Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi
Bilecik Seyh Edebali University Journal of Science
10(2), 254-272, 2023
DOİ: 10.35193/bseufbd.1064187
e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

Araştırma Makalesi - Research Article

A Novel Deep Learning Approach to Malaria Disease Detection on Two Malaria Datasets

İki Sıtma Veri Kümesinde Sıtma Hastalığı Tespitine Yönelik Yeni Bir Derin Öğrenme Yaklaşımı

İbrahim Çetiner[1*], Halit Çetiner[2]

Geliş / Received: 27/01/2022 ***Revize / Revised: 19/03/2023*** ***Kabul / Accepted: 06/06/2023***

ABSTRACT

Malaria is a contagious febrile disease transmitted to humans by the bite of female mosquitoes. It is important to diagnose this disease in a short period of time. Finding the mathematically best numerical solution to a particular problem is the most important issue for most departments. In deep learning-based systems developed, the difference between the real data and the predicted result of the model is measured using loss functions. To minimize the error rate in the predictions during the training process of deep learning models, the weight values used in the model should be updated. This update process has a significant effect on the model prediction result. This article presents a new deep learning-based malaria detection method that will help diagnose malaria in a short time. A new 21-layer Convolutional Neural Network (CNN) model is designed and proposed to describe infected and uninfected thin red blood cell images. By using thin red blood cell sample images, 95% accuracy was achieved with Nadam and RMSprop optimization techniques. The results obtained show the efficiency of the proposed method according to each optimization algorithm.

Keywords-Malaria, CNN, Optimization Algorithms, Classification

ÖZ

Sıtma, dişi sivrisineklerin ısırmasıyla insanlara bulaşan bulaşıcı bir ateşli hastalıktır. Bu hastalığın kısa sürede teşhis edilmesi önemlidir. Belirli bir probleme matematiksel olarak en iyi sayısal çözümü bulmak çoğu bölüm için en önemli konudur. Bu mantıkla geliştirilen derin öğrenme tabanlı sistemde, gerçek veriler ile modelin tahmin edilen sonucu arasındaki fark, kayıp fonksiyonları kullanılarak ölçülmektedir. Derin öğrenme modelinin eğitim sürecinde tahminlerdeki hata oranını en aza indirmek için modelde kullanılan ağırlık değerlerinin güncellenmesi gerekmektedir. Yapılan güncelleme işlemi, modelin tahmin sonucu üzerinde önemli bir etkiye sahiptir. Bu makale, sıtmayı kısa sürede teşhis etmeye yardımcı olacak yeni bir derin öğrenme tabanlı sıtma sınıflandırma yöntemi sunmaktadır. Bu amaçla, 21 katmanlı yeni bir Konvolüsyonel Sinir Ağı (CNN) modeli önerilmiştir. Önerilen bu model enfekte ve enfekte olmayan ince kırmızı kan hücresi görüntülerini sınıflandırmak için tasarlanmıştır. İnce kırmızı kan hücresi numunesi görüntüleri kullanılarak, Nadam ve RMSprop optimizasyon teknikleri ile %95 doğruluk elde edilmiştir. Elde edilen sonuçlar, önerilen yöntemin her bir optimizasyon algoritmasına göre etkinliğini göstermektedir.

Anahtar Kelimeler- Sıtma, CNN, Optimizasyon Algoritmaları, Sınıflandırma

[1]*Corresponding Author Contact: icetiner@mehmetakif.edu.tr (https://orcid.org/0000-0002-1635-6461)
Mechatronic, Burdur Mehmet Akif Ersoy University, Vocational School of Technical Sciences, Burdur, Turkey
[2]Contact: halitcetiner@isparta.edu.tr (https://orcid.org/0000-0001-7794-2555)
Computer Technology, Isparta University of Applied Sciences, Vocational School of Technical Sciences, Isparta, Turkey

I. INTRODUCTION

Malaria is a disease that occurs with the effect of the Plasmodium gene genus, threatening human health and causing death [1]. Malaria infection is transmitted to humans by a female Anopheles mosquito infected with Plasmodium parasites. When this mosquito bites a person, it injects the infected Plasmodium parasites into the person it bites. After this injection procedure, red blood cells are affected as a result of a parasitic chromosome of the genus Plasmodium [2]. There are also different species of the Plasmodium genus such as P.falciparum, P.vivax, P.malariae, P.ovale and P.knowlesi [3]. To prevent malaria, it is essential to detect the disease early before it progresses too much. If detected early, it is possible to respond to treatment and prevent it.

According to studies conducted by the World Health Organization (WHO) [4], 69,000 deaths occurred worldwide between 2019 and 2020 due to malaria infection. Furthermore, 14 million malaria cases have been reported in the last year. Diagnosis of malaria, which is threatened by WHO, can be made using various methods. In one of these methods, a visual examination of a thin and thick blood smear is performed with a microscope. The diagnosis method obtained at the end of the specified examination is called the traditional method [3]. Diagnosis with conventional microscopy, which is used in the classical method, is a time-consuming method. In addition, the accuracy of the diagnosis is largely dependent on the specialist using the microscope [5, 6]. Light microscope is used in another diagnostic method. Using the specified microscope to detect different types of malaria disease is called the gold standard method. In this method, thin and thick Giemsa stained blood smears can be examined by magnifying the image tens of times, such as 10x or 100x, with the ocular lens magnification feature of the microscope [7].

Blood smear reports state that any person can be affected by more than one malaria infection at the same time. Plasmodium genus has different species, P.falciparum, P.vivax, P.malariae, P.ovale and P.knowlesi, which requires thorough examination of blood smears in malaria screening [8]. From blood smears used for malaria parasite detection, 6050 nonoverlapping fields of view occur for a 2x1 cm sample at 1000 times magnification [9]. Due to these reasons, the diagnosis of malaria takes a lot of time. In a voluminous sample that includes different species, it is stated by the experts who examine the sample that making a diagnosis is an extremely tiring task that can cause distraction [10]. In addition, malaria cases are affected by changes in weather conditions. As a result of these changes, malaria, which is contagious, can increase rapidly. New systems are needed to accelerate diagnosis, as well as enough specialist personnel to prevent the spread of cases and ensure rapid control [11]. Delays in diagnosis can increase the severity of the disease and cause great cost to society.

Many researchers in the literature use classical machine learning techniques consisting of image segmentation, feature extraction, and classification steps [12,13]. The malaria diagnosis method based on computer vision consists of steps such as pre-processing, segmentation of objects in the image, feature extraction, and classification [14]. The most difficult process in diagnosing malaria based on the classical machine learning technique and computer vision is stated to be the feature extraction step. In this step, it is emphasized that it is important to determine the regions containing disease symptoms on the microscopic image and to determine an appropriate algorithm to map their features [15]. Identifying significant, distinctive features from the relevant region in the image directly affects the classification accuracy. Due to the difficulties in diagnosing malaria by observation, researchers have turned to develop new methods. As a result of these orientations, a research field called computer-aided detection (CAD) has been created that facilitates medical imaging of malaria cells. The main purpose of the CAD imaging systems specified is to provide malaria specialists with important ideas for the early and accurate diagnosis of malaria cells [10]. In recent studies on malaria, it is stated that models that can learn through experience like humans are deep learning models [9]. At this point, CNN-based methods are a deep learning approach that can learn through experience. In this article, unlike the deep learning models in the literature, the best combination of deep learning layers and thousands of layers and models that develop based on deep learning architectures such as ImageNet, which consists of millions of pre-trained data, have been obtained as successful as the models [16–19]. At the same time, the results obtained as a result of training the proposed deep learning model with different optimization algorithms are shared. Vijayalakshmi and Rajesh made changes to the upper layers by freezing the lower layers of the VGG19 pre-trained deep learning model [9]. In the work carried out, the features extracted with the VGG19 architecture were classified with the SVM algorithm. A success rate of 93.1% has been achieved in the classification of malaria disease.

To overcome the above-mentioned difficulties, the convolutional layer model, which is very popular recently and can be used to extract the features of malaria, is proposed. The convolution layer is one of the basic layers of Convolutional Neural Network (CNN) architectural models. Through this layer, it is possible to use windows in malaria images to allow filters of different window sizes to navigate over the image. As a result of the specified navigations, feature maps with distinctive features are created.

The main contributions to the literature in this article are given below.

- A CAD system is proposed that can accurately diagnose malaria cells at an early stage, using the powerful feature extraction capabilities of convolutional layers.
- This CAD system consists of a deep learning model proposed to assist experts.
- With the proposed deep learning model has been presented, a model that assists experts in performing automatic classification of malaria disease.
- With the proposed model, it is aimed to reduce the costs of diagnosis by reducing the difficulties experienced in the diagnosis of malaria.
- To fully evaluate performance result of the proposed model, training and testing processes were performed with optimization methods named Adam, Adamax, SGD, RMSprop, Adadelta, Adagrad, Nadam, and Ftrl, and the results were given.
- The CNN model proposed in this article is as successful as the models that develop based on deep learning architectures that consist of thousands of layers and millions of pre-trained data such as ImageNet.

The next part of the article consists of three sections. In Section 2, information is given about the data set and performance measurement techniques used in the article. In Section 3, the proposed deep learning model for malaria disease classification and the results obtained from different optimization algorithms are presented. In the last section, the net result obtained as a result of the study analysis is presented.

II. MATERIAL AND METHOD

A. Material

This article uses two datasets of infected and non-infected red blood images of general malaria. The first dataset was created by researchers working at the Lister Hill National Biomedical Communication Center contact center [20, 21]. Collected blood smears were viewed with the help of the built-in camera of an Android phone at Chittagong Medical Hospital in Bangladesh [22]. The second data set was obtained as a result of manual segmentation from cell images of plasmodium falciparum, plasmodium malariae, plasmodium ovale, plasmodium vivax malaria species [23]. Both datasets have a total of 30,000 images.

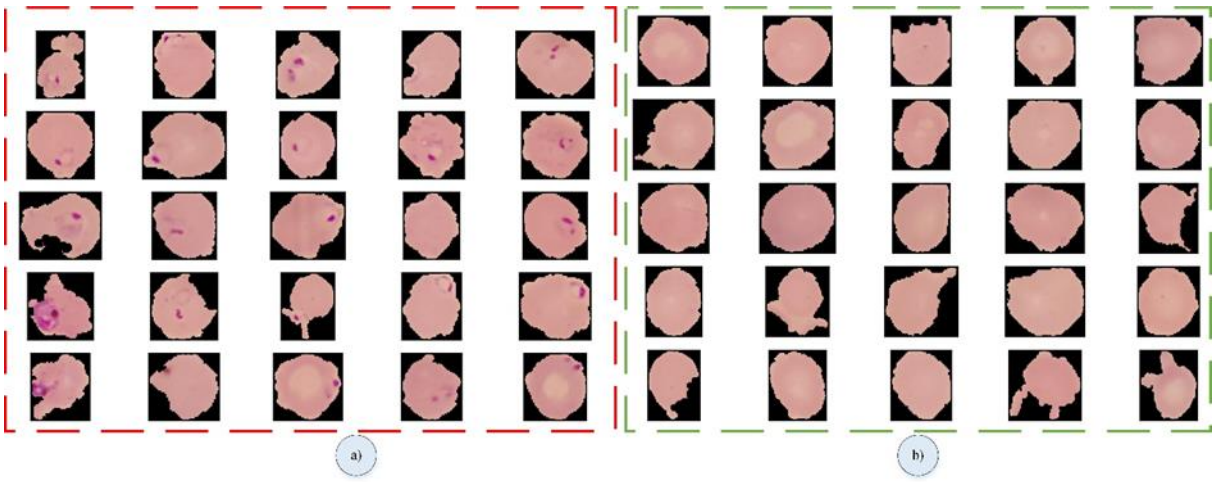

Figure 1. Number of red thin blood cell images in the data set by type: a) Parasitized, b) Uninfected

The numerical distributions of the blood cells in the data set used to evaluate the training and test results with the different optimization methods of the proposed model are shown in Figure 1. The data set used consists of two different classes, infected and uninfected, as shown in Figure 1. Class distribution numbers are the same. There is no data imbalance. The images in the data set can be given as direct input to the proposed model or by passing some pre-processing. Some pre-processing has been applied on the image due to its positive contribution to the performance of the model. The applied pre-processes consist of resizing and data duplication steps, respectively. In the resizing step, all training and test images were resized to 64x64. In the data augmentation step, data augmentation was performed using steps such as enlarging, rotating and zooming on the image. With these processes, the aim is to classify the malaria images of the proposed model from all angles and distances.

B. Hyperparameter Settings

The set of parameters that will affect the training and test results with different optimization techniques of the proposed CNN model within the scope of this study is called hyperparameter. These parameters are regulatory parameters such as image size, number of steps, number of sections, learning rate, and optimization method. As a result of different experimental studies, the images used in all model training and testing processes were determined as 64x64 with three channels, the number of epochs 50, and the number of batch sizes 12. The epoch value was determined according to the fact that the accuracy value did not increase and remained constant. The batch size value is set to 12, indicating the number of images processed before updating the model. The accuracy values obtained as a result of testing different learning rate and batch size values are presented in Table 1. The highest accuracy value of 91% was achieved in the test results performed with batch size 12 and learning rate 1e-1 values. The highest accuracy values of 89% were achieved in 64 batch size and 1e-5 learning rate values. For the stated reasons, experimental studies were carried out by choosing 1e-5 as the learning rate and 12 as the batch size value. Compared to the Vijayalakshmi and Rajesh [9] transfer learning based study, different performance results are obtained in the CNN or transfer learning based models. At the same time, when a model with different hyperparameter values is tested, changes in performance results occur.

Table 1. Experimentally tested hyperparameters and results

Batch size	Learning rate	Optimization Techniques	Type	Accuracy
12	1e-1	Nadam	Parasitized	0.91
12	1e-1	Nadam	Uninfected	0.91
12	1e-1	RMSprop	Parasitized	0.91
12	1e-1	RMSprop	Uninfected	0.91
12	1e-1	Adam	Parasitized	0.90
12	1e-1	Adam	Uninfected	0.90
64	1e-5	Nadam	Parasitized	0.89
64	1e-5	Nadam	Uninfected	0.89
64	1e-5	RMSprop	Parasitized	0.89
64	1e-5	RMSprop	Uninfected	0.88
64	1e-5	Adam	Parasitized	0.88
64	1e-5	Adam	Uninfected	0.88

In addition to these, training processes were carried out with eight different optimization methods to determine the effect of optimization methods on the performance of the proposed model. The optimization methods used are Adam, Adamax, SGD, RMSprop, Adadelta, Adagrad, Nadam and Ftrl, respectively. Training and testing results of different hyperparameters and optimization techniques are presented in Section 3. As a method, a proposed CNN-based model provides the automatic extraction of distinctive features. In classical machine learning algorithms, subjective and meaningless features cause low classification accuracy. For this reason, the CNN-based method, which provides reliable results in method selection, was preferred.

C. Performance Evaluation Criteria

The machine used in the experimental studies was a G15 5515 Amd Ryzen 7 5800H 16GB 512GB SSD RTX3060 Windows 10 Home 15.6 FHD Portable Dell computer and the images were processed in Graphics Processing Unit (GPU) mode. Implemented in CUDA version 11.4 with Keras libraries using Tensorflow-GPU. Accuracy, precision, F1 score, and recall measurement formulas were used to evaluate the performance of the proposed method [24].

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

In Equation 1 and beyond, TP defines the number of malaria blood cell samples that can be correctly classified, while FP defines the number of malaria blood cell samples that are misclassified. TN, on the other hand, defines the number of malaria blood cell samples correctly classified from other classes. FN represents the number of malaria blood cell samples misclassified from other classes under observation. The ratio of the malaria blood cell class classified as positive in Equation 2 and correctly predicted to the total number of positive malaria blood cells is called recall. The recall formula identifies the correct classes that the model finds. The higher the recall, the higher the number of correctly classified malaria blood cells. The malaria blood cell class found incorrectly in the recall equation is not of interest. However, recall results are required to understand whether all positive malaria blood cells in the model are classified correctly.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision is defined as the ratio of TP to the total number of malaria blood cells predicted as TP and positive. If some of the negative malaria blood cells are classified as positive, it is necessary to obtain a result from the precision equation to determine this. The precision formula is given in Equation 3.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

The F1 score is the weighted harmonic mean of the results obtained from the recall and precision equations. The equation presented in Equation 4 is used to measure the balance between recall and precision. If the F1 score is close to one, it means that a small number of low errors have formed a model.

$$F1 = 2x\frac{PrecisionxRecall}{Precision + Recall} \tag{4}$$

In this study, all specified performance criteria were used to correctly evaluate the performance of the proposed methods. In addition to the performance measurements used, the loss and accuracy results are given graphically. In addition, the confusion matrix obtained from the models is presented under the heading of each optimization technique. The results of each optimization technique obtained from the CNN model are given according to the equations between Equations 1-4.

D. The Proposed CNN Model for Malaria Disease Classification

Today, deep learning, which is gaining popularity in the field of image classification, can offer solutions to different problems. In this article, a system for the classification of malaria disease, which causes thousands of people to die inexorably every year, is proposed. The proposed system consists of 21 layers. Layers are designed to have the specified properties one after the other.

Figure 2. Proposed CNN model

Basic preprocessing is passed before the image is input to the proposed model. Most importantly, all of the training and test images should be brought to 64x64 width and height dimensions. In this way, the image dimensions are brought to the same format. As all images have three color channels, the pixel values of each image are divided by 255 values, enabling fast processing. After these basic pre-processes, the image is presented as an input to the proposed model shown in Figure 2 for training operations. Images with width and height of 64x64 are taken as input to the model in the first layer. In the second layer, a convolution layer with 24 filters and 5x5 window size ReLU activation function has been added. In the third layer, a second convolution layer with the features in the second layer is subsequently defined. The default maximum pooling layer has been added in the fourth layer. In the fifth and sixth layers, two interconnected layers with 32 filters with SeLU activation function in 5x5 window sizes are defined. In the seventh layer, the Batch Normalization layer has been added, which performs inter-layer normalization. The default maximum pooling layer has been added in the eighth layer. In the ninth layer, the dropout layer, which performs 0.5 neuron dropout, is defined to prevent overfitting. In the tenth and eleventh layers, two interconnected layers with 256 filters with SeLU activation function in 5x5 dimensions are defined. In the twelfth layer, Batch Normalization, which performs normalization between layers, is applied. The maximum pooling layer defined by default in the thirteenth layer has been implemented. In the fourteenth layer, the feature map obtained from the convolution layers is transformed into a one-dimensional vector. In the fifteenth layer, a dense layer with 512 neurons SeLU activation function has been added. In the sixteenth layer, the Batch Normalization layer is defined, which performs the inter-layer normalization. In the seventeenth layer, a dense layer with 128 neurons SeLU activation function is defined. With this layer, neurons are strengthened. Batch Normalization, which provides normalization between layers, is defined in the eighteenth layer. In the nineteenth

layer, the dropout layer, which provides a loss of 0.5 neurons, was applied. In the twentieth layer, the fully connected layer is defined, which prepares the model for the learning process. In the twenty-first layer, a classification layer with a two-output sigmoid activation function is defined.

In Figure 2, the layers, layer order and parameters in the proposed model are presented in detail. The parameters used were determined as a result of experimental studies. Convolution and dense layers, which are given as layers, have a high effect on obtaining results at a level that can compete with the studies in the literature.

III. RESULTS AND DISCUSSION

Finding the best numerical solution to a mathematical problem is the most important task in most departments. Mathematically, loss functions are used to measure the inaccuracy of our estimates. This loss function has an important role in the decision about fine-tuning and changing the parameters (weights) of our model to minimize the training process of the models. This has a significant effect on getting our estimate closer to the accuracy. In this sense, the optimizers deal with the weights for the loss function and ensure that our model is formatted in the most accurate way. Optimizers is an algorithm that checks for loss functions. The selection of the optimization algorithm is one of the most important design decisions in deep learning, and making this decision is very important to improve learning performance. In this sense, with the proposed model, which of the different optimization methods will lose the least has been obtained by experimental studies. During the experimental studies, the optimization methods Adam, Adamax, SGD, RMSprop, Adadelta, Adagrad, Nadam, and Ftrl with the proposed model were tested, respectively. Except for the optimization method, other hyperparameters remain constant. Therefore, when comparing model results, the difference between Adam, Adamax, SGD, RMSprop, Adadelta, Adagrad, Nadam, and Ftrl is mentioned in detail. In this section, the obtained results from each optimization method and the proposed model are presented. The optimization methods used and the comparisons of the proposed model among themselves and the comparison results according to different studies using the same data set in the literature are given.

The results of the training performance obtained in the optimization methods used are shown in Figure 3. When Figure 3 is examined, it is determined that the results of each optimization method are different from each other. The training accuracy graphs of the optimization methods are given in Figure 3a and the loss graphs in Figure 3b. According to these graphs, it has been seen that Nadam, Adam and RMSprop are more successful than other optimization methods. The success of Adamax is lower than those of Nadam, Adam and RMSprop. Since the RMSprop optimization method is more successful in nonparasitic cells than Adamax, the accuracy rate is higher.

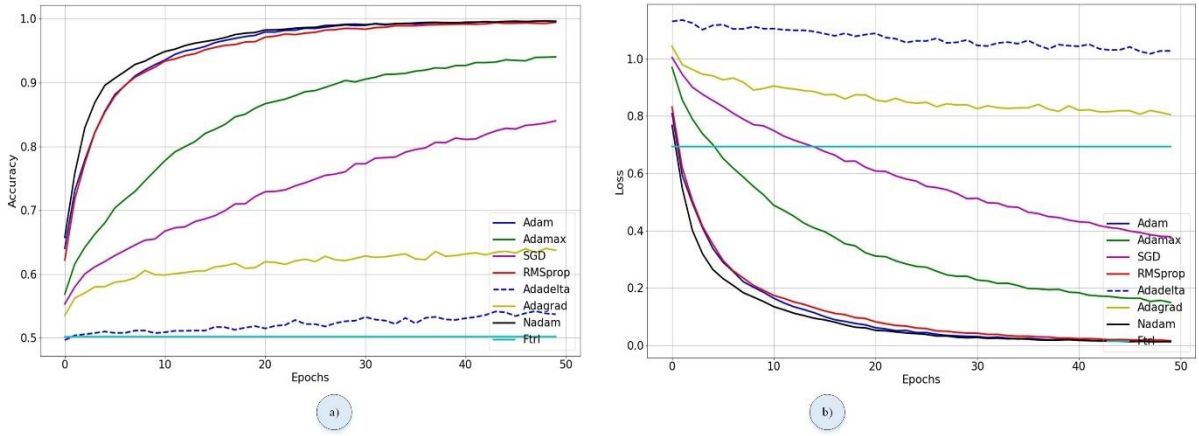

Figure 3. Training graphs of the proposed CNN model with optimization techniques a) accuracy and b) loss

In Figure 4, the obtained test results with the optimization methods of the proposed CNN model are shown. Taking into account the test results, there are differences between each optimization method. The test results have given a parallel result to the training results.

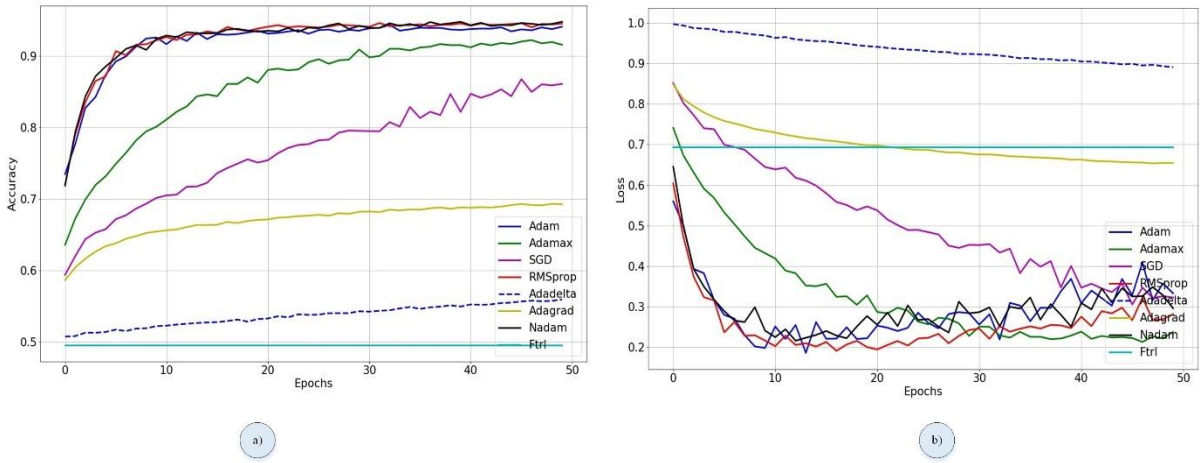

Figure 4. Test graphs of the proposed CNN model with optimization techniques a) accuracy and b) loss

Figure 3 and Figure 4 show the accuracy and loss graphs for the training and testing processes of the proposed model in Figure 2. In addition to these, confusion matrices were created in order to examine the performance results of test data not used in training in more detail. These generated matrices are shown in Figure 5 to Figure 12. In the rest of this section, the detailed results of the Adagrad, RMSprop, Adadelta, SGD, Ftrl, Adam, Adamax, and Nadam optimization methods are given, whose training and test results are given above.

A. AdaGrad

In the AdaGrad optimization method, the sums of the partial derivatives of the parameters are taken into account. These sums of partial derivatives are values proportional to the mean square slope [25]. A_i, i. represented as the sum of the parameterized values. Weight values are updated as in Equation 6 in each operation.

$$A_i \Leftarrow A_i + \left(\frac{\partial L}{\partial w_i}\right)^2 \forall i \tag{5}$$

$$w_i \Leftarrow w_i - \frac{\alpha}{\sqrt{A_i}}\left(\frac{\partial L}{\partial w_i}\right); \quad \forall i \tag{6}$$

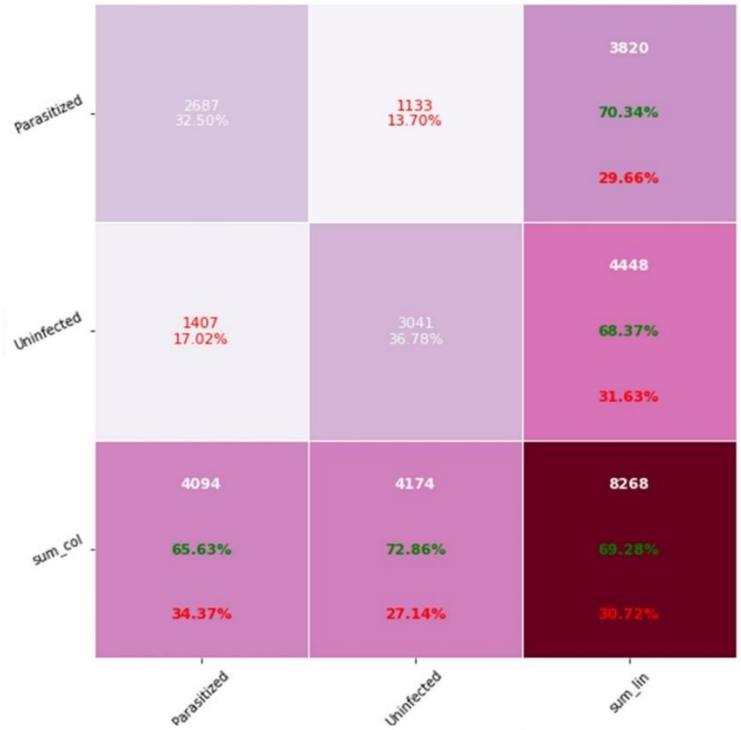

Figure 5. Adagrad optimization technique confusion matrix of the proposed CNN method

The obtained performance results with the Adagrad optimization technique used in training the proposed CNN method to detect the parasite-infected cells are shown in Figure 5. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 70%, while cells without parasites are classified with a success rate of 68%. With the Adagrad optimization technique, 1133 of 13,779 cells with noise in 27,558 images are misclassified. In this case, the loss rate of the proposed system in this technique is determined to be 13.70%. Out of 13,779 parasite-free cells, 1407 were misclassified by the adagrad-optimized model. The loss rate is determined at 17.02%.

B. RMSprop

The RMSprop optimization method is similar to the Adagrad optimization method in that it uses the A_i value, which indicates the total magnitude of the gradients. Unlike the AdaGrad optimization method, it uses the exponential mean instead of estimating the A_i values. The RMSprop method, which is one of the optimization methods commonly used in deep learning models, t. weights the square of the partial derivatives that occur before the update by the p^t value. It uses the $\rho \in (0,1)$ bias value to avoid possible error situations. The update of the A_i value showing the total magnitude of the gradients is given in Equation 7 [25].

$$A_i \Leftarrow \rho A_i + (1-\rho)\left(\frac{\partial L}{\partial w_i}\right)^2 \forall i \tag{7}$$

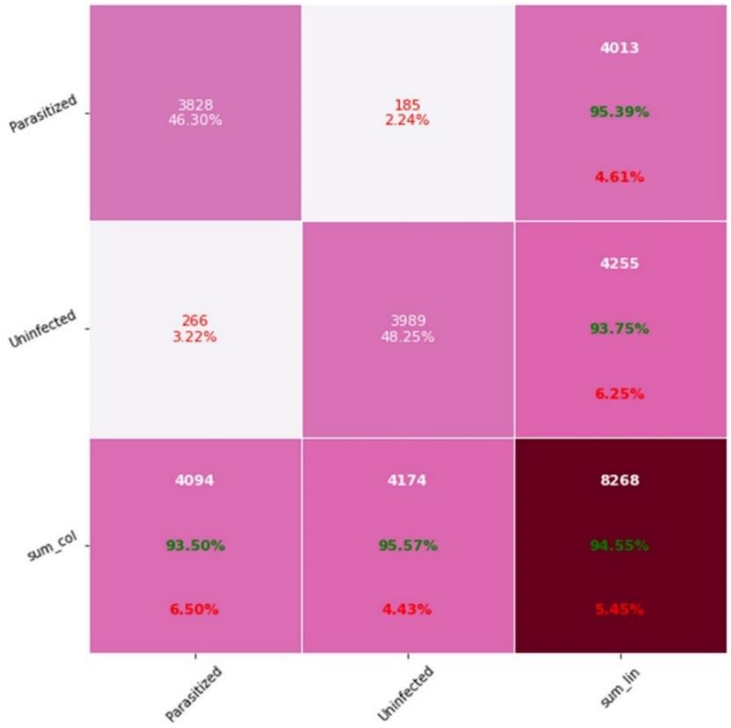

Figure 6. RMSprop optimization technique confusion matrix of proposed CNN method,

The performance results obtained with the RMSprop optimization technique used in training the proposed CNN method to detect the parasite-infected cells are shown in Figure 6. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 95%, while cells without parasites are classified with a success rate of 94%. With the RMSprop optimization technique, 185 of 13,779 cells with noise in 27,558 images are misclassified. In this case, the loss rate of the proposed system using this technique was determined at 2.24%. Out of 13,779 parasite-free cells, 266 were misclassified with the RMSprop optimized model. The loss rate is determined as 3.22%.

C. AdaDelta

AdaDelta optimization method is similar to RMSprop optimization algorithm. Unlike the RMSprop algorithm, it does not need learning parameters [25]. Equation 9 shows the weight update of the RMSprop algorithm.

$$w_i \Leftarrow w_i - \underbrace{\frac{\alpha}{\sqrt{A_i}}\left(\frac{\partial L}{\partial w_i}\right)}_{\Delta w_i} \tag{9}$$

The α parameter changes depending on the updates. The increase in w_i creates the difference value Δw_i. This process occurs with every weight update. Equation 10 holds the exponentially corrected δ_i value of the ρ parameter and the difference weight value Δw_i. These operations are as in A_i.

$$\delta_i \Leftarrow \rho\delta_i + (1-\rho)(\Delta w_i)^2 \tag{10}$$

Calculating the δ_i value allows new iterations to be made. The value of A_i can be calculated using the partial derivative in the iteration process. The difference in the calculations of A_i and δ_i creates Equation 11.

$$w_i \Leftarrow w_i - \underbrace{\sqrt{\frac{\delta_i}{A_i}}\left(\frac{\partial L}{\partial w_i}\right)}_{\Delta w_i} \quad \forall i \tag{11}$$

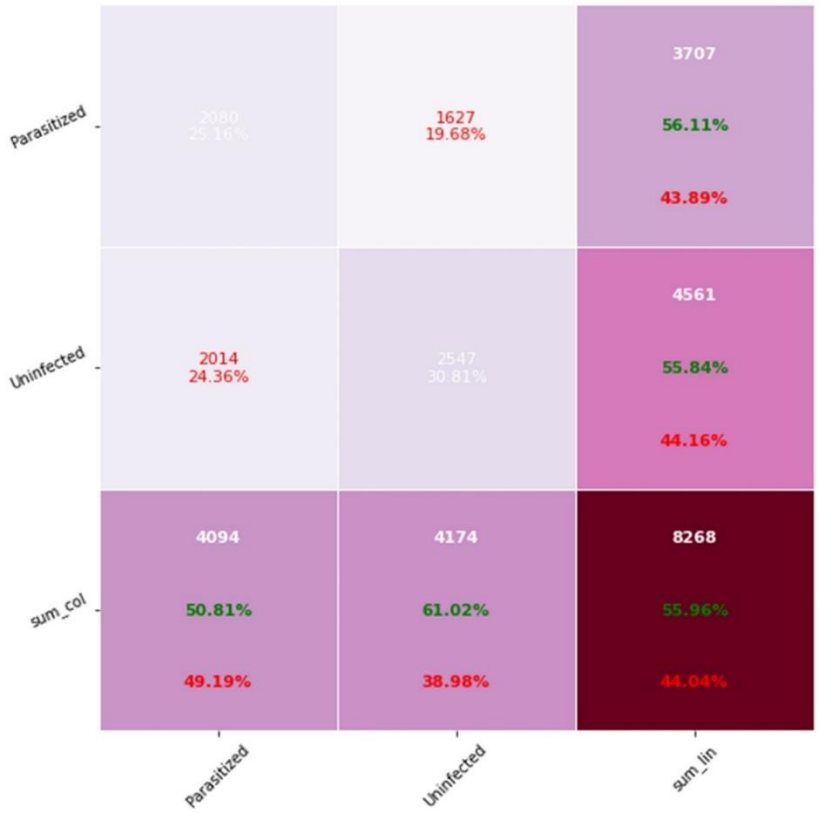

Figure 7. Adadelta optimization technique confusion matrix of proposed CNN method

The performance results obtained with the Adadelta optimization technique used in training the proposed CNN method to detect the parasite-infected cells are shown in Figure 7. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 56.11%, while cells without parasites are classified with a success rate of 55.84%. With Adadelta optimization technique, 1627 of 13,779 cells with noise in 27,558 images are misclassified. In this case, the loss rate of the proposed system in this technique is determined to be 19.68%. Out of 13,779 parasite-free cells, 2014 were misclassified with the AdaDelta optimized model. The loss rate is determined at 24.36%.

D. SGD

The objective function used in deep learning is the average of the loss functions for each sample in the training dataset [26]. According to the equation in Equation 12, given a data set with n samples, the parameter vector x is i. with respect to the training example $f_i(x)$ represents the loss function. With this equation, the objective function $f(x)$ is calculated in Equation 12.

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \tag{12}$$

The gradient of the objective function in the x parameter vector can be calculated by Equation 13.

$$\nabla f(x) = \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(x) \tag{13}$$

Stochastic gradient descent (SGD) reduces the cost of gradient descent at each iteration, which increases as the training data set grows. In order to update the x value in Equation 14, the difference gradient value $\nabla f_i(x)$ is calculated.

$$x \leftarrow x - \eta \nabla f_i(x) \tag{14}$$

The parameter η defined in Equation 14 is the learning rate. In this equation, it can be seen that the cost value decreases with each iteration. In Equation 15, a stochastic gradient calculation is performed for estimation of $\nabla f(x)$ value. The stochastic gradient value is function $\nabla f_i(x)$.

$$\mathbb{E}_i \nabla f_i(x) = \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(x) = \nabla f_i(x) \tag{15}$$

Optimization happens everywhere. Machine learning is an example, and gradient descent is probably the most famous algorithm for performing optimization. Optimization means finding the best value of a function or model. This can be a maximum or minimum according to some criteria [25]. SGD is an optimization algorithm that follows the negative gradient of an objective function to find the minimum of the function used.

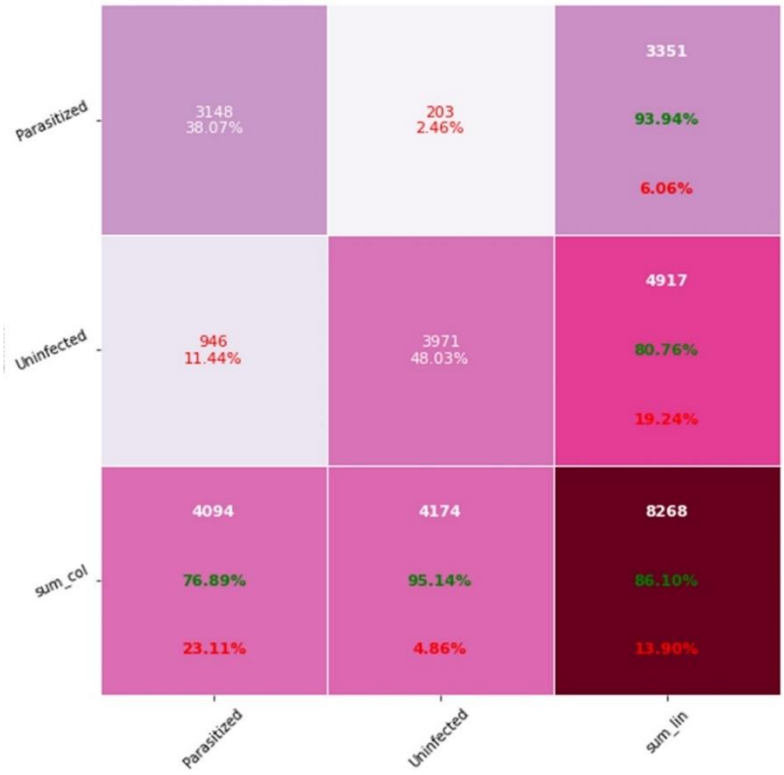

Figure 8. SGD optimization technique confusion matrix of proposed CNN method

The performance results obtained with the SGD optimization technique used in training the proposed CNN method to detect the parasite-infected cells are shown in Figure 8. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 94%, while cells without parasites are classified with a success rate of 81%. With the SGD optimization technique, 203 of 13,779 cells with noise in 27,558 images are misclassified. In this case, the loss rate of the proposed system in this technique is determined to be 2.46%. Out of 13,779 parasite-free cells, 946 were misclassified with the SGD optimized model. The loss rate is determined as 11.44%.

E. Ftrl

Ftrl or Ftrl-proximally [27] is short for follow the (proximally) regularized leader. The algorithm was used by [18] due to its ability to handle larger data sets and larger models. The aim of the algorithm is to get both sparsity and improved accuracy on predictions from the standard SGD, and has shown to do so with minimum computing resources. Ftrl is SGD with regularization and is defined as [28].

$$\theta_{t+1,i} = \arg\min(g_{1:t,i}\theta_i + t\lambda \parallel \theta_i \parallel_1 + \frac{1}{2}\sum_{s=1}^{t} \sigma_s \parallel \theta_i - \theta_{s,i} \parallel_2^2) \quad (16)$$

The first term $(g_{1:t,i}\theta_i)$ of the update is an approximation to the objection function, second term $(t\lambda \parallel \theta_i \parallel_1)$ is a non-smooth convex function and the last term $(\frac{1}{2}\sum_{s=1}^{t} \sigma_s \parallel \theta_i - \theta_{s,i} \parallel_2^2)$ is an additional strong convexity. Here, $\sigma_s^t = \frac{1}{\eta_t}$ and are generalized learning rates, while $\lambda > 0$ induces sparsity in the way zero features and less important features are removed.

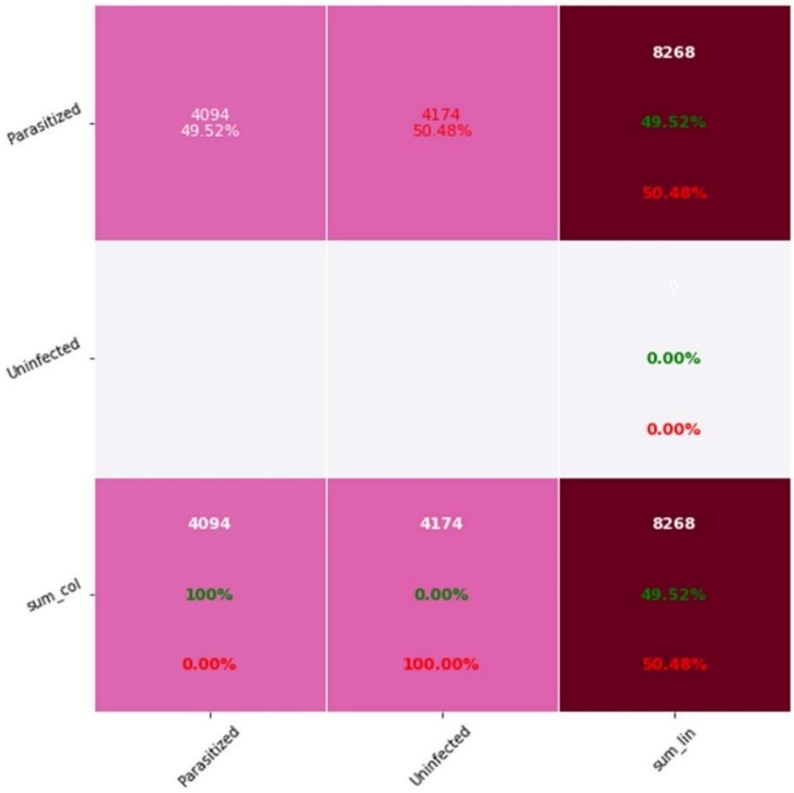

Figure 9. Ftrl optimization technique confusion matrix of the proposed CNN method

The performance results obtained with the Ftrl optimization technique used in training the proposed CNN method to detect the parasite-infected cell are shown in Figure 9. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 50%, while cells without parasites were classified with a rate of 00%. The Ftrl optimization method did not study in the classification of parasite-free cells. With the Ftrl optimization technique, 4174 of 13,779 cells with noise in 27,558 images were misclassified. In this case, the loss rate of the proposed system in this technique was determined to be 50.48%. Out of 13,779 parasite-free cells, 13,779 were misclassified with the Ftrl optimized model. The loss rate was determined to be 100%.

F. Adam

Adam optimization method uses noise normalization similar to AdaGrad and RMSProp algorithms [25]. Flattens the first-order gradient to incorporate momentum into the update. The smoothed error is addressed when the smoothed value is initialized to 0. In Equation 17, the value of A_i is the i of the w_i parameter. is the exponentially average value of the index. This value is updated with a reduction value of $\rho \in (0,1)$.

$$A_i \Leftarrow \rho A_i + (1-\rho)\left(\frac{\partial L}{\partial w_i}\right)^2 \quad (17)$$

In Equation 18, the reduction parameter is denoted by ρf, while F_i represents the exponentially smoothed value.

$$F_i \Leftarrow \rho f F_i + (1-\rho f)\left(\frac{\partial L}{\partial w_i}\right) \quad (18)$$

In Equation 19, the α_t parameter is used as the learning rate in the t th iteration. Initial values of symbols F_i and A_i defined in Equation 19 are assigned as 0. The specified assignment operation causes errors in iterations. It uses $\sqrt{A_{i+\epsilon}}$ instead of $\sqrt{A_i}$ to increase the power of weight update and reduce the margin of error.

$$w_i \Leftarrow w_i - \frac{\alpha_t}{\sqrt{A_i}} F_i \tag{19}$$

The Adam method has two different features from the RMSprop optimization method. The first is that incorporating the gradient into the momentum uses the smoothed value. The second is the use of the symbol α_t, which represents the learning rate. This usage is illustrated in Equation 20.

$$\alpha_t \Leftarrow \alpha \underbrace{\left(\frac{\sqrt{1-\rho^t}}{1-\rho^t f}\right)}_{Adjust\ Bias} \tag{20}$$

It represents the error correction factor α_t, which is defined in detail in Equation 20. The Adam optimization method has many of the useful advantages of many of the other optimization methods.

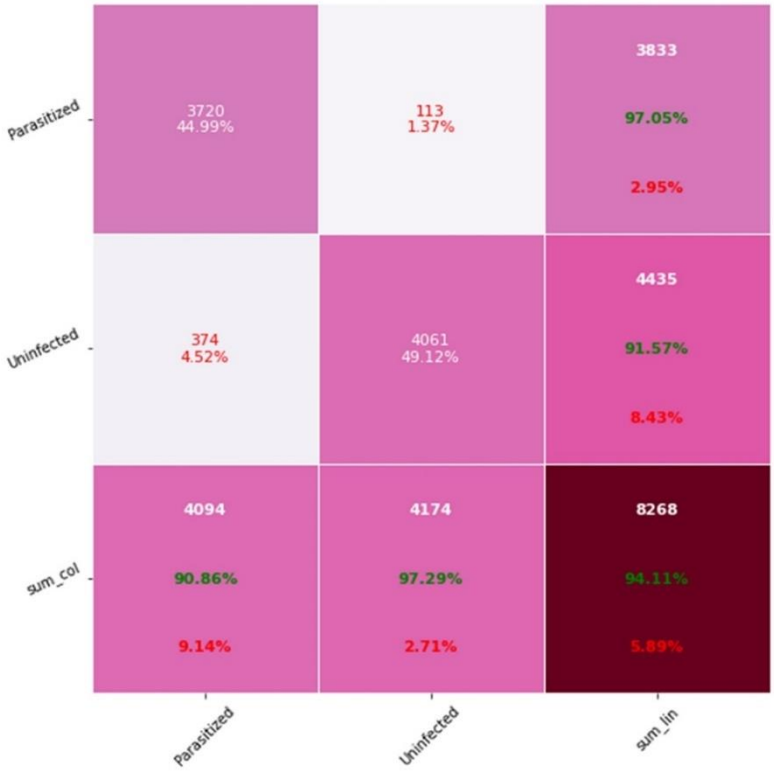

Figure 10. Adam optimization technique confusion matrix of the proposed CNN method

The performance results obtained with the Adam optimization technique used to train the proposed CNN method to detect the parasite-infected cell are shown in Figure 10. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 97%, while cells without parasites are classified with a success rate of 91%. With the Adam optimization technique, 113 of 13,779 cells with noise in 27,558 images are misclassified. In this case, the loss rate of the proposed system in this technique is determined as 1.37%. Of the 13,779 parasite-free cells, 374 are misclassified. Out of 13,779 parasite-free cells, 374 were misclassified with the Adam optimized model. The loss rate is determined to be 4.52%.

G. Adamax

The Adamax optimization method is a variation of the Adam optimization method, which has update rules to weight the models. Proportional scales the L^p norm of gradients from the current and previous networks [29]. Equation 21 shows the weight update according to the Adamax optimization method. The $\hat{m}$, $\hat{v}_t, u_t$ in this equation represent the moment estimate, error corrected moment estimate, and the second raw moment estimate, respectively [30]. $\beta_1 = 0.9$ and $\beta_2 = 0.9999$ denote the reduction ratios, while α denotes the step size. The w value represents the weights at t and $t+1$ steps.

$$w_t = w_{t-1} - \frac{\alpha}{1-\beta_1^t} x \frac{m_t}{\max\left(\beta_2 u_{t-1}, \left|g_t\right|\right)}, \tag{21}$$

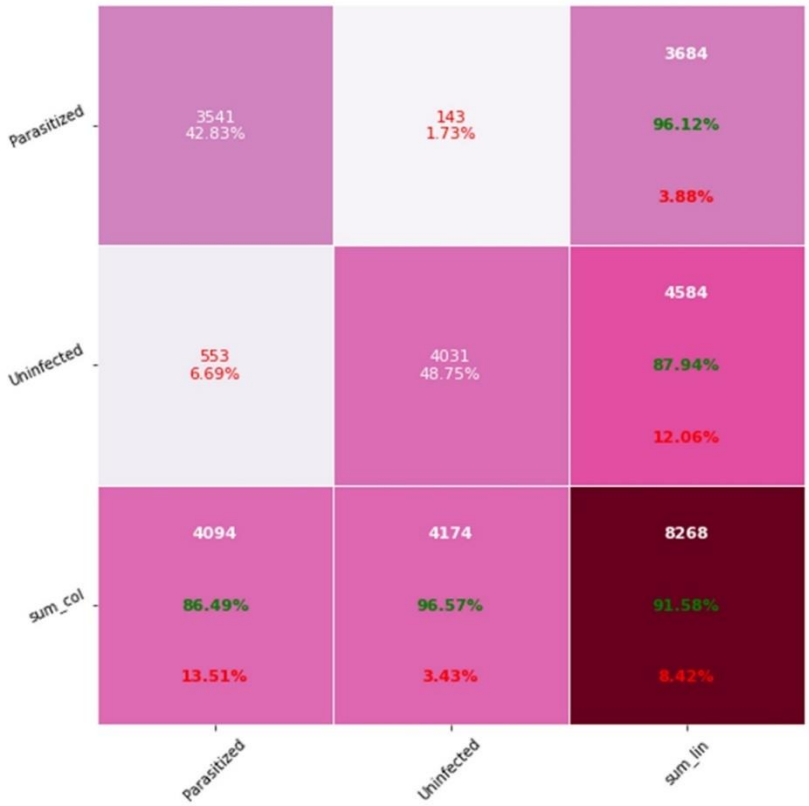

Figure 11. Adamax optimization technique confusion matrix of the proposed CNN method

The performance results obtained with the Adamax optimization technique used in training the proposed CNN method to detect the parasite-infected cells are shown in Figure 11. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 96%, while cells without parasites were classified with a success rate of 91%. With the Adamax optimization technique, 143 of 13,779 cells with noise in 27,558 images are misclassified. In this case, the loss rate of the proposed system in this technique is determined to be 1.73%. Out of 13,779 parasite-free cells, 553 were misclassified with the Adamax optimized model. The loss rate is determined to be 6.69%.

H. Nadam

Nesterov-Accelerated Adaptive Moment Estimation (Nadam) is calculated by integrating adaptive moment estimation into the optimization method Adam optimization method calculation formula [29]. It ensures that the specified insertion gradient value achieves high accuracy. Equation 22 shows the weight updates of the Nadam optimization method. Between Equations 23-25, the expansions of the parameters in Equation 22 are shown.

$$w_t = w_{t-1} - \alpha x \frac{\overline{m_t}}{\sqrt{\widehat{v_t}} + \varepsilon}, \tag{22}$$

$$\overline{m_t} = \left(1 - \beta_{1,t}\right)\hat{g}_t + \beta_{1,t}\hat{m}_t, \tag{23}$$

$$\hat{m}_t = \frac{m_t}{1 - \prod_{i=1}^{t+1} \beta_{1i}}, \tag{24}$$

$$\hat{g}_t = \frac{g_t}{1 - \prod_{i=1}^{t+1} \beta_{1i}}, \tag{25}$$

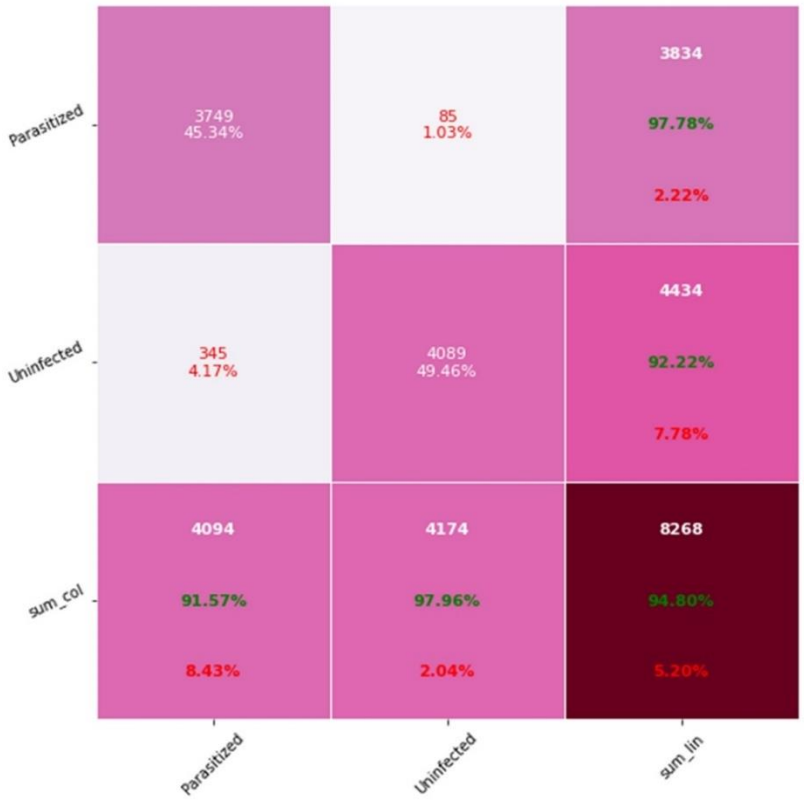

Figure 12. Nadam optimization technique confusion matrix of proposed CNN method

The performance results obtained with the Nadam optimization technique used in training the proposed CNN method to detect the parasite-infected cell are shown in Figure 12. According to the confusion matrix shown in this figure, parasitic cells could be classified with a success rate of 98%, while cells without parasites are classified with a success rate of 92%. With the Nadam optimization technique, 85 of 13,779 cells with noise in 27,558 images are misclassified. In this case, the loss rate of the proposed system in this technique is determined to be 1.03%. Out of 13,779 parasite-free cells, 345 were misclassified with the Nadam optimized model. The loss rate is determined as 4.17%.

The proposed CNN model has been tested with Adam, Adamax, SGD, RMSprop, Adadelta, Adagrad, Nadam and Ftrl optimization methods and their results are given above. Eight basic optimization methods were tested on the data set consisting of 13,779 images with and 13,779 cells without noise out of 27,558 images. The precision, recall, F1 score, and accuracy values are presented in Table 2 to correctly interpret the results.

Table 2. Proposed CNN performance results

Optimization Techniques	Type	Precision	Recall	F1 score	Accuracy
Nadam	Parasitized	0.98	0.92	0.95	0.95
Nadam	Uninfected	0.92	0.98	0.95	0.95
RMSprop	Parasitized	0.95	0.94	0.94	0.95
RMSprop	Uninfected	0.94	0.96	0.95	0.95
Adam	Parasitized	0.97	0.91	0.94	0.94
Adam	Uninfected	0.92	0.97	0.94	0.94
Adamax	Parasitized	0.96	0.86	0.91	0.92
Adamax	Uninfected	0.88	0.97	0.92	0.92
SGD	Parasitized	0.94	0.77	0.85	0.86
SGD	Uninfected	0.81	0.95	0.87	0.86
Adagrad	Parasitized	0.70	0.66	0.68	0.69
Adagrad	Uninfected	0.68	0.73	0.71	0.69
Adadelta	Parasitized	0.56	0.51	0.53	0.56
Adadelta	Uninfected	0.56	0.61	0.58	0.56
Ftrl	Parasitized	0.50	1.00	0.66	0.50
Ftrl	Uninfected	0.00	0.00	0.00	0.00

While Nadam gives better results in the parasitized class, the RMSprop method gives better results in the uninfected class.

IV. SUMMARY OF RESULTS

We evaluated the proposed CNN model with Adam, Adamax, SGD, RMSprop, Adadelta, Adagrad, Nadam and Ftrl optimization techniques.

Table 3. Loss of performance results of the proposed CNN model

Optimization Techniques	Parasitized	Uninfected	Total
Nadam	85	345	430
RMSprop	185	266	451
Adam	113	374	487
Adamax	143	553	696
SGD	203	946	1149
Adagrad	1133	1407	2540
Adadelta	1627	2014	3641
Ftrl	4174	13779	17953

The number of results of loss performance according to these optimization methods is given in Table 3. According to this table, it is seen that the Nadam optimization method has the least loss function in the malaria data set. The Ftrl optimizer could not work well. We examined 13.779 with parasitized and uninfected totally 27.558 images using the optimization methods. While a total of loss than 500 errors occurred in the Nadam, RMSprop, and Adam optimization methods, more errors have been received in the Adamax, SGD, Adagrad, Adadelta and Ftrl optimization methods. We observed that the optimization methods of Nadam, RMSprop, and Adam gave better results than other optimization methods in the proposed CNN model when evaluated in a publicly available dataset named Malaria using the same parameters in eight different optimization methods.

The results of the article up to this step were obtained using the first data set. The second dataset is used to show the robustness and flexibility of the model. No change in performance was observed. Apart from this article, two separate models based on ResNet50V2 and CNN were developed in the study [31] shown in Table 4. In this study, which was carried out differently from this developed model, a second data set was added in addition to a commonly used data set in the literature, as can be seen in Table 4. At the same time, it has been observed whether there will be a change in the results of the proposed CNN model despite the addition of the data set. In this study, a robust model has been proposed that will not change the performance results despite the enlargement of the data set. Proposed model has more distinctive feature extraction structure and features compared to [31] study. This has been demonstrated in the performance results given in Table 4.

As a result of the study, the Nadam optimization method with the least number of losses is defined as the most successful model. The performance comparison of the results obtained with similar data sets in the literature with the results of the proposed model is given in Table 4. When Table 4 is analyzed briefly, it is seen that CNN-based methods have been developed as well as pre-trained transfer learning-based methods in malaria classification. The Proposed CNN model can compete with pre-trained or CNN-based methods in the literature. In order to achieve this, the filter and window sizes used in the convolution layers in the model were determined by experimental studies. Apart from the Convolution layer, the basic layers in deep learning are arranged for malaria detection with an appropriate merger. With the Proposed CNN model, convolution, max pooling, batch normalization, dropout, dense, fully connected layers were brought together in a certain order in order to extract the most distinctive features of malaria dataset images. In addition, the number of epochs and the best optimization method for compiling the model were determined by experimental studies.

Classification performance criteria are considered to determine the optimization method in which the Proposed CNN model gives the best results. According to these criteria, the proposed CNN model can compete with the latest technology architectures such as LeNet-5, EfficientNetB0, AlexNet, VGG16, VGG19, GoogleNet, ResNet. Adagrad and Adadelta optimization methods lagged behind other optimization methods. The CNN-based malaria classification model proposed in this study was made to be lightweight. At the same time, a deep learning model with low computational cost was obtained according to the studies in the literature. With the determined targets, it was ensured that the proposed model could compete with the studies in the literature and a model suitable for real-time systems was proposed.

In Table 4, Rajaraman et al. classifies malaria images with the pre-trained ResNet50 model [11]. Vijayalakshmi and Kanna replaced the classification layer in the upper layer of VGG16, VGG19, GoogleNet, AlexNet, LeNet-5 based architectures, which are pre-trained transfer learning techniques, with the SVM method [9]. With the change made, the classification process was carried out by changing the classification layer of the existing method in the literature. They got the best performance result with VGG19 architecture. However, the VGG19 architecture consists of 138.4 million parameters. The number of parameters specified increases the computational cost considerably. Raihan and Nahid provided the classification of features extracted by a CNN and wavelet-based method with the XgBoost algorithm [32]. Khan et al. classified malaria images with three different

models, which he proposed based on random forest, decision tree and logistic regression methods from machine learning techniques [33]. Montalbo et al. performed classification of malaria images without any preprocessing, optimization, or data augmentation with the EfficientNetB0 model, which uses pre-trained weight values with fine tuning. [34]. Reddy et al. performed classification on malaria images viewed microscopically with the ResNet50 model, which is one of the transfer learning-based architectures [35]. By fixing the image size to 224x224 width and height, they performed a training process consisting of 193 steps in each epoch step. They used the SGD optimization method in their studies, claiming that it gave better results than many other optimization methods [36]. Oyewola et al. uses reinforcement learning to classify malaria, which caused 627,000 deaths in 2020 [37]. By performing data augmentation, they increased the data.

Table 4. Comparison results with similar datasets

Class	Model	F1 score (%)	Precision (%)	Recall (%)	Accuracy (%)
Parasitized	[38]	---	---	---	0.84
	[9] (VGG19)	0.91	0.89	0.93	0.93
	[9] (VGG16)	0.85	0.82	0.88	0.91
	[9] (GoogleLeNet)	0.83	0.81	0.85	0.86
	[9] (AlexNet)	0.77	0.78	0.77	0.82
	[9] (LeNet-5)	0.74	0.74	0.74	0.79
	[39]	---	---	---	0.88
	[32]	0.94	0.94	0.95	0.94
	[11]	0.95	-	-	0.95
	[33] (Random Forest)	0.84	0.82	0.86	-
	[33] (Decision Tree)	0.76	0.78	0.75	-
	[33] (Logistic Regression)	0.79	0.84	0.75	-
	[34]	0.94	0.93	-	0.94
	[35]	-	-	-	0.95
	[37]	-	-	-	0.94
	[31] (ResNet50V2)	0.80	0.81	0.78	0.80
	[31] (CNN)	0.94	0.98	0.91	0.95
	Proposed CNN (Nadam)	0.95	0.98	0.92	0.95
	Proposed CNN (RMSprop)	0.94	0.95	0.94	0.95
	Proposed CNN (Adam)	0.94	0.97	0.91	0.94
	Proposed CNN (Adamax)	0.91	0.96	0.86	0.92
	Proposed CNN (SGD)	0.85	0.94	0.77	0.86
	Proposed CNN (Adagrad)	0.68	0.70	0.68	0.69
	Proposed CNN (Adadelta)	0.53	0.56	0.51	0.56
Uninfected	[38]	---	---	---	0.84
	[9] (VGG19)	0.91	0.89	0.93	0.93
	[9] (VGG16)	0.85	0.82	0.88	0.91
	[9] (GoogleLeNet)	0.83	0.81	0.85	0.86
	[9] (AlexNet)	0.77	0.78	0.77	0.82
	[9] (LeNet-5)	0.74	0.74	0.74	0.79
	[39]	---	---	---	0.88
	[32]	0.94	0.94	0.95	0.94
	[11]	0.95	-	-	0.95
	[33] (Random Forest)	0.84	0.82	0.86	-
	[33] (Decision Tree)	0.76	0.78	0.75	-
	[33] (Logistic Regression)	0.79	0.84	0.75	-
	[34]	0.94	0.93	-	0.94
	[35]	-	-	-	0.95
	[37]	-	-	-	0.94
	[31] (ResNet50V2)	0.81	0.79	0.83	0.80
	[31] (CNN)	0.95	0.92	0.98	0.95
	Proposed CNN (Nadam)	0.95	0.92	0.98	0.95
	Proposed CNN (RMSprop)	0.95	0.94	0.96	0.95
	Proposed CNN (Adam)	0.94	0.92	0.97	0.94
	Proposed CNN (Adamax)	0.92	0.88	0.97	0.92
	Proposed CNN (SGD)	0.87	0.81	0.95	0.86
	Proposed CNN (Adagrad)	0.71	0.68	0.73	0.69
	Proposed CNN (Adadelta)	0.58	0.56	0.61	0.56

The results obtained with this proposed CNN model and the optimization methods used in this model were evaluated according to the performance criteria commonly used in the current literature, such as precision, recall, F1 score, and accuracy criteria. It is concluded that the proposed model and the optimization methods used can be a useful model for the identification and classification of malaria diseases.

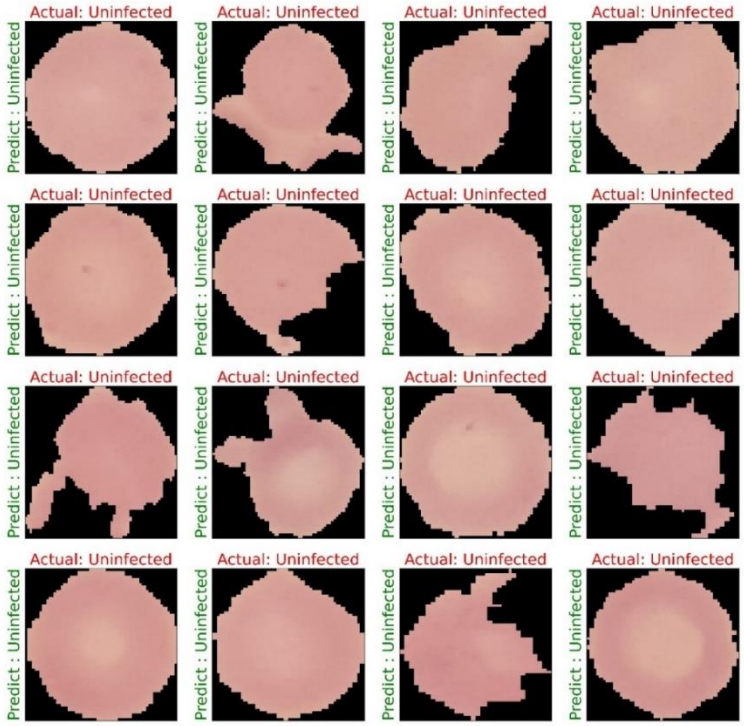

Figure 13. Test results on uninfected class of proposed model

Figure 13 shows the results obtained as a result of the tests performed. The predicted results of the model are shown in green on the y-axis. In the title, the actual class label of the malaria cell image shown is given. During the testing process, the preprocessing of the training modeling was carried out.

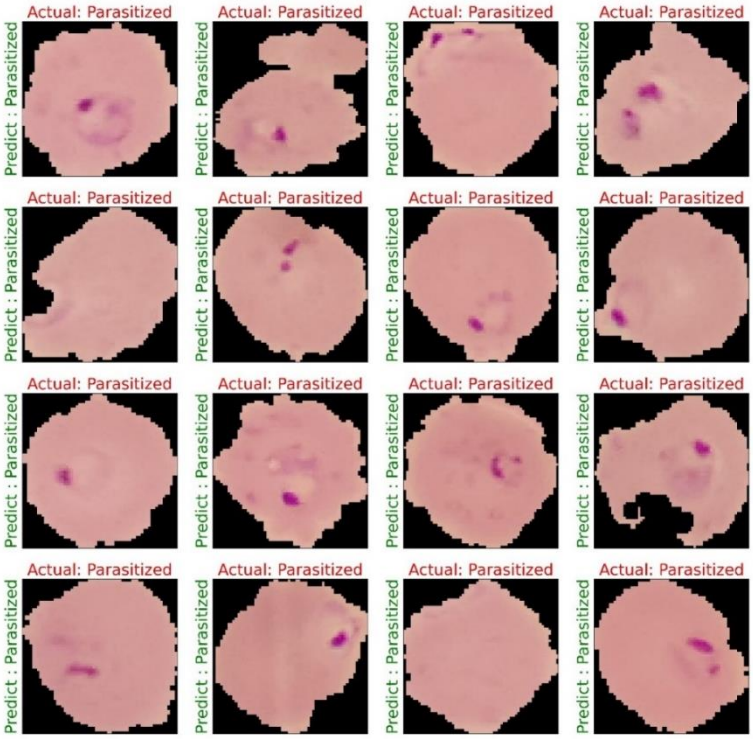

Figure 14. Test results on the parasitized class of the proposed model

Similar to Figure 13, the test results of the parasitized class are shown in Figure 14. As in Figure 13, the class predicted by the model is shown in green on the y-axis. In the header, the actual class label of the malaria cell image shown is presented.

V. CONCLUSIONS

In this study, we have helped the problem of diagnosing malaria by using image processing and deep learning architectures, which offer great opportunities in the healthcare field. Accuracies of Nadam (0.95), RMSprop (0.95), Adam (0.94), Adamax (0.92), SGD (0.86), Adagrad (0.69), Adadelta (0.56) and Ftrl (0.25) were obtained, respectively. In this study, we present a comprehensive search using basic optimization algorithms for efficient recognition and classification of malaria parasites. We use eight optimization techniques to extract meaningful distinguishing features from these images. To know whether a particular blood sample is infected or not, with each optimization technique, each sample was diagnosed and classified separately. We presented a new CNN model with optimization techniques used separately in the proposed CNN model for diagnosing malaria from images of red blood cells. This new CNN model has been proposed with eight different optimization techniques for malaria diagnosis using an up-to-date publicly available red blood cell dataset. It was achieved with a high accuracy of 95% in diagnosing malaria. It can be successfully applied in the diagnosis of malaria with optimization methods such as Nadam, RMSprop and Adam.

REFERENCES

[1] Ikerionwu, C., Ugwuishiwu, C., Okpala, I., James, I., Okoronkwo, M., Nnadi, C., & Ike, A. (2022). Application of Machine and Deep Learning Algorithms in Optical Microscopic Detection of Plasmodium Parasites: A Malaria Diagnostic Tool for the Future. *Photodiagnosis and Photodynamic Therapy*, 103198.

[2] Bonilla, J. A. (2006). Assessing the function of the aspartic proteinases of the Plasmodium falciparum digestive vacuole using gene-knockout strategies. University of Florida.

[3] Tangpukdee, N., Duangdee, C., Wilairatana, P., & Krudsood, S. (2009). Malaria Diagnosis: A Brief Review. *The Korean Journal of Parasitology*, *47*(2), 93.

[4] World Health Organization. (2021). World malaria report 2021. *World Health Organization 2021*.

[5] Das, D. K., Mukherjee, R., & Chakraborty, C. (2015). Computational microscopic imaging for malaria parasite detection: a systematic review. *Journal of Microscopy*, *260*(1), 1–19.

[6] Mitiku, K., Mengistu, G., & Gelaw, B. (2003). The reliability of blood film examination for malaria at the peripheral health unit. *Ethiopian Journal of Health Development*, *17*(3), 197–204.

[7] Chavan, S. N., & Sutkar, A. M. (2014). Malaria disease identification and analysis using image processing. *Int. J. Latest Trends Eng. Technol*, *3*(3), 218–223.

[8] Siłka, W., Wieczorek, M., Siłka, J., & Woźniak, M. (2023). Malaria Detection Using Advanced Deep Learning Architecture. *Sensors*, *23*(3), 1501.

[9] Vijayalakshmi A, & Rajesh Kanna B. (2020). Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*, *79*(21–22), 15297–15317.

[10] Poostchi, M., Silamut, K., Maude, R. J., Jaeger, S., & Thoma, G. (2018). Image analysis and machine learning for detecting malaria. *Translational Research*, *194*, 36–55.

[11] Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, *6*, e4568.

[12] Díaz, G., González, F. A., & Romero, E. (2009). A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *Journal of Biomedical Informatics*, *42*(2), 296–307.

[13] Shuleenda Devi, S., Sheikh, S. A., Talukdar, A., & Laskar, R. H. (2016). Malaria Infected Erythrocyte Classification Based on the Histogram Features using Microscopic Images of Thin Blood Smear. *Indian Journal of Science and Technology*, *9*(45).

[14] Das, D. K., Maiti, A. K., & Chakraborty, C. (2015). Automated system for characterization and classification of malaria- infected stages using light microscopic images of thin blood smears. *Journal of microscopy*, *257*(3), 238–252.

[15] Dong, Y., Jiang, Z., Shen, H., David Pan, W., Williams, L. A., Reddy, V. V. B., & Bryan, A. W. (2017). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 101–104, IEEE.

[16] Van-Quoc, V., & Thai-Nghe, N. (2023). Skin Diseases Detection with Transfer Learning. In *Proceedings of International Conference on Data Science and Applications: ICDSA 2022, 1*, 139–150, Springer.

[17] Hamedani-KarAzmoudehFar, F., Tavakkoli-Moghaddam, R., Tajally, A. R., & Aria, S. S. (2023). Breast cancer classification by a new approach to assessing deep neural network-based uncertainty quantification methods. *Biomedical Signal Processing and Control*, *79*, 104057.

[18] Gupta, K., & Bajaj, V. (2023). Deep learning models-based CT-scan image classification for automated screening of COVID-19. *Biomedical Signal Processing and Control*, *80*, 104268.

[19] Odusami, M., Maskeliunas, R., Damaševičius, R., & Misra, S. (2021). Comparable study of pre-trained model on alzheimer disease classification. In *Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part V 21*, 63–74. Springer.

[20] Rajaraman, S., Jaeger, S., & Antani, S. K. (2019). Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ*, *7*, e6977.
[21] Kassim, Y. M., Yang, F., Yu, H., Maude, R. J., & Jaeger, S. (2021). Diagnosing malaria patients with plasmodium falciparum and vivax using deep learning for thick smear images. *Diagnostics*, *11*(11), 1994.
[22] Yu, H., Yang, F., Rajaraman, S., Ersoy, I., Moallem, G., Poostchi, M., & Jaeger, S. (2020). Malaria Screener: a smartphone application for automated malaria screening. *BMC Infectious Diseases*, *20*(1), 1–8.
[23] Uddin, S. (2020). Malaria Parasite Image. Retrieved from https://www.kaggle.com/datasets/saife245/malaria-parasite-image-malaria-species
[24] Goutte, C., & Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In *Lecture Notes in Computer Science*, *3408*, 345–359.
[25] Aggarwal, C. C. (2018). Neural networks and deep learning. *Springer*, *10*, 973–978.
[26] Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into Deep Learning. *Journal of the American College of Radiology*, *17*(11), 437–516.
[27] McMahan, H. B., Holt, G., Sculley, D., Young, M., Ebner, D., Grady, J., & Kubica, J. (2013). Ad click prediction. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (Vol. 9, pp. 1222–1230). New York, NY, USA: ACM.
[28] Finsveen, L. (2018). Time-series predictions with Recurrent Neural Networks, (June), 21–22.
[29] Hoang, N.-D. (2021). Automatic Impervious Surface Area Detection Using Image Texture Analysis and Neural Computing Models with Advanced Optimizers. *Computational Intelligence and Neuroscience*, *2021*, 1–17.
[30] Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.
[31] Çetiner, İ. (2022). Konvolüsyonel Sinir Ağı Kullanılarak Sıtma Hastalığı Sınıflandırılması. *Journal*, *9*(17), 273–286.
[32] Raihan, M. J., & Nahid, A.-A. (2022). Malaria cell image classification by explainable artificial intelligence. *Health and Technology*, *12*(1), 47–58.
[33] Khan, A., Gupta, K. D., Venugopal, D., & Kumar, N. (2020). CIDMP: Completely Interpretable Detection of Malaria Parasite in Red Blood Cells using Lower-dimensional Feature Space. In *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8.
[34] Montalbo, F. J. P., & Alon, A. S. (2021). Empirical Analysis of a Fine-Tuned Deep Convolutional Model in Classifying and Detecting Malaria Parasites from Blood Smears. *KSII Transactions on Internet and Information Systems*, *15*(1), 147–165.
[35] Reddy, A. S. B., & Juliet, D. S. (2019). Transfer Learning with ResNet-50 for Malaria Cell-Image Classification. In *2019 International Conference on Communication and Signal Processing (ICCSP)*, 945–949.
[36] Keskar, N. S., & Socher, R. (2017). Improving generalization performance by switching from adam to sgd. *arXiv preprint arXiv:1712.07628*.
[37] Oyewola, D. O., Dada, E. G., Misra, S., & Damaševičius, R. (2022). A Novel Data Augmentation Convolutional Neural Network for Detecting Malaria Parasite in Blood Smear Images. *Applied Artificial Intelligence*, *36*(1), 2033473.
[38] Das, D. K., Ghosh, M., Pal, M., Maiti, A. K., & Chakraborty, C. (2013). Machine learning approach for automated screening of malaria parasite using light microscopic images. *Micron*, *45*, 97–106.
[39] Singla, N., & Srivastava, V. (2020). Deep learning enabled multi-wavelength spatial coherence microscope for the classification of malaria-infected stages with limited labelled data size. *Optics & Laser Technology*, *130*(September 2019), 106335.