History of Medicine, 2021, 7(2): 319-330

DOI:10.48047.v7.2.2021.17

MACHINE LEARNING BASED DETECTION OF MALARIA INFECTION THROUGH BLOOD SAMPLE ANALYSIS FOR MALARIA DIAGNOSIS

Parvatham Sathish[1], Senthil Kumar Murugesan[1], Akula Rajini[1]

[1]Department of Electronics and Communication Engineering

[1]Sree Dattha Group of Institutions, Sheriguda, Hyderabad, Telangana

ABSTRACT

Plasmodium parasites, which are responsible for the life-threatening disease known as malaria, are transmitted through infected mosquitoes on a global scale. Malaria continues to be a significant public health concern in many locations across the world. The diagnosis of malaria infection in its early stages and with high accuracy is essential for the timely treatment and management of the disease. The automated malaria detection method may be included into portable diagnostic instruments, which enables medical personnel to conduct malaria tests that are both quick and accurate even in settings that are resource-constrained or located in distant areas. It is possible for the system to provide academics and health organizations with assistance in tracking the prevalence of malaria and monitoring its spread, thereby helping to epidemiological studies and implementing efficient resource allocation. Traditionally, the detection of malaria is accomplished through the manual inspection of blood smears under a microscope by experienced technicians. The microscopist's skill is essential to the success of this process, which is not only time-consuming but also labor-intensive and dependent on reliability. The regression-based evaluation of blood smears presents the possibility of mistakes, which can result in results that are either falsely negative or falsely positive. Over the past few years, methods that are based on deep learning have demonstrated promising results in the process of automating the detection of malaria parasites through the examination of blood samples. In this work, an innovative machine learning-based system for the automated identification of malaria infection is presented. This method makes use of image processing techniques in order to attain high levels of accuracy and efficiency.

Keywords: Automated Malaria Detection, Plasmodium Parasites, Deep Learning, Blood Smear Analysis, Image Processing, Portable Diagnostic Devices.

1. INTRODUCTION

Malaria infection is a widespread and potentially deadly disease caused by the Plasmodium parasite, transmitted to humans through the bite of infected female Anopheles mosquitoes. Diagnosis and monitoring of malaria often rely on the analysis of blood samples, which provides crucial insights into the presence and severity of the infection. When a blood sample is obtained from a patient suspected of having malaria, it undergoes a series of laboratory tests to confirm the diagnosis and assess the level of parasitic activity. The primary diagnostic method is the examination of a thin blood smear or a thick blood smear under a microscope. Thin blood smears are used to identify the Plasmodium species responsible for the infection, while thick blood smears are employed to quantify the number of parasites present in the blood. This information is vital for determining the severity of the disease and guiding treatment decisions. Additionally, molecular techniques like polymerase chain reaction (PCR) can be employed to confirm the presence of the parasite and, in some cases, differentiate between species with high accuracy. Blood sample analysis also allows for the evaluation of other important parameters such as hematocrit levels, which help in assessing anemia, a common complication of malaria. Moreover, serological tests can be performed to detect specific antibodies against Plasmodium antigens, providing information about previous exposure to the parasite and aiding in epidemiological studies.

The timely and accurate analysis of blood samples is crucial in the management and control of malaria. Rapid and precise diagnosis enables healthcare providers to initiate appropriate treatment promptly, reducing the risk of severe complications and death. Furthermore, monitoring the parasite load in the blood over time allows healthcare professionals to gauge the effectiveness of treatment and make necessary adjustments. Therefore, blood sample analysis remains a cornerstone in the battle against malaria, contributing to both individual patient care and public health efforts to control and ultimately eliminate this devastating disease.

2. LITERATURE SURVEY

According to the World Health Organization (WHO), malaria case rates (i.e., cases per 1000 population) fell from 82 in 2000 to 57 in 2019 but rose to 59 in 2020. The WHO reported that this unusual 2020 increase in malaria case rates was related to service supply disruptions during the COVID-19 pandemic [1]. In fact, the number of malaria cases increased from 227 million in 2019 to 241 million in 2020, and the number of malaria deaths in 2020 was estimated at 627,000, a 12% increase from 2019 [2]. Moreover, in the case of malaria, the more severe problem is that the existing malaria diagnosis method relies on direct human observation, which takes much time for diagnosis, making it difficult to test many patients simultaneously. Additionally, there is a limitation in that diagnostic accuracy is greatly affected by variability between observers. In other words, the effectiveness of the conventional microscopic diagnosis is highly dependent on the expertise of parasitologists. Besides, it is common for parasitologists to work in resource-constrained environments without stringent systems to maintain their know-how or diagnostic quality [3]. This can often lead to erroneous diagnoses and inappropriate treatment, which can have fatal consequences [3-5]. There are several promising prior studies on the capabilities of ML-based techniques in detecting infectious diseases. For instance, using a machine learning framework, Colubri et al. [6] introduced an application that can predict the outcome of Ebola patients from early clinical symptoms. Smith and Kirby [7] described ML applications for analyzing different types of microbial image data, particularly progress in smear and plate interpretation.

Another notable study on ML-based infectious disease diagnosis is that of Das et al. [8], who developed a computer-aided malaria parasite characterization and classification based on light microscopy images of peripheral blood smears collected from 600 patients using an ML approach. Their proposed ML scheme applying the Bayesian approach provides 84.0% accuracy and 98.1% sensitivity by selecting the 19 most significant features, and the support vector machine (SVM) achieved 83.5% screening accuracy and 96.6% sensitivity with the 9 most significant features [8]. Similarly, there are other studies that have applied various machine learning methods to detect malaria parasites. Bibin et al. [9] proposed a deep belief network (DBN)-based trained model to classify 4100 peripheral blood smear images into parasitic or nonparasitic classes. The proposed method showed an F-score of 89.66%, a sensitivity of 97.60%, and a specificity of 95.92% [9]. Gopakumar et al. [10] used a customized CNN model operating on a focus stack of images for automated quantitative detection of Plasmodium falciparum malaria in blood smears. The detection accuracy of the CNN model was 97.06% sensitivity and 98.50% specificity [10]. Yang et al. [3] developed a method using a deep learning algorithm to detect malaria parasites in thick blood smear images, run on a smartphone. They trained and tested a deep learning method using 1819 thick smear images from 150 patients [3]. The study results showed the effectiveness of the CNN model in distinguishing positive (parasitic) image patches from negative image patches, with performance metrics of accuracy (93.46% ± 0.32%), precision (94.25% ± 1.13%), and negative predictive value (92.74% ± 1.09%) [3].

Especially in the case of the COVID-19 pandemic, Dandekar et al. [11] applied the neural network module of ML to develop a globally applicable COVID-19 diagnosis model to analyze and compare the role of quarantine control policies globally across the continents of Europe, North America, South America, and Asia. Dandekar et al. [11] also hosted quarantine diagnosis results from 70 countries around the world on a public platform: https://covid19ml.org/ (accessed on 15 March 2023). One example of a notable literature review source for ML-based infectious disease diagnosis is the work of Baldominos et al. [12]. The study performed a computer-based systematic literature review in order to investigate where and how computational intelligence (i.e., different types of machine learning techniques) is being utilized to predict patient infection [12]. Deep learning, a specific subset of machine learning, is a computational processing system composed of artificial neural networks, heavily inspired by how biological nervous systems process information and make decisions [13].

3. PROPOSED SYSTEM

The methodology leverages image processing and machine learning techniques to automate the detection of malaria parasites in blood sample images. It is a promising approach to improve the efficiency and accuracy of malaria diagnosis, particularly in resource-limited settings where access to skilled technicians may be limited. However, it's important to note that developing and fine-tuning the DLCNN model typically requires a substantial amount of labeled data and expertise in machine learning and image analysis. Additionally, the performance of the model should be rigorously evaluated to ensure its accuracy and reliability in real-world healthcare applications.

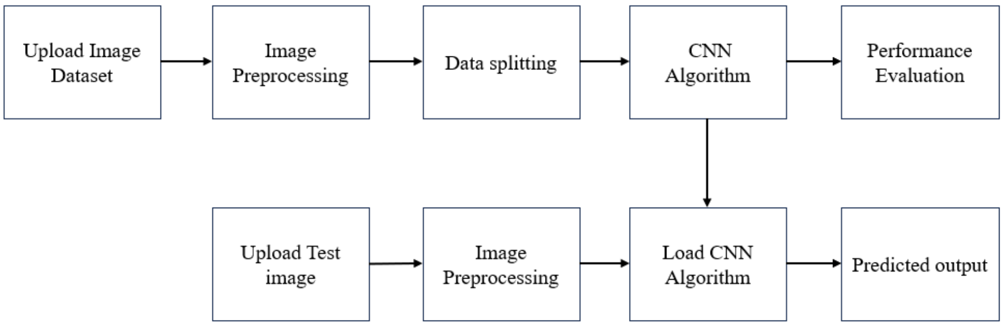

Figure 1 Proposed Block diagram of Malaria detection using Deep learning

Figure 1 shows the proposed system model. The detailed operation illustrated as follows: Step 1: Image Processing: This is the initial step where you process the blood sample images. Image processing techniques may include preprocessing steps such as noise reduction, contrast enhancement, and image segmentation to isolate the relevant features (in this case, malaria parasites) from the background and other elements in the image. This step is essential for preparing the images for further analysis.

Step 2: DLCNN Building: After image processing, the next step involves training a machine learning model, specifically a DLCNN. In this step, you would typically use a labeled dataset of blood sample images, where each image is associated with a known diagnosis (e.g., whether it contains malaria parasites or not). The DLCNN is trained to learn patterns and features in the images that distinguish between infected and uninfected samples. This classifier can handle complex relationships in the data and is capable of making predictions based on these learned patterns.

Step 3: DLCNN Prediction: Once the DLCNN model is trained, it can be used to predict whether new, unseen blood sample images contain malaria parasites or not. When a new blood sample image is input into the trained DLCNN, the model evaluates the image based on the patterns it has learned during training and produces a prediction. This prediction can help automate the process of diagnosing malaria from blood sample images, reducing the need for manual examination and potentially increasing the speed and accuracy of diagnosis.

DLCNN Algorithm

Deep neural network is gradually applied to the identification of malaria conditions. Deep neural network is designed by imitating the structure of biological neural network, an artificial neural network to imitate the brain, using learnable parameters to replace the links between neurons. Convolutional neural network is one of the most widely used deep neural network structures, which is a branch of feed forward neural network. The success of DLCNN network model also confirms the importance of convolutional neural network model. Since then, convolutional neural networks have developed vigorously and have been widely used in financial supervision, text and speech recognition, smart home, medical diagnosis, and other fields.

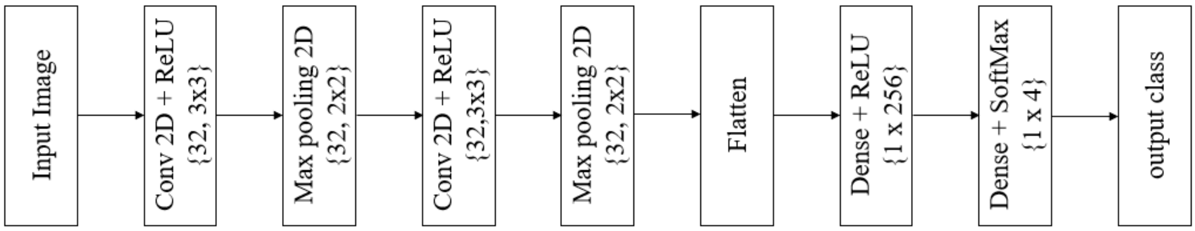

Fig. 2: Proposed DLCNN

Table.1: Layers description.

Layer Names	No. of filters	Kernel size	Feature size
Conv 2D +ReLU	32	3 x 3	62x62x32
Max pooling 2D	-	3 x 3	31x31x32
Conv 2D+ReLU	32	3 x 3	29x29x32
Max pooling 2D	-	3 x 3	14x14x32
Flatten	-	1x6272	1x6272
Dense +ReLU		1 x 256	1 x 256
Dense + SoftMax		1 x 4	1 x 4

Convolutional neural networks are generally composed of three parts. Convolution layer for feature extraction. The convergence layer, also known as the pooling layer, is mainly used for feature selection. The number of parameters is reduced by reducing the number of features. The full connection layer carries out the summary and output of the characteristics. A convolution layer is consisting of a convolution process and a nonlinear activation function ReLU. A typical architecture of CNN model for malaria condition recognition is shown in Figure 2.

The leftmost image is the input layer, which the computer understands as the input of several matrices. Next is the convolution layer, the activation function of which uses ReLU. The pooling layer has no activation function. The combination of convolution and pooling layers can be constructed many times. The combination of convolution layer and convolution layer or convolution layer and pool layer can be very flexibly, which is not limited when constructing the model. But the most common CNN is a combination of several convolution layers and pooling layers. Finally, there is a full connection layer, which acts as a classifier and maps the learned feature representation to the sample label space.

Convolutional neural network mainly solves the following two problems.

1) Problem of too many parameters: It is assumed that the size of the input picture is $50 * 50 * 3$. If placed in a fully connected feedforward network, there are 7500 mutually independent links to the hidden layer. And each link also corresponds to its unique weight parameter. With the increase of the number of layers, the size of the parameters also increases significantly. On the one hand, it will easily lead to the occurrence of over-fitting phenomenon. On the other hand, the neural network is too complex, which will seriously affect the training efficiency. In convolutional neural networks, the parameter sharing mechanism makes the same parameters used in multiple functions of a model, and each element of the convolutional kernel will act on a specific position of each local input. The neural network only needs to learn a set of parameters and does not need to optimize learning for each parameter of each position.

2) Image stability: Image stability is the local invariant feature, which means that the natural image will not be affected by the scaling, translation, and rotation of the image size. Because in deep learning, data enhancement is generally needed to improve performance, and fully connected feedforward neural is difficult to ensure the local invariance of the image. This problem can be solved by convolution operation in convolutional neural network.

Convolution layer: According to the facts, training and testing of DLCNN involves in allowing every source image via a succession of convolution layers by a kernel or filter, rectified linear unit (ReLU), max pooling, fully connected layer and utilize SoftMax layer with classification layer to categorize the objects with probabilistic values ranging from $[0,1]$.

Convolution layer as depicted in Figure 4.3 is the primary layer to extract the features from a source image and maintains the relationship between pixels by learning the features of image by employing tiny blocks of source data. It's a mathematical function which considers two inputs like source image $I(x, y, d)$ where x and y denotes the spatial coordinates i.e., number of rows and columns. d is denoted as dimension of an image (here $d = 3$, since the source image is RGB) and a filter or kernel with similar size of input image and can be denoted as $F(k_x, k_y, d)$.

4. RESULTS AND DISCUSSION

Figure 3 shows a selection of images from the dataset that are classified as belonging to the "parasitized" class. These images exhibit characteristics associated with parasitized in blood smear samples.

Figure 4 displays sample images from the dataset categorized as "uninfected." These images are examples of blood smear samples with no signs of parasitized or abnormalities.

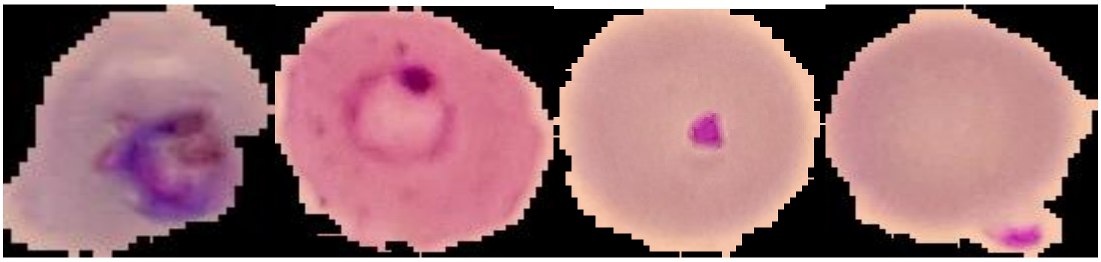

Figure 3: Sample images of dataset with parasitized class.

Figure 4: Sample images from dataset with uninfected class.

Figure 5: Sample UI used for deep learning malaria detection

Figure 5 provides a glimpse of a user interface designed for the purpose of deep learning-based malaria detection.

In Figure 6, the user interface exhibits data post image preprocessing, including various enhancements or adjustments made to the images before input into the models.

Figure 7 showcases the user interface displaying performance metrics specifically pertaining to the Gaussian naïve Bayes classifier, a machine learning algorithm often used for classification tasks.

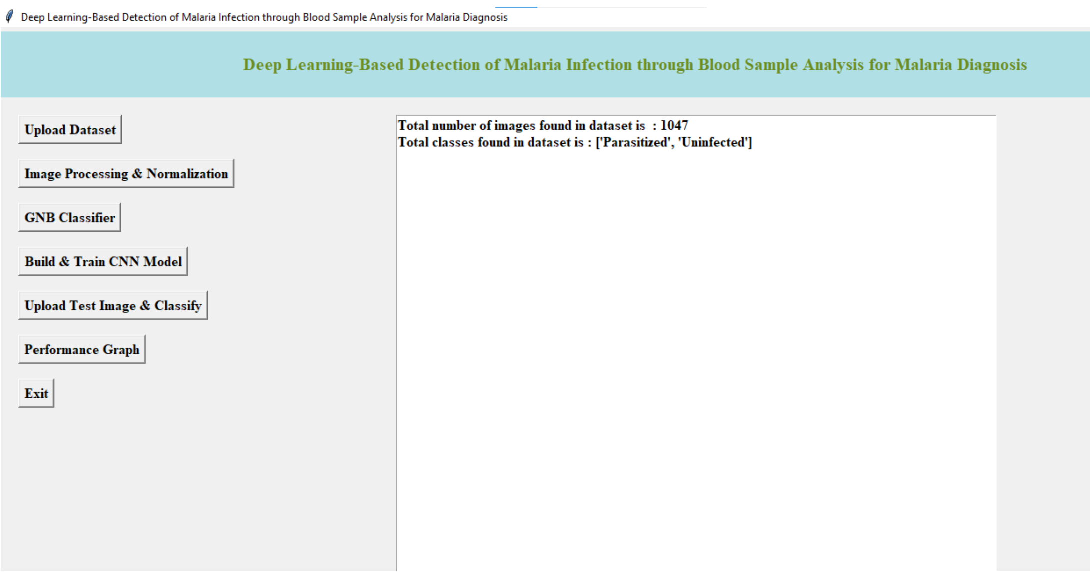

Figure 6: UI shows the data after image preprocessing

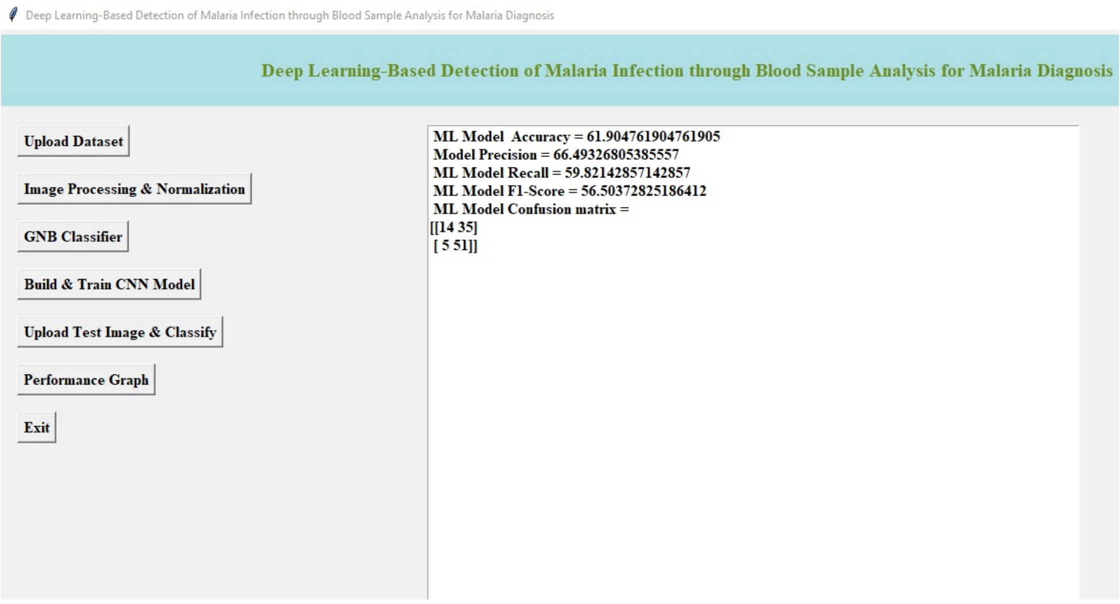

Figure 7: UI shows the performance of Gaussian naïve bayes classifier

In Figure 8, a confusion matrix is presented, detailing the performance of the Gaussian naïve Bayes classifier. Confusion matrices are commonly used to evaluate the performance of classification algorithms by visualizing the counts of true positive, false positive, true negative, and false negative predictions.

Figure 9 provides a user interface demonstrating the performance metrics of a Convolutional Neural Network (CNN) model, which is a type of deep learning architecture widely used for image recognition and classification tasks.

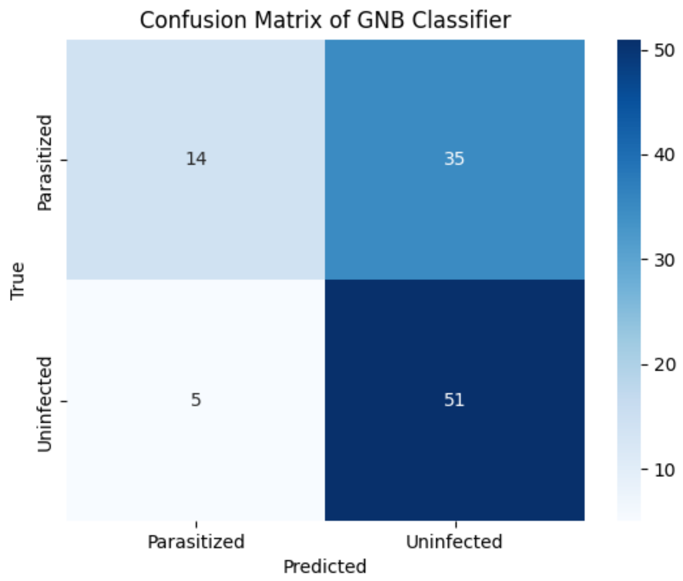

Figure 8: Confusion matrix of Gaussian naïve bayes classifier

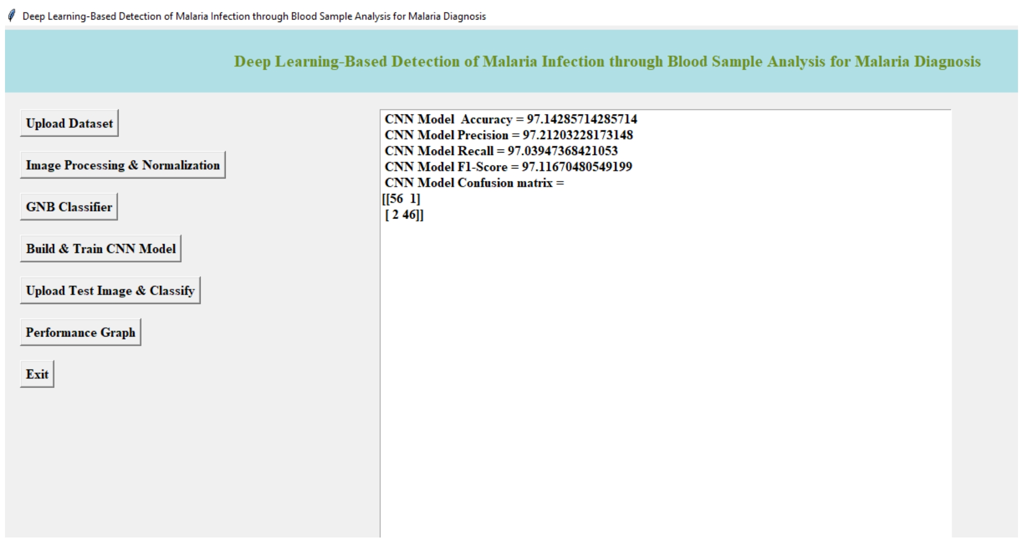

Figure 9: UI shows the performance of CNN model

Figure 10 shows the confusion matrix of the CNN algorithm, offering a detailed breakdown of its performance similar to the one seen in Figure 9 but specific to the CNN model.

In Figure 11, an accuracy and loss graph is presented, which is a common visualization used in deep learning to monitor the training process of neural networks. It likely shows the trend of accuracy and loss over the course of training epochs.

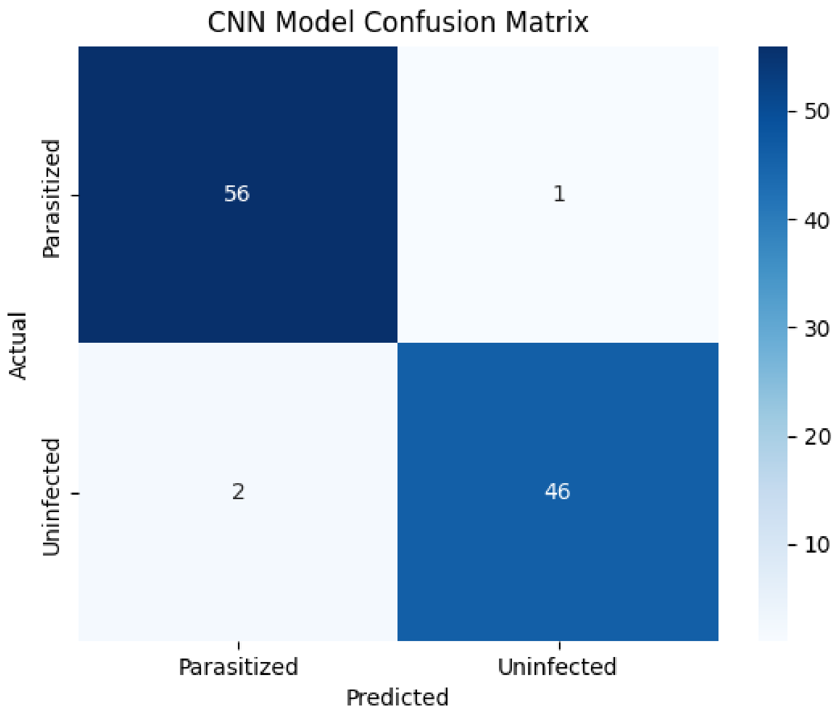

Figure 10: Confusion matrix of CNN Algorithm

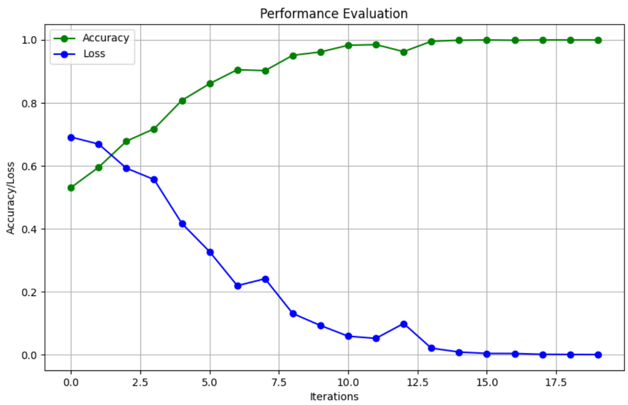

Figure 11: Accuracy and loss graph of CNN model

Figure 10 compares the performance of the Gaussian naïve Bayes classifier and the CNN model through a graphical representation, likely indicating metrics such as accuracy, precision, recall, or F1-score.

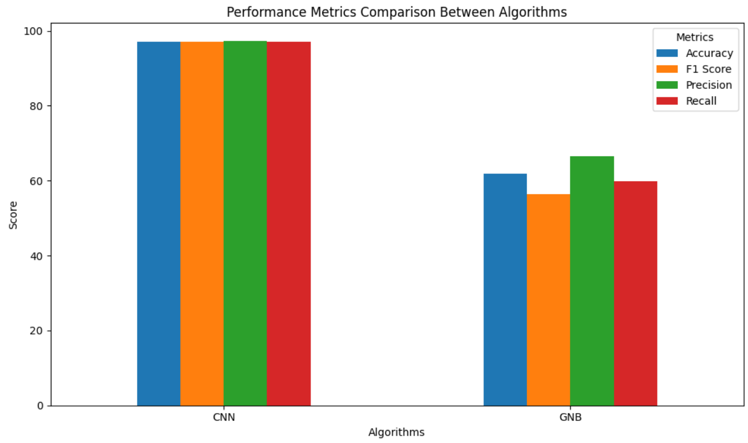

Figure 10: Performance Comparison graph of GNB and CNN Models

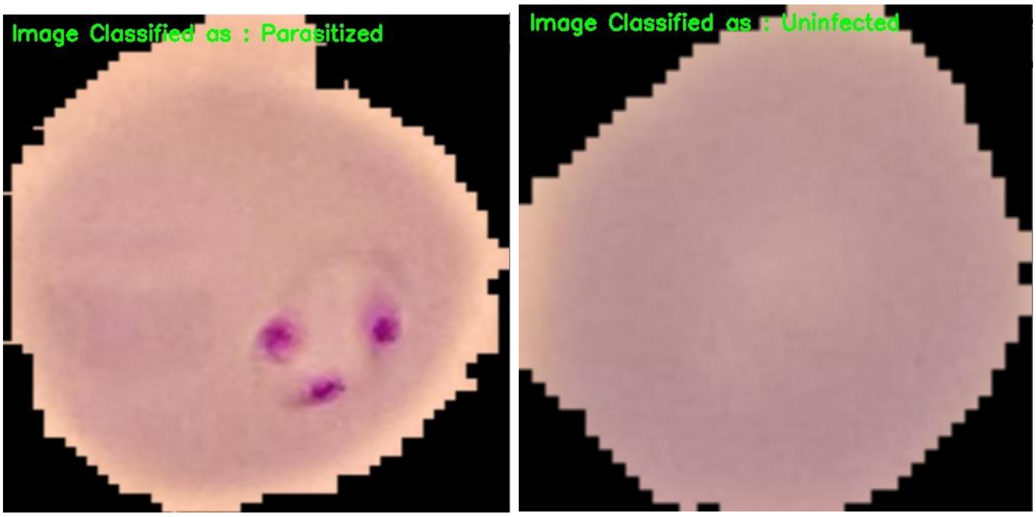

Figure 11: Predicted output using CNN model

Figure 11 presents the predicted output using the CNN model, potentially showcasing examples of correctly and incorrectly classified images.

Table 2 offers a detailed class-wise performance comparison of the proposed machine learning models, providing insights into how well each model performs on specific classes (e.g., uninfected and parasitized) in terms of precision, recall, and F1-score.

Table 2: Class-wise performance comparison of proposed ML models.

Model name	CNN		GNB classifier	
	Uninfected	Parasitized	Uninfected	Parasitized
Precision	0.95	0.94	0.64	0.61
Recall	0.96	0.94	0.59	0.56
F1-score	0.95	0.95	0.56	0.56

5. CONCLUSION

In conclusion, the methodology involving image processing followed by DLCNN building and prediction for malaria diagnosis from blood sample images represents a significant advancement in the field of healthcare and disease management. This approach addresses critical challenges related to the efficiency, accuracy, and accessibility of malaria diagnosis. By automating the analysis of blood sample images, it streamlines the diagnostic process, reducing the time required for diagnosis and treatment initiation. Additionally, it enhances diagnostic consistency, reduces the potential for human error, and offers scalability, making it suitable for both routine diagnostics and large-scale screening efforts. The integration of machine learning and image analysis technologies into healthcare systems holds promise for improving malaria control, early detection of outbreaks, and enhancing overall healthcare access. While there may be initial development costs, the long-term benefits in terms of improved healthcare delivery, reduced costs, and better disease surveillance make this methodology a valuable addition to the fight against malaria.

REFERENCES

[1]. WHO. World Malaria Report 2022. Available online: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022 (accessed on 1 March 2023).

[2]. WHO. World Malaria Report 2021: An In-Depth Update on Global and Regional Malaria Data and Trends. Available online: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021 (accessed on 1 September 2022).

[3]. Yang, F.; Poostchi, M.; Yu, H.; Zhou, Z.; Silamut, K.; Yu, J.; Maude, R.J.; Jaeger, S.; Antani, S. Deep learning for smartphone-based malaria parasite detection in thick blood smears. IEEE J. Biomed. Health Inform. 2019, 24, 1427–1438.

[4]. World Health Organization. Malaria Microscopy Quality Assurance Manual, 2nd ed.; World Health Organization: Geneva, Switzerland, 2016; Available online: https://www.who.int/docs/default-source/documents/publications/gmp/malaria-microscopy-quality-assurance-manual.pdf (accessed on 2 March 2023).

[5]. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput. 2017, 29, 2352–2449.

[6]. Colubri, A.; Silver, T.; Fradet, T.; Retzepi, K.; Fry, B.; Sabeti, P. Transforming clinical data into actionable prognosis models: Machine-learning framework and field-deployable app to predict outcome of Ebola patients. PLoS Negl. Trop. Dis. 2016, 10, e0004549. [Green Version]

[7]. Smith, K.P.; Kirby, J.E. Image analysis and artificial intelligence in infectious disease diagnostics. Clin. Microbiol. Infect. 2020, 26, 1318–1323.

[8]. Das, D.K.; Ghosh, M.; Pal, M.; Maiti, A.K.; Chakraborty, C. Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron 2013, 45, 97–106.

[9]. Bibin, D.; Nair, M.S.; Punitha, P. Malaria parasite detection from peripheral blood smear images using deep belief networks. IEEE Access 2017, 5, 9099–9108.

[10]. Gopakumar, G.P.; Swetha, M.; Sai Siva, G.; Sai Subrahmanyam, G.R.K. Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. J. Biophotonics 2018, 11, e201700003.

[11]. Dandekar, R.; Rackauckas, C.; Barbastathis, G. A machine learning-aided global diagnostic and comparative tool to assess effect of quarantine control in COVID-19 spread. Patterns 2020, 1, 100145.

[12]. Baldominos, A.; Puello, A.; Oğul, H.; Aşuroğlu, T.; Colomo-Palacios, R. Predicting infections using computational intelligence–a systematic review. IEEE Access 2020, 8, 31083–31102.

[13]. O'Shea, K.; Nash, R. An introduction to convolutional neural networks. arXiv 2015, arXiv:1511.08458.