A Project Report on

MALARIA DETECTION: COMPARISON OF MODELS

Submitted in Fulfillment of the Requirement for

"CS 438/638 APPLIED MACHINE LEARNING"

Submitted by:

1. Aadarsh Katigi
2. Aditya Shriyan
3. Emry Hankins
4. Kunal Arunkumar Singh
5. Kunal S
6. Pruthviraaj Umesh

Under the Guidance of:

Professor Wei Ding

Spring 2024

UNIVERSITY OF MASSACHUSETTS BOSTON

Massachusetts, United States Of America

ABSTRACT

Malaria remains a critical global health issue, necessitating rapid and cost-effective diagnostic solutions, especially in regions with limited access to expert medical care. This study aims to address this need through the development and comparison of four image classification models designed to distinguish parasitized from non-parasitized cell images within the malaria dataset provided by TensorFlow. The dataset comprises over 27,000 labeled cell images, evenly split between parasitized and non-parasitized types, curated by healthcare experts under the National Institutes of Health (NIH). The models explored include a Convolutional Neural Network (CNN), InceptionNet, ResNet 50, and a Feed-Forward CNN based on the Xception architecture. Extensive data preprocessing was undertaken to ensure model reliability and accuracy. Performance metrics such as the confusion matrix, F1 score, recall, and accuracy were employed to evaluate each model's effectiveness. Initial findings indicate significant variations in false positives and false negatives among the models, despite similar accuracy levels, suggesting differing levels of diagnostic efficacy. The comprehensive analysis of these models provides insights into optimal strategies for model development and highlights potential areas for further research in the automated diagnosis of malaria.

TABLE OF CONTENTS

1.Introduction

1.1 Introduction to Model Development and analysis

In the ongoing battle against malaria, a disease affecting millions annually, innovative diagnostic solutions are crucial, particularly in regions where access to expert medical care is scarce. Among the promising technologies for addressing this challenge are image classification models, which offer a potential pathway to rapid and low-cost diagnosis. This project focuses on developing and analyzing various advanced image classification models capable of differentiating between parasitized and non-parasitized cell images, using the extensive malaria dataset provided by TensorFlow.

A key component in this endeavor is the application of Convolutional Neural Networks (CNNs). CNNs are specialized kinds of neural networks for processing data that has a grid-like topology, such as images. An image is effectively a matrix of pixel values, and CNNs leverage small, square patches of an image to identify patterns that are essential for classification. This ability to autonomously extract and learn the most crucial features from images makes CNNs particularly effective for medical imaging tasks, including the identification of malaria-infected cells. The CNN architecture has been foundational in the field of deep learning, enabling the automation of predictive models that surpass traditional image processing techniques.

Throughout the development phase, various models including a basic CNN, InceptionNet, ResNet 50, and an Xception-based Feed-Forward CNN were explored to establish their efficacy in accurately classifying the provided cell images. These models were chosen for their unique architectural strengths in handling complex image data, which is pivotal in capturing the nuanced differences between parasitized and non-parasitized cells. Extensive data preprocessing and the use of sophisticated model evaluation metrics like the confusion matrix, F1 score, recall, and accuracy were integral to refining these models and enhancing their diagnostic precision.

1.2 Data Analysis:

The transition from preprocessing to model architecture is crucial. By ensuring that the models are trained on well-prepared data, they are better equipped to handle the variability and complexities of real-world diagnostic scenarios. This preparation enhances the models' generalization capabilities and accuracy when classifying unseen data. The structured approach to both preprocessing and analysis facilitates a thorough evaluation of model performance, focusing on optimizing and validating the diagnostic capabilities of each model within controlled and varied testing environments.

1.3 Data Pre-Processing

To optimize the malaria dataset for training, we undertook rigorous preprocessing measures. We adjusted image sizes according to the specific requirements of each model: the CNN and Feed-Forward CNN models used images resized to 96x96 pixels, ResNet 50 to 224x224 pixels, and InceptionV3 to 75x75 pixels. Uniform normalization was applied across all models, with pixel values scaled between 0 and 1 by dividing each by the maximum value of 255. This standardization was critical for maintaining consistency in pixel value interpretation by the models.

Additionally, data batching strategies varied among the models based on their architectural needs and training specifications. For some models, the diversity ensured by batching was essential, while for others, it was unnecessary. To combat overfitting, a robust data augmentation protocol was implemented, which included random flipping of images and adjustments to brightness and contrast. The dataset was divided into three subsets: 70% for training, 15% for validation, and 15% for testing, ensuring comprehensive coverage and evaluation capabilities during model training.

2. Model Structure:

2.1 Custom-built Convolutional Neural Network (CNN):

The model is meticulously designed to classify images into two categories: Parasitized and Uninfected cell images. It begins with an input layer that adjusts images to specified dimensions, defined by img_width and img_height, ensuring that all images are uniformly processed. This is followed by a series of convolution layers, which progressively increase in complexity from 16 to 256 filters, each with a (3,3) kernel size. These layers use ReLU activation and maintain the same padding to preserve the spatial dimensions of the image. After each convolution layer, max pooling is applied with a (2,2) window to reduce the spatial size of the representation, minimizing the number of parameters and computation in the network, thus also helping to prevent overfitting.Additionally, batch normalization follows each convolution step to stabilize and accelerate the training process by normalizing the inputs of each layer. As the convolutional layers extract and intensify features from the images, the structure then flattens these 2D feature maps into a 1D feature vector, which is essential for classification. This vector feeds into a dense network layer consisting of 64 neurons with ReLU activation, integrated with a 30% dropout rate to further mitigate overfitting risks. Finally, the model concludes with an output layer comprising a single neuron with sigmoid activation, specifically tailored for binary classification, determining whether each cell image is Parasitized or Uninfected. This architecture not only enhances the model's ability to learn detailed and complex features but also ensures high accuracy and efficiency in classifying the cell images.

2.2 ResNet50:

The architecture continues with a dense layer featuring 256 neurons, utilizing ReLU activation to maintain non-linearity in the learning process. An exceptionally high dropout rate of 80% accompanies this dense layer, significantly reducing the likelihood of overfitting by randomly omitting a large portion of the neurons during training, thereby ensuring that the model does not rely too heavily on any specific neuron and can generalize better to new, unseen data. The final component of the model is the output layer, which employs sigmoid activation to perform binary classification, determining whether cell images are parasitized or uninfected. This setup not only

exploits the advanced capabilities of the ResNet50 model but also adapts it to deliver precise and reliable classifications in the medical imaging field.

2.2 Abridged Xception:

The model utilizes an abridged version of the Xception architecture, designed to maintain efficient performance while utilizing fewer parameters, making it highly suitable for environments with limited computational resources. The initial stage of the model involves input processing, where pixel values are rescaled to a range of [0, 1] for normalization. This is followed by an initial convolution layer equipped with 128 filters, setting the stage for deeper feature extraction.

As the model progresses, it employs depthwise separable convolutions, an efficient variant of the standard convolution process. This technique separates the convolution into two layers: one for filtering and one for combining, which reduces the model's complexity and the number of parameters. The filter count increases from 256 to 512 in these layers, each followed by batch normalization and GELU (Gaussian Error Linear Unit) activation to ensure non-linearity and stability in the learning process.To further enhance the model's ability to propagate features through deep network layers, residual connections are added every few layers. These connections help to prevent the vanishing gradient problem by allowing gradients to flow through a shortcut from earlier layers to later layers. The model also incorporates max pooling with a stride of 2, effectively reducing the spatial dimensions of the feature maps gradually, thus helping to condense the information and reduce computation.

The final stages of processing include a global average pooling layer, which helps to summarize the essential features in each output channel, followed by a 30% dropout rate to minimize the risk of overfitting. This dropout ensures that the model remains robust and generalizable to new data. Depending on the application—whether it’s for multi-class or binary classification—the output layer of the model employs either softmax or sigmoid activations, respectively. This simplified yet powerful version of the Xception model ensures that it can efficiently process and classify images with high accuracy and lower computational demand.

3. Results and Model Comparison:

The evaluation of the classification models was thorough, utilizing a comprehensive set of performance metrics to ensure a holistic assessment of each model's capabilities. Notably, all models demonstrated high levels of accuracy, with ResNet50 marginally surpassing the others in this regard. This high accuracy across the board indicates that each model was effectively able to learn from the training data and generalize well to the test data.In terms of precision, recall, and F1-score, the results were consistently strong across all models. These metrics are crucial as they provide insight into the models' ability to correctly identify true positives while minimizing false positives and false negatives. Such results affirm the models' effectiveness in distinguishing between parasitized and uninfected cell images accurately.

Additionally, the ROC (Receiver Operating Characteristic) and Precision-Recall curves were particularly telling for the ResNet50 and Abridged Xception models, which displayed very close performance in these measures. The proximity of their curves in both ROC and Precision-Recall spaces highlights their similar discriminative capabilities, demonstrating their effectiveness in handling both positive and negative classes under varying thresholds.

Further insights were gained from the analysis of confusion matrices for each model, which provide a detailed breakdown of true positives, true negatives, false positives, and false negatives. These matrices are essential for understanding each model's performance in a more granular manner, allowing for targeted improvements in future iterations.

3.1 Confusion Matrix:

Each confusion matrix is divided into four parts:

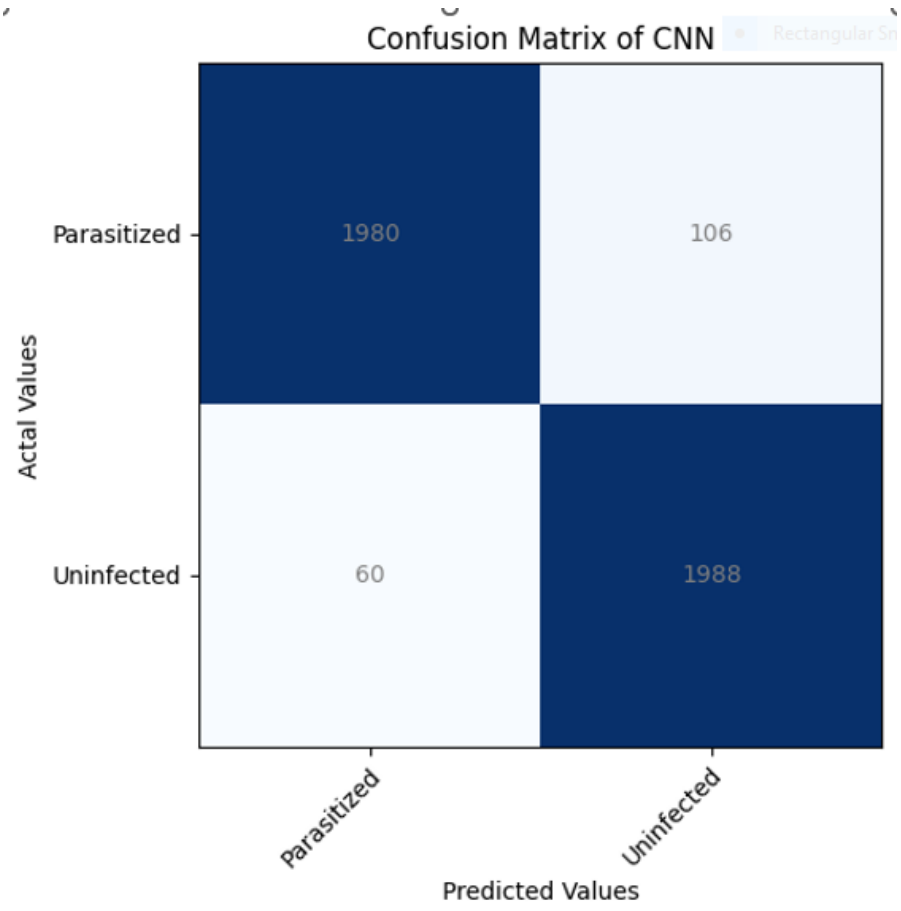
True Positive (TP): Top left cell, indicating the number of Parasitized images correctly identified as Parasitized.

False Negative (FN): Top right cell, indicating the number of Parasitized images incorrectly identified as Uninfected.

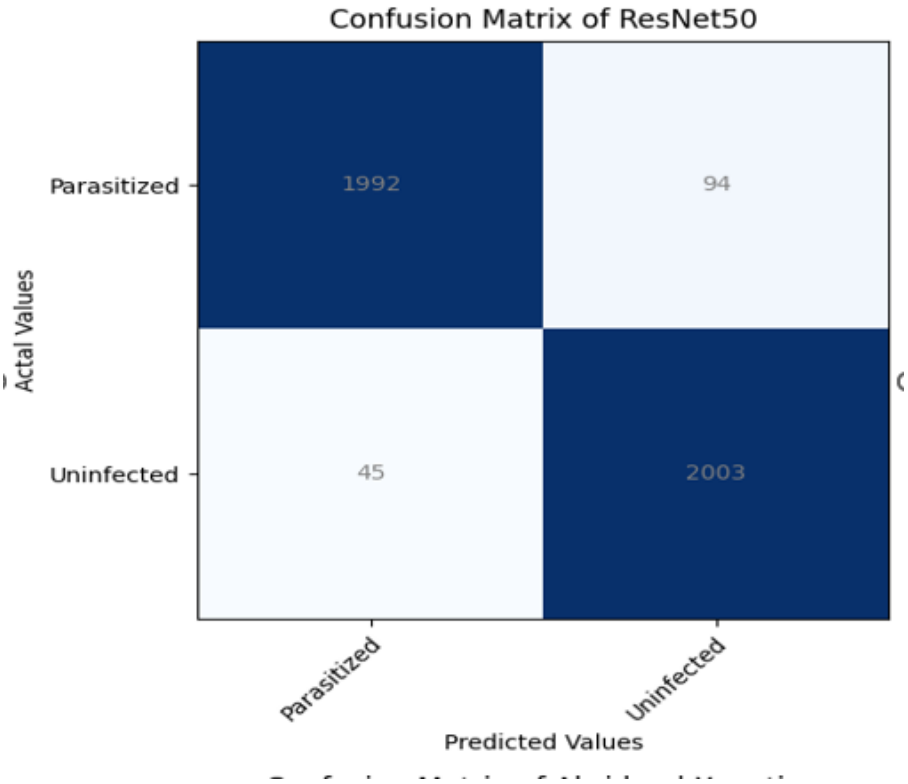
False Positive (FP): Bottom left cell, indicating the number of Uninfected images incorrectly identified as Parasitized.

True Negative (TN): Bottom right cell, indicating the number of Uninfected images correctly identified as Uninfected.

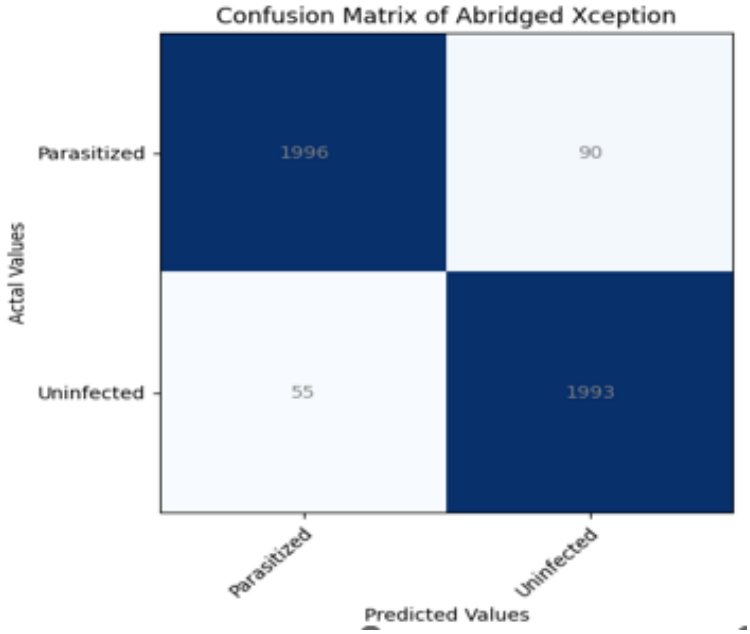
3.1.1 Confusion matrix of CNN:



3.1.2 Confusion Matrix of ResNet50:



3.1.3 Confusion matrix of Abridged Xception:



3.2 Specific Model Analysis:

3.2.1 Convolutional Neural Network (CNN)

TP: 1980 (Parasitized correctly identified)

FN: 106 (Parasitized incorrectly identified as Uninfected)

FP: 60 (Uninfected incorrectly identified as Parasitized)

TN: 1988 (Uninfected correctly identified)

3.2.2 ResNet50

TP: 1992 (Parasitized correctly identified)

FN: 94 (Parasitized incorrectly identified as Uninfected)

FP: 45 (Uninfected incorrectly identified as Parasitized)

TN: 2003 (Uninfected correctly identified)

3.2.3 Abridged Xception

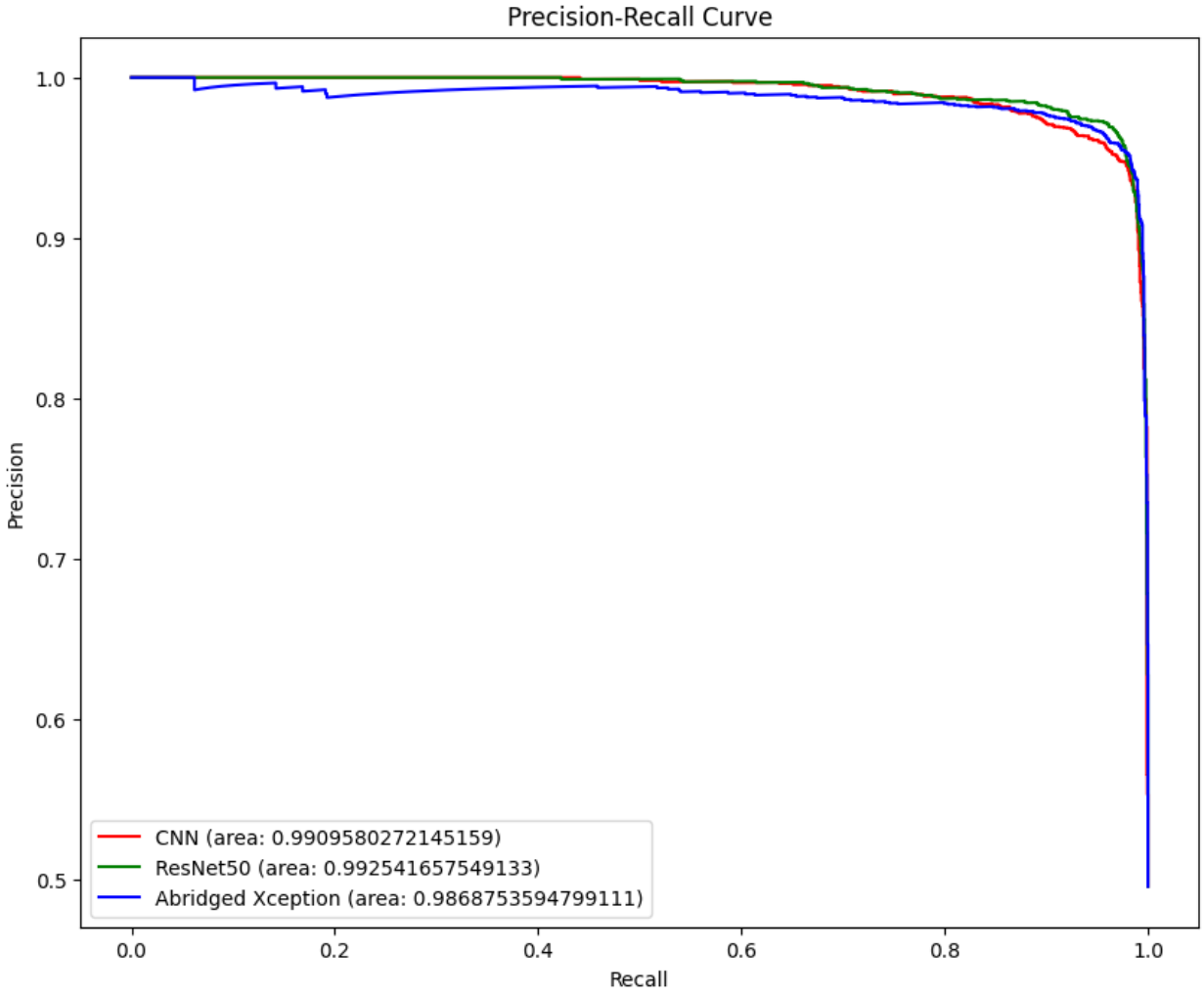
TP: 1996 (Parasitized correctly identified)

FN: 90 (Parasitized incorrectly identified as Uninfected)

FP: 55 (Uninfected incorrectly identified as Parasitized)

TN: 1993 (Uninfected correctly identified)

3.3 Precision Recall curve:

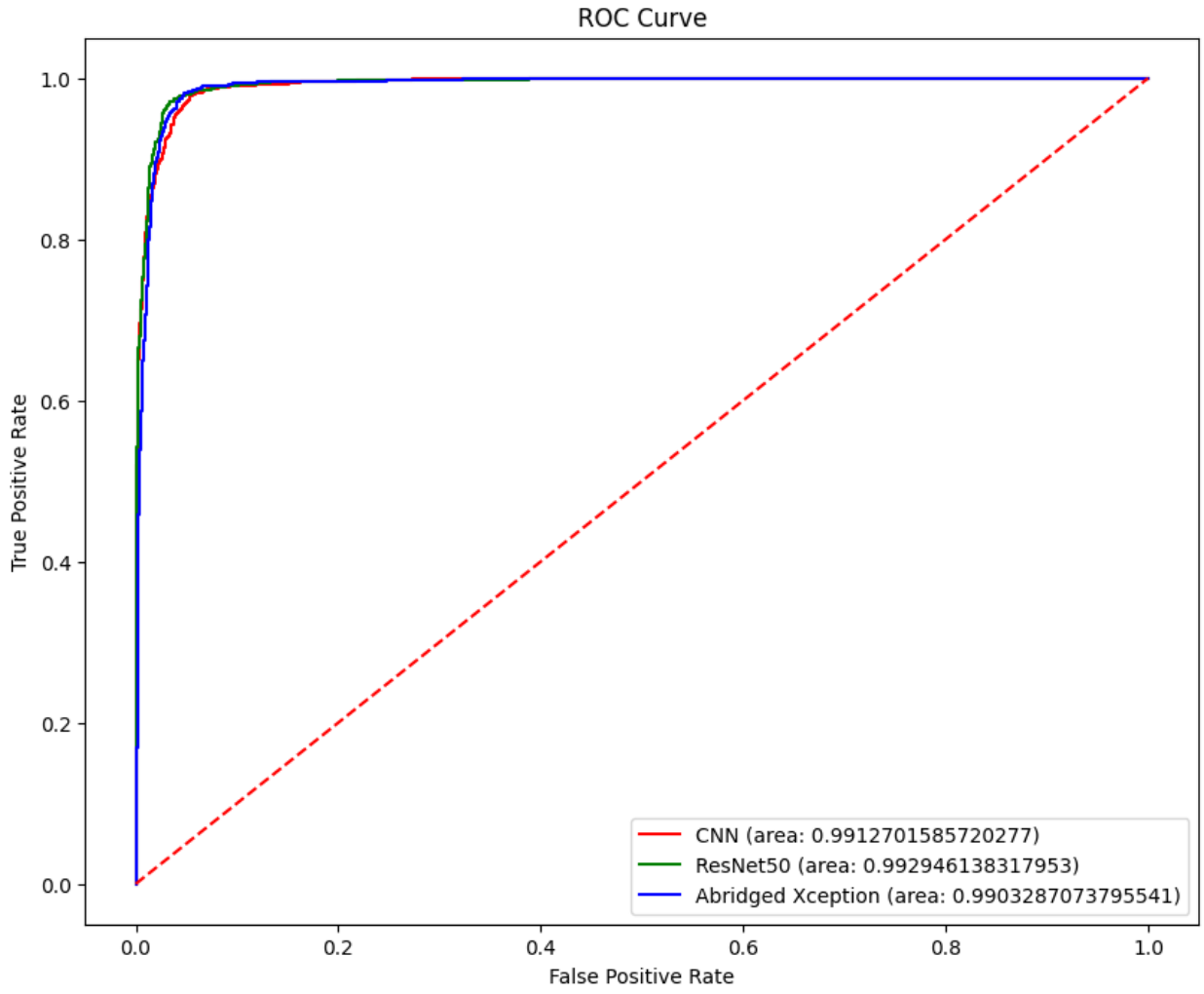


3.4 Model Performance:

In the comparative evaluation of three machine learning models—CNN, ResNet50, and Abridged Xception—across a spectrum of performance metrics, ResNet50 emerges as the optimal choice for applications demanding a delicate balance between precision and recall. The CNN model impresses with its robust precision across a majority of recall levels, boasting an impressive area under the curve (AUC) of approximately 0.99095, indicative of its proficiency in accurately identifying positive samples while effectively minimizing false positives. ResNet50, delineated by the green curve, exhibits a notable edge over both CNN and Abridged Xception, showcasing consistently superior precision across the entire range of recall. This performance is underscored by an AUC of about 0.99254, signifying its efficacy in efficiently identifying positive cases with

minimal incidence of false positives. Abridged Xception, represented by the blue curve, maintains competitive performance with an AUC of approximately 0.98687, albeit slightly lagging in precision compared to ResNet50. Noteworthy is the stable precision observed in all three models as recall increases up to around 0.9, beyond which a discernible decline is noted. This decline signals an escalated misclassification of negative cases as positive when striving to capture more positive instances. Given its marginal but consequential advantage in sustaining higher precision for a given recall level, ResNet50 emerges as the preferred model choice. This superiority is particularly pertinent in contexts where the cost of false positives holds significant implications, such as medical diagnostics, where both the accurate identification of true cases and the avoidance of false diagnoses are paramount considerations.

3.5 ROC curve:



The Receiver Operating Characteristic (ROC) curve serves as a pivotal tool for assessing the performance of machine learning models—CNN, ResNet50, and Abridged Xception—in categorizing cell images as Parasitized or Uninfected. This graphical representation delineates the diagnostic prowess of binary classifier systems across varying discrimination thresholds, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at different threshold settings.

3.5.1 Performance Analysis by Model Curve:

CNN (Red Curve): Evinces an exceptionally high area under the curve (AUC) of approximately 0.99172, indicative of its outstanding ability to differentiate between Parasitized and Uninfected classes while maintaining minimal error in terms of false positives.

ResNet50 (Green Curve): Marginally surpasses CNN with an AUC of about 0.99294, suggesting a slightly superior capability in distinguishing between the two classes across various thresholds.

Abridged Xception (Blue Curve): Exhibits an AUC of approximately 0.99032, marginally lower than the other two models yet still showcasing robust discriminative power.

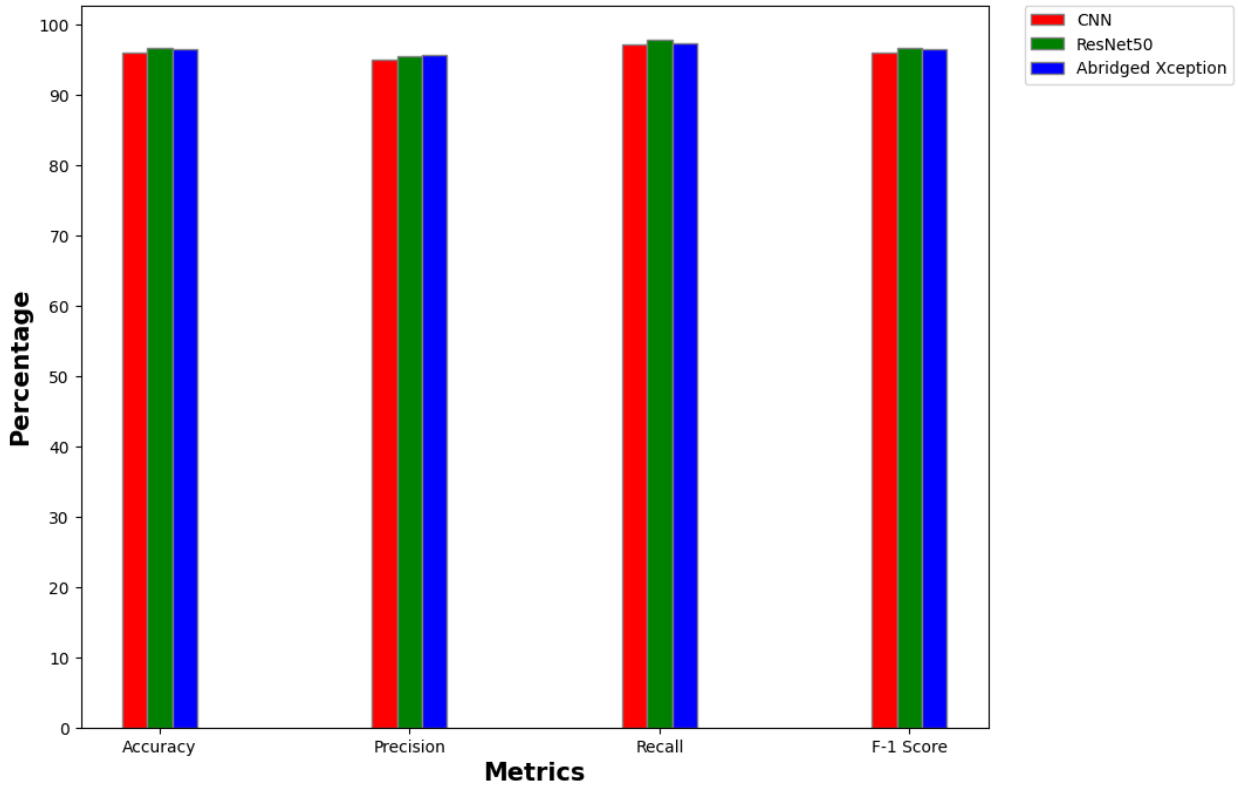
3.5.2 Interpretation of AUC Values:

The AUC values for all models hover close to 1, underscoring very high classification performance. Models with AUC values closer to 1 are deemed excellent, whereas those closer to 0.5 imply no better performance than random guessing.

3.5.3 Diagnostic Ability:

The ROC curves for all three models exhibit a rapid ascent towards the top-left corner, indicative of a high true positive rate coupled with a low false positive rate. This swift ascent is desirable as it underscores the models' capacity to achieve high sensitivity (recall) without a substantial increase in the false positive rate.

3.5.4 Bar Graph Analysis:



The bar graph serves as a comprehensive tool for comparing the performance metrics—accuracy, precision, recall, and F1-score—of three image classification models: CNN, ResNet50, and Abridged Xception, in categorizing microscope images of cells as Parasitized or Uninfected. Each model, represented by a distinct color, showcases its efficacy across the metrics, quantified as percentages, offering insights into their overall correctness, exactness, completeness, and the balance between precision and recall. This structured analysis provides a clear and concise overview of how each model performs in the intricate task of medical image classification.

3.6 Accuracy:

As a measure of overall correctness, accuracy signifies the percentage of correct predictions made by the models. Notably, all three models exhibit comparable performance, indicating a high accuracy in correctly identifying both Parasitized and Uninfected cells.

3.7 Precision:

Precision, serving as a measure of exactness, indicates the proportion of positive identifications that were correct. Across all models, the precision metric demonstrates similar performance, suggesting that when a model predicts a cell image as Parasitized, it is likely to be correct.

3.8 Recall and F1-Score:

Recall, a measure of completeness, reflects the proportion of actual positives that were correctly identified. Consistently high across the models, the recall metric implies that the models effectively identify most of the Parasitized cells present in the dataset. The F1-score, a balance between precision and recall, represents a harmonic mean used as a composite measure of a model's accuracy. Remarkably, the F1 scores exhibit equally high values across all models, showcasing a robust balance between precision and recall in their predictive capabilities.

3.9 Comparative Analysis:

CNN, represented by the red bars, demonstrates robust performance across all metrics, suggesting effective tuning and a model architecture well-suited for the classification task at hand. ResNet50, depicted by the green bars, consistently matches or slightly exceeds the performance metrics of the other models, implying that its depth and utilization of residual connections may confer slight advantages in handling complex image features. Abridged Xception, illustrated by the blue bars, competes closely with the other models, displaying minor variations in some metrics. Its streamlined architecture appears to maintain high performance while potentially offering efficiencies in computation. Despite these nuances, all three models exhibit strong performance, albeit with slight differences in the counts of false positives and false negatives, indicating variations in sensitivity and specificity between them.

4. Conclusion:

All three models performed exceptionally well on the task of classifying parasitized and uninfected cell images. ResNet50 displayed a slight advantage in terms of overall metrics, which could be attributed to its deeper and more complex architecture enabling better feature extraction. Abridged Xception also performed comparably, offering a more efficient alternative with slightly reduced computational demands. The CNN, while slightly less effective than the other two in some metrics, still proved to be a robust choice for image classification tasks.

The analysis demonstrates the effectiveness of using advanced convolutional networks for medical image analysis and provides insights into selecting appropriate models based on the trade-offs between accuracy, efficiency, and computational resources.

This report serves as a comprehensive document of the methodologies, results, and insights derived from the experiments, fulfilling the requirements of the Machine Learning Team Project Phase 3. The findings will be further discussed in the final presentation scheduled for May 8, 2024, where additional details and demonstrations of the models' capabilities will be showcased.