JHSS (Journal of Humanities and Social Studies) *Volume 10, Number 03, Page 849-856*
https://jhss-lppm.unpak.ac.id/index.php/jhss *e-ISSN: 2598-120X; p-ISSN: 2598-117X*

MALARIA BLOOD CELL IMAGE CLASSIFICATION USING RESNET50V2 BASED ON TRANSFER LEARNING WITH WEB APPLICATION IMPLEMENTATION

Rika Agisha Siti Nurazizah [a*)], Adhi Kusnadi [a)], Risnandar [a)]

[a)] *Universitas Nusa Putra, Sukabumi, Indonesia*

[*)] *Corresponding Author: rika.agisha@nusaputra.ac.id*

Article history: received 01 July 2026; revised 16 July 2026; accepted 24 July 2026 *DOI: https://doi.org/10.33751/jhss.v10i02.270*

Abstract. Malaria remains an infectious disease that requires a fast and accurate diagnostic process, particularly in areas with limited access to skilled microscopists. Microscopic examination as a conventional diagnostic method still has several limitations, including dependence on the expertise of laboratory analysts, relatively long examination time, and the potential for interpretation errors. This study aims to develop a malaria blood cell image classification system using a *deep learning*-based *transfer learning* approach with the ResNet50V2 architecture and to compare its performance with MobileNetV2 as the baseline model. The dataset used in this study was the *NIH Malaria Cell Images Dataset*, consisting of 27,558 red blood cell images categorized into two classes: *Parasitized* and *Uninfected*. The research stages included image preprocessing, data augmentation, dataset splitting with a 70:15:15 ratio, model training, performance evaluation, and implementation of the best-performing model into a web-based application. The evaluation was conducted using *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, and ROC-AUC metrics. The results showed that ResNet50V2 achieved an *accuracy* of 93.76%, *precision* of 93.81%, *recall* of 93.76%, *F1-score* of 93.76%, and an AUC of 0.985. Meanwhile, MobileNetV2 obtained an *accuracy* of 93.06%, *precision* of 93.22%, *recall* of 93.05%, *F1-score* of 93.05%, and an AUC of 0.980. ResNet50V2 also produced a lower number of *false negatives* than MobileNetV2, making it more suitable for supporting the detection of malaria-infected blood cell images. The best-performing model was then implemented into a web-based application using Flask API. The findings indicate that ResNet50V2 has the potential to be used as a decision support system for early malaria screening based on digital images, although further clinical validation is still required before implementation in real medical environments.

Keywords: image classification, malaria, ResNet50V2, transfer learning, deep learning, web application.

I. INTRODUCTION

Malaria remains one of the infectious diseases that contributes significantly to the global health burden, particularly in tropical and subtropical regions. This disease is caused by *Plasmodium* parasites, which are transmitted through the bites of infected female *Anopheles* mosquitoes. Although various malaria control programs have shown progress over the past two decades, malaria transmission persists in many endemic countries and continues to pose a high risk of morbidity and mortality, especially among vulnerable populations and communities with limited access to healthcare services (Cibulskis et al., 2016; WHO, 2025). This condition indicates that early detection and accurate diagnosis remain essential components in malaria control.

Conventional malaria diagnosis is generally performed through microscopic examination of thick and thin blood smears. Microscopic examination allows laboratory personnel to identify the presence of *Plasmodium* parasites, calculate parasitemia levels, and distinguish parasite species based on their morphological characteristics. However, this method has several limitations, including dependence on the expertise of analysts, relatively long examination time, vulnerability to variations in smear quality, and the potential for interpretation errors due to observer fatigue or poor microscopic image quality (Moody, 2002; Tangpukdee et al., 2009; Wongsrichanalai et al., 2007). These challenges become more significant in healthcare facilities with limited availability of trained microscopists.

The development of digital image processing and artificial intelligence has opened opportunities to develop medical image-based diagnostic support systems. In this context, deep learning has become a widely used approach because it can automatically extract visual features without requiring manual feature engineering. Convolutional Neural Network (CNN) has been proven effective in various medical image classification tasks, including lesion detection, skin disease classification, organ segmentation, and microscopic image-based abnormality identification (Esteva et al., 2017; LeCun et al., 2015; Litjens et al., 2017; Topol, 2019; Yu et al., 2018). The ability of CNN to learn layered visual patterns makes it relevant for distinguishing infected and uninfected red blood cells.

Several previous studies have applied deep learning to malaria classification based on blood smear images. Poostchi et al. (2018) emphasized that image analysis and machine learning have the potential to support automatic malaria detection. Rajaraman et al. (2018) showed that the use of

Volume 10, Number 03, Page 849-856
e-ISSN: 2598-120X; p-ISSN: 2598-117X

pretrained CNNs as feature extractors could improve parasite detection performance in thin blood smear images. Gopakumar et al. (2018) also demonstrated that the CNN approach can be used for malaria diagnosis based on microscopic images. In addition, Bibin et al. (2017) and Dong et al. (2017) showed that deep learning models are capable of recognizing malaria infection patterns in blood cell images with competitive classification performance. These findings strengthen the assumption that deep learning is feasible to be developed as a digital image-based malaria diagnostic support system.

Nevertheless, the development of deep learning models for malaria image classification still faces several methodological challenges. First, variations in color, lighting, focus, and image quality can affect the consistency of features learned by the model. Second, very deep CNN models may encounter performance degradation and *vanishing gradient* problems if the training process is not properly designed. Third, model evaluation often emphasizes accuracy only, whereas medical classification requires more comprehensive measurements such as precision, recall, F1-score, ROC-AUC, and confusion matrix so that prediction errors can be analyzed more clearly (Bradley, 1997; Chicco & Jurman, 2020; Fawcett, 2006; Saito & Rehmsmeier, 2015).

One architecture that is relevant for addressing problems in deep networks is the Residual Network (ResNet). ResNet introduces a *skip connection* mechanism that enables information flow to bypass several layers directly, making the optimization process more stable in deeper networks (He et al., 2016a). Further development through *identity mapping* and *pre-activation* in ResNetV2 helps improve gradient flow during the training process (He et al., 2016b). In this study, ResNet50V2 was used as the main architecture because it has strong feature extraction capability, stable training performance, and suitability for the transfer learning approach. Transfer learning itself enables a model to utilize initial weights from large-scale datasets such as ImageNet, making the training process on relatively limited medical datasets more efficient (Deng et al., 2009; Pan & Yang, 2010).

In addition to selecting the main architecture, the use of a comparison model is necessary to evaluate the effectiveness of the proposed model more objectively. MobileNetV2 was selected as the baseline because it is a lightweight CNN architecture designed with *inverted residuals* and *linear bottlenecks*, making it computationally efficient and commonly used in resource-constrained applications (Sandler et al., 2018). The comparison between ResNet50V2 and MobileNetV2 is important to determine whether a model with deeper feature extraction capacity can provide meaningful performance improvement compared to a lightweight model. Furthermore, preprocessing and data augmentation are also important stages because they can improve input consistency, enrich training data variation, and help reduce the risk of overfitting (Shorten & Khoshgoftaar, 2019).

Based on the above discussion, there remains a research gap in the development of malaria classification systems that not only focus on model accuracy but also integrate preprocessing, data augmentation, transfer learning, multi-metric evaluation, model comparison, and structured web-based system implementation. Several previous studies have demonstrated the potential of deep learning in malaria classification; however, studies that specifically use ResNet50V2 as the main model, compare it with MobileNetV2, and implement the best-performing model into a web-based system still need to be further developed.

Therefore, this study aims to develop a malaria blood cell image classification system using the ResNet50V2 architecture based on transfer learning. This study also analyzes the effect of preprocessing and data augmentation on model performance, evaluates classification results using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC, and compares the performance of ResNet50V2 with MobileNetV2 as the baseline model. The final outcome of this study is expected to contribute to the development of an accurate, efficient, and applicable digital image-based malaria diagnostic support system.

II. RESEARCH METHODS

This study used an experimental quantitative approach based on *deep learning* to develop a malaria blood cell image classification system. The study focused on binary classification, namely distinguishing between malaria-infected red blood cell images (*parasitized*) and uninfected red blood cell images (*uninfected*). In general, the research stages included dataset collection, image preprocessing, data augmentation, dataset splitting, ResNet50V2 model design, model training, performance evaluation, comparison with the MobileNetV2 baseline model, and implementation of the model into a web-based application.

Research Dataset

The dataset used in this study was the *Malaria Cell Images Dataset* from the National Institutes of Health (NIH). This dataset consists of 27,558 microscopic red blood cell images, which are evenly divided into two classes: 13,779 *parasitized* images and 13,779 *uninfected* images. This dataset was used because it has been widely utilized in image-based malaria classification studies and has undergone class labeling, making it suitable for the development of an automatic classification model based on *deep learning* (Rajaraman et al., 2018).

The entire dataset was divided into three subsets, namely training data, validation data, and testing data. The dataset was split into 70% training data, 15% validation data, and 15% testing data. Based on this division, the training data consisted of 19,290 images, while the validation and testing data each consisted of 4,134 images. The data split was performed using a *stratified split* approach to maintain a balanced distribution of the *parasitized* and *uninfected* classes in each subset.

Data Preprocessing and Augmentation

The preprocessing stage was carried out to standardize the image format before being fed into the model. Each image was resized to 224 × 224 pixels with three RGB color channels to match the input size required by the ResNet50V2 and MobileNetV2 architectures. The pixel values were then normalized to the range of 0–1 to support a more stable and efficient model training process.

In addition to preprocessing, this study also applied data augmentation to the training data. The augmentation techniques

Volume 10, Number 03, Page 849-856
e-ISSN: 2598-120X; p-ISSN: 2598-117X

used included rotation, *flipping*, *zooming*, and brightness adjustment. Augmentation was performed to enrich the visual variation of the images without changing their original class labels. This strategy aimed to reduce the risk of *overfitting* and improve the model's generalization ability on new images. Data augmentation is a commonly used approach in image model training because it can increase training data variation and strengthen model robustness against visual changes in testing data (Shorten & Khoshgoftaar, 2019).

ResNet50V2 Model Design

The main model used in this study was ResNet50V2 with a *transfer learning* approach. ResNet50V2 was selected because it has a *residual learning* mechanism that can address performance degradation problems in deep neural networks. The ResNet architecture uses *skip connections* to maintain the flow of information and gradients during the training process, allowing the model to learn more stably (He et al., 2016a). ResNet50V2 is an improved version of ResNet that applies *identity mapping* and *pre-activation*, which are designed to improve the effectiveness of deep network training (He et al., 2016b).

In this study, ResNet50V2 was used as a *feature extractor* with the include_top=False configuration, meaning that the original classification layer was removed. Input images with a size of 224 × 224 × 3 were processed through the ResNet50V2 base model to extract visual features. The extracted features were then passed to a GlobalAveragePooling2D layer to reduce feature dimensions. After that, a Dense layer with 128 neurons and the ReLU activation function was added to learn advanced feature patterns. A Dropout layer with a rate of 0.5 was used to reduce the risk of *overfitting*, while the final layer used the *softmax* activation function to generate classification probabilities for the two classes, namely *parasitized* and *uninfected*. Dropout was applied as a regularization technique to reduce dependency among neurons during the training process (Srivastava et al., 2014).

MobileNetV2 Comparison Model

This study also used MobileNetV2 as a comparison model or baseline. MobileNetV2 was selected because it is a lightweight CNN architecture designed for computational efficiency through *inverted residuals* and *linear bottlenecks* (Sandler et al., 2018). The use of MobileNetV2 as a baseline aimed to determine the extent to which ResNet50V2, as the main model, performed compared to a lighter and more efficient model.

To ensure a fair comparison, ResNet50V2 and MobileNetV2 were trained using the same dataset, preprocessing procedure, augmentation scheme, data split, and training parameters. Therefore, the performance differences obtained could be attributed to the characteristics of each model architecture rather than differences in data treatment or training configuration.

Model Training

The model training process was carried out using TensorFlow and Keras. The models were trained for a maximum of 50 epochs with a *batch size* of 32. The optimizer used was Adam with a *learning rate* of 0.001. Adam was selected because it can adaptively adjust the learning rate based on the first and second moment estimates of the gradients, making it effective for optimizing *deep learning* models (Kingma & Ba, 2015).

The loss function used was *categorical crossentropy* because the classification task involved two classes with labels represented using *one-hot encoding*. During training, model performance was monitored based on loss and accuracy values on the training and validation data. To prevent *overfitting*, an *early stopping* technique was applied with a *patience* value of 5 based on the *validation loss*. If the *validation loss* did not decrease for five consecutive epochs, the training process was automatically stopped.

Model Evaluation

Model evaluation was conducted using testing data that were not used during the training or validation process. The evaluation aimed to objectively measure the model's ability to classify new images. The evaluation metrics used included accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. The use of multiple metrics was important because evaluation of medical classification models should not rely solely on accuracy but should also consider positive and negative prediction errors (Bradley, 1997; Fawcett, 2006; Saito & Rehmsmeier, 2015).

Accuracy was used to measure the proportion of correct predictions among all testing data. Precision was used to measure the correctness of positive predictions made by the model. Recall was used to determine the model's ability to detect all actual positive cases. F1-score was used to evaluate the balance between precision and recall. The confusion matrix was used to examine the distribution of correct and incorrect predictions in each class, while ROC-AUC was used to assess the model's ability to distinguish between the *parasitized* and *uninfected* classes.

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

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

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

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Figure 1. Evaluation formulas for the classification model

In Figure 1, TP (*True Positive*) refers to the number of malaria-infected blood cell images correctly predicted as infected. TN (*True Negative*) refers to the number of uninfected blood cell images correctly predicted as uninfected. FP (*False Positive*) refers to the number of uninfected blood cell images incorrectly predicted as infected. FN (*False Negative*) refers to the number of malaria-infected blood cell images incorrectly predicted as uninfected.

Web-Based System Implementation

The best-performing trained model was implemented into a web-based application using a client-server architecture. The system backend was developed using Flask API, while the user interface was built using HTML, CSS, and JavaScript. Flask

served as a bridge between the web interface and the trained *deep learning* model.

The system workflow begins when a user uploads a blood cell image through the web interface. The uploaded image is then sent to the Flask backend using the HTTP POST method. The backend performs image preprocessing, including resizing, normalization, and input format adjustment to match the model requirements. After that, the image is processed by the model to generate a class prediction. The prediction result is then sent back to the web interface in the form of a class label and a *confidence score*.

This system was designed as an artificial intelligence-based support tool for malaria blood cell image classification. The classification results displayed by the system are not intended to replace medical diagnosis but rather to serve as a decision support system that can assist in early screening based on digital images.

III. RESULTS AND DISCUSSION

Results

This study used the *Malaria Cell Images Dataset* from the National Institutes of Health (NIH), which consists of microscopic red blood cell images categorized into two classes: *Parasitized* and *Uninfected*. The *Parasitized* class represents red blood cell images infected with malaria parasites, while the *Uninfected* class represents red blood cell images without malaria infection. The number of images in both classes was balanced, with 13,779 images in the *Parasitized* class and 13,779 images in the *Uninfected* class, resulting in a total of 27,558 images.

The balanced distribution of data between the two classes is an important aspect of this study because it can reduce the risk of model bias toward one particular class. An imbalanced dataset may cause the model to recognize the majority class more dominantly than the minority class. In the context of malaria classification, balanced data distribution supports the learning process by enabling the model to recognize visual patterns from both classes more proportionally.

The dataset was then divided into three subsets: training data, validation data, and testing data, using a ratio of 70%:15%:15%. This division resulted in 19,290 images for training, 4,134 images for validation, and 4,134 images for testing. The distribution of data in each subset was maintained equally between the *Parasitized* and *Uninfected* classes so that the training, validation, and testing processes could be conducted objectively.

Table 1. Dataset split

Subset	Parasitized	Uninfected	Total
Training	9,645	9,645	19,290
Validation	2,067	2,067	4,134
Testing	2,067	2,067	4,134
Total	13,779	13,779	27,558

Before the model training process, all images underwent a preprocessing stage. This stage included resizing images to 224 × 224 pixels, normalizing pixel values to the range of 0–1, and improving image quality. The image size of 224 × 224 pixels was used because it corresponds to the input format required by the ResNet50V2 and MobileNetV2 architectures. Normalization was performed to stabilize the training process, while image enhancement was applied to help the model recognize relevant visual patterns in blood cell images.

In addition to preprocessing, data augmentation was also applied to the training data. The augmentation techniques included rotation, position shifting, *zooming*, and brightness adjustment. The augmentation results showed that each transformation produced image variations different from the original images. Combined augmentation produced the highest variation, with a *Mean Absolute Difference* (MAD) value of 0.393, while rotation produced the smallest change, with an MAD value of 0.060. These results indicate that augmentation successfully enriched the visual variation of the training data without changing the image class labels.

The training stage was conducted using two transfer learning models: ResNet50V2 as the main model and MobileNetV2 as the comparison model. Both models were trained using the same configuration to ensure a fair performance comparison. The main parameters included an input image size of 224 × 224 × 3, a maximum of 50 epochs, a *batch size* of 32, the Adam optimizer, a *learning rate* of 0.0001, *categorical crossentropy* as the loss function, and *early stopping* with a patience value of 10.

During the training process, ResNet50V2 showed more stable performance than MobileNetV2. ResNet50V2 reached its optimal performance at the 11th epoch, with a *training accuracy* of 92.63%, a *validation accuracy* of 95.23%, a *training loss* of 0.1978, and a *validation loss* of 0.1348. Meanwhile, MobileNetV2 reached its optimal performance at the 15th epoch, with a *training accuracy* of 92.38%, a *validation accuracy* of 93.86%, a *training loss* of 0.2115, and a *validation loss* of 0.1664.

Table 2. Comparison of optimal epochs between models

Model	Optimal Epoch	Training Loss	Training Accuracy (%)	Validation Loss	Validation Accuracy (%)
ResNet50V2	11	0.1978	92.63	0.1348	95.23
MobileNetV2	15	0.2115	92.38	0.1664	93.86

These results show that ResNet50V2 was able to achieve its best performance in fewer epochs than MobileNetV2. The lower *validation loss* of ResNet50V2 indicates that the model had better generalization ability on the validation data. In addition, the higher *validation accuracy* suggests that the feature representations learned by ResNet50V2 were more effective in distinguishing between malaria-infected and uninfected blood cell images.

The final evaluation was conducted using 4,134 testing images. The testing results showed that ResNet50V2 achieved an accuracy of 93.76%, precision of 93.81%, recall of 93.76%, and F1-score of 93.76%. Meanwhile, MobileNetV2 achieved an accuracy of 93.06%, precision of 93.22%, recall of 93.05%, and F1-score of 93.05%.

Table 3. Comparison of model evaluation metrics

Metric	ResNet50V2 (%)	MobileNetV2 (%)
Precision	93.81	93.22
Recall	93.76	93.05
F1-Score	93.76	93.05
Accuracy	93.76	93.06

Based on Table 3, ResNet50V2 achieved higher scores across all evaluation metrics compared to MobileNetV2. The accuracy difference between the two models was 0.70%, with ResNet50V2 achieving 93.76% and MobileNetV2 achieving 93.06%. Although the difference was not substantial, the performance improvement of ResNet50V2 was consistent across precision, recall, F1-score, and accuracy. This indicates that ResNet50V2 demonstrated more stable classification performance than the comparison model.

A more detailed analysis was conducted using a confusion matrix. In the ResNet50V2 model, 1,905 *Parasitized* images were correctly classified as *Parasitized*, while 1,971 *Uninfected* images were correctly classified as *Uninfected*. The model produced 96 *false positives* and 162 *false negatives*. Thus, the total number of correct predictions obtained by ResNet50V2 was 3,876 out of 4,134 testing images.

Table 4. Confusion matrix of ResNet50V2

Actual	Predicted Parasitized	Predicted Uninfected
Parasitized	1,905	162
Uninfected	96	1,971

In MobileNetV2, 1,861 *Parasitized* images were correctly classified as *Parasitized*, while 1,986 *Uninfected* images were correctly classified as *Uninfected*. This model produced 81 *false positives* and 206 *false negatives*. The total number of correct predictions obtained by MobileNetV2 was 3,847 out of 4,134 testing images.

Table 5. Comparison of confusion matrix results between ResNet50V2 and MobileNetV2

Component	ResNet50V2	MobileNetV2
True Positive (TP)	1,905	1,861
True Negative (TN)	1,971	1,986
False Positive (FP)	96	81
False Negative (FN)	162	206
Total Correct Predictions	3,876	3,847
Accuracy (%)	93.76	93.06

Based on Table 5, ResNet50V2 produced fewer *false negatives* than MobileNetV2, with 162 compared to 206 images. In the context of malaria detection, *false negatives* are a critical type of error because they indicate blood cell images that are actually infected but are predicted as uninfected. The lower number of *false negatives* in ResNet50V2 indicates that this model had better infection detection capability than MobileNetV2.

In addition to the confusion matrix, model performance was also evaluated using ROC-AUC. ResNet50V2 obtained an AUC value of 0.985, while MobileNetV2 obtained an AUC value of 0.980. Both values indicate very strong class discrimination ability because they are close to the maximum value of 1.00. However, the slightly higher AUC value of ResNet50V2 shows that this model had better capability in distinguishing between the *Parasitized* and *Uninfected* classes across various classification thresholds.

The ResNet50V2 model, which achieved the best performance, was then implemented into a web-based application. The system was developed using Flask API as the backend and HTML, CSS, and JavaScript as the frontend interface. This application allows users to upload blood cell images through a browser. The system then performs preprocessing, model inference, and displays the classification result as either *Parasitized* or *Uninfected*, along with the *confidence score*. The implementation testing results showed that the system was able to provide real-time predictions with an inference time of approximately 1–2 seconds per image, depending on server conditions and the uploaded file size.

Discussion

The results of this study indicate that ResNet50V2 with a transfer learning approach was able to achieve strong classification performance on the malaria blood cell image dataset. An accuracy of 93.76%, precision of 93.81%, recall of 93.76%, and F1-score of 93.76% show that the model not only achieved a high rate of correct predictions but also maintained a balance between the accuracy of positive predictions and the ability to detect truly infected images. In the context of medical image classification, the balance between precision and recall is important because the model is not only expected to achieve high accuracy but also to minimize classification errors in the positive class.

This finding is consistent with the study by Rajaraman et al. (2018), which showed that the use of pretrained CNNs as feature extractors could improve malaria parasite detection performance in thin blood smear images. The results of this study also support the review by Poostchi et al. (2018), which stated that the combination of image analysis and machine learning has strong potential to support automatic malaria detection. Therefore, the use of ResNet50V2 in this study strengthens empirical evidence that transfer learning can be effectively applied to medical image classification, particularly in microscopic malaria blood cell images.

The higher performance of ResNet50V2 compared to MobileNetV2 can be explained by the characteristics of the residual network architecture. ResNet uses a *skip connection* mechanism that allows information and gradients to flow more effectively in deep networks (He et al., 2016a). In ResNet50V2, the use of *identity mapping* and *pre-activation residual blocks* helps improve training stability and gradient propagation (He et al., 2016b). These characteristics enable the model to learn more complex visual features, such as texture, shape, color intensity, and parasite patterns in red blood cells.

In contrast, MobileNetV2 was designed as a lightweight architecture that prioritizes computational efficiency through *inverted residuals* and *linear bottlenecks* (Sandler et al., 2018). Therefore, the slightly lower performance of MobileNetV2 compared to ResNet50V2 is understandable because this model was designed to balance accuracy and efficiency. In this study, MobileNetV2 still demonstrated good performance with an accuracy of 93.06%, but ResNet50V2 outperformed it in terms of accuracy, recall, F1-score, and AUC. This indicates that for

Volume 10, Number 03, Page 849-856
e-ISSN: 2598-120X; p-ISSN: 2598-117X

malaria image classification tasks where performance is the main priority, ResNet50V2 is more suitable as the main model.

The preprocessing and data augmentation stages also played an important role in model performance. Resizing images to 224 × 224 pixels ensured input compatibility with the model architecture, while normalization helped stabilize the learning process. Augmentation through rotation, shifting, zooming, and brightness adjustment enriched the variation of the training data. This is in line with Shorten and Khoshgoftaar (2019), who explained that image augmentation can increase data diversity, reduce the risk of overfitting, and improve model generalization. In this study, combined augmentation produced the greatest image variation based on MAD values, allowing the model to learn from broader visual variations during training.

The confusion matrix results show that ResNet50V2 produced 162 *false negatives*, fewer than MobileNetV2, which produced 206 *false negatives*. This finding is important because a *false negative* in malaria detection means that a blood cell image that is actually infected is classified as uninfected. In the context of a diagnostic support system, this type of error is more critical than a *false positive* because it may cause an infection to go undetected. Therefore, the higher recall of ResNet50V2 compared to MobileNetV2 indicates that the main model had better sensitivity in detecting malaria-infected images.

The AUC value of 0.985 obtained by ResNet50V2 further strengthens the evaluation results, indicating that the model had excellent class discrimination capability. According to Fawcett (2006), ROC-AUC is an important metric for evaluating the ability of a model to distinguish classes across various classification thresholds. In addition, Saito and Rehmsmeier (2015) emphasized that classification model evaluation should not rely solely on accuracy, particularly when errors in certain classes have important consequences. Therefore, the use of accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC in this study provides a more comprehensive overview of model performance.

Compared with previous studies, the performance of ResNet50V2 in this study is in a good category and is consistent with the trend of using deep learning for malaria classification. Gopakumar et al. (2018) showed that CNNs can be used for malaria diagnosis based on microscopic images, while Dong et al. (2017) demonstrated that deep convolutional neural networks can automatically identify malaria-infected cells. This study extends those findings by using ResNet50V2 as the main model, MobileNetV2 as the baseline model, and adding a web-based implementation as a form of system deployment.

Nevertheless, this study has several limitations. First, the dataset used was obtained from the public NIH dataset, so the model's performance may not directly represent image conditions from different healthcare facilities with variations in microscope devices, staining procedures, and lighting quality. Second, this study only performed binary classification between *Parasitized* and *Uninfected*, so it did not include the identification of *Plasmodium* species or parasitemia levels. Third, the web implementation remains at the prototype stage and has not yet undergone clinical validation with healthcare professionals.

Overall, the results of this study demonstrate that ResNet50V2 is an effective architecture for malaria blood cell image classification. This model achieved better performance than MobileNetV2 in almost all evaluation metrics, particularly in reducing the number of *false negatives*. The web-based implementation also shows that the model can be applied to an accessible and responsive system. Therefore, the developed system has the potential to be used as a prototype of a digital image-based early malaria screening support tool, although further validation using independent datasets and clinical testing is still required before being applied in real medical practice.

IV. CONCLUSIONS

Based on the research findings, the application of *deep learning* based on *transfer learning* using the ResNet50V2 and MobileNetV2 architectures proved capable of effectively classifying malaria blood cell images on the *NIH Malaria Cell Images Dataset*. Both models were able to distinguish between malaria-infected blood cell images (*Parasitized*) and uninfected blood cell images (*Uninfected*) with an accuracy level above 93%. These results indicate that the *transfer learning* approach can be optimally utilized in medical image classification, particularly in malaria detection based on microscopic images.

The evaluation results show that ResNet50V2 outperformed MobileNetV2 in almost all evaluation metrics. ResNet50V2 achieved an *accuracy* of 93.76%, *precision* of 93.81%, *recall* of 93.76%, *F1-score* of 93.76%, and an AUC value of 0.985. Meanwhile, MobileNetV2 achieved an *accuracy* of 93.06%, *precision* of 93.22%, *recall* of 93.05%, *F1-score* of 93.05%, and an AUC value of 0.980. The superiority of ResNet50V2 was also reflected in its lower number of *false negatives*, indicating that this model performed better in detecting blood cell images that were truly infected with malaria.

The implementation of *preprocessing* and data augmentation also contributed to improving the generalization capability of the model. Image resizing, pixel value normalization, and augmentation techniques such as rotation, shifting, *zooming*, and brightness adjustment helped the model learn visual variations more effectively and reduce the risk of *overfitting*. In addition, the implementation of the model into a web-based application demonstrated that the research outcomes were not limited to computational experiments but could also be applied as an accessible and usable prototype for malaria image classification.

Overall, this study shows that ResNet50V2 is a more suitable architecture than MobileNetV2 for malaria blood cell image classification tasks, particularly when the main objective is to achieve more accurate classification performance and higher sensitivity toward infection cases. The developed system has the potential to serve as a digital image-based early malaria screening support tool; however, it is not intended to replace medical diagnosis by professional healthcare personnel.

Recommendations

This study still has several limitations that need to be considered. The dataset used was limited to the *NIH Malaria Cell Images Dataset*, so the image characteristics may not fully

Volume 10, Number 03, Page 849-856
e-ISSN: 2598-120X; p-ISSN: 2598-117X

represent variations from different healthcare facilities, microscope types, staining techniques, and lighting conditions. In addition, the developed system was limited to binary classification, namely *Parasitized* and *Uninfected*, and therefore has not yet been able to identify *Plasmodium* species, parasite developmental stages, or parasitemia levels. The model used also remains a *black box*, meaning that its decision-making process cannot yet be explained in detail without additional interpretability methods.

Based on these limitations, future research is recommended to use more diverse datasets from multiple clinical sources so that the generalization ability of the model can be evaluated more broadly. Multi-class classification should also be developed so that the system can not only distinguish infected and uninfected images but also identify *Plasmodium* species or malaria infection stages. Furthermore, future studies may evaluate more advanced *deep learning* architectures, such as EfficientNet, DenseNet, ConvNeXt, Vision Transformer, or Swin Transformer, to explore potential improvements in classification performance.

The use of *Explainable Artificial Intelligence* methods such as Grad-CAM, LIME, or SHAP is also recommended so that the image regions used as the basis for the model's decisions can be visualized and interpreted more clearly. This is important for improving model transparency, especially if the system is further developed as a support tool in the healthcare field. In addition, clinical validation involving doctors, laboratory analysts, or healthcare professionals should be conducted before the system is implemented in real medical environments. Further development into a mobile application, cloud-based service, or integration with health information systems may also be considered to make the system more applicable and accessible.

REFERENCES

Bibin, D., Nair, M. S., & Punitha, P. (2017). Malaria parasite detection from peripheral blood smear images using deep belief networks. *IEEE Access, 5*, 9099–9108. https://doi.org/10.1109/ACCESS.2017.2705642

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*, Article 6. https://doi.org/10.1186/s12864-019-6413-7

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., ... Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface, 15*(141), Article 20170387. https://doi.org/10.1098/rsif.2017.0387

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251–1258). https://doi.org/10.1109/CVPR.2017.195

Cibulskis, R. E., Alonso, P., Aponte, J., Aregawi, M., Barrette, A., Bergeron, L., Fergus, C. A., Knox, T., Lynch, M., Patouillard, E., Schwarte, S., Stewart, S., & Williams, R. (2016). Malaria: Global progress 2000–2015 and future challenges. *Infectious Diseases of Poverty, 5*, Article 61. https://doi.org/10.1186/s40249-016-0151-8

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). https://doi.org/10.1109/CVPR.2009.5206848

Dong, Y., Jiang, Z., Shen, H., Pan, W. D., Williams, L. A., Reddy, V. V. B., Benjamin, W. H., & Bryan, A. W. (2017). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 101–104). https://doi.org/10.1109/EMBC.2017.8036814

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118. https://doi.org/10.1038/nature21056

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Gopakumar, G. P., Swetha, M., Siva, G. S., & Sai Subrahmanyam, G. R. K. (2018). Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. *Journal of Biophotonics, 11*(3), Article e201700003. https://doi.org/10.1002/jbio.201700003

He, K., Zhang, X., Ren, S., & Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

He, K., Zhang, X., Ren, S., & Sun, J. (2016b). Identity mappings in deep residual networks. In *Lecture Notes in Computer Science: Vol. 9908* (pp. 630–645). Springer. https://doi.org/10.1007/978-3-319-46493-0_38

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708). https://doi.org/10.1109/CVPR.2017.243

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*, Article 195. https://doi.org/10.1186/s12916-019-1426-2

Volume 10, Number 03, Page 849-856
e-ISSN: 2598-120X; p-ISSN: 2598-117X

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*. https://arxiv.org/abs/1412.6980

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A. W. M., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health, 1*(6), e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

Moody, A. (2002). Rapid diagnostic tests for malaria parasites. *Clinical Microbiology Reviews, 15*(1), 66–78. https://doi.org/10.1128/CMR.15.1.66-78.2002

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Poostchi, M., Silamut, K., Maude, R. J., Jaeger, S., & Thoma, G. (2018). Image analysis and machine learning for detecting malaria. *Malaria Journal, 17*, Article 177. https://doi.org/10.1186/s12936-018-2492-9

Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., Jaeger, S., & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ, 6*, Article e4568. https://doi.org/10.7717/peerj.4568

Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C., McCague, C., Beer, L., Weir-McCall, J. R., Teng, Z., Gkrania-Klotsas, E., Rudd, J. H. F., Sala, E., & Schönlieb, C. B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence, 3*, 199–217. https://doi.org/10.1038/s42256-021-00307-0

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), Article e0118432. https://doi.org/10.1371/journal.pone.0118432

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4510–4520). https://doi.org/10.1109/CVPR.2018.00474

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*, Article 60. https://doi.org/10.1186/s40537-019-0197-0

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.

Tangpukdee, N., Duangdee, C., Wilairatana, P., & Krudsood, S. (2009). Malaria diagnosis: A brief review. *The Korean Journal of Parasitology, 47*(2), 93–102. https://doi.org/10.3347/kjp.2009.47.2.93

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine, 25*(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

Wongsrichanalai, C., Barcus, M. J., Muth, S., Sutamihardja, A., & Wernsdorfer, W. H. (2007). A review of malaria diagnostic tools: Microscopy and rapid diagnostic test. *The American Journal of Tropical Medicine and Hygiene, 77*(6 Suppl), 119–127. https://doi.org/10.4269/ajtmh.2007.77.119

World Health Organization. (2025). *World malaria report 2025*. World Health Organization.

Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering, 2*(10), 719–731. https://doi.org/10.1038/s41551-018-0305-z