

# Deep Learning and Neural Networks for Malaria Classification

Saheel Ramji  
[ramji.s@northeastern.edu](mailto:ramji.s@northeastern.edu)

Jainil Shah  
[shah.jainil@northeastern.edu](mailto:shah.jainil@northeastern.edu)

Rutuja Doiphode  
[doiphode.r@northeastern.edu](mailto:doiphode.r@northeastern.edu)

## *Abstract—*

Malaria is one of the most important health problems in tropical and subtropical regions, causing millions of infections and deaths every year. Its timely and accurate detection is very important for effective treatment and control. This paper uses deep learning to automate malaria diagnosis by classifying microscopic blood cell images as parasitized or uninfected. Three models—CNN, ResNet50, and InceptionV3—were evaluated to determine the most effective approach. The CNN model had the best performance among the compared models, reaching a test accuracy of 94.54%, with precision, recall, and F1-score at 91.82%, 97.78%, and 94.71%, respectively. InceptionV3 came close, while ResNet50 presented limitations for recall and stability in validations. This work illustrates well the potentiality of deep learning in medical diagnostics, especially within resource-constrained settings, as it eliminates dependency on skilled professionals, advances diagnosis speed, and enhances diagnosis accuracy. The results point out the transformational role of AI in responding to global health challenges and create room for developing low-cost, accessible, and reliable systems for malaria detection

**Keywords—**Malaria Detection, Deep Learning, Convolutional Neural Networks (CNNs), ResNet50, InceptionV3, Microscopic Images, Parasitized and Uninfected Cells.

## 1. INTRODUCTION

### *I. Overview of project*

Malaria is a severe disease caused by Plasmodium parasites that are very common in tropical and subtropical areas. It is conveyed through the bites of female Anopheles mosquitoes and results in millions of infections and considerable mortality every year. Traditional diagnosis based on microscopic examination of blood smears, though effective, is extremely time-consuming and requires skilled personnel, hence prone to many human errors. These challenges, however, are amplified by the inadequacy of health facilities in rural and resource-poor settings. This paper proposes a deep learning-based solution for automating malaria diagnosis by developing a system capable of classifying blood cell images into parasitized and uninfected. In this work, three advanced architectures—custom CNN, ResNet50, and InceptionV3—are trained on a publicly available malaria dataset consisting of 27,558 images. The main keys to detecting malaria-infected cells with high accuracy and minimum human intervention were transfer learning and fine-tuning of pre-trained models. Among the models tested, the highest accuracy achieved was by the custom CNN, with 95%, followed closely by

InceptionV3. ResNet50 performed on the lower side compared to these.

The study points to the transformative potential that AI has in the diagnosis of diseases for faster, more correct, and accessible health care. It automates the process, hence reducing the workload of the medical professionals involved and managing diseases in under-resourced settings—a factor that goes toward contributing to the reduction of mortality caused by malaria.

### *II. Motivation*

Malaria diagnosis remains a challenge, especially in the rural areas where health infrastructures are poor and skilled professionals are few. The usual conventional techniques of diagnosis include the manual examination of blood smears, which is extremely time-consuming and requires a great deal of expertise. Most of the time, misdiagnosis or late identification of the disease results in serious health consequences and increased mortality rates.

This project is motivated by the demand for faster, more accurate, and scalable diagnostics. Deep learning models have a promising future in medical imaging, whereby an automated system could process thousands of samples in mere seconds. This will reduce some of the major challenges in healthcare delivery, such as diagnosis errors, delays resulting in late treatment, and the loss of lives. It is hoped that this work furthers the cause of AI in narrowing health disparities, most especially in resource-constrained settings, adding to the fight against this killer disease..

### *III. Approach*

This project will use deep learning to automate malaria diagnosis from microscopic blood cell images. Three CNN architectures have been evaluated in this project: a custom CNN, ResNet50, and InceptionV3. A dataset of 27,558 images was pre-processed for compatibility with the models. The preprocessing involved resizing all images to 128x128 pixels to meet the input requirements of the models and normalizing pixel values to a range of [0, 1]. The dataset was then split into training (60%), validation (20%), and testing (20%) subsets using stratified sampling to maintain class balance.

In contrast to common practices of image-based deep learning, neither flipping, rotation, nor zooming data augmentation techniques were done for this project. Here, the models were to be trained on raw data, simply leveraging diversity and balance in the data for good generalization performance.

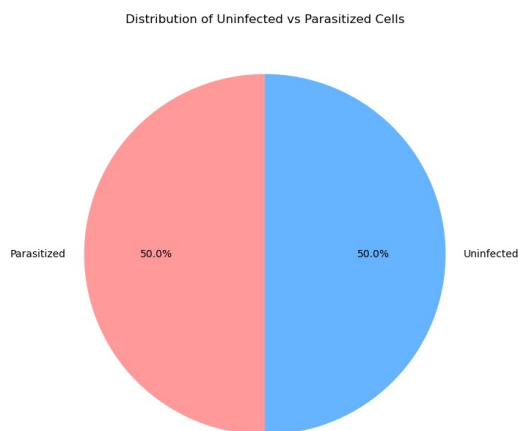
In the process of transfer learning, the initial layers in ResNet50 and InceptionV3 were frozen to apply pre-trained features from ImageNet, while fine-tuning the later layers for the task of malaria classification. Each model was trained with a binary cross-entropy loss function and the Adam optimizer, using early stopping and scheduling the learning

rate to avoid overfitting and optimize their performances. The performance metrics used for evaluation include accuracy, precision, recall, and F1-score. The best performance was observed for the custom CNN, which was closely followed by InceptionV3, while ResNet50 resulted in relatively low recall and F1-scores, and thus needs further improvements.

#### IV. Dataset Used

The dataset used in this work is the NIH Malaria Dataset, from National Library of Medicine – National Institutes of Health, consisting of 27,558 labelled images of parasitized and uninfected blood cells. This dataset is quite balanced, having 13,779 images for each class to avoid any bias during binary classification tasks. This highly qualitative dataset, with the process, served as the best avenue for assessing the deep learning models. The dataset can be downloaded from the following link:

[https://data.lhncbc.nlm.nih.gov/public/Malaria/cell\\_images.zip](https://data.lhncbc.nlm.nih.gov/public/Malaria/cell_images.zip)



## 2. BACKGROUND

According to ResNet50, introduced by He et al. [1], deep learning was revolutionized by the introduction of residual connections in order to solve the vanishing gradient problem. This allows deeper networks to train effectively by allowing information to bypass layers, thus facilitating robust feature extraction. Among such techniques, ResNet50 can easily manage complex architectures for hierarchical feature extraction and therefore is an important choice for this work. In the case of this project, this neural network was pre-trained further with ImageNet and then fine-tuned on the malaria dataset. This model shows a good possibility that ResNet50 holds for automatic malaria detection in blood smear images.

InceptionV3 is proposed by Szegedy et al. [2], which adopts factorized convolutions and batch normalization to improve the computational efficiency and enhance performance. This architecture is also suitable for multiscale feature extraction, thus appropriate for locating minute patterns in medical images. Guided by its excellence in dealing with high-resolution data, this work applied InceptionV3 to classify parasitized and uninfected cells. The architecture performed well and complemented the results from ResNet50 in highlighting subtle variations in malaria-infected cells.

Rajaraman et al. [3] estimated the capability of deep learning for medical diagnosis and pointed out the effective work done by CNN, which replaces the role traditionally done under a microscope. Therefore, this work considered benchmarking ResNet50, InceptionV3, and a custom CNN on the malaria dataset. Since most works place a significant emphasis on the metrics of precision, recall, and F1-score, this work puts extra emphasis on these during its evaluation to ensure strong and reliable results.

Consequently, Zhu et al. 4 and Mahajan et al. 5 have presented practical applications using ResNet50 and InceptionV3 for the automation of medical image analyses, thus further justifying the choice of these architecture options as ideal for the diagnosis of malaria infection. Results from their findings have been considered in designing these models and training methodologies used during their approach to this task. The dataset used consists of a balanced set of blood smear images from the National Institutes of Health [7], labeled parasitized or uninfected, hence serving as a perfect benchmark for evaluating deep learning models in this domain.

## 3. APPROACH

Three deep CNN architectures like CNN, ResNet50, and InceptionV3 have been employed to solve the problem of malaria detection using deep learning. First, preprocessing-division into training, validation, and test sets for robust performance evaluation of the dataset of the parasitized and uninfected blood cell images was performed. Each model was prepared for binary classification.

CNN contains simple but effective layers: convolution, pooling, and dense connectivity that captures image patterns hierarchically. ResNet50 used residual blocks to avoid the vanishing of gradients and thus enabled better training of deeper networks. InceptionV3 extracted multi-scale features through its inception modules, which enabled the network to learn detailed and abstract features simultaneously. These models are trained using the Adam optimizer, which includes early stopping to avoid overfitting, and their performance is checked using metrics like accuracy, precision, recall, and F1 score.

### I. Data Preparation and Splitting

The dataset is split in 60-20-20 for training, validation, and testing, respectively. Stratified sampling has been done such that class balance across the splits was maintained with an equal number of parasitized and uninfected cells in each subset. All images were resized to the same size, 128x128 pixels, and their pixel values were normalized within the range of [0, 1]. These steps preprocessed the data well enough for training and testing of the deep learning models.

### II. CONVOLUTIONAL NEURAL NETWORK

In regard to deep learning, CNN is a very essential part; using CNN in the analysis of images, for example, will apply to medical image analysis. In addition, it has been noticed that CNN automatically learns feature hierarchies of representations from input images; because of this, CNNs easily take hold of intricate details that represent visual content. As explained before, CNN will be great at detecting and differentiating between parasitized and uninfected blood

cells by extracting very minute features of this class that the human naked eye could hardly identify manually.

A CNN designed for this project utilized three convolutional layers, each with a  $3 \times 3$  kernel size to leave space for the detection of the spatial features. Each convolutional layer followed a ReLU activation function to handle linearity and batch normalization for regularization and normalization of weights and inputs. On top of these convolution layers came the MaxPooling ones for progressive reduction in spatial dimensions with maintenance of essential information in a feature map. In order to avoid overfitting, Dropout layers were added after each convolutional block, with rates increasing progressively from 20% to 40%. Further, a Global Average Pooling layer was applied on top of the feature maps to summarize the spatial information, followed by a dense layer of 128 neurons serving as a classifier. The final layer, with the sigmoid activation function, provided binary classification for parasitized versus uninfected blood cells.

The architecture was designed to process images hierarchically, extracting low-level features such as edges and high-level features such as cellular structures. This custom CNN combined convolutional, pooling, and fully connected layers to ensure efficiency in feature extraction and classification, thus being suitable for automated malaria diagnosis.

### III. ResNet

ResNet or Residual Convolutional Network, is a deep learning breakthrough that resolves such problems as the vanishing gradient problem and accuracy degradation of very deep networks. The core of its novelty is residual learning, which allows shortcut connections to skip one or more layers. It enables the network to learn residual mappings instead of direct mappings, which refers to small refinements based on the identity mapping. Such mechanism provides an efficient optimization, even for very deep-layered networks, and makes ResNet very effective in complex image analysis tasks. In this work, the 50-layer variant of ResNet, ResNet50, is utilized for malaria cell classification. The depth of ResNet50, combined with its residual blocks, allows it to extract detailed hierarchical features from low-level textures to high-level patterns while avoiding overfitting. The remaining blocks combine identity mappings with known transformations. This ensures efficient feature propagation and better gradient flow through the network. This architecture is ideal for tasks that require microscopic analysis, such as distinguishing parasites from uninfected red blood cells. Pre-trained weights from ImageNet were used to tune ResNet50 using transfer learning. The first 140 layers of ResNet50 are frozen, retaining general feature extraction capabilities. The other layers were fine-tuned so that the model could start specializing in malaria detection and adapt its learned representation to the dataset. A Global Average Pooling layer followed the convolutional backbone to condense spatial features, and a Dense layer with 256 neurons was added to capture task-specific patterns. A 50% dropout layer prevents over-fitting. And the last dense layer provides a binary classification with a sigmoid activation

function.

Unlike standard CNNs, ResNet's architecture allows training much deeper networks without affecting performance. Therefore, it is more flexible for complex datasets such as malaria blood test images. Hierarchical performance and fine-tuning capabilities make it an ideal choice in this application. It provides a scalable and efficient solution for the automatic diagnosis of malaria.

### IV. Inception V3

InceptionV3 is a high-performance deep learning architecture optimized for processing efficiency and accuracy in image classification tasks. The feature definition of the Inception module uses multiple enhancement filters of different sizes simultaneously to allow spatial feature extraction at different parameters. In this project, InceptionV3 was fine-tuned for malaria detection using pre-trained ImageNet weights. The first 249 layers of the model were frozen to maintain general feature extraction capability. The remaining layers are tuned for task-specific learning. Multiscale feature extraction and computational efficiency. Its architecture makes it ideal for searching for micropatterns in parasitic and non-infected red blood cells. It excels at capturing features at multiple levels. of abstract things together. By leveraging these unique strengths, InceptionV3 has demonstrated its strength in medical image analysis. This is critical to the development of a reliable and scalable malaria diagnosis system.

## 4. RESULTS

### I. Convolutional Neural Network

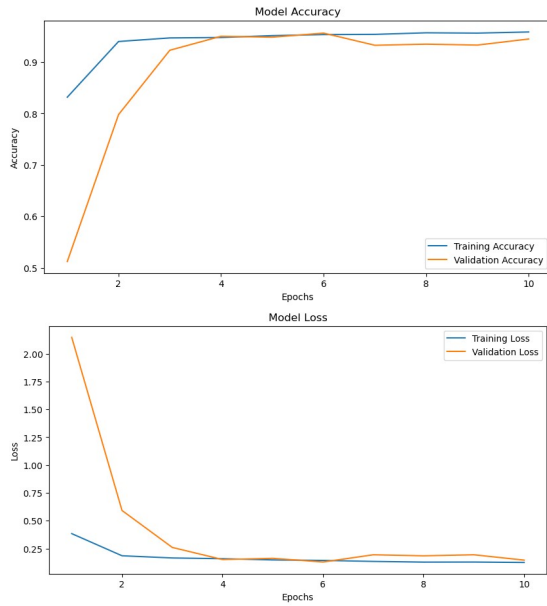
The custom CNN architecture consists of three alternating layers with ReLU activation followed by MaxPooling and Dropout layers to reduce overfitting. The model was trained for 10 epochs with stop-start timing and learning rate using the ADAM optimizer. Performance indicators and visualizations are as follows:

#### 1. Performance Metrics:

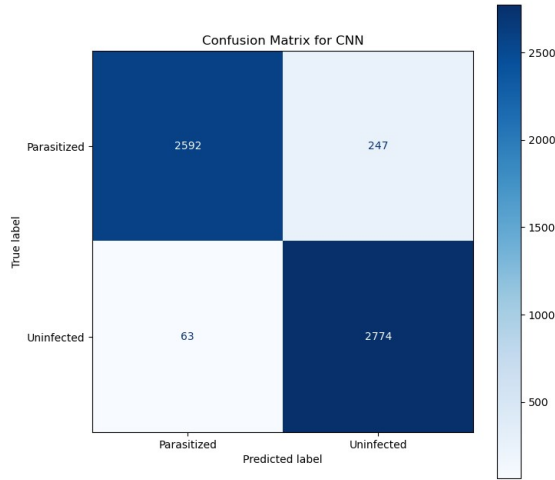
- **Accuracy:** 95.6%
- **Precision:** 92%
- **Recall:** 98%
- **F1-Score:** 95%

#### 2. Graphical Representations:

- **Figure 2:** Model Accuracy and Model Loss during training and validation. These graphs indicate continued improvement in training and validation accuracy. with minimal over-installation This can be seen from the loss graph.
- **Figure 3:** Confusion Matrix for the CNN model. The matrix illustrates the number of true positives, true negatives, false positives, and false negatives, showcasing the model's high classification accuracy.



**Fig. 2.** Model Accuracy and Model Loss during training and validation



**Fig. 3.** Confusion Matrix for the CNN model

The custom CNN was trained in 10 epochs with fast stopping and learning rate scheduling. In the first epoch, the model started with a training accuracy of 73.01% and validation accuracy. Must be 51.22% at the age of six. It has a training accuracy of 95.18% and a verification accuracy of 95.63%, which shows good learning and generalization. In the 10th generation, verification accuracy is stable at 94.48% and accuracy in Training as high as 95.59%

The training loss has continuously gone down from 0.5358 to 0.1280, while the validation loss also decreased significantly from 2.1460 to 0.1467, proving the effective reduction of error. The learning rate also tumbled down from an initial 0.0010 gradually to 0.00053144 in the final epoch to let the latter help refine the model periodically.

Figure 1 and Figure 2 present the consistent training progress with very slight overfitting in the loss curves of accuracy. Figure 3 shows the confusion matrix with high sensitivity and accuracy regarding the classification of cells that are parasite non-infected. Permission granted for

verification of the performance and reliability of a custom-made CNN for malaria diagnosis by.

## II. ResNet50 Model

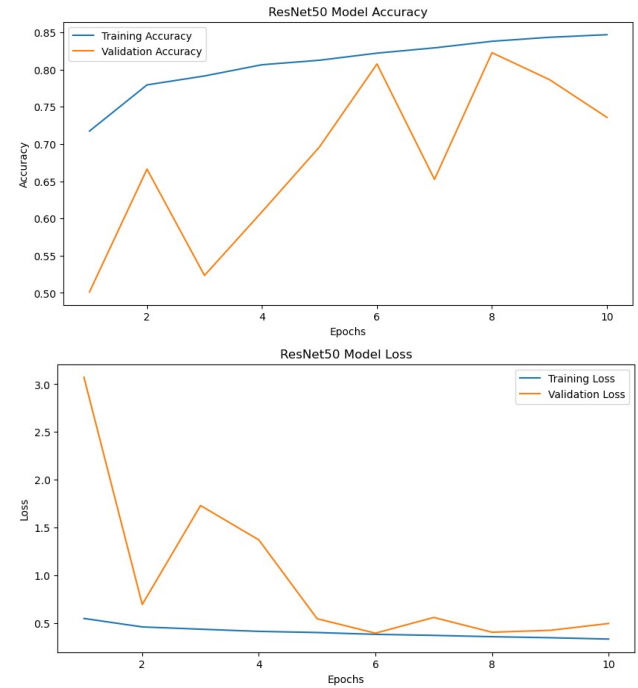
ResNet50 with 50 layers and residual pooling was fine-tuned for this task with pre-trained weights from ImageNet. The first 140 layers were frozen. and the remaining class was trained on the malaria dataset. Training was stopped early and the learning rate was fixed over 10 epochs

### 1. Performance Metrics:

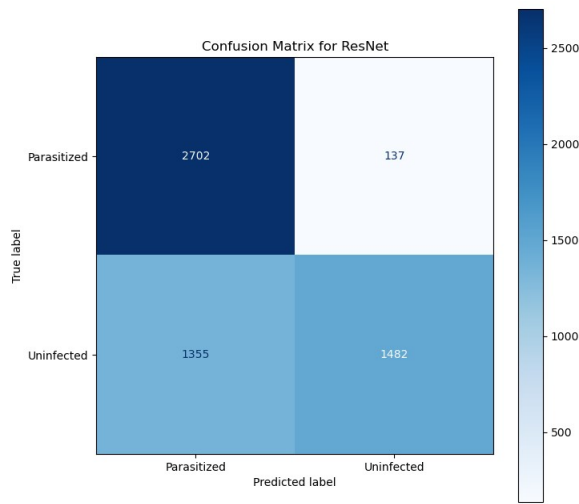
- **Accuracy:** 73.7%
- **Precision:** 91.5%
- **Recall:** 52.2%
- **F1-Score:** 66.5%

### 2. Graphical Representations:

- **Figure 4:** Model Accuracy and Model Loss for ResNet50. These graphs show some divergence between training and validation metrics, suggesting overfitting during training.
- **Figure 5:** Confusion Matrix for ResNet50. The matrix highlights misclassifications, particularly in detecting parasitized cells, leading to a lower recall score.



**Fig. 4.** Model Accuracy and Model Loss for ResNet50



**Fig. 5.** Confusion Matrix for ResNet50

10 epochs in transfer learning of the ResNet50 was trained with early stopping and learning rate scheduling such that the best performance could be achieved. The initial training accuracy of the model stood at 67.613% against a validation accuracy of 50.127%. This suggests how difficult it is to retune pre-trained weights with that of a malaria dataset. The peak validation accuracy achieved during the 6 epochs was at 80.76%. The persistent fluctuations across all subsequent epochs have the effect of settling it down to 73.58% by the 10th epoch in comparison to training accuracy, which consistently ascended to 84.99%, indicating a lucrative learning process but possible overfitting.

This statement is further confirmed by that of loss metrics, which all reflected the same fact, with loss for training decreasing from 0.5996 to 0.3259 and validation loss starting from 3.0720 to 0.3932 in the 6th epoch and gradually rising to 0.4947 by the end. Starting from 0.0001 in increments down to 0.00006561, learning rate optimizes the perfection of the weight in the latter epochs.

The training and validation trends are depicted in the combined model accuracy and loss curves (Fig. 4), with validation performance heaping instability. Performance metrics in Fig. 5 summarizes that there was high precision but also low recall. This indicates that the model was biased towards uninfected as compared to parasitized samples. But the depth of ResNet50 provided sufficient power to extract appropriate features; the results indicated that fine-tuning such a mysterious architecture is not easy for any specific medical imaging task.

### III. Inception V3 Model

Fine-tuned to extract multiscale features, InceptionV3 employs weights originating from ImageNet. The first 249 layers were held fixed while the rest were trained on the malaria dataset. Model training continued for 10 epochs with an early stopping procedure to avoid overfitting. The fine-tuning stage results generated the following performance metrics:

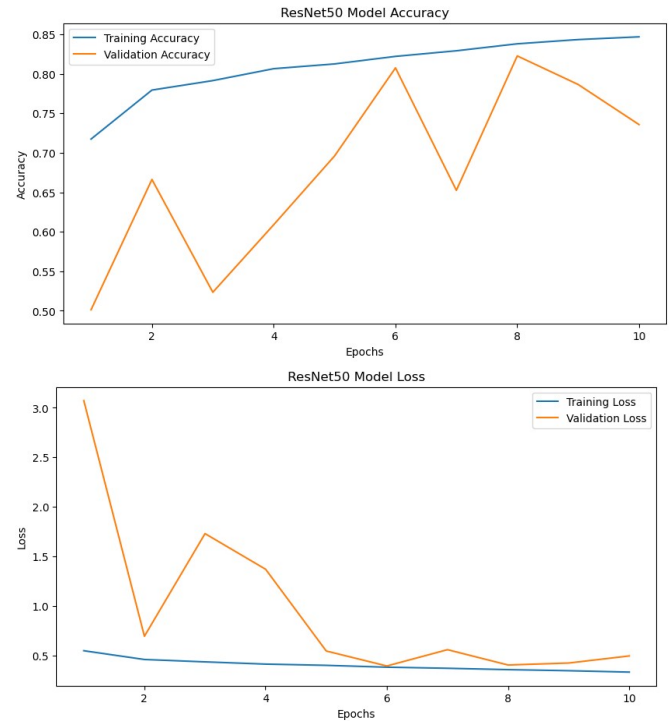
#### 1. Performance Metrics:

- **Accuracy:** 93.4%

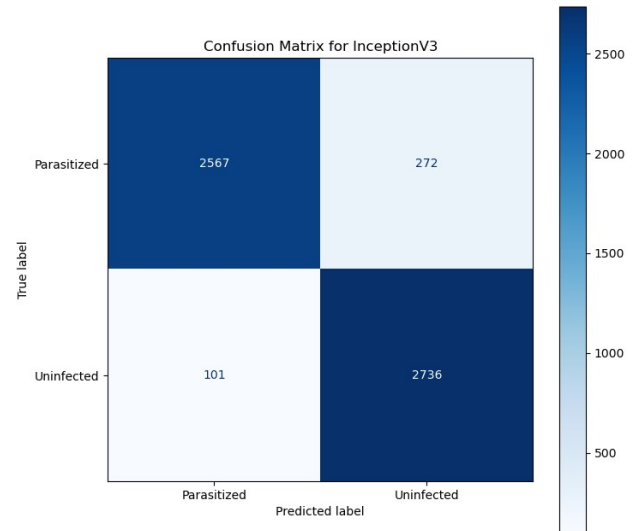
- **Precision:** 91%
- **Recall:** 96%
- **F1-Score:** 93.6%

#### 2. Graphical Representations:

- **Figure 6:** Model Accuracy and Model Loss for InceptionV3. These graphs indicate stable training and validation curves with good generalization.
- **Figure 6:** Confusion Matrix for InceptionV3. Between their precision and recall, the matrix promotes a good balance with minimal misclassified outputs.



**Fig. 6.** Model Accuracy and Model Loss for InceptionV3



**Fig. 7.** Confusion Matrix for InceptionV3

The InceptionV3 model has undergone a training of about 10 epochs, with early stopping whereby the training is



ceased as soon as the validation loss starts to increase, resulting in preventing overfitting. As observed from the training convergence, the model initially had 87.92% training accuracy with a 93.95% validation accuracy after the first epoch. By the time of the second epoch, the training accuracy had increased to 95.32% with the validation accuracy at its peak of 94.23%. Early stopping was done after the 4th epoch when validation loss increased towards finished training.

The plot illustrated a constant decrease in training losses, from 0.2962 in the first epoch to a lower 0.0507 by the 4th epoch, and only slight fluctuations were observed in the validation loss from 0.1575 to 0.2072. The above trend could be taken to mean that though the model learned effectively, it faced a bit of difficulty on the generalization scope in the later epochs.

Both training and validation accuracy/loss plots (below Fig. 6) demonstrate rapid improvement and introduction of stabilization due to early stopping. Performance metrics served to corroborate the strength of InceptionV3 as demonstrated by high accuracy scores, precision, recall, and F1 scores (Fig. 7). Since the confusion matrix shows that the two classes are well separated, it shows that InceptionV3 is very strong in detecting parasitized from uninfected cells due to its multiscale feature extraction design into efficient computations to be a good candidate for the automated diagnosis of malaria.

#### IV. Comparison

Evaluates the three models CNN, ResNet50, and InceptionV3 on their performances in training, validation, and test datasets with accuracy, precision, recall, and F1-score metrics for effective classification of parasitized and uninfected blood cells. A summary of performances of all models is made in Table 8, which shows the strengths and weaknesses of each.

##### I. Training and Validation Results

###### 1. CNN:

- Training accuracy got to 95.59% and validation accuracy at 94.48% after 10 epochs.
- Clearly showed very strong convergence on stable training and validation curves.
- Little observation of overfitting as there was a near match between training and validation performance.

###### 2. ResNet50:

- Achieved a training accuracy score of 84.99% and a rather fluctuating profile with respect to validation data with a final mean value of 73.58%.
- Have issues with overfitting; there was an observed instability of the validation loss which implied a poor generalization from pre-trained weights.

###### 3. InceptionV3:

- Training accuracy peaked at **98.26%**, with validation accuracy stabilizing at **94.23%**.
- Early stopping prevented overfitting, and training and validation metrics remained closely aligned.

##### II. Test Data Results

The test data results further confirmed the models' relative strengths and weaknesses:

1. CNN scored highest in the entire test with a test accuracy of 94.54%, precision of 91.82%, recall of 97.78%, and F1-score of 94.71%.
2. Quite close was InceptionV3 with a test accuracy of 93.42%, precision of 90.96%, recall of 96.44%, and an F1-score of 93.62%.
3. ResNet50 did not do so well as the other models in performance, with the test accuracy being 73.71%, precision of 91.54%, recall of 52.24%, and an F1-score of 66.52%.

Model	Accuracy	Precision	Recall	F1 Score
CNN	0.95	0.92	0.98	0.95
ResNet50	0.74	0.92	0.52	0.67
InceptionV3	0.93	0.91	0.96	0.94

**Fig.8.** Performance of all models

Almost all metrics taken into account showed that CNN was the most effective model as compared to ResNet50 and InceptionV3. The architecture of the CNN being simpler kept it away from the overfitting that affected ResNet50, as seen in its consistent training and validation metrics. It was close to InceptionV3, as far as measures of precision and recall were concerned, but slightly behind this model in terms of accuracy and F1 score. Recall and validation stability problems were begotten by ResNet50 due to being high in depth and fine-tuning difficulties.

This comparison demonstrates that sometimes simpler, task-specific architectures such as the CNN are better than more complicated, pre-trained models, particularly when a dataset is well-prepared and balanced.

#### 5. CONCLUSION

This study shows the efficacy of deep learning models in performing automatic malaria diagnosis. Out of the three models evaluated, the CNN yielded the best performance with respect to accuracy, precision, recall, and F1 score, confirming its suitability for the task at hand. InceptionV3 was fairly promising in doing efficient multi-scale feature extraction, but lagged slightly behind the CNN in performance. The hierarchical feature extraction capabilities of ResNet50, however, did not generalize well, indicating that there needs to be a balance between model complexity and fine-tuning for a given dataset.

It shows deep learning reshaping medical diagnostics in underserved communities. Future research could explore hybrid architectures tapping the best of several models with data augmentation to increase generalization and real-world deployment for evaluating the realistic usefulness of these systems. All these could vastly enhance malaria diagnosis for global health equity augmentation purposes.

#### 6. REFERENCES

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for

Computer Vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2818–2826. DOI: 10.1109/CVPR.2016.308.

- [3] R. Rajaraman, S. Jaeger, and S. Antani, "Performance Evaluation of Deep Neural Networks for Malaria Parasite Detection," in *Computerized Medical Imaging and Graphics*, vol. 69, pp. 1–13, 2018. DOI: 10.1016/j.compmedimag.2018.07.001.
- [4] H. Zhu et al., "InceptionV3-Based Neural Network for Malaria Parasite Detection," in *IEEE Access*, vol. 7, pp. 158161–158170, 2019. DOI: 10.1109/ACCESS.2019.2949566.
- [5] A. Mahajan, A. Mitra, and D. Ganguly, "Automated Malaria Detection in Blood Smear Images Using ResNet-Based Convolutional Neural Networks," in *Biomedical Signal Processing and Control*, vol. 59, pp. 101879, 2020. DOI: 10.1016/j.bspc.2020.101879.
- [6] National Institutes of Health, "Malaria Dataset," accessed Dec. 2024. [Online]. Available: [https://data.lhncbc.nlm.nih.gov/public/Malaria/cell\\_images.zip](https://data.lhncbc.nlm.nih.gov/public/Malaria/cell_images.zip)

1

---

<sup>1</sup> All authors contributed equally to the project design, implementation, evaluation, and documentation.