New ensemble learning approach to detect malaria from microscopic red blood smear images

POULAMI BERA 20BCE1305 | TANUSHREEE GORAI 20BCE1269 | VANSH SHAN 20BCE1533

BACHELOR OF TECHNOLOGY IN VELLORE INSTITUTE OF TECHNOLOGY , CHENNAI

School of Computer Science

ABSTRACT :
Malaria is a life-threatening parasitic disease spread by infected female Anopheles mosquitoes. After analysing it, microscopists detect this disease from the sample of microscopic red blood cell images. A professional microscopist is required to conduct the detection process, such an analysis may be time-consuming and provide low quality results for large-scale diagnoses. This paper develops an ensemble learning-based deep learning model to identify malaria parasites from red blood cell images. VGG16(Retrained), VGG19(Retrained), and DenseNet201(Retrained) are three models that are used in developing the adaptive weighted average ensemble models. To reduce the dispersion of predictions, a max voting ensemble technique is then applied in combination with adaptive weighted average ensemble models. A variety of image processing techniques are utilised including the data augmentation technique to increase the number of data and solve the overfitting problem of the model. Some other approaches of custom CNN, Transfer Learning, and CNN-Machine Learning (ML) classifier techniques are also implemented for

comparing their performance with the ensemble learning model. The proposed ensemble learning model provides the best performance among all with an accuracy of 97.92% to classify parasitized and uninfected cells. Therefore, the deep learning model has the potential to diagnose malaria more accurately and automatically.

KEYWORDS:
Machine learning
Deep learning
CNN Transfer learning
Ensemble Malaria

1. INTRODUCTION:

Need for the system-Malaria is a serious disease that can be fatal. An infected Anopheles mosquito bite is the most common way to contract the disease. Infected mosquitos carry plasmodium parasites. Malaria is transferred when a mosquito bites an infected person and then bites a non-affected person. When parasites get mature, they travel to the liver and enter the bloodstream, where they begin to infect red blood cells after a few days. Malaria can also be transmitted through organ transplants, blood transfusions, and the use of blood-contaminated syringes and needles. Malaria seems to be the most common disease in tropical and subtropical regions, where the parasites proliferate. According to the most recent World Malaria Report, estimated 241 million malaria cases in 2020, slightly higher compared to 227 million in 2019. It was estimated that Malaria killed 627,000 individuals in 2020. Children under age 5 are in the most dangerous position. In 2020, 80% of children worldwide died from malaria [20]. Plasmodium parasites are the cause of malaria. Humans can be infected by five different malaria parasites. These are Plasmodium falciparum, P. vivax, P. ovale, P. malariae, and P. knowlesi. Malaria signs typically appear 10 to 4 weeks after the infection has occurred. It can take several months for symptoms to appear in some cases. Some common signs of malaria disease include moderate to severe shivering chills, high fever, muscle pain, profuse sweating, headache, vomiting, bloody feces, abdominal pain, diarrhea, nausea, and convulsions [5]. Because of the loss of red blood cells, malaria can induce jaundice (yellow skin and eyes) and anemia. If not treated

promptly, the infection can cause kidney failure, mental confusion, coma, convulsions, and death

Current status –The most common and accurate test is the thick and thin blood smear test. Microscopists determine the parasites from red blood cells using a microscope. This test must require trained and experienced microscopists. Another is a Rapid diagnostic test or antigen test. When microscopy is not available, a rapid diagnostic test can be performed instead of blood smears. This test identifies malaria antigens from a person's blood and changing the color of the testing result gives a positive result. It also has no way of knowing whether the infection is minor or severe. Another is the Molecular test or Polymerase chain reaction (PCR). The PCR is a laboratory tool to detect and identify Plasmodium species from parasite DNA. In laboratories where there is a limitation of training and expertise in microscopic examination of malaria, a PCR test might be a preferable option to begin the treatment. The two other tests are the Antibody test and the Susceptibility test [4]. Therefore, there are several benefits of using an automated malaria detection process. It provides much higher accurate results compared to the manual process. It reduces the workload to serve more patients. Deep learning has that much potential to provide an accurate and faster result for parasites.

Platform/Application-All the experiments are performed in the Google Colab platform with 13 GB Random Access Memory (RAM), 69 GB Colab Disk, Keras® 2.4.3 API with Tensorflow® 2.5.0 backend, Python® 3.7.10, Nvidia® Tesla T4 Graphical Processing Unit (GPU), CUDA V11.2 dependencies for GPU acceleration and matplotlib® 3.2.2 library

Data preprocessing- Preprocessing methods he mentioned mean filter, Laplacian filter, wiener filter, median filter, adaptive histogram equalization, contrast enhancement, etc. Segmentation techniques are morphological operation, Hough transformation, k-means clustering, Fuzzy segmentation, etc. Features that are mentioned in his paper are a color feature, texture feature, and morphologic feature.

The sizes of malaria images are varying from 110 pixels to 150 pixels. In this experiment, the cell images are scaled to 64 pixels in height and 64 pixels in width. Fig. 3 includes the resized cell images of malaria. Dataset splitting is the general part of model building. The amount of the data impacts how the dataset is split [3]. In this research, the dataset is split into 70% of the training set, 10%

of the validation set, and 20% of the testing set using scikit learn library. The model is trained to find patterns in the data by utilizing the training set. The validation set is used to test the model's performance during training. Finally, the model is tested with a test set when it has completed training.

Table 1
Dataset splitting into training, validation, and testing sets.

Splitting Type	Percentage
Training	70%
Validation	10%
Testing	20%

Classification-In the classification method, he included supervised and unsupervised learning algorithms. Deep learning is the most prominent way to detect the disease automatically and more correctly. Nowadays, the most often used approach is the convolutional neural network (CNN) [16]. It has the potential to extract features from images automatically. In Ref. [22], the author proposed a fast Convolutional Neural Network of six convolutional layers, one fully connected layer, and one classification layer. He also evaluated different transfer learning techniques such as AlexNet [15], VGG16 [31], ResNet50 [11], DenseNet121 [12]. Finally, he provided an overall accuracy of 96.7% for all models using the test dataset. Transfer learning is a way of using pre-trained networks. He found an accuracy of 95.9% using the transfer learning model.

Contributions-In Ref. [29], the author developed a shallow CNN model that performs similarly to the VGG16 [31] and ResNet50 [11]. In deep learning, high computational complexity and high computational cost is a big issue. His proposed custom CNN has lower computational complexity and lower computational run time. Evaluation metrics used in his research are accuracy, sensitivity, specificity, F1 score, and MCC (Mathew's Correlation coefficient). He found an accuracy of 95.32% in his research. In Ref. [7], the author proposed a three-stage pipeline that is segmentation, crop & musk, and classification. In the segmentation stage, a segmentation neural network (SNN) [2] has been built for red blood cell (RBC) segmentation. Another 13 layered

CNN architecture has been created to classify malaria disease. He got an overall 93.72% accuracy from his custom CNN architecture. In Ref. [9], the author introduced a new approach to classify malaria disease that is CNN-SVM and CNN–KNN. In this approach, CNN is employed as a feature extractor and SVM/KNN is applied as a classifier. In addition, he demonstrated a new autoencoder technique. Autoencoder is a form of neural network that is trained to regenerate output from its input. Again in Ref. [8], the author proposed a pre-trained network technique. He employed LeNet [17,18], AlexNet [15], and GoogLeNet [34] in his research. A different dataset was collected from the pathologists of the University of Alabama at Birmingham. Pre-trained networks were used as a feature extractor. Using KUllback-Leibler (KL) distance, the best seven features were chosen from 76 features. He applied an SVM classifier to classify malaria. In Ref. [21], the author proposed a custom CNN architecture. Bilateral filter & data augmentation techniques were applied as data preprocessing steps. The preprocessed input images were then fed into a custom-built CNN model. He found an accuracy of 96.82%. In Ref. [25], the author introduced a unique CNN architecture called Attentive Dense Circular Net (ADCN) in their research. ADCN was inspired by ResNet [11] and DenseNet [12]. There are three parts to the ADCN architecture. In the first part, there is a custom CNN model with two dense blocks and one attention module. The dense block has six dense conv block that consists of two convolutional layers. Every dense conv block is connected with another dense conv block. Finally, an attention module is split into two parts. One is a 1x1 convolutional layer, while the other is an attention branch for extracting the attention feature map. To ensure that the branch's output size is consistent, the attention branch uses downsampling and upsampling techniques. In Ref. [26], the author proposed three different techniques that are custom CNN network, transfer learning as VGG16 [31], and CNN-SVM. Then he used the ensemble learning method to reduce prediction variance and generalisation error. The results from each of the three networks are combined, and the final prediction is calculated using a weighted average ensemble. In Ref. [1], the author proposed a novel method known as Incremental Modular Networks (IMNets) to classify malaria. This method combines multiple SubNet to provide additional information. Each SubNets module is added iteratively, either serially or parallelly, to the current architecture. In Ref. [10], the author discussed some protein sequence formulation strategies such as discrete methods, biochemical, physiochemical, and natural language processing techniques. These are used to convert protein

sequences into numerical descriptors. Four classification methods are applied, and the predicted outcomes of these classifiers are then merged to develop an ensemble model using majority and genetic algorithm.

2. LITERATURE REVIEW-

1) Predicting Infectious Disease Using Deep Learning and Big Data

Sangwon Chae, Sungjun Kwon and Donghyun Lee

With the expansion of the Internet, search query data, social media big data, and meteorological data were used in this study to review the factors influencing the prevalence of infectious diseases. Also, it created conventional prediction models like as OLS, ARIMA, and deep learning prediction models like DNN and LSTM and evaluated their prediction abilities in order to establish that the deep learning models are the most effective in predicting infectious diseases. Deep learning-based infectious illness prediction models are thought to be able to complement existing infectious disease surveillance systems while also predicting future trends in infectious disease. It is anticipated that quick responses to infectious disease will become possible and costs to society will be kept to a minimum if this can lessen the time gaps in reporting systems so that infectious disease trends can be understood promptly.

2)Social media based surveillance systems for healthcare using machine learning: A systematic review

Aakansha Gupta, Rahul Katarya

The study identifies the various social media or web-based platforms utilised for surveillance in the healthcare domain, as well as the health topic(s), based on the corpus of 148 selected articles. 26 articles that used machine learning techniques were found in the corpus of the chosen articles. In these articles, analysed to identify frequently employed ML methods. The bulk of research (24%) concentrated on monitoring influenza and influenza-like illnesses (ILI). Twitter (64%) is the most often utilised social media text data source for surveillance research, while Support Vector Machine (SVM) (33%) is the most widely used machine learning technique for text classification.

Compared to conventional syndromic surveillance systems, the capacity to anticipate diseases has increased with the incorporation of internet data in surveillance systems. The drawbacks and difficulties with social media-based monitoring systems, however, include noise, demographic bias, privacy concerns, etc. Future directions are discussed in our study, which may be helpful to specialists. Researchers can expand social media-based surveillance systems in the healthcare industry by including the future research we discuss in this study and using this publication as a reference library.

3)Performance Analysis of Deep Learning Algorithms in Diagnosis of Malaria Disease

K. Hemachandran,Areej Alasiry, Mehrez Marzougui , Shahid Mohammad Ganie, Anil Audumbar Pise, M. Turki-Hadj Alouane and Channabasava Chola

In this study, we examine three deep learning methods for identifying malaria, including CNN, MobileNetV2, and ResNet50. Conclusions were drawn based on a comparison of the developed models to determine which was superior. We might therefore say that Environment-related factors are essential for facilitation. Existence and transmission of malaria. ResNet50 outperformed the other models and delivered the best outcomes for the diagnosis of malaria. To validate the findings, statistical metrics including precision, recall, f1-score, roc curve, etc. were computed. When compared to previous research, it can be said that this study exhibits state-of-the-art findings. For better outcomes, the work might be expanded to investigate more deep learning methodologies with other image processing pretreatment techniques.

4)Leveraging Deep Learning Techniques for Malaria Parasite Detection Using Mobile Application

Mehedi Masud,Hesham Alhumyani,Sultan S. Alshamrani,Omar Cheikhrouhou, Saleh Ibrahim,Ghulam Muhammad,M. Shamim Hossain, and Mohammad Shorfuzzaman

First, a proprietary CNN-based end-to-end deep learning model was assessed in the paper to enhance the identification of malaria on thinblood smear

images. We demonstrated how cyclical Using a learning rate schedule with an automatic learning rate finder along with a regularisation method that is frequently used, including batch normalisation and dropouts, leads to encouraging outcomes in the classification of malaria. In categorising parasitized and uninfected cell pictures with a high level of precision and sensitivity, our best model obtains an accuracy of 97.30%. The model also produces a high MCC value (94.17%) when compared to all other models already in use, which shows a significant connection between predicted and actual labels. Also, we found that the proposed improved model outperformed previously developed CNN models (pretrained models like VGG-16 and ResNet-50) [4] in terms of accuracy, precision, sensitivity, and MCC when distinguishing malaria-free and infected cells. For easier and quicker malaria detection, we integrated our top-performing model into an android-based mobile application. We therefore expect that the outcomes of this research will contribute to the creation of worthwhile mobile-based solutions that will address issues with treatment reliability and a lack of medical competence. This work will be immediately extended by applying picture augmentation on the training data in an effort to further reduce the overfitting issue and by experimenting with various adaptive SGD optimizer variants to see how they affect performance outcomes. We also intend to use ensemble approaches and model stacking in the future to improve prediction.

5)Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security

Xin Guo, Muhammad Arslan Khalid, Ivo Domingos, Anna Lito Michala, Moses Adriko, Candia Rowel, Diana Ajambo, Alice Garrett, Shantimoy Kar, Xiaoxiang Yan, Julien Reboud,Edridah M. Tukahebwaand Jonathan M. Cooper

A smartphone-based end-to-end platform for multiplexed DNA-based lateral flow diagnostic tests in far-off, rural areas has been disclosed by us. poor resource conditions. Our decision-support tool offers automatic results detection and analysis, supplementing human expertise, and blockchain technology is used to encrypt, trust, and authorise data handling transactions. We built our platform to support the following features in anticipation of future AI guidelines for the healthcare industry: explainability, accuracy to

enable AI decision trust, ethical data use through privacy-preserving blockchain networks, interoperability to enable wider connectivity with varied standards and policies, and data formatting for standardisation and provenance. We will enhance user-friendliness in the future for practitioners in many sub-Saharan nations.

6)The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review

RaynerAlfred, Joe HenryObit

The purpose of this literature review is to identify and analyse various approaches, types of datasets, parameters or variables, individual models, ensemble models, performance measures, and approaches used in the prior works on utilising machine learning approaches to control the spread of deadly disease outbreaks. In this study, six online digital libraries were used to find all relevant peer-reviewed articles. Just 47 research were chosen from among those published between 2010 and 2020, and their quality was evaluated using seven key questions. Based on the earlier listed seven questions, this SLR was carried out to evaluate and choose all pertinent research studies relating to the detection and prediction of disease outbreaks using machine learning. The following is a summary of the contributions made by this paper: •The types of databases and variables used are defined, and it is discovered that meteorological and epidemiological data are the most helpful datasets for predicting and detecting disease outbreaks. •Data from multiple sources helps to enhance predictions of disease epidemics. •When compared to other linear and non-linear machine learning techniques, neural network family algorithms have been found to perform better. •Hybrid and ensemble methods fared better and should be used for predicting and detecting disease outbreaks. •Exploring unstructured data, such as that found in news articles, blogs, and search phrase trends, may help predict and detect disease outbreaks more accurately.

7)Predicting malaria epidemics in Burkina Faso with machine learning

David HarveyID, Wessel Valkenburg, Amara Amara

We develop a test set with three possible outcomes: first, a situation where cases increase (similar to an epidemic); second, a scenario where case

numbers remain unchanged; and third, a scenario where case numbers decrease. We discover that the instance where our method is least sensitive is when figures increase and become more exact when the number of cases decreases. This is because the fast, exponential rise in the case rate is challenging to model, in contrast to the decrease, which is frequently linear and simpler to forecast. A five-tier epidemic alert system is calibrated and put to the test using the algorithm's lower limit threshold. The lower-bound threshold alert is based on a scenario in which we issue a warning whenever the lower limit of our expected case rate (at a specific confidence interval) exceeds the threshold for an epidemic (the five-year mean plus two standard deviations). The 95% confidence lower-bound, for instance, represents the upper limitation at which we are 95% certain that the case rate will exceed this prediction. We will be very confident that an epidemic will happen in the next 13 weeks if this limit is higher than the level that signals the start of one. We evaluate and test this warning system. We discover that our system's accuracy in predicting epidemics for the 95th, 68th, 32nd, and 5th percentiles, respectively, is 32%, 51%, 83%, and > 99%. The recall rate, however, is > 99%, 90%, 66%, and 5%, respectively, for the same lower-limits. We discuss potential biases in the algorithm and emphasise the need for additional characteristics to be added in order to break any potential societal degeneracies, such as strikes and malaria intervention programmes. Also, there has to be more research done to identify any potential demographic biases in the system. We conclude by pointing out the algorithm's limitations. Predictions must be scaled to the entire population because the existing IeDA database only includes consultations of infants under the age of five. Solving the scaling issue is not an easy task. In addition, the data is It is only three years old, hasn't covered many malaria seasons, and doesn't currently cover the full country of Burkina Faso. As a result, it might not be as accurate a predictor in its current form as modern epidemiological techniques (such as explicitly modelling disease transmission). In the future, such a comparison would be intriguing. Despite Not withstanding these drawbacks, this work represents the first attempts to create a data-driven malaria predictor in sub-Saharan Africa.

8)Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system

Christopher L. Merkord, Yi Liu, Abere Mihretie, Teklehaymanot Gebrehiwot, Worku Awoke,Estifanos Bayabil, Geoffrey M. Henebry, Gebeyaw T. Kassa, Mastewal Lake and Michael C. Wimberly

In order to provide nearly real-time malaria forecasts in the environment, the EPIDEMIA system has simplified the integration of malaria surveillance data and environmental monitoring data. Region Amhara in Ethiopia. As a result, it has been possible to involve end users in a continuous process of feedback and improvement and to share malaria forecasts to public health partners for an extended length of time. The creation and execution of EPIDEMIA have brought to light a number of factors that anyone looking to create such a system should take into account. The need for software tools and an enabling environment to deliver timely harmonised epidemiological and environmental data, the significance of ongoing stakeholder input throughout the system's design, implementation, and operation, and the requirement to be adaptable to changes in the input data are among the critical points. The evaluation and improvement of forecasting models as new data is added to the system are ongoing challenges, as are the creation of more complex reporting features like interactive web-based visualisations, a better integration of early detection and early warning results into public health and emergency management decision making, and ultimately the transfer of the tools and knowledge necessary to run the system to Ethiopia's public health sector. After achieving these longer-term objectives, the system should be adaptable to changes in the larger social and environmental contexts of malaria as well as extendable to other diseases and regions.

9)Internet-based surveillance systems for monitoring emerging infectious diseases

Gabriel J Milinovich, Gail M Williams, Archie C A Clements, Wenbiao Hu

Public health officials and governments face a challenging problem from emerging infectious diseases, which has been made worse by the globalisation and fast evolving patterns of human behaviour. Calls for new technology and methods for detection, tracking, reporting, and treatment of developing infectious diseases have increased as a result. response. Internet-based surveillance systems provide a cutting-edge and evolving way to keep an eye on public health issues, such as newly emerging infectious

diseases. We examine studies that have used internet usage and search patterns to track the progression of dengue and influenza. Internet-based surveillance systems and conventional surveillance methods are quite consistent. Internet-based methods are also advantageous logistically and financially. Yet, they are unable to take the position of conventional surveillance systems; rather, they should be seen as an addition rather than a substitute. The capacity of conventional surveillance systems for newly developing infectious illnesses could be improved by the use of data produced by internet-based surveillance and response systems, according to future research.

10)Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health Classification

Ovidiu Șerban,Nicholas Thapen, Brendan Maginnisa, Chris Hankina, Virginia Football

Now, the system gathers, analyses, and saves over 1.8 million tweets every day. It produces events and related situational awareness reports every day. It conducts nowcasting and gets CDC data on a weekly basis. The entire system was created using dependable open-source technologies that represent the cutting edge of software engineering. The analysis of our findings demonstrates the strength and accuracy of the health classifiers, with the selected classifier providing an F1 of 0.852 for the Twitter classification job and 0.939 for the news classification challenge. These classifiers performed better than the baselines, proving the value of deep learning in this area of text classification. Although there is space for improvement, our news crawler is retrieving a sizable number of health-related articles that are connected to a sizable portion of observed occurrences. The evaluation of event detection demonstrates the value of the tools provided on the Event Details page of the App for event evaluation. With the right filter parameters, almost one-third of detected events were significant enough to be verified by the ground truth data currently available. Last but not least, the Nowcasting evaluation revealed that using our Twitter data increased Nowcasting accuracy by 13% over the baseline. • Including epidemiological models: Extending nowcasting to include data from the specific disease module and epidemiological models. Forecasting disease levels in the future would be considerably more possible as a result. • Better News Linkage: The primary

topic of each news story could be determined using topic modelling techniques like Latent Dirichlet Allocation (LDA). In order to ensure that only articles that specifically discuss the symptom or disease in issue are connected, this might then be utilised to improve the linking of news articles and events.

3. SYSTEM ARCHITECTURE

An ensemble learning-based system is developed in this research to reduce the burden on microscopists. Ensemble learning is a technique where multiple models are trained on the same dataset and combined with their results to achieve better accuracy and reduce the variance of the model. The input data is initially resized into 64 pixels of height and 64 pixels of width. Then, the dataset is divided into 70% training, 10% validation, and 20% testing sets. After that data augmentation technique is applied with a variety of domains to increase the number of training data which enhances the model's performance. VGG16(R), VGG19(R), and DenseNet201(R) models are fine-tuned using several hyperparameters such as optimizer, learning rate, activation function, etc.* Then, these models are trained to learn patterns in the data using a training set. From the individual model, two of the models are taken and applied to an adaptive weighted average ensemble to achieve better performance by reducing the variance of the model. This method finds the best outcome with the appropriate weights automatically. VGG16(R) and VGG19(R) achieve a higher result with 0.7 and 0.6 wt respectively. With 0.6 and 0.5 wt, VGG19(R) and DenseNet201(R) produce a better result. When compared to other weights, VGG16(R) and DenseNet201(R) perform better with 0.4 and 0.5 wt, respectively. Three ensembled models are obtained after performing an adaptive weighted average technique. After that, another ensemble technique known as max voting is applied using three ensembled models with hard voting to improve the model's performance and reduce the dispersion of the predictions.

Custom CNN architecture

The custom CNN architecture is developed with a total of 16 layers which has 3 blocks of convolutional, activation, max pooling, and batch normalization layers, 1 flatten, and dropout layer, and the last 2 dense or fully connected layers. The first convolutional layer performs convolved operations with 32 kernels to extract feature maps from the input images. The kernel is applied with a 5x5 matrix including the same padding and 1-pixel stride to maintain the

same dimension of the feature map. The output of a convolution layer is processed through a ReLu activation function to introduce non-linearity. It speeds up the training process and converges faster [30]. It also allows neurons to learn arbitrary complex transformations. The ReLu activation function is shown in Equation (1), where negative values become zero and positive values remain the same.

$$R(x) = \max\ (0, x)$$

Here, R is the result after adding non-linearity to matrix x. Then a max-pooling layer is applied with a 3x3 matrix and a stride of 2 to reduce the dimension of images from the output of a convolutional layer. It also reduces the computational cost by minimizing the number of parameters. After that, a batch normalization layer is used to convert numerical data to a common scale without changing the shape of the data. It solves the overfitting problem and makes neural network faster as well as more stable.

$$\mu = \frac{1}{n}\sum_{i} N^{(i)}$$

$$\sigma = \frac{1}{n}\sum_{i} \left(N^{(i)} - \mu\right)$$

$$N_{norm}^{(i)} = \frac{N^{(i)} - \mu}{\sqrt{\sigma^2 - \varepsilon}}$$

$$N = \gamma^{*} N_{norm}^{(i)} + \beta$$

each convolutional layer. The feature extraction part is completed in the upper section and now this section will describe classifying malaria parasites by learning parameters. Previously, the data is stored as a three-dimensional feature vector. It must be converted into a one-dimensional array to classify parasitized and uninfected cells. So, flatten layer is used to make the three-dimensional array into a one-dimensional linear array to feed data in a fully connected layer. It consists of 4096 neurons and this is the first fully connected layer or dense layer of the model. The output from flatten layer is passed to the next dense layer. The hidden dense layer has 1024 neurons with a ReLu activation function to transform linear input to nonlinear. L2 regularization is applied in that layer

to minimize the value of weighted metrics. Therefore, it solves the overfitting problem by reducing the weighted values. These values update the cost function which is shown in equation (6)

CNN-ML classifier

In this section, the custom CNN architecture is used as a feature extractor and several machine learning algorithms are employed as a classifier to compare their performance with the softmax classifier [32]. Some of the machine learning algorithms such as SVM, KNN, Decision Tree, and Random Forest are applied in this research. SVM is a supervised learning algorithm, that gives effective results when the number of dimensions exceeds the number of samples. Radial Basis Function (RBF) is applied in SVM classifier as a kernel whose value depends on the distance from the origin. Equation (8) shows RBF kernel mathematical expression.

$$K(X_1, X_2) = exp\left(- \gamma ||X_1 - X_2||^2 \right)$$

Where, X1 and X2 are the variables and jjX1 X2jj is Euclidean distance between X1 & X2. A Random Forest classifier that combines multiple decision trees because of that 100 trees are applied to solve a complex problem. KNN and Decision Tree are also used to evaluate the model's performance.

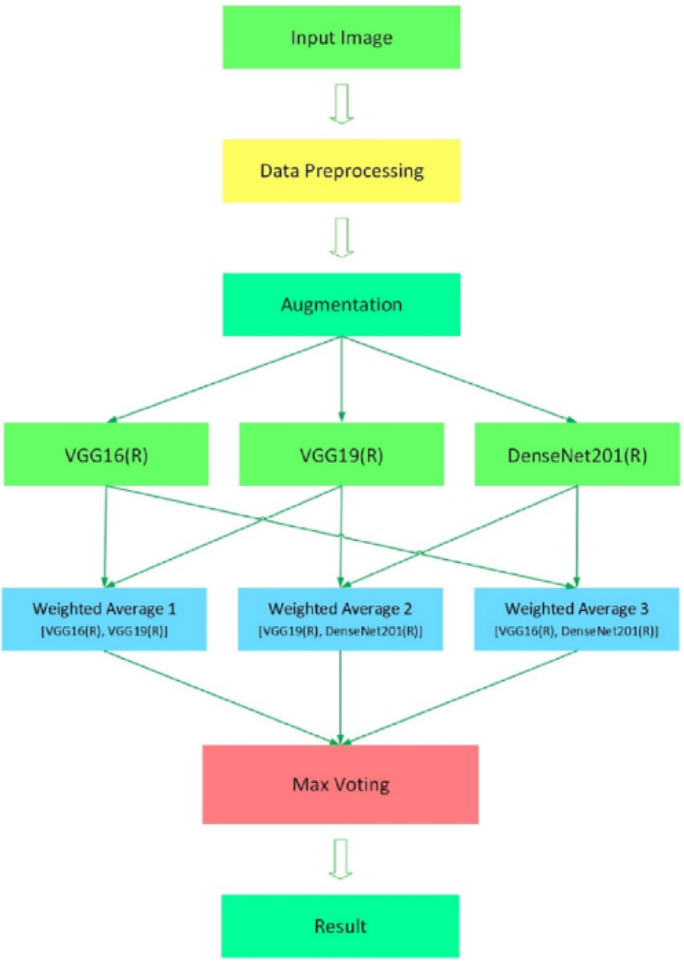


Fig. 1. Proposed methodology.

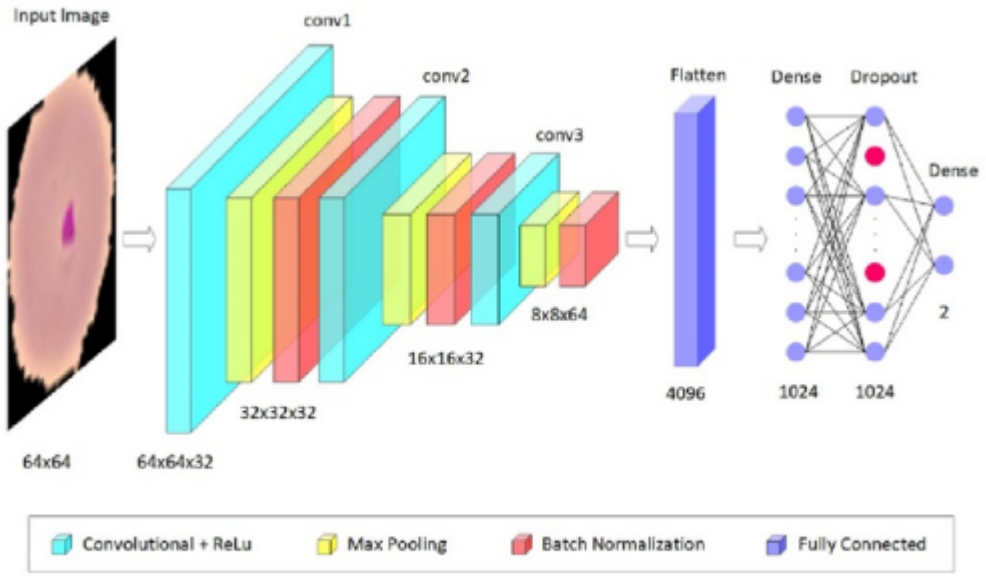


Fig. 7. Custom CNN architecture.

Table 4
Layer's name, shape, and parameter values in Custom CNN.

Layer	Output Shape	Parameter
Conv2D	64 x 64 x 32	2432
ReLu	64 x 64 x 32	0
Max pooling	32 x 32 x 32	0
Batch Normalization	32 x 32 x 32	128
Conv2D	32 x 32 x 32	9248
ReLu	32 x 32 x 32	0
Max pooling	16 x 16 x 32	0
Batch Normalization	16 x 16 x 32	128
Conv2D	16 x 16 x 64	18496
ReLu	16 x 16 x 64	0
Max pooling	8 x 8 x 64	0
Batch Normalization	8 x 8 x 64	256
Flatten	4096	0
Dense	1024	4195328
Dropout	1024	0
Dense	2	2050
Total Parameter		4228066

Table 5
Hyperparameter and their values in Custom CNN.

Parameter Name	Type/Value
Input Size	64 x 64
Batch Size	32
Epochs	100
Regularization	0.001
Dropout	0.2
Activation Function	Softmax
Optimizer	Adam
Learning Rate	1e-04
Loss Function	Sparse Categorical Cross entropy

4. PROPOSED WORK

1.1 Background-

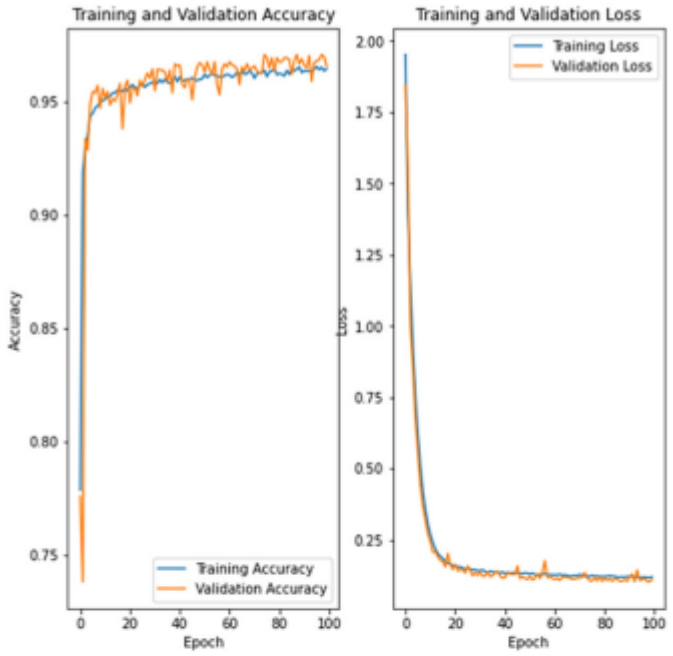
An ensemble learning-based system is developed in this research to reduce the burden on microscopists. Ensemble learning is a technique where multiple models are trained on the same dataset and combined with their results to achieve better accuracy and reduce the variance of the model. The input data is initially resized into 64 pixels of height and 64 pixels of width. Then, the dataset is divided into 70% training, 10% validation, and 20% testing sets. After that data augmentation technique is applied with a variety of domains to increase the number of training data which enhances the model's performance. VGG16(R), VGG19(R), and DenseNet201(R) models are fine-tuned using several hyperparameters such as optimizer, learning rate, activation function, etc.* Then, these models are trained to learn patterns in the data using a training set. From the individual model, two of the models are taken and applied adaptive weighted average ensemble to achieve better performance by reducing the variance of the model. This method finds the best.outcome with the appropriate weights automatically. VGG16(R) and VGG19(R) achieve a higher result with 0.7 and 0.6 wt respectively. With 0.6 and 0.5 wt, VGG19(R) and DenseNet201(R) produce a better result. When compared to other weights, VGG16(R) and DenseNet201(R) perform better with 0.4 and 0.5 wt, respectively. Three ensembled models are obtained after performing an adaptive weighted average technique. After that, another ensemble technique known as max voting is applied using three ensembled models with hard voting to improve the model's performance and reduce the dispersion of the predictions. The whole technique of this study is depicted in Fig. 1

1.2 Proposed Model

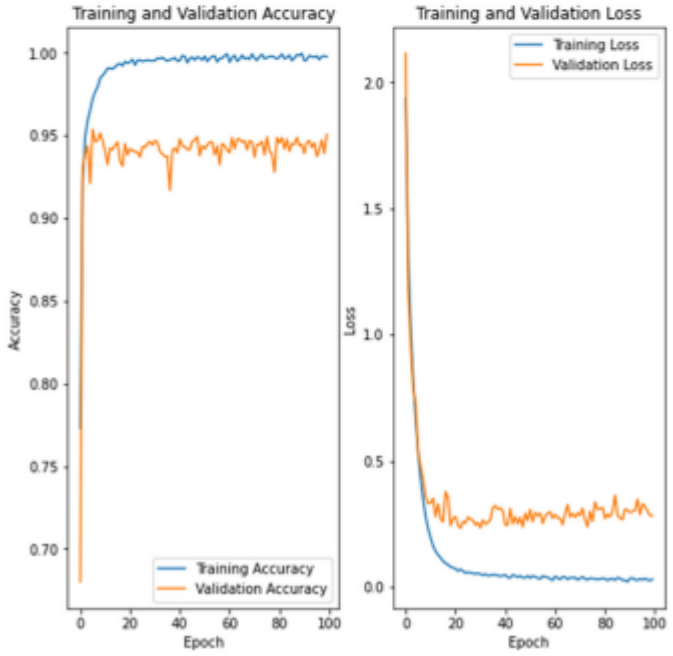
Data augmentation

Data augmentation is the process of creating new artificial data from the existing data. It avoids unwanted features from being learned and improves overall performance. Deep learning requires a larger dataset for training so that the model can learn more patterns from images and make accurate predictions. The data augmentation technique increases the quantity of the training dataset with different transformations and enhances the performance of the model. Transforms of data include some

operations such as rotating, zooming, flipping, rescaling, shearing, etc. In this research, the data augmentation technique is applied to overcome the overfitting problem where a model learns effectively in the training set but not in the test set. Fig. 4 shows the transformation of data augmentation techniques that are applied to the training dataset and Table 2 shows their respected values. Random rotation is applied with the value of 0.2 which describes a result of random rotation with upper and lower bound in the range of [20% * 2pi, 20% * 2pi] [27]. The resulting image is rotating randomly both clockwise and counterclockwise. The value of 0.2 is used for the zoom transformation which denotes random zooming vertically with upper and lower bound in the range of 20% [28]. The resulting image is represented in the form of zooming in and zooming out transformations. Horizontal and vertical flips are used to shift images from left to right and top to bottom. Images are rescaled to map features in the range of 0 to 1 by passing scale ¼ 1/255 to get faster convergence.



(a) With augmentation



(b) Without augmentation

Transfer learning models

Transfer learning models such as VGG16(R), VGG19(R), and DenseNet201(R) are employed to apply the adaptive weighted average ensemble in the proposed method. Here models are retrained so that the

weights are updated and assist the network in learning and identifying features associated with new images and categories. In this approach, the final dense layer is eliminated as transfer learning models are created for 1000 categories. A new output dense layer is added with 2 neurons to classify parasitized and uninfected cells. To achieve the best outcomes, hyperparameters are used to fine-tune the transfer learning models. Adam optimizer is one of the hyperparameters which is used with a learning rate of 1e-04 to reduce the error by updating weights and biases. As a loss function, sparse categorical cross-entropy is applied to calculate the error between the ground and predicted values. The batch size and the epochs number are 32 and 100 respectively. The architecture is given with all of the transfer learning models in Fig. 5 and their hyperparameters value is listed in Table 3. 2.6. Ensemble of ensemble method In this step, an ensemble of ensemble method is applied to get the best performance. After taking three individual transfer learning models, two models are combined every time to apply the adaptive weighted average method. As a result, three different adaptive weighted average models are discovered. Weighted average model 1 is formed by combining VGG16(R) and VGG19(R), while weighted average model 2 is developed by combining VGG19(R) and DenseNet201(R). Weighted average model 3 is created by using VGG16(R) and DenseNet201(R). In a normal weighted average method, weights need to be initialized to get the best result. However adaptive weighted average method automatically chooses appropriate weights based on the data. In weighted average model 1, VGG16(R) and VGG19(R) produce a better outcome with 0.7 and 0.6 wt respectively. With 0.6 and 0.5 wt, VGG19(R) and DenseNet201(R) achieve a higher result in weighted average model 2. In weighted average model 3, VGG16(R) and DenseNet201(R) perform better with 0.4 and 0.5 wt, respectively. Finally, the max voting ensemble method is applied to the three weighted average models where the majority class is the final output.

Ensemble method

In this step, an ensemble method is applied to get the best performance. After taking three individual transfer learning models, two models are combined every time to apply the adaptive weighted average method. As a result, three different adaptive weighted average models are discovered.

Weighted average model 1 is formed by combining VGG16(R) and VGG19(R), while weighted average model 2 is developed by combining VGG19(R) and DenseNet201(R). Weighted average model 3 is created by using VGG16(R) and DenseNet201(R). In a normal weighted average method, weights need to be initialized to get the best result. However, the adaptive weighted average method automatically chooses appropriate weights based on the data. In weighted average model 1, VGG16(R) and VGG19(R) produce a better outcome with 0.7 and 0.6 wt respectively. With 0.6 and 0.5 wt, VGG19(R) and DenseNet201(R) achieve a higher result in weighted average model 2. In weighted average model 3, VGG16(R) and DenseNet201(R) perform better with 0.4 and 0.5 wt, respectively. Finally, the max voting ensemble method is applied to the three weighted average models where the majority class is the final output.

5. RESULTS AND DISCUSSION

1.1 Experimental setup-

All the experiments are performed in the Google Colab platform with 13 GB Random Access Memory (RAM), 69 GB Colab Disk, Keras® 2.4.3 API with Tensorflow® 2.5.0 backend, Python® 3.7.10, Nvidia® Tesla T4 Graphical Processing Unit (GPU), CUDA V11.2 dependencies for GPU acceleration and matplotlib® 3.2.2 library.

1.2 Dataset Description -

The dataset is collected from the National Institutes of Health (NIH) for this research [23]; LHNCBC). There are 27,558 cell images in total. It is classified into two sections: 13,779 parasitized and 13,779 uninfected cells. The image resolutions are varying between 110 and 150 pixels. In the preprocessing section, the images are scaled into a fixed range of 64 pixels. Researchers at the Lister Hill National Center for Biomedical Communications (LHNCBC), which is a part of the National Library of Medicine (NLM), have built an Android mobile application attached to a conventional light microscope. Giemsa-stained thin blood smear slides

were provided by 150 Plasmodium falciparum-infected patients and 50 healthy people. Then, images were captured from those samples by using a smartphone at Chittagong Medical College Hospital, Bangladesh. A professional slide reader annotated the images manually at the Mahidol-Oxford Tropical Medicine Research Unit in Bangkok, Thailand (NIH Malaria Dataset, LHNCBC). Fig. 2 shows the parasitized and uninfected microscopic red blood cell images of malaria.

According to Ref. [9], there is some mislabeled data in the dataset. Some of the data are labelled as parasitized but they are uninfected. In addition, some of the data from uninfected cells are labelled as parasitized. The expert also verified that some of the data is incorrectly labelled. There are found 647 falsely labelled parasitized data and 750 falsely uninfected data [9]. Around 5% of data is mislabeled in the total dataset which may have an impact on the performance of the model.

1.3 Evaluation parameter

In this research, different deep learning methods are tested to find a more accurate and efficient model for identifying malaria from red cell images. The results are evaluated based on different optimizers, learning rates, transfer learning models, ML classifiers, and ensemble learning methods. All of the experiments are run with a learning rate of 1e-04, a batch size of 32. Precision, recall, F1 score, and accuracy are utilized as performance metrics to evaluate the model's performance. There are four important parameters. They are true positive (TP), false positive (FP), true negative (TN), and false negative (FN). True positive denotes infected cells classified correctly as infected. False positive defines infected cells classified as uninfected. True negative describes uninfected cells correctly classified and false negative is the opposite of this. In equations (9)–(12), various evaluation metrics are shown.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

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

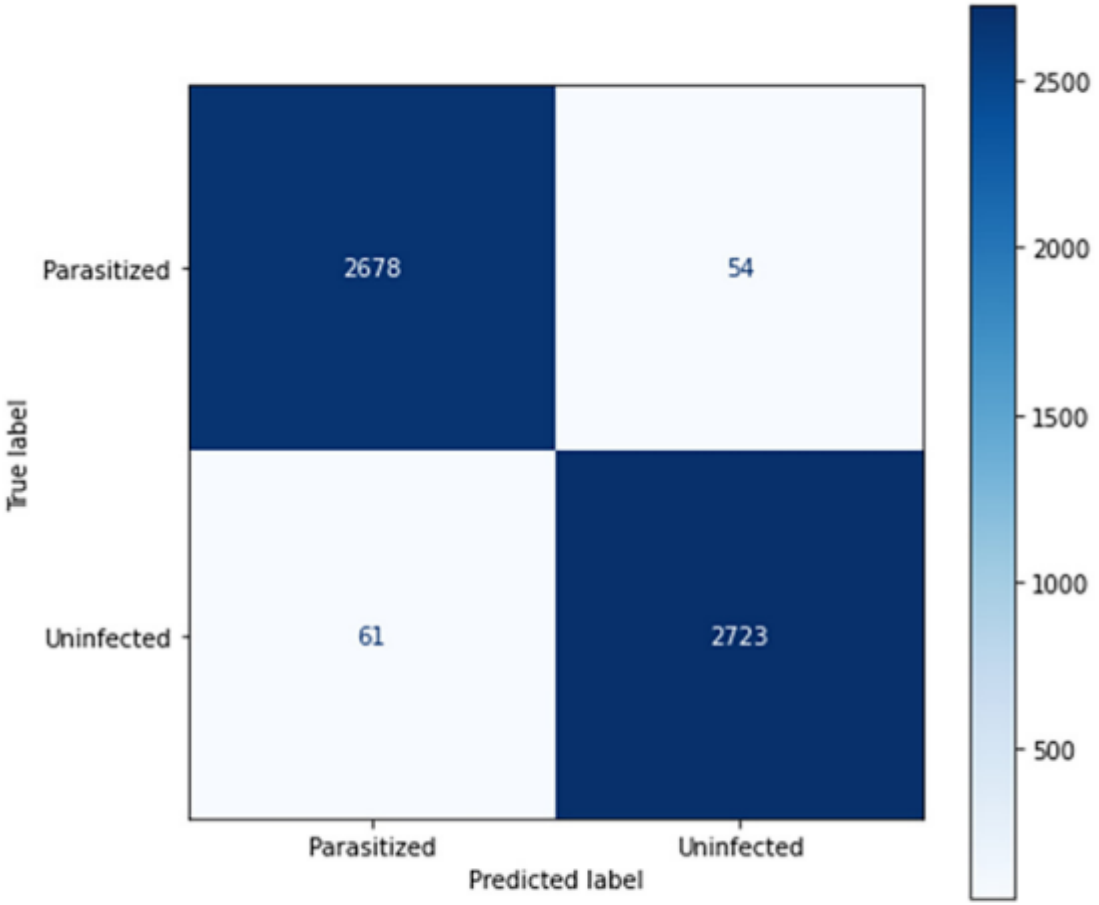


Fig. 14. Confusion matrix of the proposed ensemble method.

1.4 Experimental results

Result performance of different ensemble learning methods.

Ensemble method	Model	Weights	Precision	Recall	F1 score	Accuracy
Adaptive Weighted Average	VGG16(R), VGG19(R)	0.7, 0.6	0.9759	0.9759	0.9759	0.9759
	VGG19(R), DenseNet201(R)	0.6, 0.5	0.9784	0.9784	0.9784	0.9784
	VGG16(R), DenseNet201(R)	0.4, 0.5	0.9781	0.9781	0.9781	0.9781
	VGG16(R), VGG19(R) & DenseNet201(R)	0.1, 0.9, 0.8	0.9786	0.9786	0.9786	0.9786
Max Voting	VGG16(R), VGG19(R) & DenseNet201(R)	–	0.9773	0.9773	0.9773	0.9773
Adaptive Weighted Average & Max Voting	**VGG16(R), VGG19(R) & DenseNet201(R)**		**0.9791**	**0.9792**	**0.9792**	**0.9792**

Custom CNN performance

The custom CNN model achieves 96.43% of training accuracyand 96.56% of validation accuracy. There is no overfitting problem as the data augmentation technique is applied. Without data augmentation, the training accuracy is found at 99.77% and the validation accuracy is 95.06%. As can be seen, the training and validation accuracy results are vastly different. This is called the overfitting problem where the model performs well in training data but not in test data. Fig. 9 shows the performance of the custom CNN model with and without augmentation. There are a series number of optimizers available such as Adadelta, Adagrad, Ftrl, SGD, Adam, RMSprop, etc. Fig. 10 shows the performance of those optimizers in the custom CNN model. Adam outperforms other optimizers because it is computationally efficient and uses less memory. It is invariant to diagonal rescaling of gradients and is ideally suited to large data/parameter problems [14]. Table 6 represents the performance results of the custom CNN model at different optimizers. Fig. 11 demonstrates the performance of the custom CNN model at different learning rates. An optimal learning rate of 1e-04 is required to reach in global minima point. At this point, the loss is minimum, and identify malaria parasites more accurately. Because the two other learning rate values are too high, they are unable to achieve global minima point. Table 7 shows the performance results of the custom CNN model at different learning rates.

Result performance of custom CNN model at different Learning rates.

Learning Rate	Precision	Recall	F1 score	Accuracy
1e-02	0.2476	0.5000	0.3312	0.4953
1e-03	0.9550	0.9547	0.9545	0.9545
1e-04	0.9723	0.9718	0.9719	0.9719

CNN-ML classifier performance

The outcomes of various machine learning algorithms are now being examined. Custom CNN is used as a feature extractor and machine learning algorithms are applied as a classifier. CNN-SVM is performing better than the three other CNN-ML classifiers. The test accuracy is found 81.67% for the SVM classifier. As compared to the softmax classifier, these ML classifiers are not performing very well

to classify malaria parasites. Table 8 shows the performance results for different CNN-ML classifiers.

Result performance of different CNN-ML classifiers.

Model	Precision	Recall	F1 score	Accuracy
CNN-SVM	0.8171	0.8165	0.8166	0.8167
CNN–KNN	0.6458	0.6155	0.5958	0.6177
CNN-Decision Tree	0.7386	0.7386	0.7386	0.7386
CNN-Random Forest	0.8132	0.8123	0.8119	0.8120

Transfer learning performance

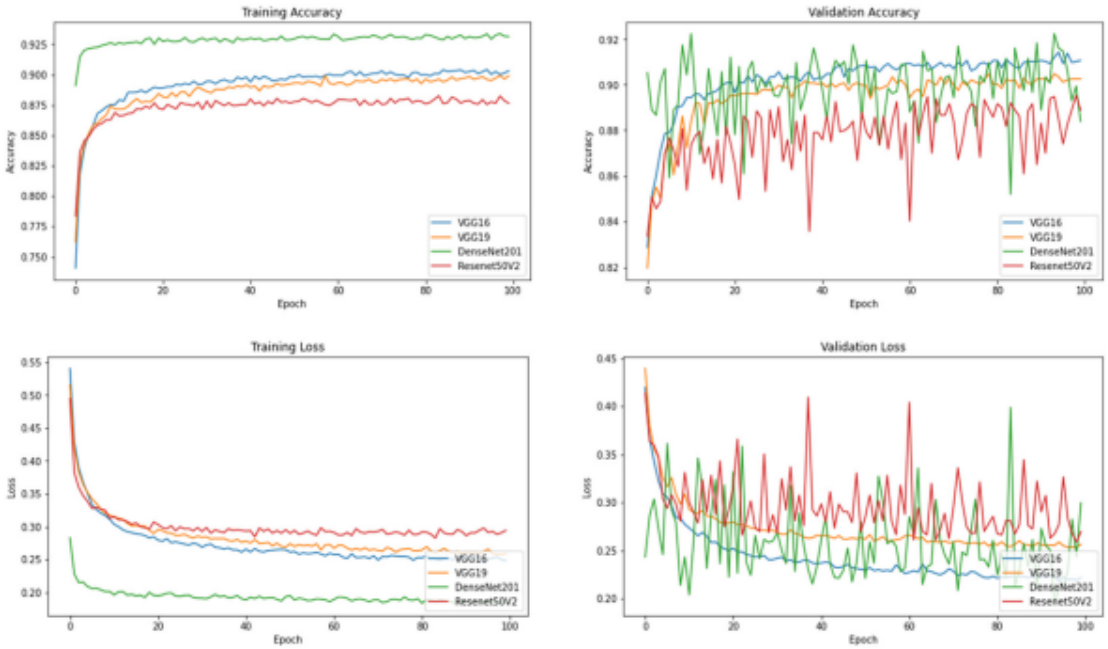


Fig. 12. Transfer learning models performance.

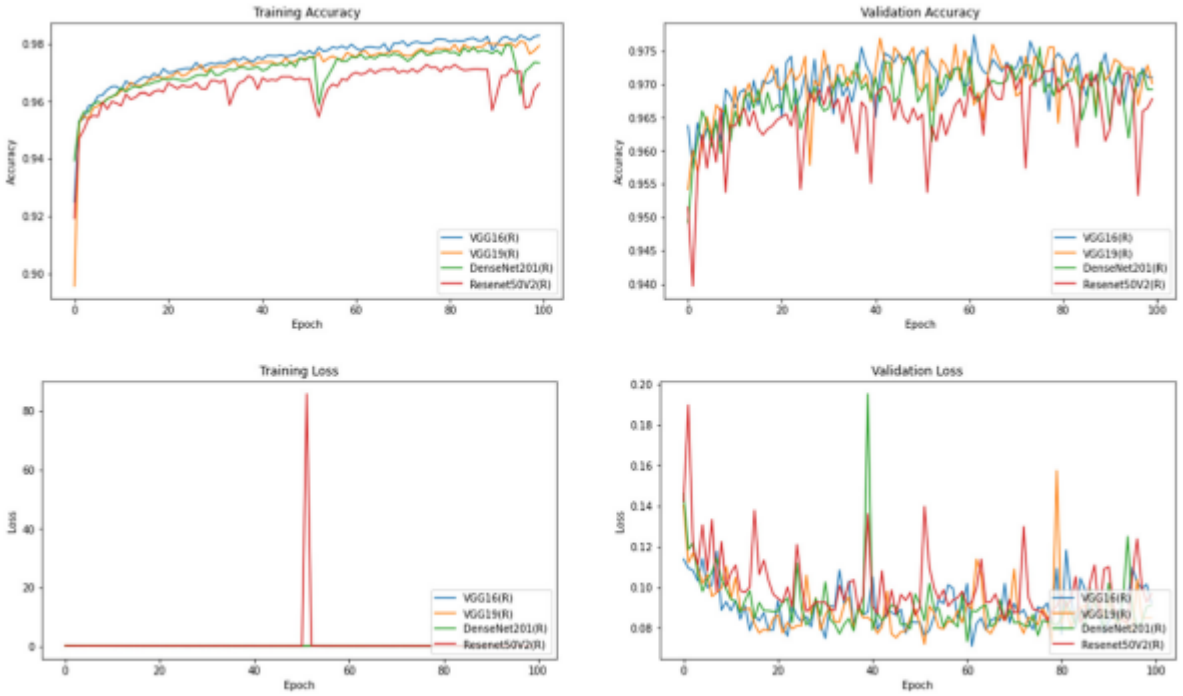


Fig. 13. Retrained transfer learning models performance.

From the figures, models that use pre-trained weights do not perform very well. However, models that have been retrained, perform incredibly well since all layers are trained and the networks learn new attributes associated with new images and categories. VGG19 (R) outperforms the others by achieving a testing accuracy of 97.52%. Table 9 shows the performance results for different transfer learning models.

1.5 Comparative analysis

Result comparison with different models

The outcomes are now compared to other deep learning models, including the proposed ensemble learning method. The ensemble model is the best model among others and achieves a high testing accuracy of 97.92%. Table 11 shows the performance comparison between the proposed ensemble model and other models using several evaluation metrics.

Result comparison between proposed ensemble model and other models.

Model	Precision	Recall	F1 score	Accuracy
Custom CNN	0.9723	0.9718	0.9719	0.9719
VGG16 (R)	0.9752	0.9752	0.9752	0.9752
CNN-SVM	0.8171	0.8165	0.8166	0.8167
Ensemble	**0.9791**	**0.9792**	**0.9792**	**0.9792**

Result comparison with existing methods

In this section, the result is compared with existing works basedon different performance metrics. Many researchers have proposed the transfer learning model as well as the CNN model in their papers. However, the results are not as good as the proposed method. The proposed ensemble model obtains a significant testing accuracy of 97.92% and detects malaria more accurately and efficiently. The performance comparison between the proposed ensemble method and existing methods is shown in Table .

Result Comparison between proposed ensemble model and existing models.

Model	Dataset	Precision	Recall	F1 score	Accuracy
Otsu segmentation, K-means clustering [33]	NIH	0.9607	0.93	0.945	0.946
CNN [22]	NIH	–	–	–	0.96
VGG16 [29]	NIH	–	–	–	0.9615
CNN [21]	NIH	0.9682	0.9633	0.9682	0.9682
VGG16 [26]	NIH	0.9719	0.9720	0.9709	0.9777
VGG16 [35]	NIH	–	0.956	0.956	0.96
CNN [19]	NIH	0.9773	0.9699	0.9736	0.9737
Proposed Ensemble model	**NIH**	**0.9791**	**0.9792**	**0.9792**	**0.9792**