Journal of
Electronics Technology Exploration

Homepage: https://shmpublisher.com/index.php/joetex

p-ISSN 3025-3470 | e-ISSN 3026-1066

Malaria Disease Detection System in Humans Using Convolutional Neural Network (CNN)

Natasya Siska Fitri Yana[1], **Achmad Rozin Shabaha**[2], **Jumanto**[3]
[1,2,3]Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received Jan 18, 2026
Revised Jan 25, 2026
Accepted Jan 25, 2026

Keywords:

Plasmodium parasites
Blood smear
Microscopy
CNN

ABSTRACT

Malaria is a deadly disease transmitted by the Plasmodium parasite. Detection is performed by trained microscopists who analyze microscopic images of blood smears. This analysis can be done automatically using modern deep learning techniques. The need for skilled labor can be significantly reduced by developing accurate and efficient automated models. In this article, we propose a fully automated convolutional neural network (CNN)-based model for diagnosing malaria from microscopic images of blood smears. Various techniques including knowledge distillation, data augmentation, autoencoder, feature extraction with CNN model to optimize and improve model accuracy and reasoning performance. Our deep learning model can detect malaria parasites from microscopic images with 95% accuracy requiring more than 27,600 images. This shows that the mode is able to provide more accurate predictions compared to malaria disease detection models using other algorithms such as in previous studies with an accuracy of 90%. By using CNN algorithm, this article can contribute novelty in the development of effective malaria detection methods for malaria disease.

Corresponding Author:

Natasya Siska Fitri Yana,
Department of Computer Science,
Universitas Negeri Semarang,
Semarang, Jawa Tengah, Indonesia.
Email: natasyasiska@students.unnes.ac.id

https://doi.org/10.52465/joetex.v3i2.646

1. INTRODUCTION

Malaria is still one of the most widespread and least protected diseases affecting the world's population and is caused by a parasite transmitted through the bite of an infected mosquito. It is also considered an endemic disease and a leading cause of death in humans especially in the tropics. A total of 3.2 billion people live in 97 countries worldwide, areas at risk of malaria of which 1.2 billion live in high-risk areas (214 million people have been exposed to malaria according to a WHO report in 2015 with approximately 438,000 deaths) [1]–[4]. Malaria is a contagious infection caused by the Plasmodium parasite. Malaria can be spread through the bite of Anopheles mosquitoes infected with Plasmodium parasites. Poor socioeconomic conditions make treatment and prevention of the disease difficult [5]. Malaria that affects humans has three stages: trophozoites, schizonts, and gametocytes. Infected people have a range of symptoms from mild to severe, serious enough to cause death [6]. After entering the human body, Plasmodium parasites will spread into red blood cells and divide themselves in them. This process will cause the body's immune response, which can produce symptoms such as fever, headache, nausea, vomiting, and anemia. Therefore, it is important to treat malaria quickly and appropriately to prevent more serious infections.

Malaria parasites have a complex life cycle because they require a host or habitat, both for humans and anopheles mosquitoes. There are several ways to treat malaria such as clinical diagnosis to look for symptoms experienced, as well as laboratory diagnosis to look at the patient's blood [6]. Many methods are used to test for malaria, using diagnostic methods commonly used in malaria health laboratories, such as biochemical testing (parasite lactate dehydrogenase LDH), fluorometric methods (Plasmodium falciparum fluorochrome Hoechst 33458-AND), or the addition of radioactive labeled precursors. (Tritiated hypoxanthine, the true gold standard in antimalarial drug research). One method commonly used in health facilities for malaria research is microscopic examination. Although this method is considered the most accurate as well as cost-effective, it requires expert expertise in verifying the results [7].

Rapid and accurate malaria detection remains a major challenge, especially in areas with limited health facilities. In recent years, advances in artificial intelligence, particularly deep learning, have shown great potential in the field of medical image analysis. For example, Convolutional Neural Networks (CNNs) are widely used in image classification tasks due to their ability to automatically extract spatial features from image data without the need for manual feature extraction. A number of studies have shown that CNN-based approaches are able to provide better performance than conventional machine learning methods in detecting malaria parasites in blood smear images, with high accuracy and sensitivity and greater resistance to variations in coloring and lighting. For example, Rajaraman et al. showed that CNN architecture can significantly improve the accuracy of malaria parasite detection compared to artificial feature-based methods [8]. In addition, Poostchi et al. reported that CNNs are effective in identifying Plasmodium-infected erythrocytes and have the potential to be applied in resource-limited settings [9]. Furthermore, various studies in the field of medical imaging also state that the deep learning approach is a reliable solution for automatic malaria diagnosis due to its scalability and ability to handle large amounts of data [10].

Based on the description above, this study uses a Convolutional Neural Network (CNN) as the main approach in improving the accuracy of the malaria detection system. CNN was chosen based on its ability to recognize patterns in blood images better than conventional methods, so it is expected to provide more accurate classification results. In this study, the model training and testing processes were carried out using the Google Colaboratory platform, utilizing a malaria image dataset obtained from Kaggle. The use of this platform facilitated the data processing and model training processes efficiently. Through this approach, this study is expected to produce a more reliable malaria detection system that can be used as a tool to support the accurate early detection of malaria, particularly in Indonesia.

2. METHOD

2.1 Deep Learning

It belongs to the branch of machine learning that has artificial neural network (ANN) algorithms that use hidden layer processed metadata and non-linear data modification to calculate output values [11]. Deep learning algorithms have the unique feature of automatically extracting data. The algorithms are used to solve problems. Deep learning has applications in many fields such as science, business, government, and medicine. These machine learning techniques can perform image recognition and speech recognition, predict actions and analyze data [11].

2.2 Conventional Neural Network

The method used in the implementation of this research is Conventional Neural Network. Neural Network is a set of algorithms that attempt to find underlying relationships in data sets through a process that mimics the way the human brain works. In this sense, a neural network refers to a system of organic and artificial neurons. Neural networks can adapt to changing inputs in this way, the network provides the best results without having to redefine the exit criteria. Neural networks work similar to neural networks in the human brain "neurons" of a neural network are mathematical functions that collect and classify information according to a specific architecture. Networks are very similar to statistical methods such as curve fitting and regression analysis. A neural network contains layers of interconnected nodes. Each node is a perceptron and is similar to multiple linear regression. The perceptron integrates the signals generated by multiple linear regression into an activation function, which can be non-linear.

2.3 Research Data

.

In this study, the dataset used was obtained from the Kaggle Malaria Cell Images Dataset (Luca, 2020). The total images that have been used are 27,600, which are divided into two classes of images infected with malaria parasites (*parasitized*) and images not infected with malaria parasites, namely (*uninfected)*. Based on the total data obtained, 75% composition is used as training data and the rest is used for test data. Below are shown some images for each class in the Malaria Cell Images Dataset in Figure 1.

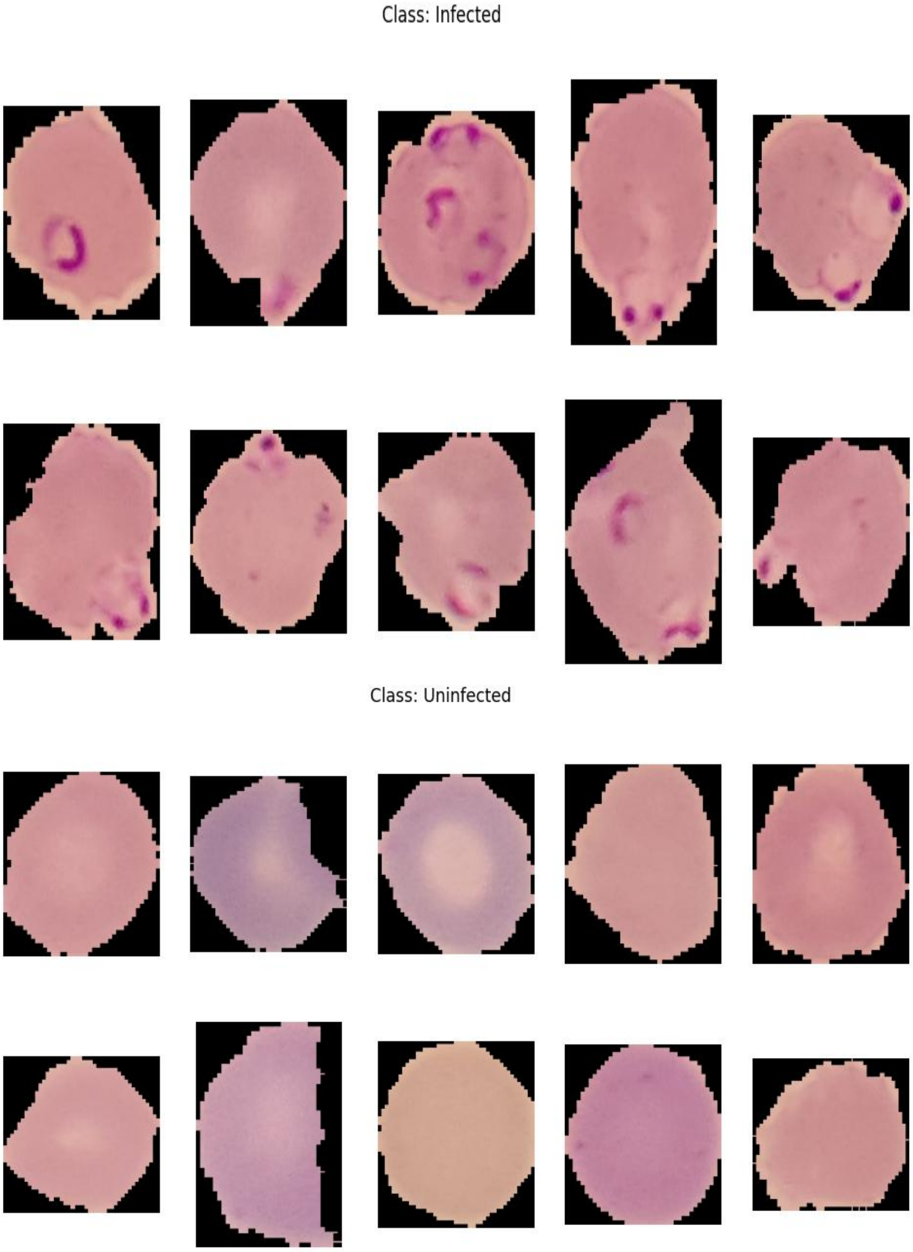

Figure 1. Example images in the infected (top) and uninfected (bottom) classes

2.4 System Architecture Design

The system architecture design used in testing the image classification system is as shown in Figure 2 below.

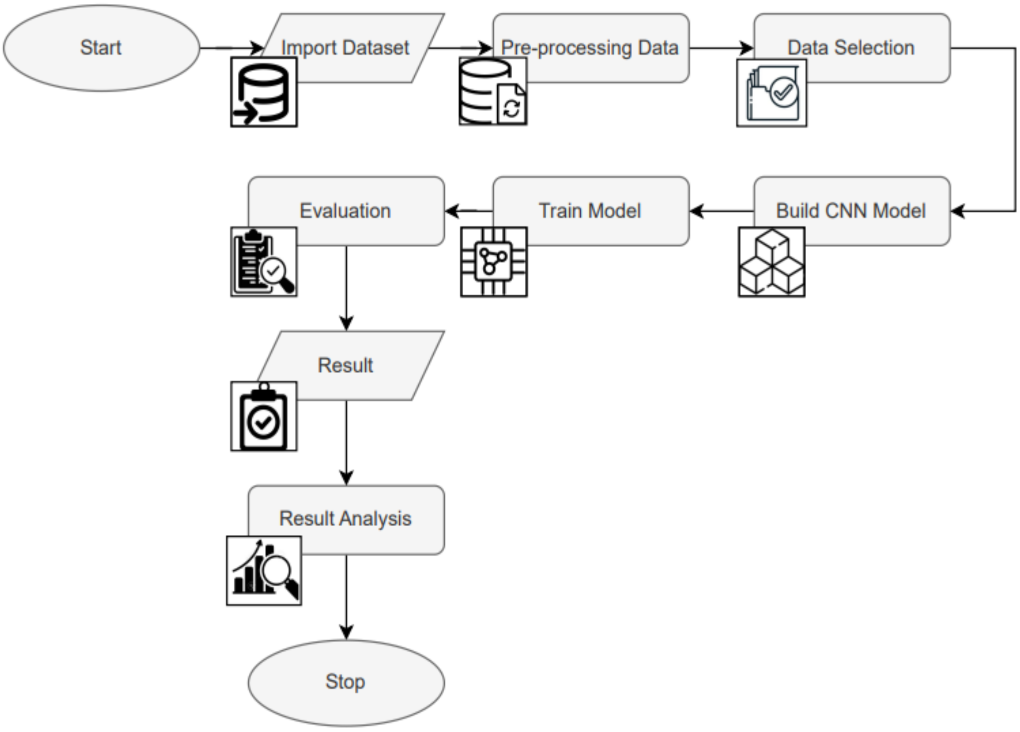

Figure 2. Flowchart of the proposed system

1. Data collection

The data collection stage in this research begins by utilizing the Malaria Cell Images Dataset which can be accessed through kaggle. This dataset is the main data source of information containing images infected with malaria parasites (*parasitized*) and images not infected with malaria parasites, namely (*uninfected*). By collecting data from the Malaria Cell Images Dataset, this research has a representative database to classify malaria using Convolutional Neural Network.

2. Data pre-processing

Data pre-processing is carried out to prepare data that will be used in the classification of images of images infected with malaria parasites (parasitized) and images not infected with malaria parasites, namely (uninfected) using Convolutional Neural Network. This data pre-processing stage aims to improve performance and predict malaria disease. By doing proper data pre-processing, it can prepare optimal data for testing CNN models in predicting malaria disease.

3. Data Selection

The selection of appropriate data on malaria in the Convolutional Neural Network method is very important to ensure the reliability, accuracy and generalization of the model. With a representative dataset that can meet the needs and to achieve a goal well as well as complete, accurate, and unbiased data so that the data selection carried out produces valid results in diagnosing malaria disease better.

4. Model formation

The model building stage is an important step in this research for the Convolutional Neural Network approach to the malaria disease detection system. To build the model, we will use the selection data generated from the data selection stage. This formation process is carried out by dividing the data into training data and testing data. Training data is used to train the model and test data is used to test the accuracy of the model. The CNN model created will use several layers, namely Conv2D, MaxPooling2D, Batch Normalization, DropOut, Flatten, and Dense. An overview of the CNN architecture created in this research will be shown in Figure 3 below.

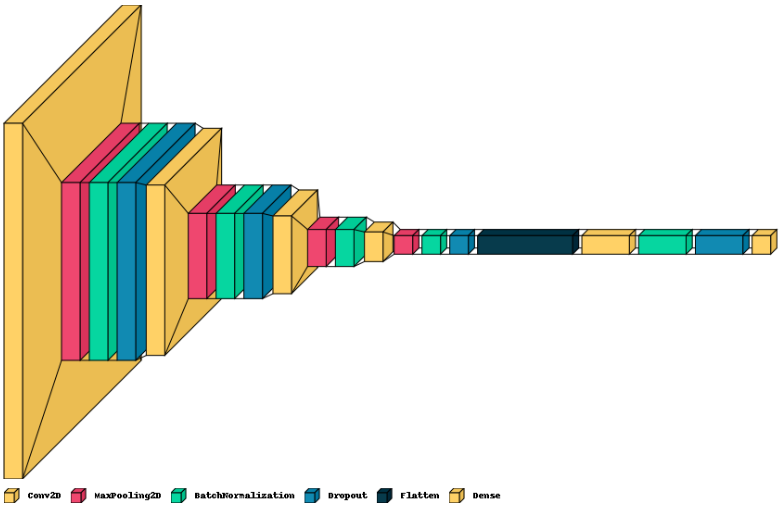

Figure 3. CNN model architecture

5. Performance Test

After the formation of the Convolutional Neural Network model, a performance test is conducted to determine how well the model performs in predicting a person's potential for malaria. The CNN performance test on malaria will measure the accuracy and effectiveness of the algorithm in recognizing and predicting the presence of malaria parasites in blood samples, so that it can provide information for research and technology development in the medical field.

6. Analysis of Results

The result analysis stage is a step that must be taken in this study to measure the performance of Convolutional Neural Network in malaria images. After the model is tested using testing data that has previously been separated. Testing data is used to test the model's ability to classify malaria images. This analysis stage is carried out with the intention of knowing the results of the contribution of accuracy in the Convolutional Neural Network model in predicting a person affected by malaria.

2.5 Classification Testing

At the classification testing stage is the stage to measure the performance value of the system. In this study, testing has been carried out on the *Confusion Matrix* algorithm. *Confusion Matrix* includes *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), and *False Negatives* (FN) which present the number of correct predictions in the positive class, the number of incorrect predictions in the positive class, the number of correct predictions in the negative class, and the number of incorrect predictions in the negative class in a dataset. The *Confusion Matrix* used to evaluate the performance of the model in the study is as follows:

Accuracy, the accuracy of a classifier in categorizing data points into their respective classes can be represented by accuracy. Formally, accuracy is defined as the ratio between the total number of correctly classified samples and the total number of samples. In mathematical notation, accuracy can be represented by equation (1).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

3. RESULTS AND DISCUSSIONS

The optimization results using both PSO and GOA for PD, PI, and PID controllers were analyzed based on compressor power consumption and battery temperature control. The system was simulated in MATLAB/Simulink, and the performance was evaluated through convergence graphs, objective function values, and time-domain response.

In data pre-processing, one image from each class will be selected to be displayed as an overview of each class. for its implementation can be seen from Figure 4. Then in the data selection stage, the data will be divided into two parts, namely training data and testing data. In this model, the data is divided into 80% training data and 20% testing data.

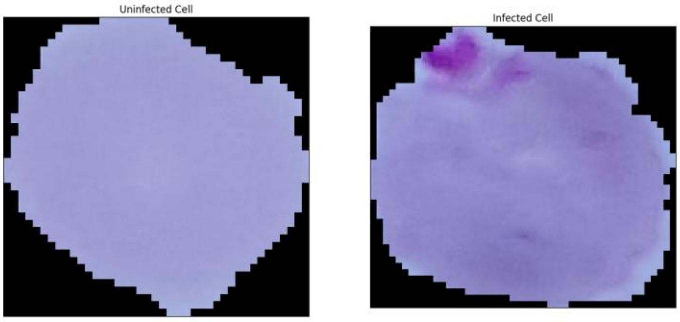

Figure 4. Class uninfected (left) and infected (right)

After being divided, the pre-processed images from the dataset will be stacked together with the original BGR image into a 5D image (width, height, 5). The output of this stacked result will be displayed in Figure 5 below.

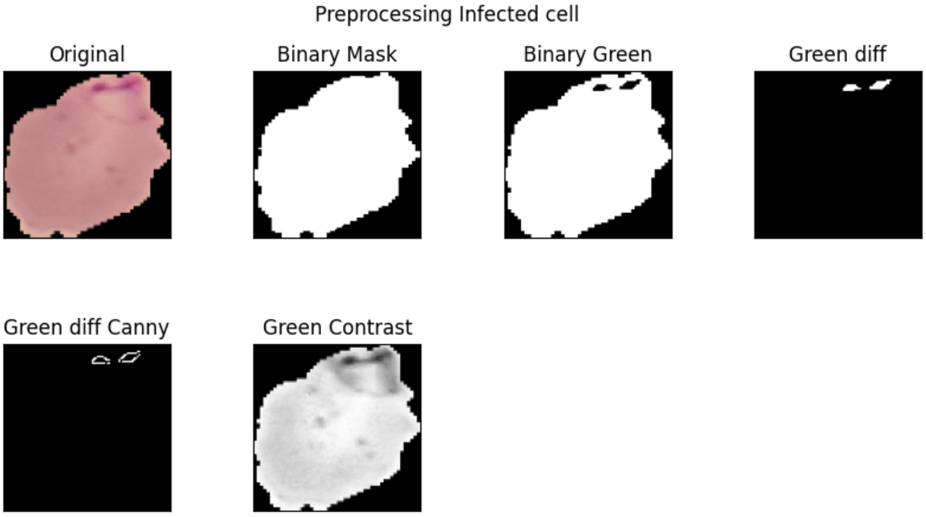

Figure 5. Pre-processed image after stacked

After being stacked, the list of dataset images will be converted into several arrays which will then be normalized. The arrays and normalization results are shown in Figure 6 below.

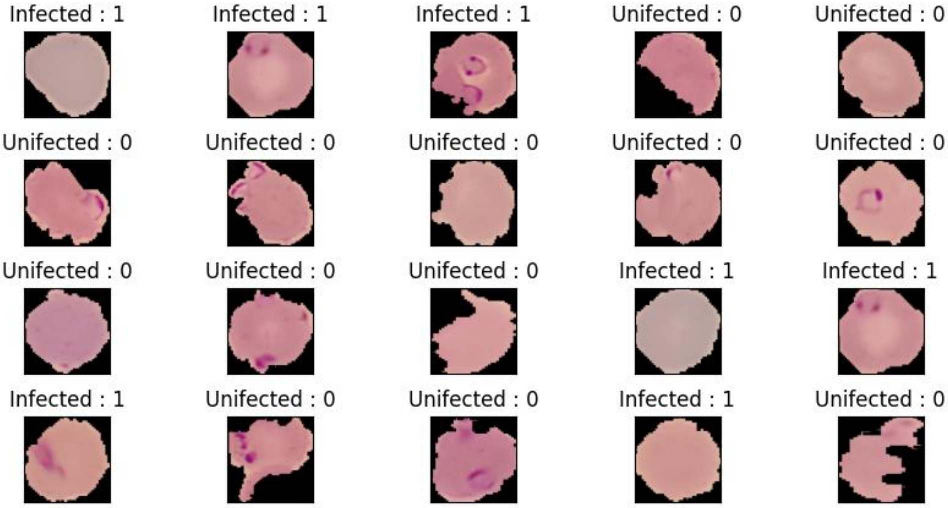

Figure 6. Arrays and normalization

In the CNN modeling section, the model used in the research uses 20 layers filled with Conv2D, MaxPooling2D, Batch Normalization, and DropOut layers. The following shows the contents of the layers used in the form of a table.

Model: “sequential”

Tabel 1. Layer from CNN model

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 96, 96, 16)	2016
max_pooling2d (MaxPooling2D)	(None, 48, 48, 16)	0
batch_normalization (BatchNormalization)	(None, 48, 48, 16)	64
dropout (Dropout)	(None, 48, 48, 16)	0
conv2d_1 (Conv2D)	(None, 46, 46, 32)	4640
max_pooling2d_1 (MaxPooling2D)	(None, 23, 23, 32)	0
batch_normalization_1 (BatchNormalization)	(None, 23, 23, 32)	128
dropout_1 (Dropout)	(None, 23, 23, 32)	0
conv2d_2 (Conv2D)	(None, 21, 21, 32)	9248
max_pooling2d_2 (MaxPooling2D)	(None, 10, 10, 32)	0
batch_normalization_2 (BatchNormalization)	(None, 10, 10, 32)	128
conv2d_3 (Conv2D)	(None, 8, 8, 64)	18496
max_pooling2d_3 (MaxPooling2D)	(None, 4, 4, 64)	0
batch_normalization_3 (BatchNormalization)	(None, 4, 4, 64)	256
dropout_2 (Dropout)	(None, 4, 4, 64)	0
flatten (Flatten)	(None, 1024)	0
dense (Dense)	(None, 512)	524800
batch_normalization_4 (BatchNormalization)	(None, 512)	2048
dropout_3 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 1)	513

Total params: 562,337
Trainable params: 561,025
Non-trainable params: 1,312

The accuracy obtained during the training process showed fluctuations in the early epochs, particularly in epochs 1 to 5. This is a common condition in the deep learning model training process, where the model is still adjusting its initial weights and parameters. After several iterations, the model's performance began to stabilize and reached its highest accuracy in epochs 21 and 22, at 99.3%. This shows that the CNN model is able to learn image patterns more optimally as the number of epochs increases.

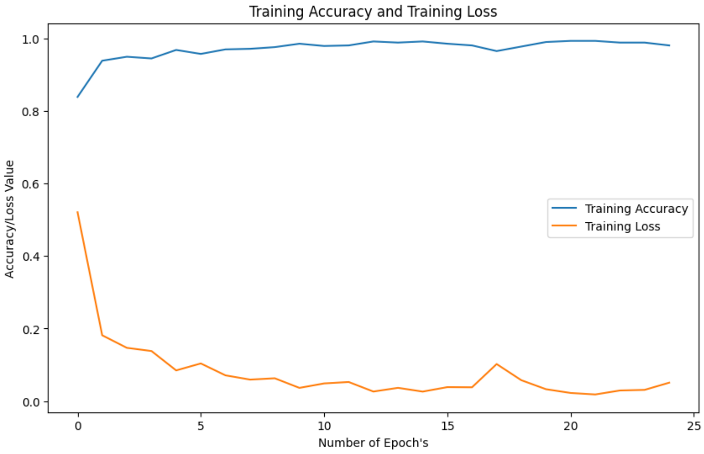

Figure 7. CNN model accuration graph

During the evaluation stage, the dataset was divided into 80% training data and 20% test data, with 25 epochs and a batch size of 32. The test results showed that the model achieved an accuracy rate of 95%, indicating that the model has good generalization capabilities for data it has never seen before. These results show that the model is able to learn malaria image patterns effectively without experiencing significant overfitting.

When compared to a number of previous studies using similar datasets, the results obtained in this study show competitive performance. Several previous studies reported varying levels of accuracy, depending on the method, model architecture, and preprocessing techniques used. As shown in Table 2, these studies generally produced accuracy below or close to the values obtained in this study. This shows that the proposed CNN approach is capable of providing comparable, even better performance than some previously reported methods.

Table 2. Performance comparison with previous studies

Author	Year	Method	Accuracy
[4]	2018	CNN	94.1%
[1]	2020	Deep CNN	94.8%
This Study	2026	CNN	95.0%

Based on the comparison results in Table 2, it can be seen that the method proposed in this study is capable of producing higher accuracy compared to several previous studies that used similar datasets. Despite differences in the amount of data, preprocessing techniques, and model architecture used, the results obtained show that the CNN approach applied has competitive performance.

This performance improvement is thought to be influenced by the selection of a more suitable CNN architecture and an optimal training process, enabling the model to extract important features from malaria images more effectively. Thus, the results of this study indicate that the proposed method has the potential to be applied as an automatic malaria detection support system, especially in environments with limited experts and medical facilities.

4. CONCLUSION

This study examines the application of Convolutional Neural Networks (CNN) in detecting malaria based on blood smear images. The test results show that the proposed model is capable of achieving an accuracy rate of 95%, indicating that CNN is effective in recognizing malaria image patterns. When compared to several previous studies using similar datasets, the results obtained show competitive performance. However, this study still has several limitations. The dataset used is still limited to one data source, and testing was only conducted on one model architecture. In addition, the evaluation of model performance still focuses on accuracy metrics. Therefore, further research is expected to use more diverse datasets, apply various deep learning architectures, and add other evaluation metrics such as precision, recall, and F1-score to obtain more comprehensive results. Further development is also expected to improve the reliability of the system in supporting automatic malaria detection, especially in resource-limited environments.

REFERENCES

[1] A. Vijayalakshmi and B. R. Kanna, “Deep learning approach to detect malaria from microscopic images,”

Multimed. Tools Appl., vol. 79, no. 21--22, pp. 15297–15317, 2020.

[2] A. Bashir, Z. A. Mustafa, I. Abdelhameid, and R. Ibrahem, "Detection of malaria parasites using digital image processing," in *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, 2017, pp. 1–5.

[3] H. A. Nugroho, S. A. Akbar, and E. E. H. Murhandarwati, "Feature extraction and classification for detection malaria parasites in thin blood smear," in *2015 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2015, pp. 197–201.

[4] Y. Dong *et al.*, "Evaluations of Deep Convolutional Neural Networks for Automatic Identification of Malaria Infected Cells," *IEEE J. Biomed. Heal. Informatics*, 2018.

[5] D. Liu and Jianyu, "Otsu Method and K-means," in *International Conference on Hybrid Intelligent Systems*, 2009, pp. 344–349.

[6] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," *SN Appl. Sci.*, vol. 2, no. 3, 2020.

[7] M. Tripathi, "Analysis of Convolutional Neural Network based Image Classification Techniques," *J. Innov. Image Process.*, vol. 3, no. 2, pp. 100–117, 2021.

[8] S. Rajaraman, S. Jaeger, and S. K. Antani, "Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images," 2019.

[9] M. Poostchi, K. Silamut, R. J. Maude, and S. Jaeger, "Image analysis and machine learning for detecting malaria," *Transl. Res.*, vol. 194, pp. 36–55, 2016.

[10] "World malaria report 2021," *World Health Organization*, 2021. [Online]. Available: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021.

[11] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.