Abstract

This study addresses the pressing need for efficient and accessible malaria detection and classification systems, particularly in regions with limited resources and high disease prevalence. With millions affected annually, accurate diagnosis and classification are essential for effective treatment and transmission control. Traditional methods like rapid test kits and microscopic blood smear analysis have limitations, prompting the exploration of automated solutions using pretrained convolutional neural network (CNN) models. However, challenges arise due to the scarcity of reliably labelled data. This research aims to develop a preprocessing methodology and identify the most suitable pretrained CNN models to accurately classify malaria parasites in blood smears as Uninfected, Falciparum, or Vivax, thus facilitating improved diagnosis and treatment in resource-constrained settings.

Introduction

Malaria is a serious, life-threatening disease that affects millions every year. According to the consensus of WHO in the year 2023, a total of 167 million people were infected with the disease. Africa, being one of the most severely affected nations, accounted for 75% of these cases. The disease affects those with weak or compromised immune systems to a greater degree, which can be seen from the fact that out of the 700,000 to 2.3 million deaths from malaria each year, 75% of them are young children and women. Detection and treatment of malaria are vital not only to cure the disease but also to reduce its high transmission. Malaria has four types of viral species: Falciparum, Vivax, Ovale, and Malariae, with Falciparum and Vivax being the most common ones. Early parasite detection and classification in the infected can ensure not only the appropriate course of treatment but also guards against anti-malarials being used injudiciously, which is crucial in co-endemic areas.

Malaria detection is usually performed with the help of rapid test kits or through microscopic blood smear image analysis by trained professionals. While Rapid test kits can provide parasite detection, they cannot classify the virus into one of the four categories. Blood Smear detection can be performed with i) Thin Blood Smears and ii) Thick Blood Smear. Thick blood smears are a concentrated level of dehemoglobinized red blood cells; this high density of cells allows for easier and more efficient parasite detection but does not provide a good environment for studying parasite morphology. Thin blood smears analysis can provide parasite classification in the positively detected samples.

Due to the high skill demand and Malaria being a prevalent disease in developing countries, it is important to develop a system that can process blood smear data with accuracy and proficiency by anyone with basic knowledge of technology. Pretrained CNN models like VGG16, ResNet50, etc., can provide ways to address the problem; however, the study suffers from a problem with the collection of reliably labelled data. The purpose of this study is to tackle the problem of finding an efficient preprocessing methodology and to figure out the most efficient pretrained models to build a reliable system to classify parasites accurately with limited availability of data and accurately label blood smears as Uninfected, Falciparum, and Vivax.

Literature Review

A good amount of research has been done in the field of Malaria detection and classification which used different image processing models and techniques to detect malaria parasites in blood smear images, [1] classified samples into different species of malaria using 3 convolutional layers, and a convolution2D convolution operation, ReLU for activation function, and a dilation rate of 2 as hyper-parameters. For these layers a dilation rate of 2 was used, the study used a publicly available dataset of 27699 for detecting parasites and performed with an accuracy of 99.9%. The study was able to classify the virus at an accuracy of 99% for Falciparum, 64.6% for Malariae, 39.1% for Ovale, and 37.3% for Vivax. The study had low and unreliable levels of accuracy for Vivax which is also one of the two most common forms of Malaria, also the study does not experiment with different pre-processing to make the dataset more reliable and remove any unnecessary features from the dataset.

Studies like [3] used transformer models for malaria parasite detection, and the study used transformer models and generative adversarial networks for multi-class plasmodium classification and malaria detection. A Residual Convolutional Neural Network utilizing a Bayesian method was used to classify blood cell images as infected or uninfected. The study deals with new ways of dealing with the issue of lack of data by using GANs but falls short in ensuring data reliability, as labeling of the generated data is not cross-verified by biologists or medical professionals models trained on such data cannot be relied on for real-world usage. The data available of labeled blood smears also contains wrongly labelled data thus generating images using that data cannot be relied on.

[2] experimented with deep convolutional networks such as DenseNet121, DenseNet201, ResNet152V2, NasNetLarge, MobileNetV2, and the hybridized DenseNet201 and ResNet152V2. The study showed good results with MobileNetV2 computing the highest accuracy, and ResNet152V2 having the best loss value by 0.005. For this study data was collected from three sources. Watershed segmentation was used which is a digital image processing technique used for segmenting objects in an image based on the topography of the image intensity. The study had good accuracy with these deep CNN models, but training these models takes a large amount of time and processing power and the lack of a big enough dataset can lead to overfitting.

[4] investigates the problem of choosing the best pre-processing methodology for malaria blood smear images. With microscopic images regular image pre-processing involving sharpening and noise reduction can lead to the loss of certain features and the exaggeration of others. The study runs CNN models on three iterations of the same data once with no processing, then with Colour Normalization: Gray-world Normalization, and lastly with Colour Normalization: Comprehensive Normalization. The study uses data of thin smear blood images from researchers at Lister Hill National Center for Biomedical Communications (LHCBC), on this data the study post-pre-processing does binary classification to do parasite detection and label the blood smears as infected or uninfected. Through this study, it was determined that Gray World Normalization provides better accuracy. The study provides a good insight into the preprocessing techniques and their working, but it does not determine their effect on parasite classification.

[5] the study researches the shortcomings of the traditional malaria detection and classification techniques. To resolve this, it experiments with Convolutional Neural Networks, specifically the VGG-16 pre-trained model. The study uses this model for feature extraction and classifies malaria into four subspecies. After feature extraction, SVM is used for training and testing. This three-phase model structure provided with 93% accuracy in detecting the virus type. With this accuracy, the study provided with a good insight into model structures that can be used to tackle the issue, but left room for improvement in making the results more reliable.

Chromatin found in Eukaryotic cells is a complex of DNA and protein, Chromatin profiling can help us access this chromatinized DNA. [6] does a deeper analysis of the effect of malaria on genetic material to study how it's regulated and organized. For this regulatory sequence patterns were used including TF binding, chromatin accessibility, and histone modification profiles. The study developed a framework MalriaSED, which was tested to be utilized to look into epigenetic regulatory implications of reported noncoding genetic variations. In conclusion, it was observed how malaria exploits the immune system and leads to drug resistance. This technology studied further can help scientists develop drugs for personalized attacks to the virus in the way it best counters the effect it has on the body. Malaria Detection, classification, and studying the virus all together can provide a way for fast and efficient recovery of the patients.

As researched in [6], understanding Malaria's effects on Chromatin is an important part of understanding the virus, but another vital step is to analyze the stage of the disease. Malaria has 3 stages, and in [7] a dataset comprising of the three classes(stages) of Falciparum is used for developing a model trained to detect the stage of the disease. For the study, the dataset was first put through image enhancement and augmentation and used deep convolutional networks like GoogleNet and VGG-19. The primary focus of the disease is to detect malaria at its schizont stage, which is an early stage and makes recovery easier. To achieve this goal a 10-fold cross-validation method was used, and fine-tuning was used to alter the pre-trained model to better fit the needs of the study. It concluded that GoogleNet gave better performance than VGG-19 with an accuracy of 97.48%. Building a reliable model for accurately predicting the stage of disease can revolutionize the recovery approach and can also be vital for better drug management. To facilitate a system that can predict the stage of a disease, we also need a fast and reliable methodology to detect the virus and classify it into one of the four categories. Unless we establish this by the time detection and classification are performed by traditional methods, the detecting stage will become futile as it may have progressed.

[8] is another study that aimed at classifying blood smear images into falciparum, malariae, and vivax. For the study in total, a dataset of 90 images was used. Three SVM models were then compared by training and testing on this dataset. Out of the three, Linear SVM, Polynomial SVM, and Gaussian SVM, Gaussian SVM outperformed with an accuracy of 86.67%. The study lacks in credibility as the dataset it trained the models on is not diverse enough. This can lead to overfitting and an unreliable prediction system. Also, having a dataset this small can lead to the model not being trained to identify a diverse range of geographically modified strains of the virus. Due to a small training data resulting in poor approximation, the model needs to be enhanced by training on a larger set of blood smear images.

In traditional diagnosis of Malaria, staining the blood smear film is an important part of the process, it is specifically critical for the identification of the malaria species. For malaria, Geimsa stain is used commonly and is also considered the most reliable. It can be used for staining both thick and thin blood smears. Giemsa solution which is used for this staining technique is composed of eosin and azure (methylene blue). [9] investigates the importance and advantage of staining in the process of malaria classification, different components of the dye help in enhancing features of the blood smear and making the identification process easier. Eosin stains the parasite nucleus red, and methylene blue stains the cytoplasm blue. The paper focuses on taking a dataset of stained blood smears of Falciparum and Vivax-infected samples. A seven-stage algorithm was derived, and the dataset was pre-processed to reduce noise and improve image differentiation. This would allow for an easier feature extraction. The seven stages included Image Acquisition, Extraction of RBC, Detection of edge, Binary Image, RBC Counting, Thresholding, and Extraction of the parasite. This study gives important insight into valuable pre-processing for better parasite detection. However, to attain this type of data with enough geographical, stage, and parasite diversity we would require a large amount of sample collection and a large number of trained professionals to prepare the stained blood smear films.

[10] studies to build a system that can accurately detect the virus, as well as determine the virus's kind and stage of infection. Automating the whole process makes the treatment process more efficient and faster. The study faces the huddle of detecting malaria in the early stage due to its complex structure. To facilitate the process the database images were pre-processed with GM and median filter, this would help make the feature extraction process simpler. Segmentation was done using the U-Net model, specifically focusing on segmenting red blood cells (RBC) in blood smear images for malaria diagnosis. With the VGG model, the study got an accuracy of 95.5%, an accuracy of 95% in classifying the cells as infected or uninfected, and an accuracy of 90.17% in classifying the species.

Methodology and Dataset

1. Dataset

The study uses two datasets, one with a total of 27,558 which had 13,779 samples of both infected and uninfected thin blood smear samples. This dataset was for parasite detection. The other dataset was collected from a variety of sources and has 2895 images of Uninfected, Vivax, and Falciparum, with 965 images in each category. The images are of dimensions 5312 X 2988pixels, with optical zooms to optical microscope images.

1. Preprocessing

Image preprocessing can help in better identification of the features which can lead to easier and more reliable detection and classification of the parasite. For the study the models were trained and tested on three forms of the dataset:

1. Unprocessed
2. Noise reduction, sharpening, and Histogram Equalization
3. Gray World Normalization

2.2.1 Gaussian blur is an image processing technique used to reduce noise and smooth images. In the project, it is applied during preprocessing to minimize high-frequency noise by averaging pixel values around each pixel, emphasizing the central pixel. By removing fine details and noise, Gaussian blur enhances image quality, making it easier for subsequent steps to focus on vital cell features. The application of Gaussian blur in malaria classification is crucial for improving accuracy and reliability by reducing noise and artifacts in blood cell images.

2.2.2 Sharpening filtering is an image processing method that accentuates edges and fine details by enhancing rapid changes in pixel intensity. This technique is valuable for highlighting subtle image features, improving local contrast, and aiding in precise feature recognition, such as identifying specific characteristics in malaria-infected blood cell images. Sharpening filters are beneficial in malaria classification as they enhance the visibility of critical features within blood cell images. By accentuating edges and fine details, these filters make it easier for machine learning models to recognize specific characteristics of infected and uninfected cells, like staining patterns and cell morphology.

2.2.3Histogram equalization is an image processing technique used to enhance the contrast and overall visibility of details in an image. It works by redistributing the intensity values of an image's pixels to cover the entire available range. This process can help bring out fine details in both dark and light regions of the image, making it particularly useful when dealing with images that have uneven lighting or limited contrast. Histogram equalization contributes to improved image quality, making it easier to identify and analyse features within the image, which is valuable in various image processing and computer vision applications. Histogram equalization enhances image contrast and feature visibility, which is crucial in malaria classification. By addressing uneven lighting and low contrast issues in blood cell images, it ensures that staining patterns and cell texture variations are discernible. This improvement aids machine learning models in accurately distinguishing between infected and uninfected cells, ultimately contributing to more reliable malaria classification results.

2.3.1 As suggested in [4] Gray world normalization leads to better outcomes in detecting malaria from thin smear blood sample images. In gray scale normalization it is initially assumed the the global average colour of the image is a shade of gray, this works on the assumption that the intensity in any image of the red, green and blue channel are equal. To process the image the process starts with calculating the average intensity of each of the aforementioned color channels. By calculating this the scaling factor for making the intensity of these channels equal can be calculated. The scaling factor is applied to each of the pixel in the image and thus a volor balanced version of the image is attained.

Further the study also uses ImageDataGenetaror It is used to create data generators for both the training and testing datasets. These generators apply various transformations and augmentations to the image data to improve model generalization and performance. Two generators are used one for train data and one for test data with parameters such as rescale, zoom\_range, horizontal\_flip and rotation\_range. After configuring the data generators, the flow method is used to generate batches of preprocessed and augmented data. For the training data generator, it takes the training features X\_train and labels y\_train and specifies a batch size of 64. The purpose of using data generators is to efficiently feed batches of data to the neural network during training and testing, making it possible to work with large datasets without loading all data into memory at once. Data augmentation techniques applied during training can help improve the model's ability to generalize and perform well on new, unseen data.

1. Parasite Detection

The study performed parasite detection on the dataset with 27,558 images. In this stage binary classification is performed on a model trained on this infected and uninfected data. The process starts with image acquisition, after which the image data is pre-processed with noise reduction, filtering and Histogram equalization. This augmented data is used to train and test two models:

* 1. CNN

CNN stands for Convolutional Neural Networks, they come in the class of deep neural networks , and are an important tool for image classification and analysis. There are multiple components to a CNN model the core one being the Convolutional layer, the operations performed in this block involves calculating the dot product of the sliding filter and the local region of the input image. After this calculation the next step is the Activation function this step is responsible for introducing non-linearity, some of the commonly used activation functions include ReLu which is also the one used in the study, other than which there are Sigmoid and Tanh. Being a deep neural network it is important to control the chances of overfitting in the model which is done with the help of Pooling Layer. The final two layers are Fully Connected Layer and Flattening layer which function to produce the final output by connecting each neuron of the layer to the previous layer which gives a way for complex mapping. The flattening layer is used to convert the output of previous layers into a 1D vector for FC layer.

* 1. VGG-19

VGG-19 was proposed by the Visual Geometry Group, similar to CNN it is a deep convolutional neural network. The model is very useful for large scale image recognition. As its name suggests the model in total has 19 layers. Out of these 19, 16 are convolutional layers and the rest three are fully connected layers. The model makes use of 3X3 convolutional filters in the convolutional layer and for max pooling it employs a 2X2 filter with a stride of 2 which is used for down sampling the spatial dimensions. Similar to CNN the study uses ReLu as the activation function for VGG-19 which introduces non-linearity. The fully connected layers of VGG-19 comprise of 4096 neurons each which is followed by a softmax layer for the purpose of classification.

Multiple factors are used to verify the working of the model which include: training accuracy, training loss, validation accuracy, validation loss. These help in making sure that the model gives a reliable accuracy and is also not subjected to over or under fitting.

1. Parasite Classification

For the purpose of parasite classification, the study divided the process in two parts, one on the dataset without any pre- processing and one on the dataset with guassian blur and sharpening. Multiple Deep convolutional networks were trained and tested on these two forms of a dataset containing 956 images per each class of Falciparum, Vivax and Uninfected. This classification would lead to models being able to predict whether a patient has malaria or not, and if the sample is infected it would classify it into one of the two leading types of malaria: Falciparum and Vivax.

The models used for this part of the study are:

4.1 VGG-19

4.2 Xception:

Xception which stands for Extreme Inception in another deep convolutional neural networks model used in the study to determine it’s effectiveness in classifying the blood samples. Xception as the name suggests is a variation of Inception architecture. To make it more computationally effective the modules of a regular Inception model are changed with Depth-Wise separable convolutions. With these modifications it becomes possible to achieve great efficiency with reduced number of parameters.

4.3 InceptionV3

Another model used for the purpose of the study is InceptionV3, it’s a convolutional neural network developed by Google. The model makes the task of image classification simpler by increasing efficiency and effectiveness. Inception modules, batch normalization, and factorization are some of the techniques used for the purpose of effective image classification. These layes also help in speeding us the classification process without compromising on reliability. The convolutional filters in such a model are used parallelly, which results in better capturing of features.

4.4ResNet152

ResNet152 is a deep convolutional neural network architecture which was developed by Microsoft. The model was developed with the purpose of addressing gradient problem. The architecture of the mosel includes deep stack of residual blocks, and each of these consists of multiple convolutional layes which are followed by skip connections which serve the purpose of letting the gradients flow more directly through the network. The model provised an efficient and effective architecture for image classification, and is an excellent model of the ResNet family.

Multiple factors are used to verify the working of the model which include: training accuracy, training loss, validation accuracy, validation loss. These help in making sure that the model gives a reliable accuracy and is also not subjected to over or under fitting.

1. Hybrid Model for Feature Extraction with Ensemble Learning for Classification

The study uses a hybrid InceptionResnetV2 model for feature extraction. This is done on all three forms of data i) Unprocessed ii) Gaussian blur and sharpening iii) Gray world normalization. Using Feature Extraction simplifies the complexity of extracting features from the image data as it can be too complex to directly analyse the voluminous pixel values. Feature Extraction also reduces the issue of overfitting by doing the same. This happens as this reduces the chances of the model learning from the noise or spurious correlations present in the image data.

5.1 InceptionResNetV2

InceptionResNetV2 is a hybrid deep convolutional neural networks architecture. It is a combination of Inception and ResNet family architectures which was developed by Google. It takes the residual networks from Resnet architectures and incorporates with them the inception modules which help in reducing the training speed and performance while the residual networks help Gradient problem.

With the features extracted by InceptionResNetV2, ensemble learning models: Random Forest and XG Boost are trained by splitting the features into training and testing.

5.2 Random Forest

Random Forest is a type of an ensemble learning model which can be used for the purposes of classification, regression etc. the architecture of the model includes multiple decision trees which are trained on different subsets of the data. The prediction of this architecture depends on the votes of each individual tree, the final outcome depends on majority voting for classification.

5.3 XGBoost

XGBoost stands for extreme gradient boosting, the architecture is used for supervised learning tasks. XGBoost is a type of an ensemble learning model and is based on gradient boosting. Gradient boosting is a technique of combining a few weak learners which in turn creates a powerful model. The weak learners in XGBoost are a sequence of decision trees which are laid out in a stage wise manner. All the wrong predictions made by a specific tree are fed to the next one.

Multiple factors are used to verify the working of the model which include: accuracy, precision, recall, confusion matrix. These help in making sure that the model gives a reliable accuracy and is also not subjected to over or under fitting.

Results and Discussion

1. Parasite Detection

For the purpose of parasite detection, the dataset of thin blood smear images is processed with Noise reduction, sharpening, and Histogram Equalization. This allows for the models to run more efficiently and avoid extracting features such as noise.

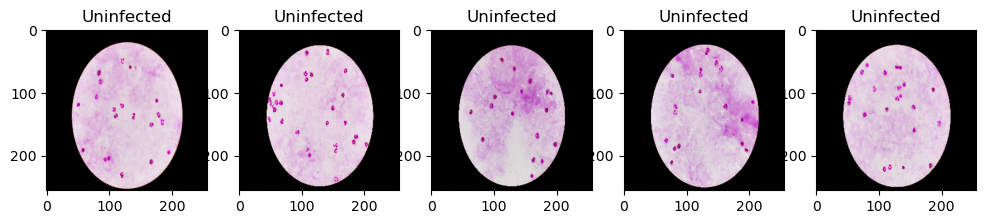


Figure 1 Before Pre-processing

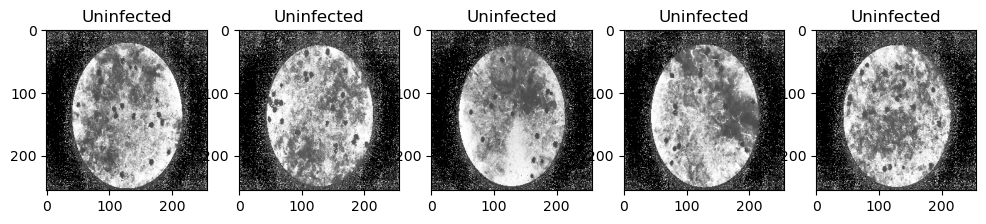


Figure 2 After Pre-processing

Once the data is pre-processed CNN and VGG-19 are used for the purpose of training and testing. For this purpose the dataset is split: 70% for training and 30% testing. ImageDataGenerator is used for this purpuso which provide as an easy solution for image augmentation. Different techniques like standarizattion, rotation, shifts, flips, brightness change etc are used to do so.

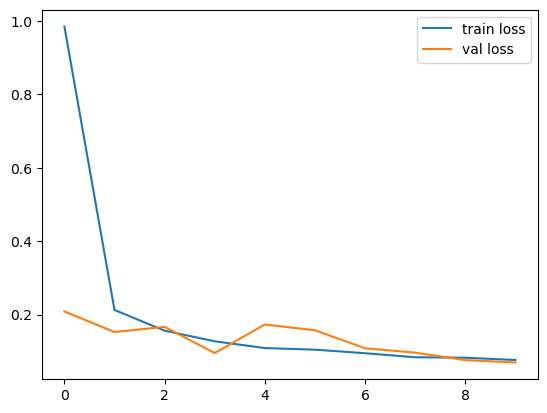


Figure 3 Training and Validation Loss for CNN

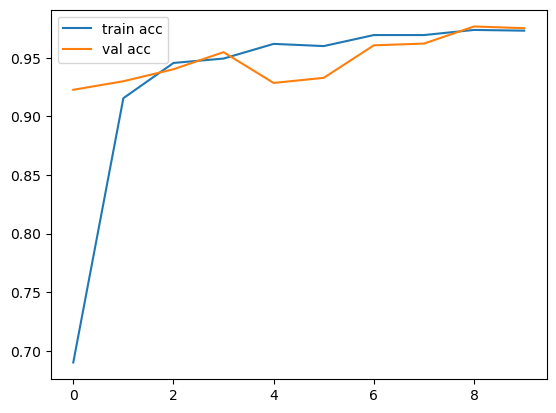


Figure 4 Training and Validation accuracy for CNN

The training process has been carried out for a total of 10 epochs, implying that the dataset has been iterated over 10 times during the training phase. From the training and validation accuracy and loss we can see how with each epoch the model provides with better accuracy. After ten epochs the model has a testing accuracy of 97.31% and validation accuracy of 97.52%, as the difference between the testing and validation accuracy is minimal it indicates that the model generalizes well to unseen data.

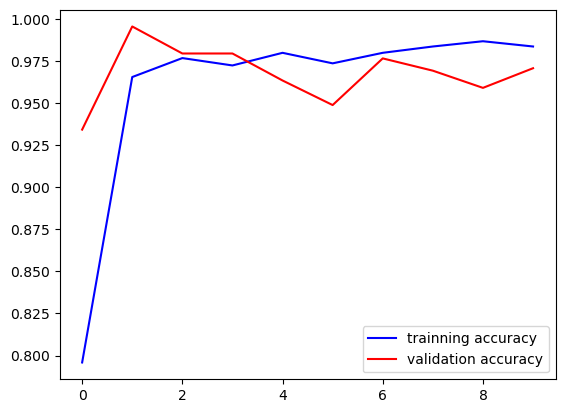


Figure 5 Training and Validation accuracy for VGG-19

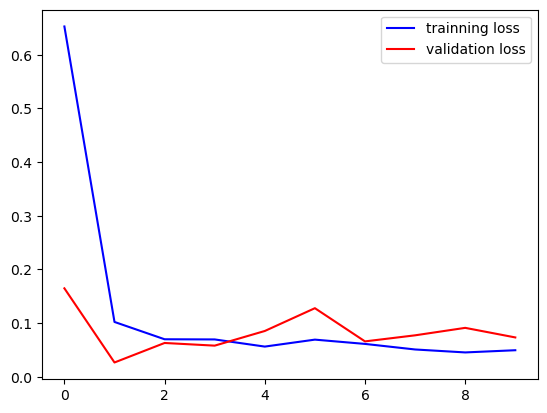


Figure 6 Training and Validation loss for VGG-19

The dataset is further iterated over 10 times for training the VGG-19 model. The model gives an accuracy of 98.37% and validation accuracy of 97.08%. The loss values are decreasing over epochs, which indicates that the model is learning to minimize its error. The validation accuracy is reasonably high, indicating that the model generalizes well to unseen data.

|  |  |  |
| --- | --- | --- |
| Model | Testing Accuracy | Validation Accuracy |
| CNN | 97.31% | 97.52% |
| VGG-19 | 98.37% | 97.08% |

1. Parasite Detection

For parasite detection the study first generates two types of data i) Unprocessed ii) Processed with Gaussian blur and sharpening, these data sets had three categories of data: i) Falciparum ii) Vivax iii) Uninfected. On these two sets of data several Image processing models are used to be trained and tested for parasite detection and classification. Gaussian blur and sharpening help reduce noise and smoothen the image, this give the data depth perception. The preprocessing enhances the details and features of the images making training more effective.

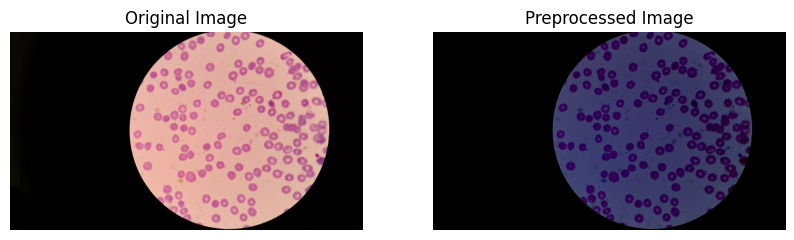


Figure 7 Before and after Pre-processing

The models used for training are: VGG-19, InceptionV3, Xception, and Resnet-152. For all these models the training process is carried in 10 epochs. After the epochs Testing and training validation and loss are plotted to observe the behaviour of the models with each epoch.

2.1 Unprocessed Data

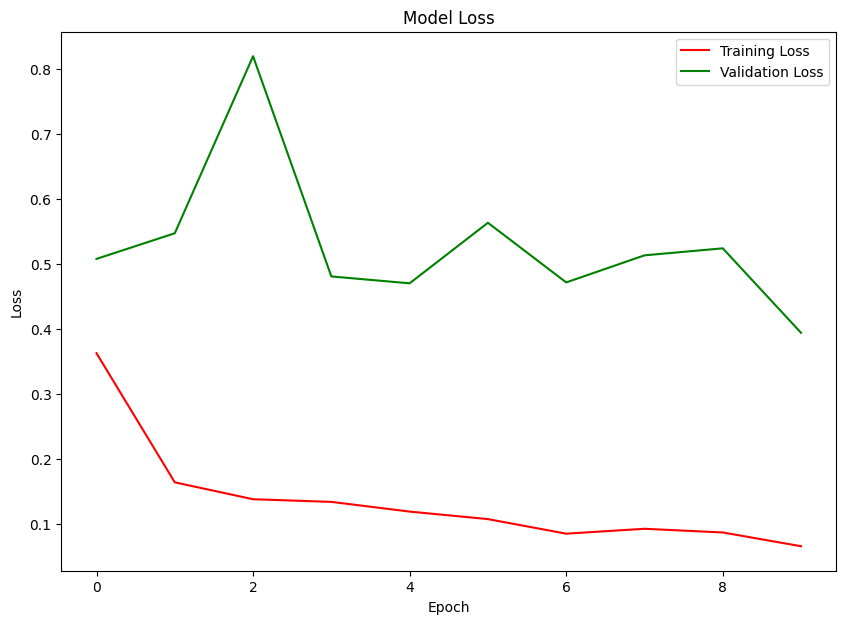


Figure 8 Training and validation loss: VGG-19

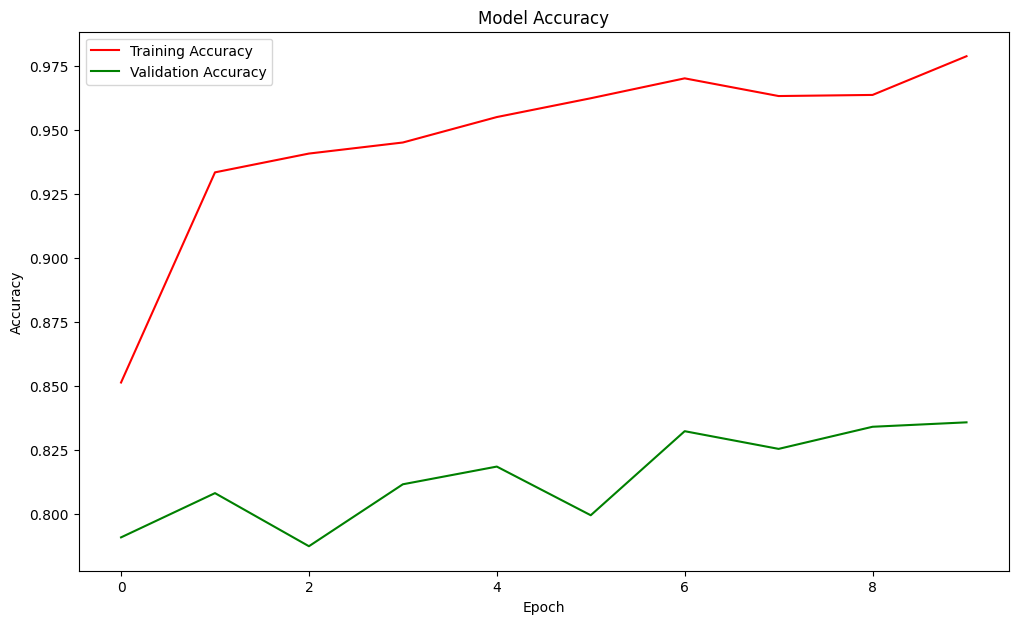


Figure 9 Training and testing Accuracy: VGG19

Initially, with a loss of 0.3623 and an accuracy of 85.15% on the training set, the model exhibits decent performance, yet it struggles slightly on unseen data, reflected in the validation loss of 0.5074 and accuracy of 79.10%. As training progresses, both training and validation losses gradually decrease, indicating the model's ability to better fit the data. The increasing disparity between training and validation accuracies in later epochs, suggests potential overfitting. Despite this, the model achieves a validation accuracy of around 83-84%, demonstrating its capacity to generalize to new data.

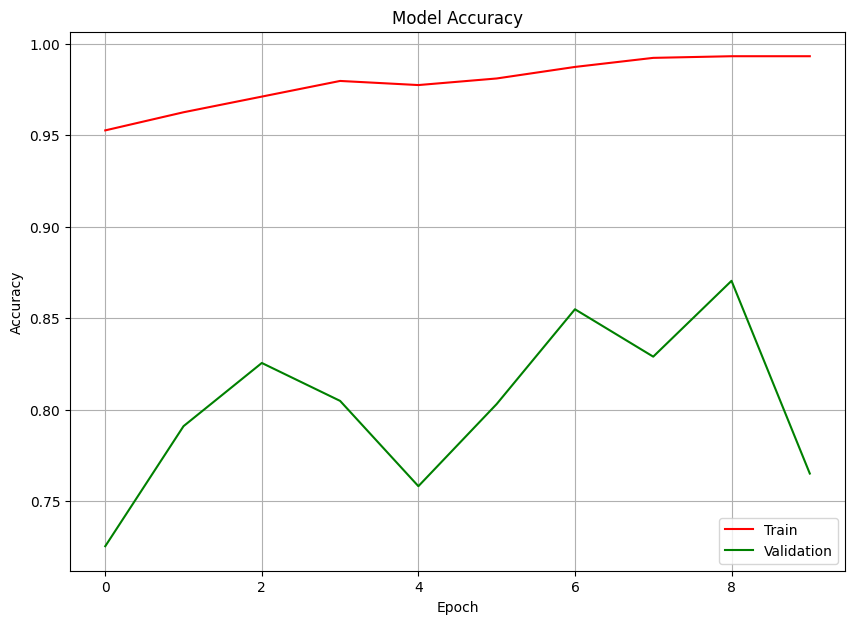


Figure 10 Training and testing accuracy: InceptionV3

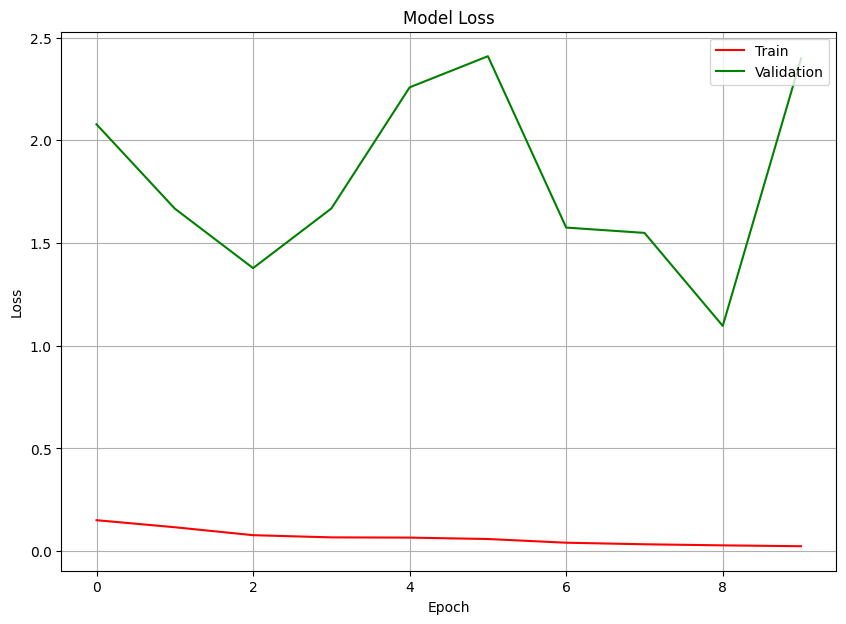


Figure 11 Training and testing loss: InceptionV3

For inceptionV3 it can be observed that there is a vast difference between the training and validation accuracy, which suggests overfitting. Same can be observed for loss, as the training goes on the loss for training set reduces but for validation increases. The trained model does not perform reliably for unseen data.

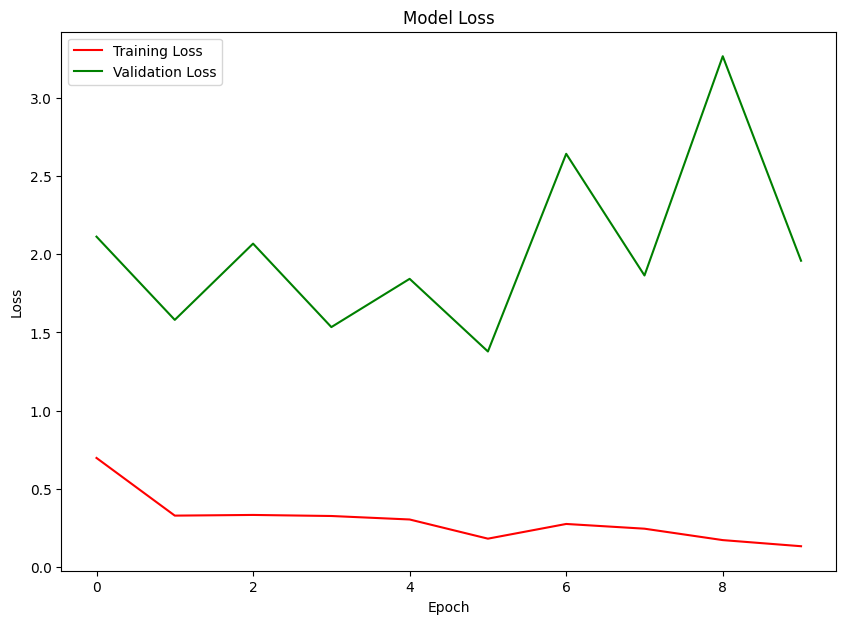


Figure 12 Training and testing loss: Xception

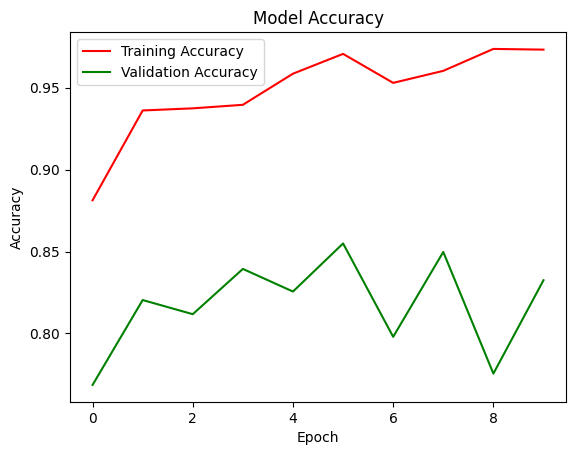


Figure 13 Training and Testing accuracy: Xception

The model starts with promising accuracy on both training (88.13%) and validation (76.86%) sets, although the validation loss is notably higher. While the model's accuracy improves over epochs, fluctuations in metrics suggest potential instability in learning. Notably, validation loss varies significantly, indicating potential challenges in convergence or overfitting. Further analysis, including regularization techniques and model adjustments, may be necessary to enhance stability and generalization.

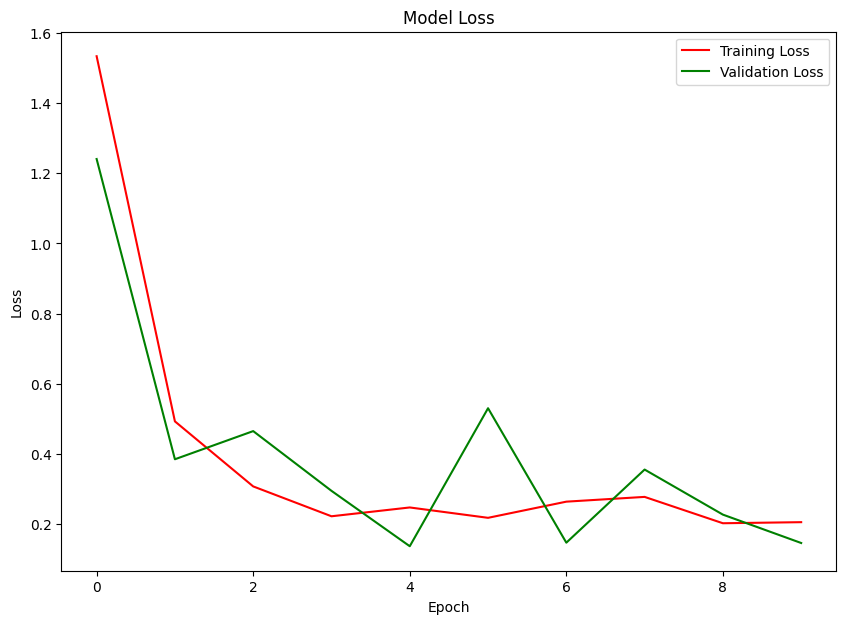


Figure 14 Training and testing loss: ResNet152

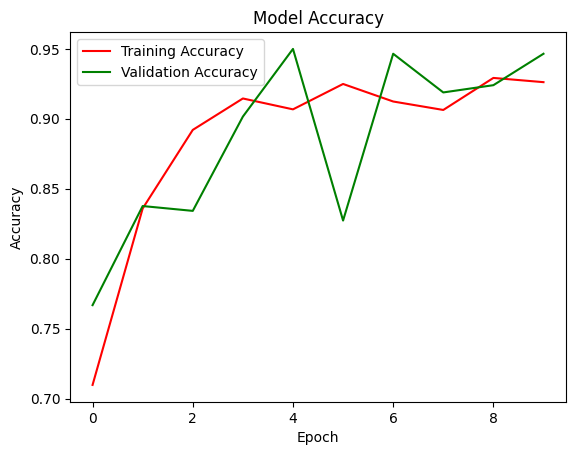


Figure 15 Training and testing accuracy: ResNet152

Initially, the model achieves moderate accuracy on both the training (70.98%) and validation (76.68%) sets, with corresponding losses of 1.5331 and 1.2404, respectively. Subsequent epochs witness improvements in both accuracy and loss metrics, with accuracy peaking at 92.49% on the training set and 94.99% on the validation set. Fluctuations in these metrics across epochs suggest potential instability in the learning process. The validation accuracy displays variability, indicating possible challenges in model generalization.

2.2 Pre-processed Data

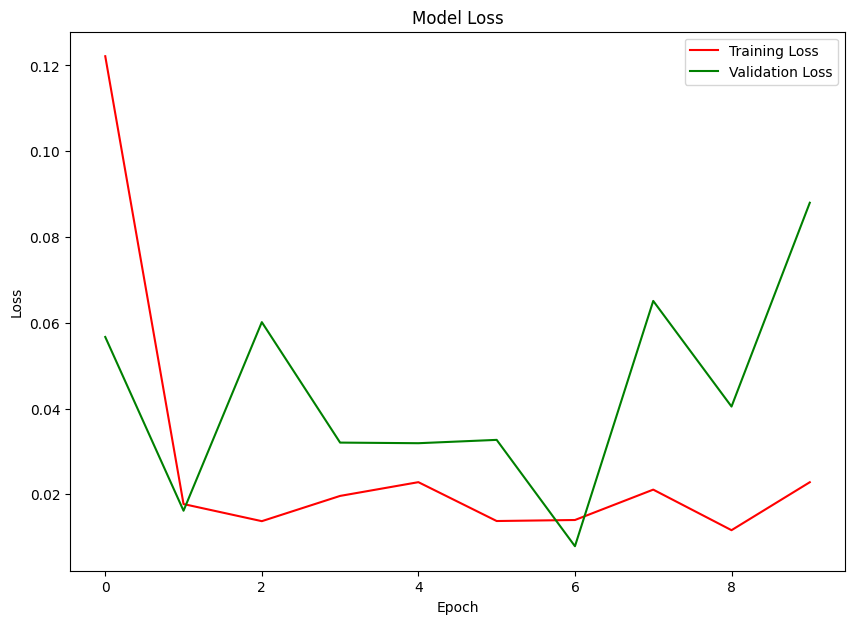


Figure 16 Training and testing loss: VGG-19

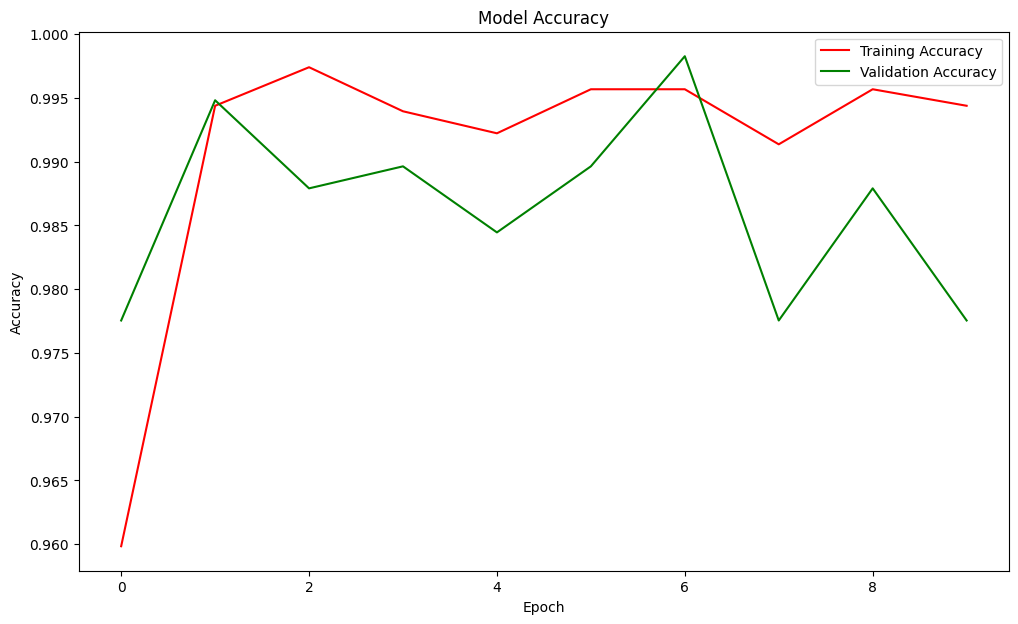


Figure 17 Training and testing accuracy: VGG-19

Initially, the model achieves a high training accuracy of 95.98% and validation accuracy of 97.75%, with corresponding losses of 0.1222 and 0.0567, respectively. The model consistently improves, reaching peak accuracies of 99.74% on the training set and 99.83% on the validation set by the third epoch. Validation accuracy occasionally surpasses training accuracy, indicating effective generalization. Fluctuations in accuracy and loss metrics in later epochs suggest potential instability or overfitting.

References

[1] Garba, S., Abdullahi, M.B., Bashir, S.A. and Abisoye, O.A., 2022, November. Implementation of Malaria Parasite Detection and Species Classification Using Dilated Convolutional Neural Network. In *2022 5th Information Technology for Education and Development (ITED)* (pp. 1-6). IEEE.

[2] Kakkar, B., goyal, M., Johri, P. and Kumar, Y., 2023. Artificial Intelligence-Based Approaches for Detection and Classification of Different Classes of Malaria Parasites Using Microscopic Images: A Systematic Review. *Archives of Computational Methods in Engineering*, pp.1-20.

[3] Tan, D. and Liang, X., 2023. Multiclass malaria parasite recognition based on transformer models and a generative adversarial network. *Scientific Reports*, *13*(1), p.17136.

[4] Swastika, W., Kristianti, G.M. and Widodo, R.B., 2021, April. Effective preprocessed thin blood smear images to improve malaria parasite detection using deep learning. In *Journal of Physics: Conference Series* (Vol. 1869, No. 1, p. 012092). IOP Publishing.

[5] Ohdar, K. and Nigam, A., 2023, July. A Robust Approach for Malaria Parasite Identification with CNN Based Feature Extraction and Classification using SVM. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

[6] Wang, C., Dong, Y., Li, C., Oberstaller, J., Zhang, M., Gibbons, J., Pires, C.V., Xiao, M., Zhu, L., Jiang, R.H. and Kim, K., 2023. MalariaSED: a deep learning framework to decipher the regulatory contributions of noncoding variants in malaria parasites. *Genome Biology*, *24*(1), p.231.

[7] Jusman, Y., Aftal, A.A., Tyassari, W., Kanafiah, S.N.A.M., Hayati, N. and Mohamed, Z., 2023, August. Classification of Parasite Malaria in Schizon Stage with GoogleNet and VGG-19 Pre-Trained Models. In *2023 10th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* (pp. 219-223). IEEE.

[8] Jusman, Y., Ardiyanto, Y., Kanafiah, S.N.A.M., Mohamed, Z. and Hassan, R., 2022, November. Classification System of Malaria Disease with Hu Moment Invariant and Support Vector Machines. In *2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)* (pp. 365-368). IEEE.

[9] Kundu, T.K. and Anguraj, D.K., 2023, March. Using Image Classification-based RBC Extraction for Malarial Parasite Identification in Peripheral Blood Smear. In *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 1554-1559). IEEE.

[10] Sifat, M.M.H. and Islam, M.M., 2020, June. A fully automated system to detect malaria parasites and their stages from the blood smear. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 1351-1354). IEEE.

Top of Form