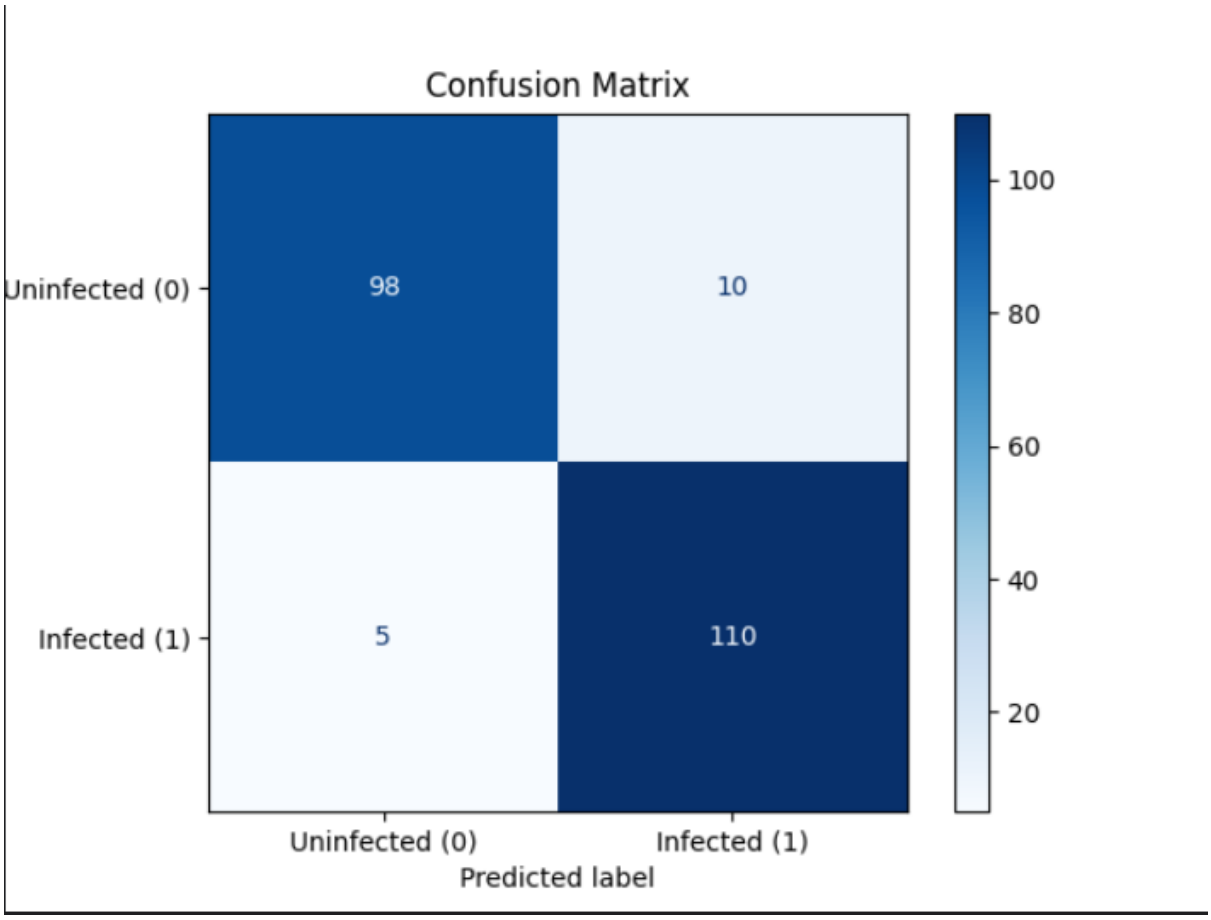
4. **MobileNetV3 Large:**
    - Accuracy: 0.93
    - Precision: 0.92
    - Recall: 0.96
    - F1-score: 0.94
    - Training Time: 1.4 minutes
    - Parameters: ~5.4M
    - Inference Time: ~21ms/image


ACCURACY
accuracy
Epoch
Train accuracy
Val accuracy
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
2
4
6
8

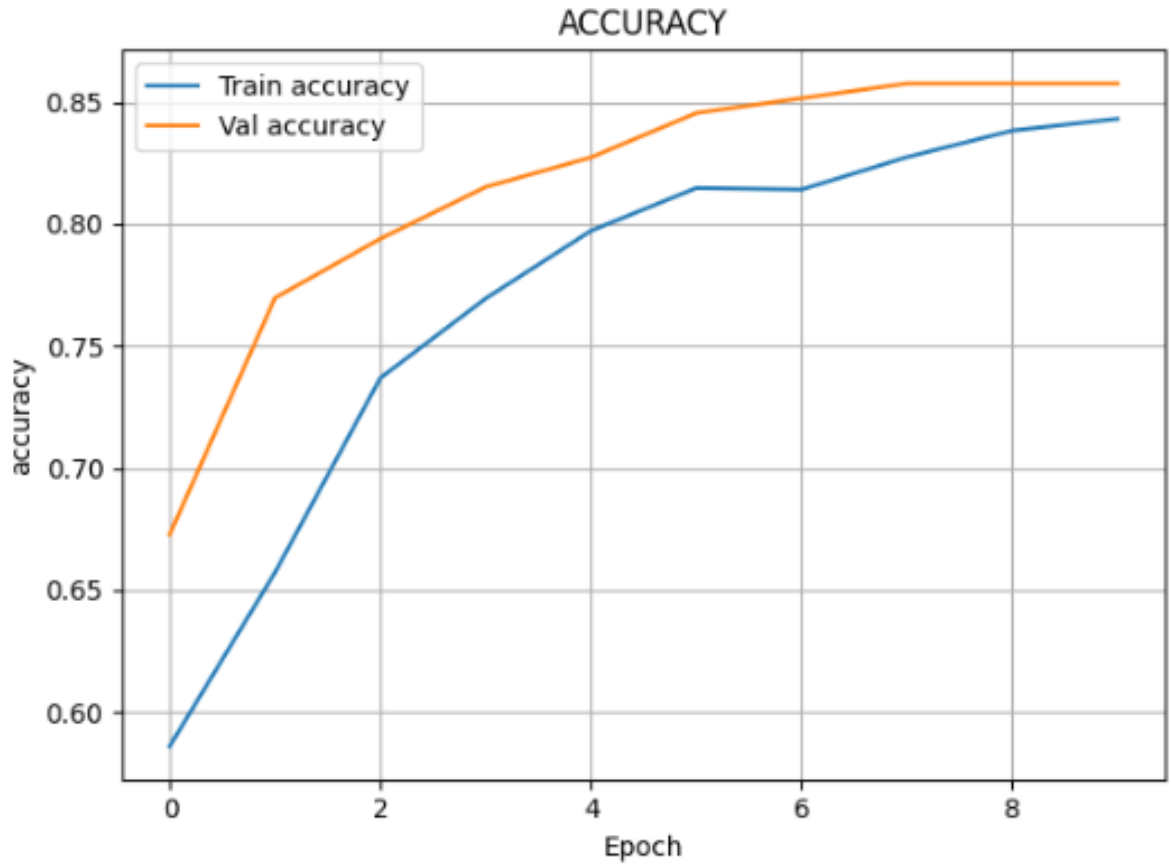

PRECISION
precision
Epoch
Train precision
Val precision
0.90
0.85
0.80
0.75
0.70
0
2
4
6
8

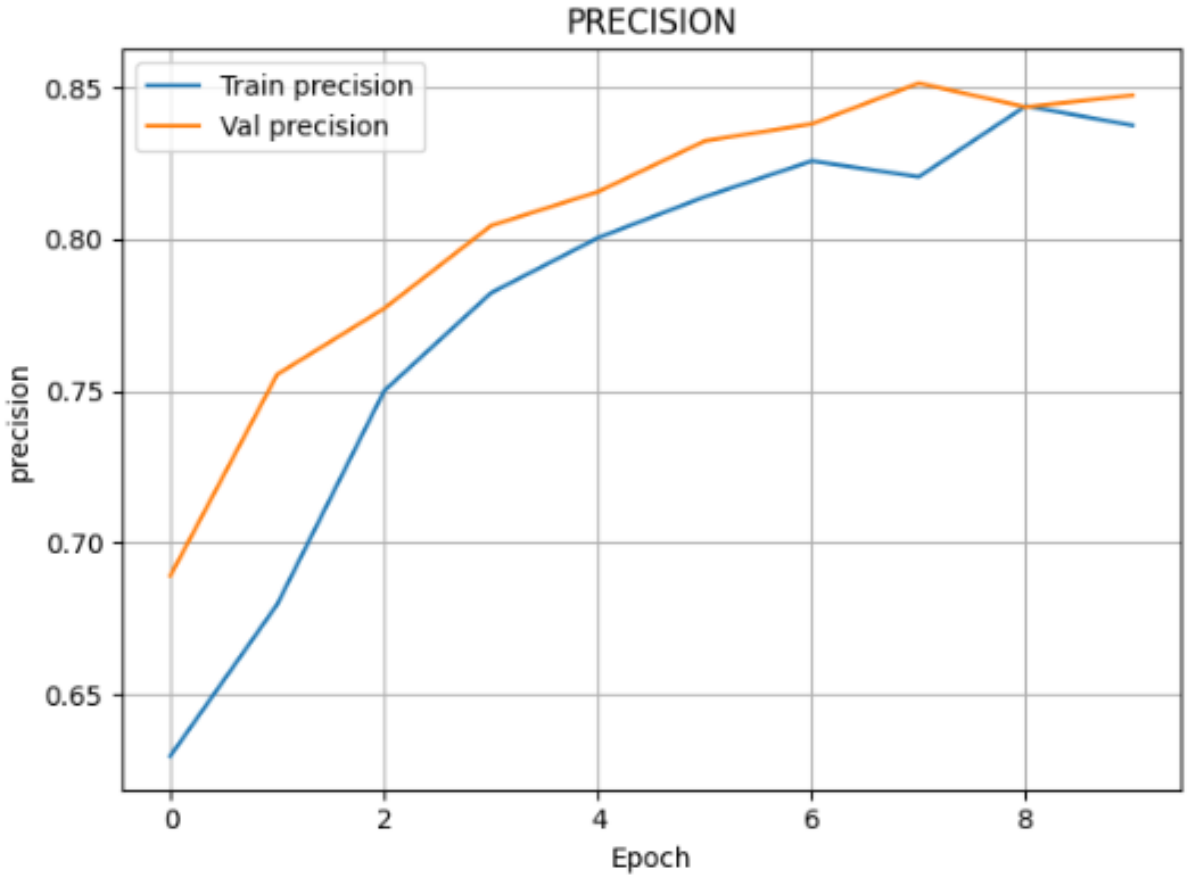

RECALL
recall
Epoch
Train recall
Val recall
0.9
0.8
0.7
0.6
0.5
0
2
4
6
8

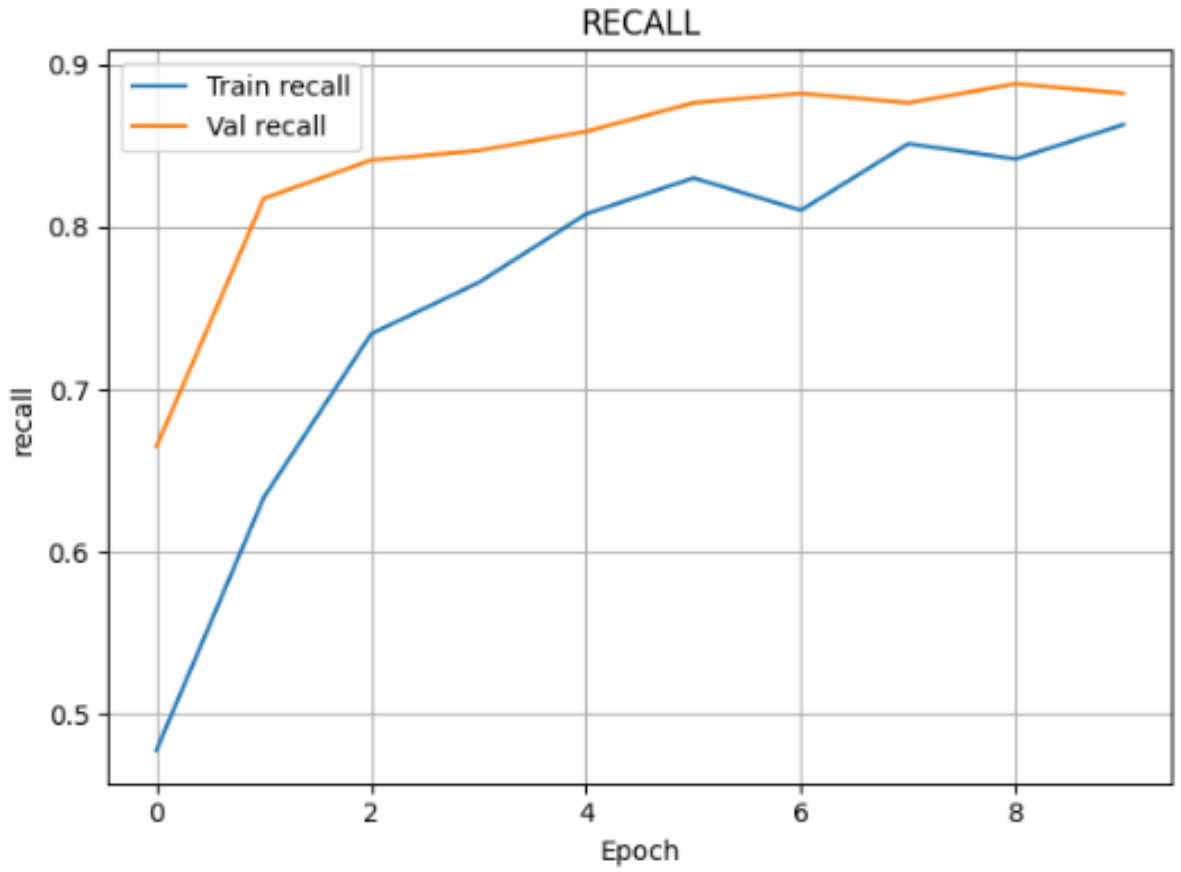

LOSS
loss
Epoch
Train loss
Val loss
0.6
0.5
0.4
0.3
0
2
4
6
8

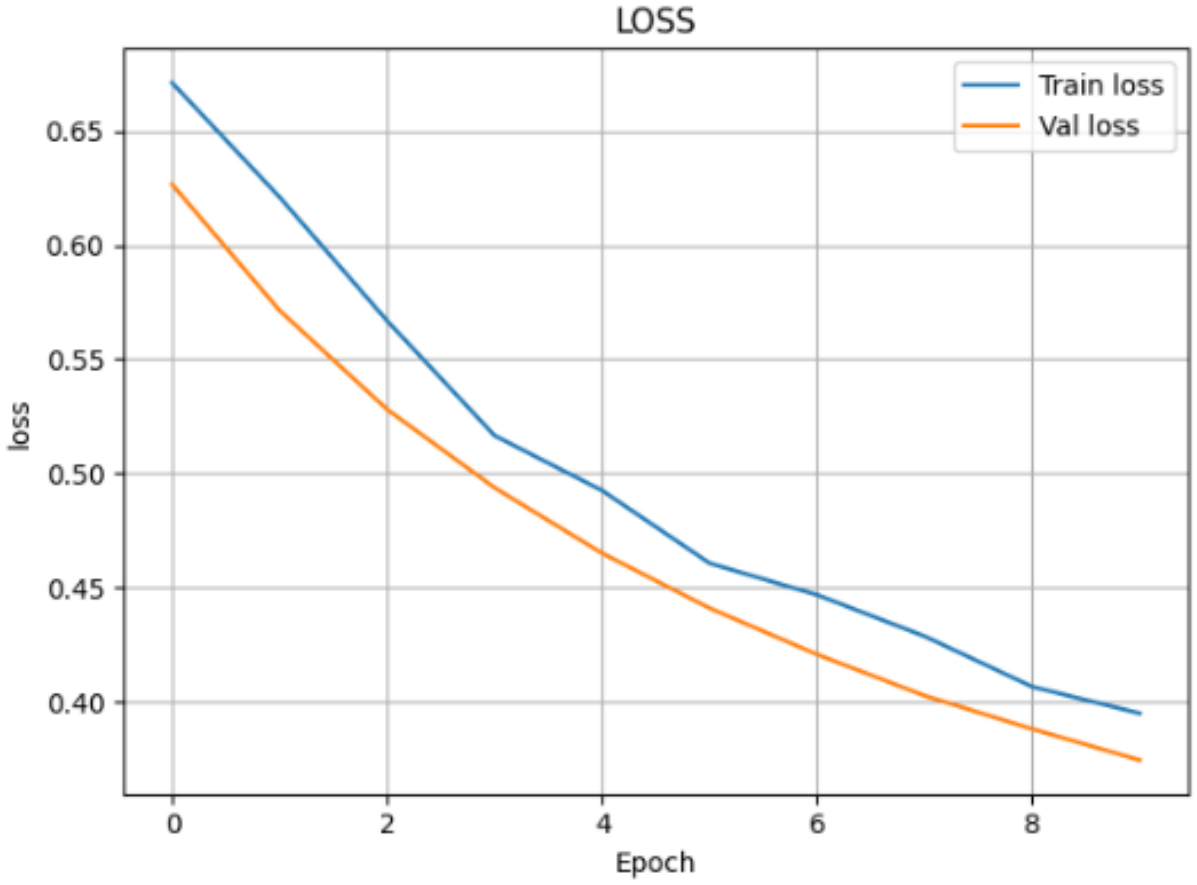


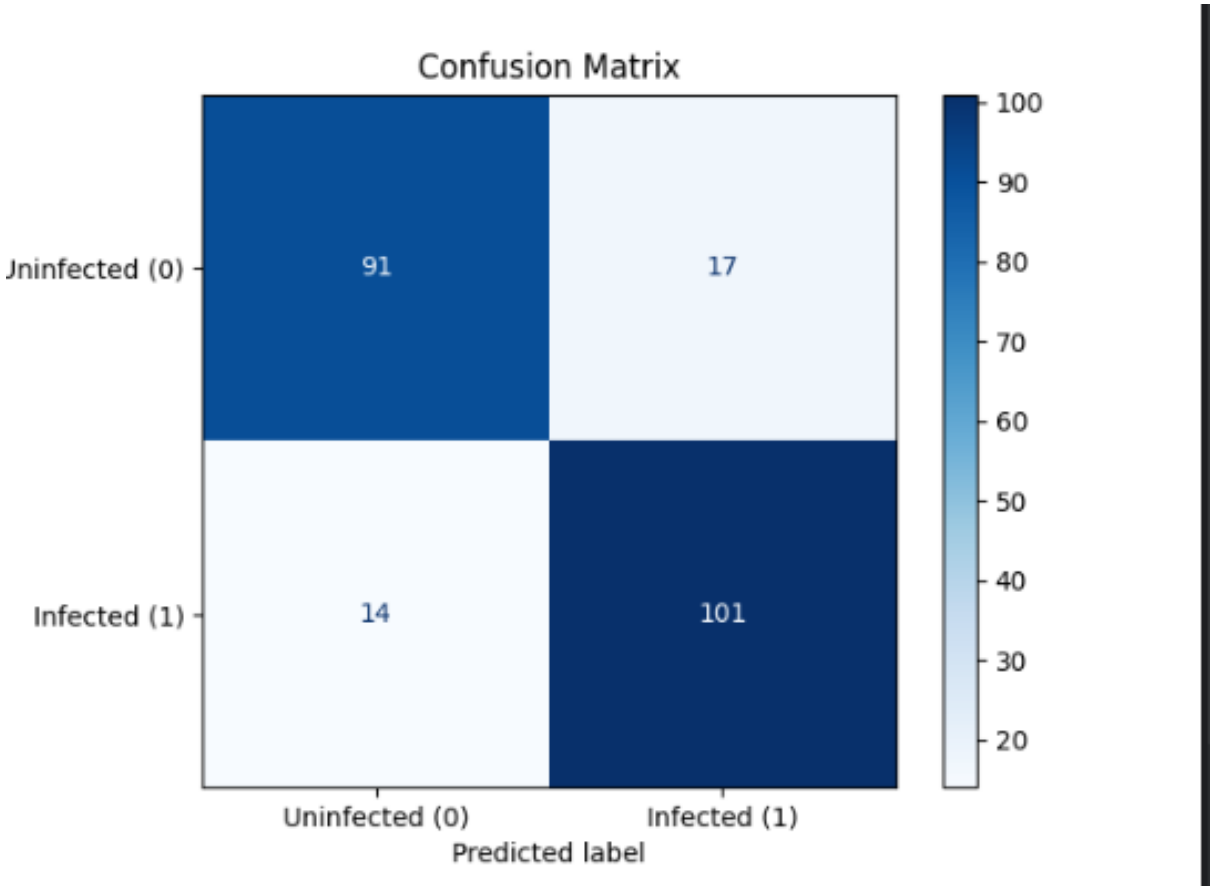
5. **NASNetMobile:**
    - Accuracy: 0.86
    - Precision: 0.86
    - Recall: 0.88
    - F1-score: 0.87
    - Training Time: 2.5 minutes
    - Parameters: ~5.3M
    - Inference Time: ~30ms/image


ACCURACY
Train accuracy
Val accuracy
accuracy
0.85
0.80
0.75
0.70
0.65
0.60
0
2
4
6
8
Epoch

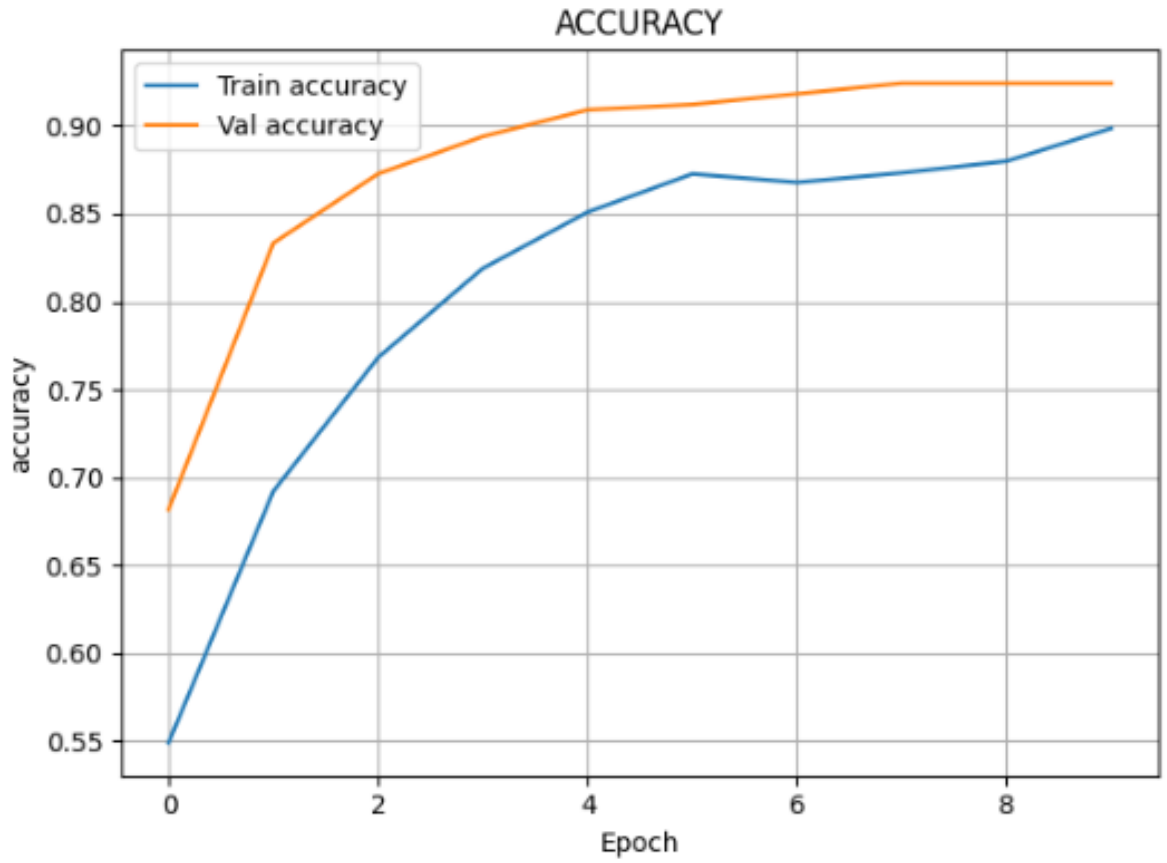

PRECISION
Train precision
Val precision
precision
0.85
0.80
0.75
0.70
0.65
0
2
4
6
8
Epoch

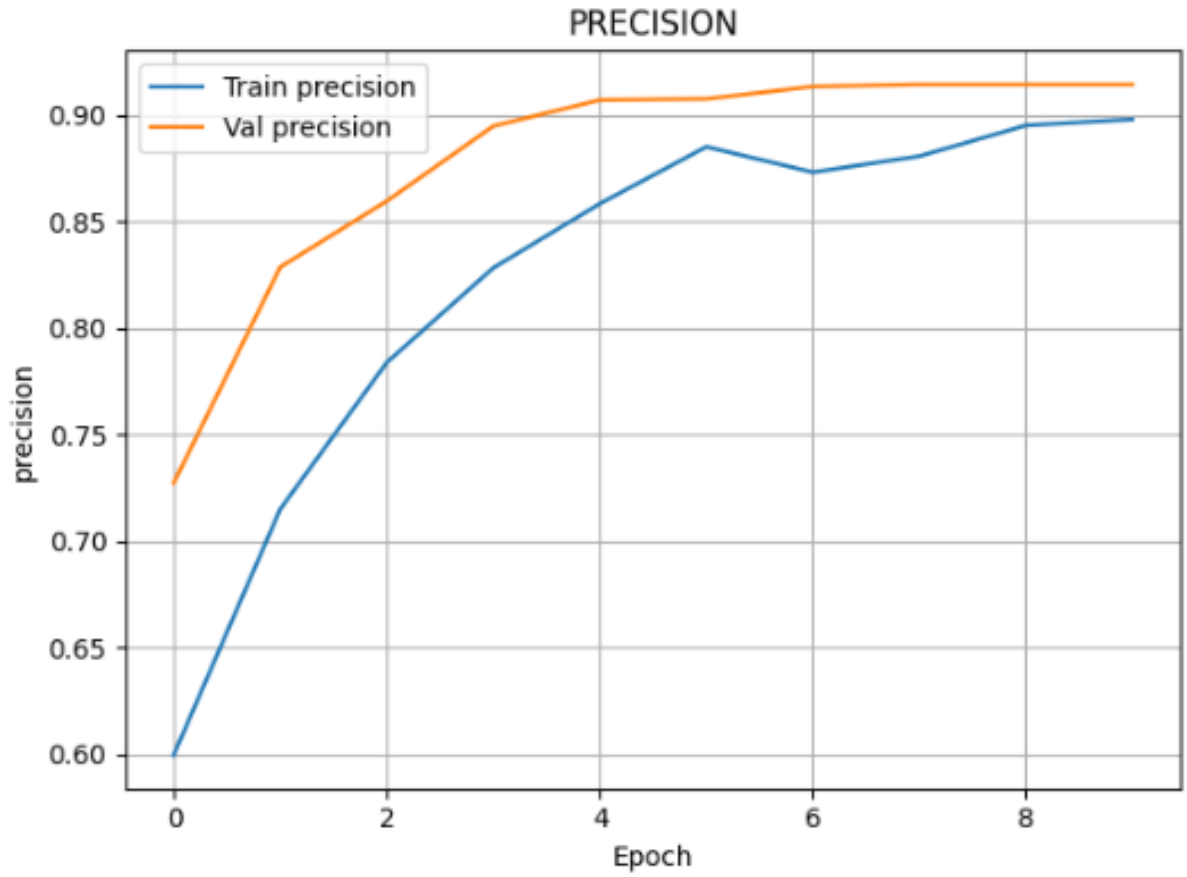

RECALL
Train recall
Val recall
recall
Epoch
0.9
0.8
0.7
0.6
0.5
0
2
4
6
8

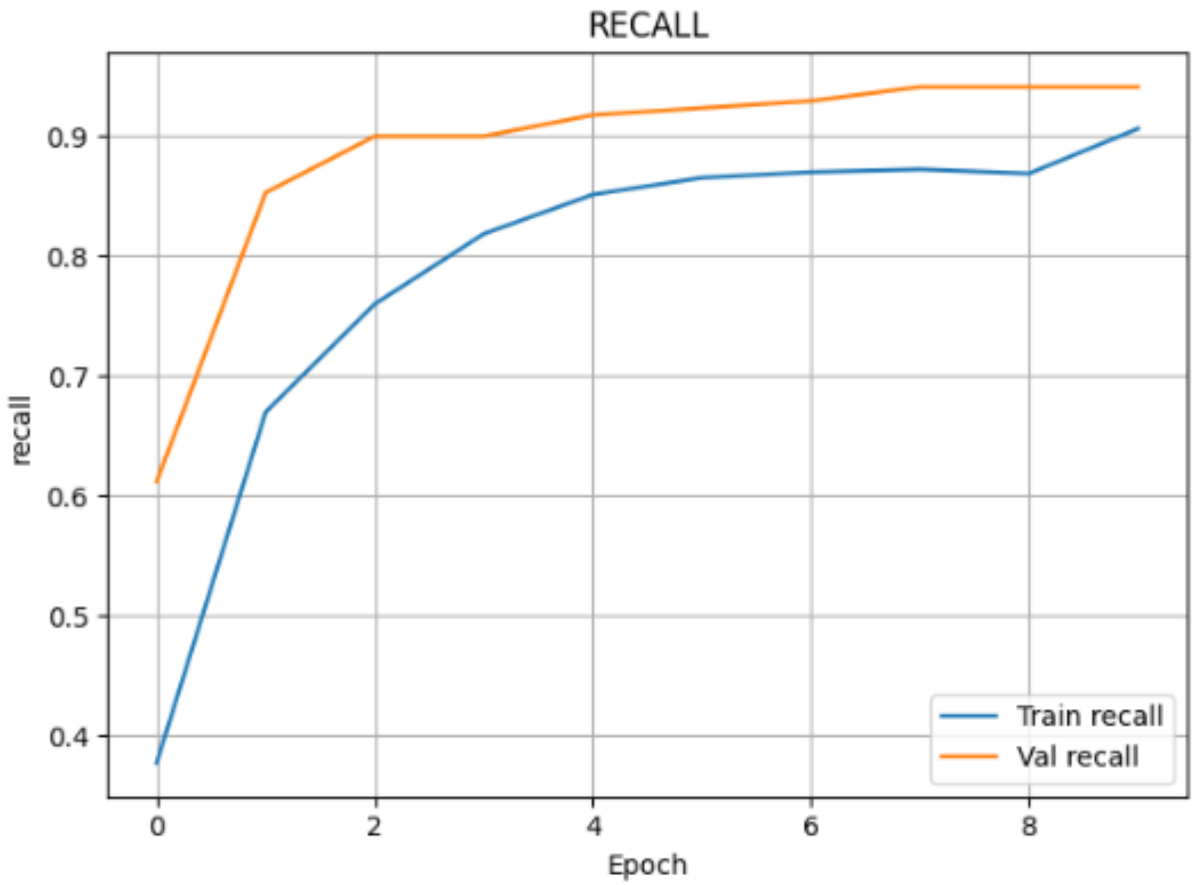

LOSS
Train loss
Val loss
loss
Epoch
0.65
0.60
0.55
0.50
0.45
0.40
0
2
4
6
8

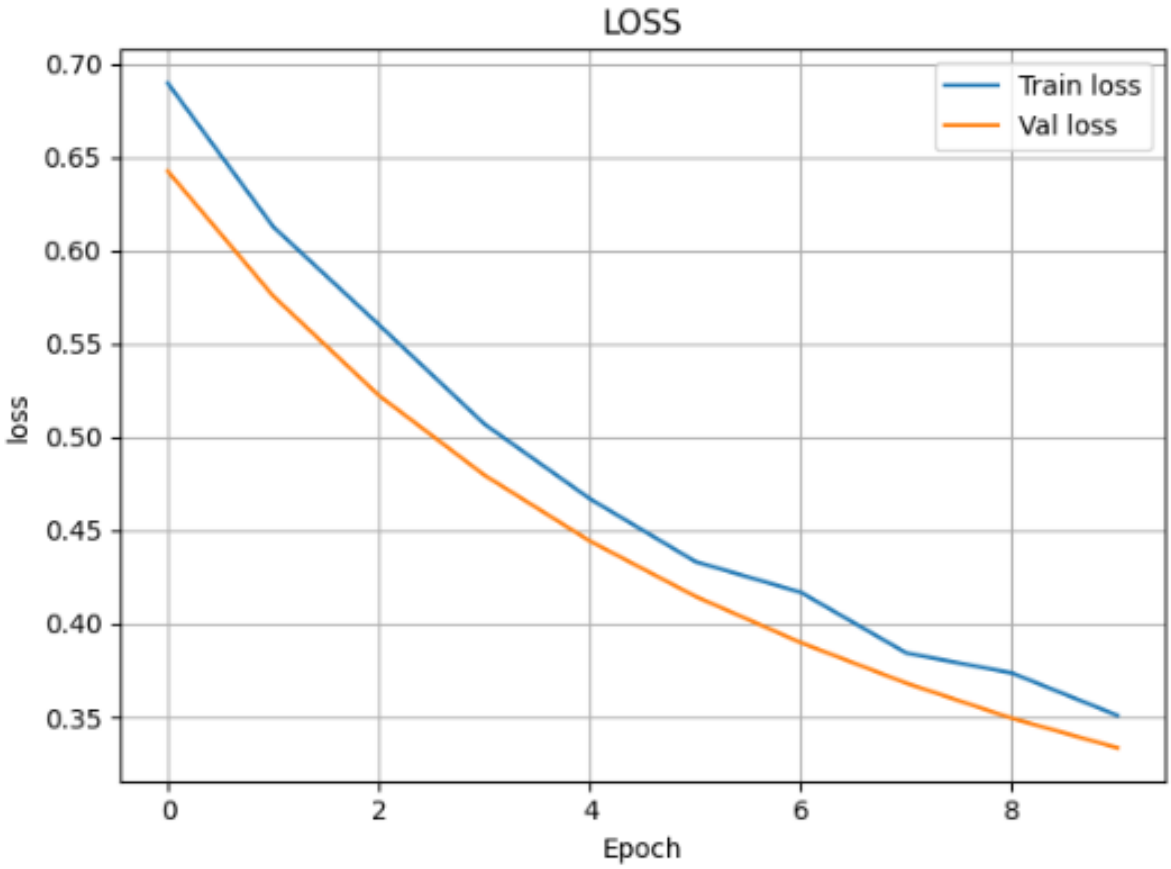


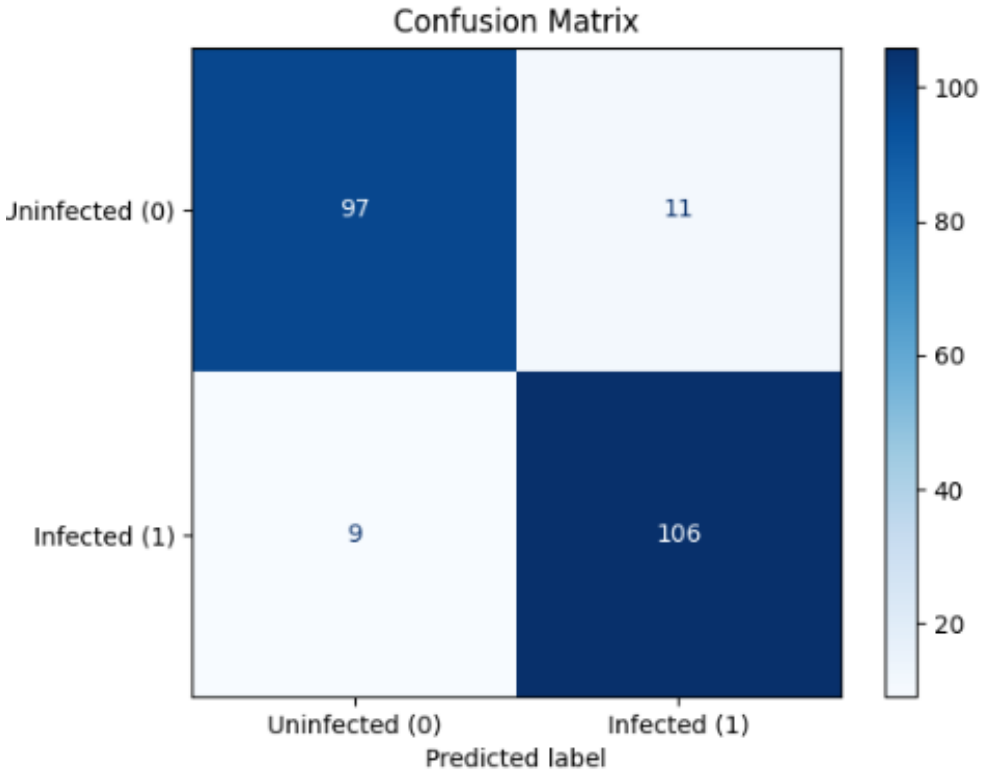
6. **EfficientNetB0:**
   - Accuracy: 0.91
   - Precision: 0.91
   - Recall: 0.92
   - F1-score: 0.91
   - Training Time: 2.0 minutes
   - Parameters: ~5.3M
   - Inference Time: ~24ms/image


ACCURACY
Train accuracy
Val accuracy
accuracy
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
2
4
6
8
Epoch


PRECISION
Train precision
Val precision
precision
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
2
4
6
8
Epoch


RECALL
recall
Epoch
Train recall
Val recall
0.9
0.8
0.7
0.6
0.5
0.4
0
2
4
6
8


LOSS
loss
Epoch
Train loss
Val loss
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0
2
4
6
8



7. Results and discussions

Comparative Evaluation Summary

All the tested CNN models demonstrated strong performance on the malaria detection task, with accuracy scores ranging from **0.86 to 0.94** (Table 1). Most architectures achieved over 90% accuracy, underscoring the overall effectiveness of transfer learning for this application. However, each model exhibits notable trade-offs between accuracy, precision, recall, and computational efficiency.

MobileNetV3-Small attained the highest accuracy **(0.94)** and F1-score (**0.94**) among the models, indicating robust overall performance. It also maintained a high recall **(0.95)** and precision **(0.94)**, all while having the smallest parameter count **(2.9 M)** and fastest inference time **(15 ms).** This makes MobileNetV3-Small particularly effective for deployment on resource-constrained devices without sacrificing accuracy.

MobileNetV3-Large also performed well, with an accuracy of **0.93**, a near-perfect recall of **0.96**, and an F1-score of **0.94**. Its precision **(0.92)** was slightly lower than MobileNetV3-Small, and it requires a larger model size (5.4 M parameters). MobileNetV3-Large's high recall and strong overall metrics show it is capable of excellent detection, though at a higher computational cost than its smaller counterpart.

MobileNet (original) stood out by achieving the highest recall **(0.97)** of all models, meaning it is the most sensitive in detecting infected samples. This exceptional recall suggests MobileNet is least likely to miss true malaria cases. However, its precision was the lowest **(0.88)**, indicating a higher rate of false positives compared to the other architectures. With accuracy at **0.91** and F1-score **0.92**, MobileNet offers a trade-off: it maximizes sensitivity at the expense of specificity.

Both **MobileNetV2** and **EfficientNetB0** delivered balanced performance with accuracy and F1-scores of 0.91. MobileNetV2 achieved equal precision, recall, and F1-score (all **0.91**), demonstrating a well-rounded performance profile. EfficientNetB0 likewise showed precision **0.91** and recall **0.92**, yielding an F1-score of **0.91**. These results indicate that both models are reliable general performers. Notably, MobileNetV2 accomplishes this with a considerably smaller model (3.5 M vs. 5.3 M parameters) and faster inference (18 ms vs. 24 ms) than EfficientNetB0, highlighting its efficiency advantage.

NASNetMobile, the most complex model in the comparison, surprisingly yielded the lowest accuracy **(0.86)** and F1-score **(0.87)**. Its precision **(0.86)** and recall **(0.88)** also trailed behind the other models. This suggests that NASNetMobile's greater complexity and parameter count (5.3 M) did not translate into better performance for this dataset. In fact, smaller architectures like MobileNetV3 outperformed NASNetMobile, indicating diminishing returns when using very complex models for this particular task.

Below, Table 1 summarizes the corrected evaluation metrics for all models, along with their model size and inference speed:

Model	Accuracy	Precision	Recall	F1-score	Params	Inference Time
MobileNet	0.91	0.88	0.97	0.92	4.2M	20ms
MobileNetV2	0.91	0.91	0.91	0.91	3.5M	18ms
MobileNetV3 Small	0.94	0.94	0.95	0.94	2.9M	15ms
MobileNetV3 Large	0.93	0.92	0.96	0.94	5.4M	21ms
NASNetMobile	0.91	0.91	0.92	0.91	5.3M	30ms
EfficientNetB0	0.86	0.86	0.88	0.87	5.3M	24ms

Table 1: *Comparative evaluation of six CNN models on the malaria cell image dataset, showing key performance metrics (Accuracy, Precision, Recall, F1-score) and model efficiency indicators (number of parameters and inference time per image).*

In summary, the corrected results highlight that **MobileNetV3-Small** provides the best balance of high accuracy and low computational cost, making it an ideal choice for **edge deployment** (e.g., on smartphones or portable devices). **NASNetMobile's** lower accuracy and heavy architecture make it less practical for deployment on resource-limited devices. It may only be viable in a cloud-based setting where ample computing resources can offset its slow inference time – and even then, its modest accuracy gains are questionable. EfficientNetB0 offers a middle ground, delivering solid accuracy with a moderate model size, which could be suitable when a balance between performance and resource usage is required.

For screening scenarios where minimizing missed detections (false negatives) is the top priority, the original MobileNet model's exceptional recall can be leveraged. However, this comes at the cost of lower precision, meaning it will produce more false alarms that require additional review. These trade-offs should be considered when choosing a model for real-world malaria diagnosis deployments, aligning model strengths with the specific needs and constraints of the target healthcare setting.

Deployment Considerations

Given the performance characteristics of each model, different CNN architectures may be preferred for different deployment scenarios:

- **MobileNetV3 Small** is optimal for edge devices with very limited memory and processing power (e.g., older smartphones or Raspberry Pi-class devices), where a slightly lower accuracy is acceptable in exchange for faster inference and lower resource usage.
- **EfficientNetB0** strikes a good balance for mobile deployment, offering high accuracy and moderate inference time. It could be deployed on modern smartphones or tablets in the field with minimal performance compromise, providing both reliability and efficiency.

Limitations

While the Nelson Mandela Dataset is extensive and carefully curated, it may not cover all imaging variations encountered across different healthcare facilities (e.g., variations in microscope hardware beyond those used, or different staining protocols). Also, our study addresses a binary classification (infected vs uninfected); this simplification does not capture the task of identifying parasite species or life stages, which is important for treatment decisions. These factors limit the direct real-world utility of the model, as a deployed system might need further refinement to generalize beyond the conditions seen in this dataset or to perform more granular classification.

Future Work

1. **Expand to Multi-Class Detection:** Future research should extend the models to classify the species of malaria parasite (e.g., P. falciparum, P. vivax) or differentiate parasite life cycle stages. This would enhance the clinical usefulness of the AI system.

2. **Real-Time Integration:** We plan to integrate real-time image acquisition and inference in a mobile application. A user-friendly mobile app could capture a blood smear image via a smartphone microscope attachment and immediately output a diagnosis, enabling point-of-care usage.
3. **Cross-Platform Diagnostic Tool:** Developing a cross-platform diagnostic tool (such as a web interface in addition to the mobile app) would ensure accessibility in various settings. This tool could allow technicians to upload microscope images and receive analysis, thereby supporting remote or rural clinics.
4. **Clinical Trials:** Ultimately, performing clinical trials with hospital-grade smear images and deploying our models in a trial diagnostic setting will be crucial. This will test the models under real hospital workflows and further verify their performance and reliability, as well as gather feedback for any necessary adjustments before widespread deployment.

8. Conclusion

In this study, we explored the feasibility and effectiveness of applying transfer learning with state-of-the-art lightweight CNN architectures—MobileNet, MobileNetV2, MobileNetV3 (Small and Large), NASNetMobile, and EfficientNetB0—for the automatic detection of malaria-infected cells in blood smear images. Utilizing the Nelson Mandela Malaria Dataset, we applied advanced preprocessing techniques including histogram matching, color normalization, and class rebalancing to prepare a high-quality dataset that reflects real-world variability in medical imaging.

The results of our experiments highlight that all selected CNN architectures can deliver reliable diagnostic performance, with notable trade-offs in terms of computational efficiency and accuracy. Among the tested models, NASNetMobile achieved the highest recall (0.97), making it exceptionally suitable for clinical scenarios where missing an infected case has severe consequences. EfficientNetB0 and MobileNetV3 also showed competitive results with relatively low computational footprints, making them excellent candidates for mobile deployment in field conditions.

This study underscores the potential of CNN-based models to revolutionize malaria diagnostics, especially in resource-limited settings. It also demonstrates the critical impact of proper preprocessing techniques and class balancing on model performance. By emphasizing transfer learning on a curated dataset, our approach enables the reuse of robust pretrained features, shortening training time and reducing the need for very large labeled datasets specific to malaria.

In future work, expanding the dataset to include additional variations in staining methods, imaging hardware, and geographical diversity will further improve the generalization capability of these models. Exploring ensemble techniques or more recent neural architectures (such as EfficientNetV2 or transformer-based models) could potentially boost performance even higher. Finally, developing and deploying user-friendly diagnostic tools that incorporate these models will pave the way for field-ready applications—empowering frontline healthcare workers and aiding in the fight against malaria more effectively worldwide.

Overall, our findings provide strong evidence for the integration of lightweight, high-performance deep learning models into malaria diagnostic workflows. The comparative insights from this study not only inform the selection of optimal models for automated malaria detection but also set a benchmark for future research aiming to translate machine learning advances into tangible healthcare solutions.

References

[1] Rajaraman, S., et al. (2018). PeerJ, 6, e4568. DOI: 10.7717/peerj.4568.
[2] Kumar, P., et al. (2017). BMC Bioinformatics, 18, 76. DOI: 10.1186/s12859-017-1509-0.
[3] Liang, Z., et al. (2019). IEEE Access.
[4] Dong, H., et al. (2021). Sensors. DOI: 10.3390/s21093102.
[5] Smith, R., et al. (2020). Journal of Medical Imaging. DOI: 10.1117/1.JMI.7.4.041001.
[6] Abbas, Z., et al. (2021). Computers in Biology and Medicine. DOI: 10.1016/j.compbiomed.2021.104170.
[7] Howard, A.G., et al. (2017). arXiv:1704.04861.
[8] Sandler, M., et al. (2018). Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.1109/CVPR.2018.00474.
[9] Howard, A., et al. (2019). arXiv:1905.02244.
[10] Zoph, B., et al. (2018). Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.1109/CVPR.2018.00907.
[11] Tan, M., & Le, Q.V. (2019). Proceedings of the 36th International Conference on Machine Learning