IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1316~1326
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1316-1326

Hybrid kernel support vector machine with cuckoo search optimization for malaria detection from blood smear images

Sri Huning Anwariningsih[1], Indrarini Dyah Irawati[2]
[1]Department of Informatics, Faculty of Sains, Technology, and Health, Universitas Sahid Surakarta, Surakarta, Indonesia
[2]School of Applied Science, Telkom University, Bandung, Indonesia

Article Info

Article history:

Received Jun 23, 2025
Revised Jan 7, 2026
Accepted Jan 22, 2026

Keywords:

Cuckoo search algorithm
Histogram intersection
Hybrid kernel
Malaria detection
Support vector machine

ABSTRACT

Microscopic image-based malaria detection still struggles to capture complex features due to variations in lighting and color. The support vector machine (SVM) method is often used in medical image detection, but its performance depends heavily on the selection of optimal kernel and hyperparameters (C and gamma). Conventional approaches, with single kernels and manual tuning, have limitations in capturing both spatial information and color distribution simultaneously. Therefore, this research proposes hybrid kernel support vector machine-cuckoo search algorithm (HKSVM-CSA) method that combines the radial basis function (RBF) kernel and histogram intersection for SVM, along with hyperparameter optimization using the CSA. The dataset used is malaria cell images, which contains parasitized and uninfected images of blood cells. The proposed method comprises five main steps: dataset preparation, feature extraction, HKSVM, hyperparameter optimization, and model evaluation. Experiments demonstrate that the proposed model achieves 94% accuracy, 93% sensitivity, 94% specificity, and area under the curve (AUC) of 0.98, which is significantly better than standard SVM, SVM-genetic algorithm (GA), and k-nearest neighbors (KNN). These results show that combining kernel and CSA significantly improves detection accuracy. This approach is promising for image-based automatic systems for infectious disease diagnosis.

Corresponding Author:

Sri Huning Anwariningsih
Department of Informatics, Faculty of Sains, Technology, and Health, Universitas Sahid Surakarta
St. Adi Sucipto, No 154, Jajar, Surakarta, Central Java, Indonesia
Email: huning1706@gmail.com

1. INTRODUCTION

Malaria is a deadly infectious disease that is caused by protozoan parasites. These protozoan parasites belong to the genus *Plasmodium* and are transmitted to humans via the bite of a female *Anopheles mosquito* [1]. Several species from both genera are responsible for malaria transmission across different endemic regions, that is, *Plasmodium ovale*, *P. malariae*, *P. falciparum*, *P. vivax*, and *P. knowlesi* [2].

The WHO predicts 230 million malaria cases worldwide in 2023. This is equivalent to 60.4 cases per 1,000 individuals. This is an increase compared to 2022, when there were 58.6 cases per 1,000 individuals. In 2023, approximately 94% of all global cases were reported in Africa. Similarly, 95% of the projected 597,000 global malaria fatalities in 2023 occurred in the African Region [3]. Malaria symptoms appear when the *Plasmodium* parasite attacks and multiplies within the red blood cells. Inside red blood cells, the *Plasmodium* parasite undergoes several developmental stages. Initially, in its ring form, it develops into a trophozoite and ultimately forms a schizont [4]. The schizont divides into many merozoites, which are released when the red blood cell bursts, infecting other cells and initiating the

infection cycle that causes fever. This cycle will continue to repeat [5]. The invasion of malaria parasites into red blood cells alters the morphology and structure of these cells. These changes can be observed microscopically, providing an opportunity to use red blood cell images for the automatic detection of malaria infections [6].

Accurate early detection of malaria infection is crucial because it informs decisions on when to administer treatment, thereby reducing the risk of disease transmission. Generally, malaria infection detection relies on traditional methods, including rapid diagnostic tests (RDTs) and microscopy. However, these methods have limitations, including the need for expert skills, variations in interpretation results, and the inability to automate [7], [8]. Recent research has demonstrated that machine learning and computer vision can identify malaria in blood smear images automatically [9], [10].

Over the past decade, many automatic detection methods for malaria parasites based on blood smear images have been developed. Some traditional image processing techniques, such as thresholding [11], morphological operations [12], and segmentation to isolate infected cells [13] are used before classification. However, these methods struggle to handle variations in staining, lighting, and cell morphology, leading to inconsistent performance in real-world applications.

Machine learning-based approaches have demonstrated the ability to tackle these challenges. Several conventional classifiers, including k-nearest neighbors (KNN), decision trees, and naive Bayes, have been widely used in various studies. However, the performance of these conventional methods tends to decline when facing high-dimensional feature spaces and overlapping classes for non-linear data [13].

Among various classifiers, support vector machine (SVM) has demonstrated superior or comparable performance to other machine learning methods, particularly in disease diagnosis and prognosis. The use of kernel functions (e.g., radial basis function (RBF) and polynomials) can efficiently handle high-dimensional feature spaces and nonlinear relationships in medical data, making it particularly useful for large-scale applications [13]. SVMs perform well on datasets with a small number of samples and a large number of features (e.g., microarray gene data), where traditional models may struggle due to overfitting [14].

The SVM method has been widely used to enhance the accuracy of malaria disease detection and classification [15]–[17]. Convolutional neural networks (CNNs) and other deep learning algorithms have proven more effective at classifying malaria cells [18]–[20]. However, deep learning methods require not only large amounts of annotated data but also significant computational resources. This limits their practical application in resource-constrained environments. In contrast, using an SVM combined with dimensionality reduction techniques has proven both helpful and computationally efficient [17].

However, using a single kernel, such as the RBF kernel, in SVM often makes it insufficient for capturing the full diversity of visual patterns in blood cell images. Additionally, SVM's performance is influenced by hyperparameters (C and gamma) that affect accuracy [21]. Generally, both parameters are determined through manual tuning or grid search. Both methods require a high computational time and do not guarantee optimal values.

Considering these limitations, this research proposes an improved SVM framework that integrates RBF and histogram intersection kernels within a hybrid kernel design. Superiority hybrid kernels allow the classifier to capture both local and global structures in the data by combining multiple similarity measures [22]. The proposed method will utilize a hybrid kernel and apply the cuckoo search algorithm (CSA) to determine the C and gamma parameters, thereby obtaining the optimal parameters for use in the SVM model. The CSA method proposed by Yang and Deb [23] has several advantages, including a simple structure, global search capabilities via Levy flights, and high convergence efficiency. Previous studies have shown that CSA performs well across diverse optimization problems [24]. By combining the advantages of multiple kernel functions with the global search capabilities of CSA, the proposed method overcomes the limitations of traditional SVMs, thereby achieving higher accuracy while maintaining computational efficiency.

This study aims to detect malaria by classifying microscopic blood smear images into parasitized and uninfected cell categories based on visual features. The proposed method was tested on the public malaria cell image dataset and compared with the baseline performance of standard SVM, SVM optimized by a genetic algorithm (GA), and the KNN classifier. The model pipeline includes the image preprocessing, the feature extraction, the hybrid kernel-based SVM classification, the hyperparameter optimization via CSA, and the model evaluation. There are several contributions from this study, namely: i) proposing a new framework of SVM that integrates multi-kernel, namely RBF and histogram intersection, so that it can capture both texture and color features in blood smear images, ii) determining the optimal parameters of C and gamma in SVM automatically using CSA, and iii) combine color histogram and texture-based features for feature extraction.

The paper is organized as follows. Section 2 outlines the proposed method. Section 3 presents the dataset and experimental results. Section 4 concludes the paper.

2. METHOD

This study proposes a framework, hybrid kernel support vector machine (HKSVM)-CSA, for an improved SVM model that integrates a hybrid kernel function and hyperparameter optimization via the CSA to enhance classification performance for malaria-infected red blood cells in microscopic images. The methodological pipeline comprises five key stages: dataset preparation, feature extraction, construction of a HKSVM, CSA-based hyperparameter tuning, and comprehensive model evaluation. The overall framework of the proposed approach is shown in Figure 1.

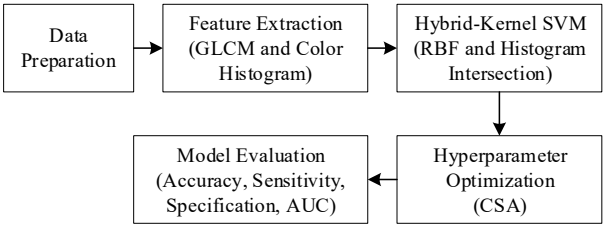
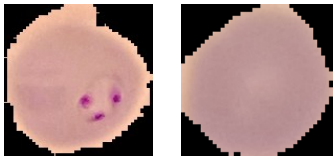


Figure 1. Workflow of the proposed method

2.1. Data preparation

This study used a subset of the publicly available malaria cell images dataset from Kaggle [25]. From the 27,558 images, this study uses only 3,000 images per class (parasitized and uninfected) to create a balanced, computationally manageable dataset (Figure 2). This subset size provides sufficient variability for reliable feature extraction and classification while keeping the CSA-based hyperparameter optimization computationally feasible. The images were chosen based on their order in the dataset directory, as the primary objective at this stage was to establish a consistent, representative subset for initial experimentation. To ensure consistency and reduce computational load, several preprocessing steps were applied. Each image was scaled to 128×128 pixels before further processing. Color information is preserved in the RGB format for color-based feature extraction, while a grayscale version is also generated for texture analysis. No additional denoising or contrast enhancement is applied, ensuring that the natural character of blood smear samples is maintained.

Figure 2 shows the visual appearance of blood smear images; Figure 2(a) displays a parasitized image, a red blood cell containing malaria parasites (*Plasmodium*), marked by small colored structures. This small structure often appears as a ring/band/crescent (depending on the stage of development). Figure 2(b) shows an uninfected image of a healthy, normal red blood cell, generally uniform in biconcave disk shape and showing no foreign structures or deformities.



(a) (b)

Figure 2. Blood smear images of (a) parasitized image and (b) uninfected image

2.2. Feature extraction

The proposed method uses a combination of color histogram features and texture-based features obtained from the gray-level co-occurrence matrix (GLCM). The combination of the two methods was chosen to capture the chromatic and structural variations that indicate *Plasmodium* infection. The steps of feature extraction are shown in Figure 3. In extracting the color histogram feature, the input RGB image is split into three channels (R, G, and B). Each channel is then calculated and normalized to 32 bins, resulting in a total of 96 color features. Based on these color features, staining patterns, and cytoplasmic changes in infected cells, infected cells will be identified, as reflected in the global intensity distribution.

Meanwhile, in the texture feature extraction step, the input image is converted to grayscale. Then, the GLCM is calculated with a pixel distance of 1 and an angular direction of 0°, using the feature descriptors contrast, correlation, energy, and homogeneity. These four features will provide a comprehensive overview of the local structure's texture. The combination of the two feature extraction methods produces a final

feature vector with 100 dimensions, consisting of 96 colors from the feature color histogram stage and four textures from the texture feature extraction stage. The final feature vector will later serve as the input to the SVM classifier.

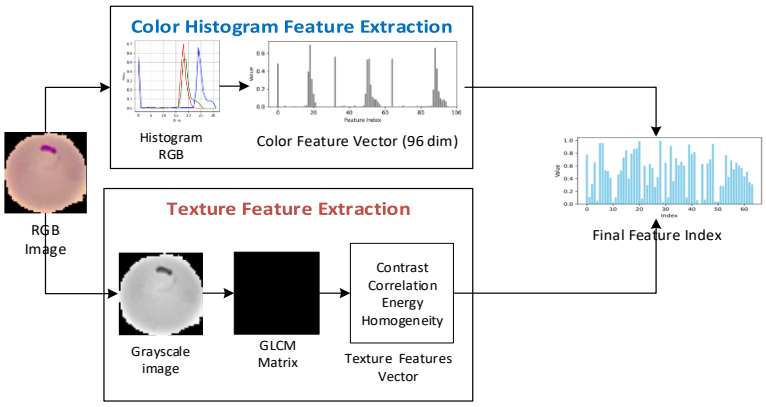


Figure 3. Step of feature extraction

2.3. Hybrid kernel support vector machine

The proposed method employs a hybrid kernel function that integrates the RBF and histogram intersection kernels, as shown in Figure 4. The proposed method starts with the final feature index values obtained from the feature extraction stage, which will be used in the RBF kernel and the histogram intersection kernel. The results of both will then be combined to obtain a kernel combination.

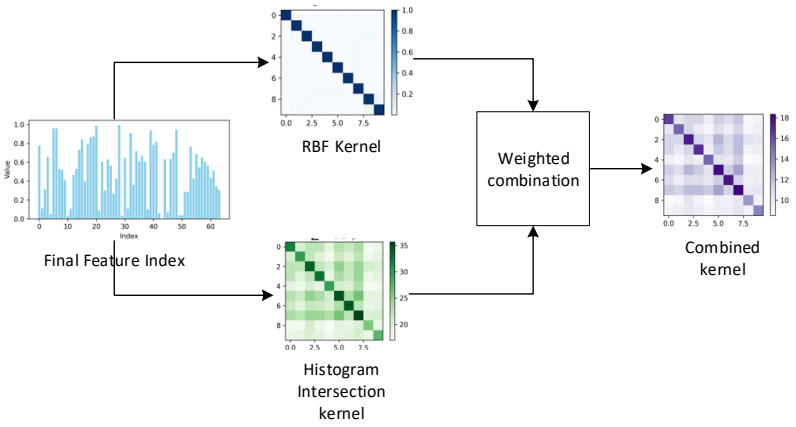


Figure 4. Step of HKSVM

The final feature vector resulting from the previous step will serve as the input for calculating each kernel. The RBF kernel will be calculated using (1) while the kernel histogram intersection will be calculated using (2). The results of the two calculations will then be combined using (3), which employs a weighted combination approach. The proportions of the weights of both are equal to enhance the discriminative capacity of SVM classifiers. Combining the two kernels results in a single kernel.

$$K_{RBF}(x,y) = exp\,(-\gamma\|x-y\|^2) \quad (1)$$

$$K_{HI}(x,y) = \sum_{i=1}^{d} min\,(x_i,y_i) \quad (2)$$

Where $x,y \in \mathbb{R}^d$ are feature vectors and γ is the kernel width parameter optimized using the CSA.

$$K_{hybrid}(x,y) = \alpha.K_{RBF}(x,y) + (1-\alpha).K_{HI}(x,y) \quad (3)$$

Where $\alpha \in [0,1]$ is weight coefficient, α, set to 0.5 for each kernel, indicating that each kernel contributes equally.

2.4. Hyperparameter optimization

In this study, the CSA is used to optimize the SVM hyperparameters (C and gamma). The steps performed at this stage are presented in Figure 5. The first step is the initialization of the nest, where the solution sought is a pair of parameter values C and gamma. The fitness value will then be calculated. If the fitness value is worse, then CSA will update them via Lévy flights and replace the worst solutions using a discovery probability mechanism [23]. This process will be repeated until the optimal solution is achieved, specifically, the best C and gamma value pairs.

The objective fitness function is the SVM model's classification accuracy on the validation data. Each solution pair (C and gamma) will be trained on the training data, and its accuracy in the validation data will then be calculated as a fitness value. Accuracy determination is an objective function that determines hyperparameter values based on the model's generalization capabilities. The objective function is expressed as (4). In this study, CSA was run with 15 nests, a discovery rate of 0.25, and 10 iterations.

$$Fit_{(C,\gamma)} = \frac{1}{N}\sum_{i=1}^{N} \delta(\hat{y}_i = y_i) \quad (4)$$

Where N denotes the number of validation samples, $\hat{y}_i$ is the predicted label, y_i is the actual label, and δ(.) is the Kronecker delta function, which returns 1 if the condition is true and 0 otherwise. This function directly estimates classification accuracy, guiding the CSA in searching for optimal parameter values.

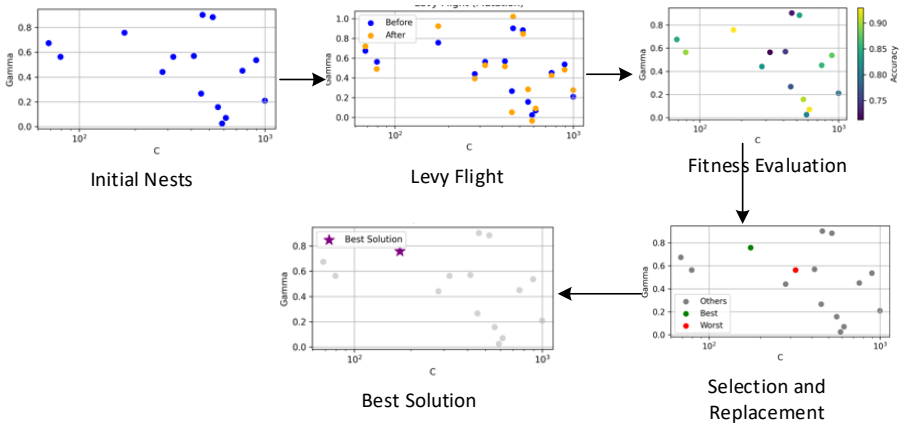


Figure 5. Step of hyperparameter optimization

2.5. Model evaluation

After the optimal hyperparameters are identified via CSA, the improved SVM model is trained using the entire training set. The model is then evaluated on a hold-out test set using standard metrics [26], namely accuracy, sensitivity/recall, specificity, and the area under the curve (AUC). Sensitivity is the ratio of true-positive predictions to the total number of true-positive samples; in other words, it measures the model's ability to accurately identify parasitized cells. Sensitivity can be calculated using (5).

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

Specificity is the ratio of true-negative predictions to the number of truly negative samples. In other words, specificity measures the model's ability to correctly identify uninfected cells. Specificity can be calculated using (6).

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

Accuracy refers to the proportion of correctly detected positive and negative results. Accuracy serves as a benchmark for evaluating and comparing diagnostic methods. A high accuracy value indicates that the model is highly effective in classification [26]. Accuracy can be calculated using (7).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Where TP indicates correct positive predictions, FP denotes incorrect positive predictions, FN represents missed positive cases, and TN indicates correct negative predictions. AUC is a widely used metric for evaluating binary classification models, reflecting the trade-off between the true positive rate (TPR) and the false positive rate (FPR) in distinguishing positive and negative classes [27].

Additionally, confusion matrices and receiver operating characteristic (ROC) curves are plotted to provide visual insights into classification performance. All evaluations are conducted under identical conditions to ensure a fair comparison with the baseline model. The baseline classifier employed in this study is an RBF-SVM with scikit-learn's default parameters (C =1.0, γ ='scale') implemented through a StandardScaler–SVM pipeline. Other baseline models include a KNN classifier (k =3) and SVM-GA, both implemented through StandardScaler–based pipelines. The GA used a population of 15, 10 generations, and a mutation rate of 0.1.

3. RESULTS AND DISCUSSION

3.1. Dataset description

This study utilizes 6,000 RGB images from the malaria cell images dataset developed by the National Library of Medicine. The dataset is evenly split into two classes, namely parasitized and uninfected. Each image is 128×128 pixels in size and was captured using standard light microscopy.

3.2. Experimental setup

All experimental procedures were conducted in Python 3.12.0 and the scikit-learn and OpenCV libraries. The following results were obtained on a PC with an Intel Core i5-8250U CPU running at 1.80 GHz, 8 GB DDR4 RAM, and an NVIDIA GeForce 930MX GPU. The dataset was randomly split into training and testing sets at an 80:20 ratio. Additionally, 20% of the training set was further divided into a validation set for hyperparameter optimization using CSA. Before extracting features, images were uniformly resized to 128×128 pixels. The CSA was configured with following parameters, as shown in Table 1.

Table 1. CSA's parameters

Parameters	Value
Number of nests (n)	15
Discovery rate (pa)	0.25
Number of iterations	10
Search bounds for C	[0.1…1000]
Search bounds for γ	[10^{-5}...1]

3.3. Performance analysis

To validate the performance of the proposed framework, a comparative analysis was conducted between the proposed method and the comparator method (the baseline standard SVM using an RBF kernel, KNN, and SVM-GA). The performance metrics used in this comparison include accuracy, sensitivity, specificity, and AUC. Sensitivity represents the ability of the proposed method to accurately identify cells infected with malaria. Specificity evaluates the capability of the proposed method to accurately classify uninfected cells. AUC is a metric that reflects how effectively a model differentiates between positive and negative classes. AUC values range from 0 to 1, with higher values corresponding to better classification performance. A value under 0.5 indicates poor performance. Additionally, 5-fold cross-validation testing was performed to evaluate model performance, specifically to assess stability and ensure that the performance estimates were more representative of the overall data.

The proposed method employed a hybrid kernel (RBF and histogram intersection kernels), with α=0.5. The proposed method was trained and tested on the same data split. The comparative performance between the proposed method and the comparator methods (SVM standard, SVM-GA, and KNN) for 5-fold cross-validation is presented in Table 2.

Table 2. Comparative performance analysis between the proposed method and comparator methods

Model	Accuracy	Sensitivity	Specificity	AUC
SVM standard	0.92	0.94	0.91	0.97
KNN	0.84	0.84	0.84	0.90
SVM-GA	0.74	0.54	0.95	0.83
The proposed method (HKSVM-CSA)	0.94	0.93	0.94	0.98

Table 2 summarizes the comparative performance of the proposed model (HKSVM-CSA) and the comparator methods on four key performance metrics: accuracy, sensitivity, specificity, and AUC. The experimental findings showed improved performance for the HKSVM-CSA classification compared to the comparator model. HKSVM-CSA achieves 94% accuracy, significantly outperforming the baseline. This demonstrates that integrating hybrid kernel functions with CSA-based hyperparameter optimization significantly enhances the classification capabilities of blood smear images.

Based on the sensitivity measurement, HKSVM-CSA achieved a 0.93, surpassing SVM-GA and KNN. This improvement is crucial in medical diagnosis, where failure to detect positive cases has a profound impact on patient care. In the specificity measurement, HKSVM-CSA achieved 0.94, indicating that the proposed method effectively reduces false positives and thereby minimizes the risk of misdiagnosing healthy individuals as infected.

Additionally, in the AUC metric, HKSVM-CSA achieved 0.98, outperforming KNN and SVM-GA while comparable with SVM standard (Figure 6). An AUC value of 0.98 indicates excellent classification performance, indicating that the proposed model can effectively distinguish between images of infected and uninfected red blood cells at various decision thresholds. The AUC results further validate the superiority of the HKSVM-CSA method. Based on performance measurements, the proposed method remains superior because it offers a more balanced sensitivity–specificity trade-off, yields higher accuracy, exhibits lower fold-to-fold variability, and benefits from adaptive hyperparameter optimization via CSA. These characteristics make it more reliable for real-world malaria detection tasks.

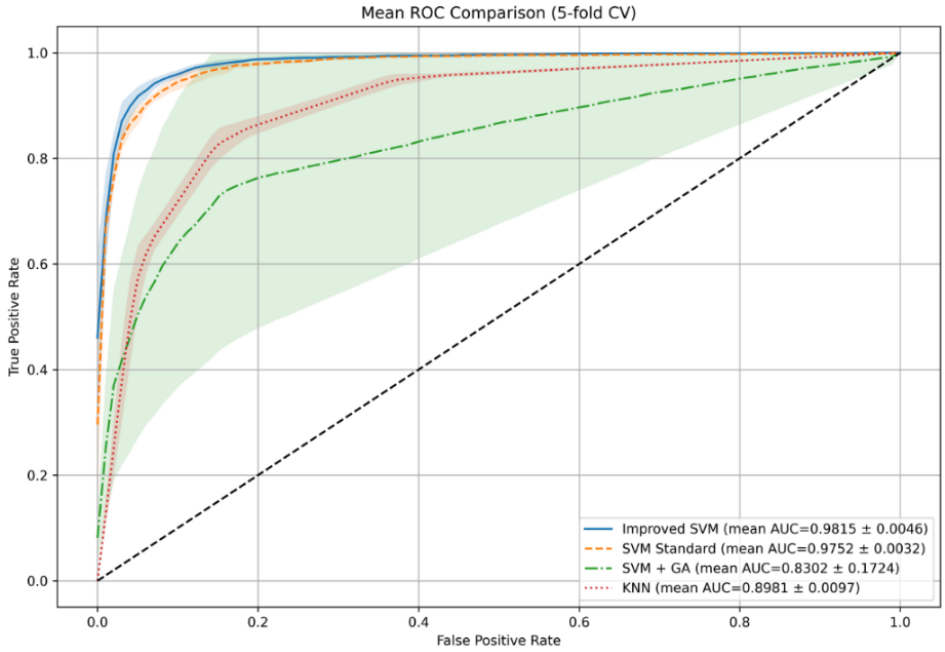


Figure 6. ROC curve comparison

Based on the mean ROC across all folds (Figure 6), the proposed method's curve is at the top for almost the entire range of false-positive rates. The AUC is very high, with an average accuracy of 0.9815 across 5-fold cross-validation and a standard deviation of 0.0046. The results confirm that the model shows robust stability and is not significantly affected by changes in the data. The narrowest shaded area shows that the folds are very stable. In contrast, SVM-GA shows a wide range of shading, indicating high variance across folds (sometimes good and sometimes bad). Based on the mean ROC value, HKSVM-CSA demonstrates the strongest discriminative ability, with the highest mean AUC and the smallest fold-to-fold variance, indicating robust and consistent performance.

The confusion matrix, as shown in Figure 7, can show the comparison of true positive, false positive, false negative, and true negative values between the proposed method and the comparator methods (SVM standard in Figure 7(a), KNN in Figure 7(b), and SVM-GA in Figure 7(c)). The proposed method (Figure 7(d)) showed a significant increase in the true positivity and true negative values. This indicates that,

compared to the other, HKSVM-CSA is more accurate in detecting cases of malaria that are truly infected (parasitized) and those that are completely uninfected. Additionally, there was a significant decrease in the false positive and false negative values. This demonstrates that HKSVM-CSA is highly robust and reliable in classification compared to the standard SVM. This performance improvement ensures that malaria cases are accurately detected while minimizing the risk of misdiagnosis in uninfected individuals.

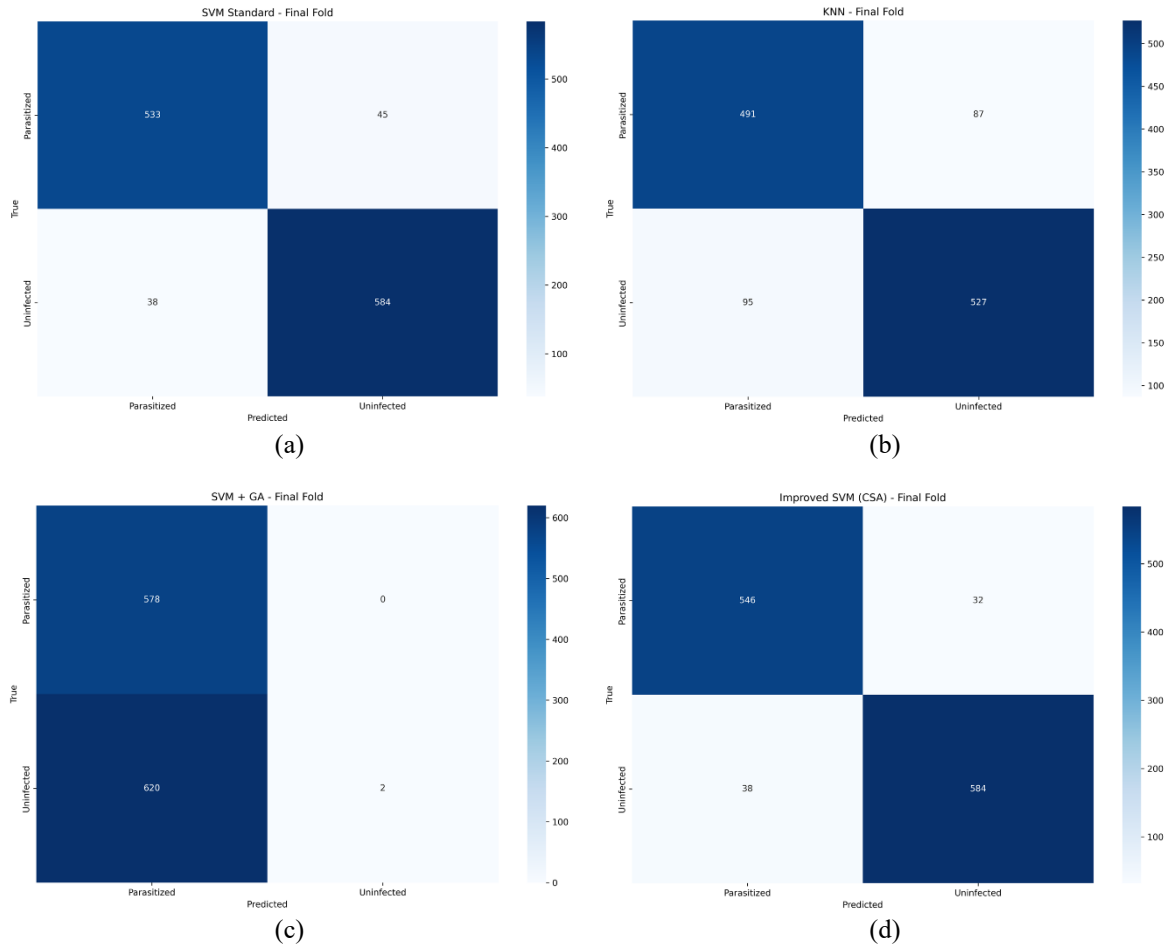


Figure 7. Confusion matrix for (a) SVM standard, (b) KNN, (c) SVM-GA, and (d) HKSVM-CSA

This study also conducted statistical testing using the paired t-test (Table 3). Both were used to test the statistical significance of differences in performance between models. This study also compares training and testing times across all models (Table 4). In a paired t-test, the t-value indicates the difference in average performance between two models. The larger the t-value, the more significantly different the two models' performance. A positive t-value means the first model performs better than the second, while a negative t-value indicates the opposite. Meanwhile, the p-value indicates the likelihood that the difference occurred by chance. A p-value less than 0.05 indicates that the difference is statistically significant, suggesting it is unlikely due to random fluctuations. Conversely, a p-value above 0.05 indicates that the observed difference is not statistically significant, and both models can be said to perform similarly [28].

Table 3. Statistical testing using the paired t-test

Model 1	Model 2	t_stat	p-value
HKSVM-CSA	SVM standard	4.619	0.010
HKSVM-CSA	SVM-GA	2.974	0.041
HKSVM-CSA	KNN	21.500	0.000
SVM-standard	SVM-GA	2.760	0.051
SVM-standard	KNN	14.864	0.0001
SVM-GA	KNN	-1.555	0.195

Based on Table 3, a significant statistical difference exists between HKSVM-CSA and the comparison methods. A positive t-value indicates that HKSVM-CSA has a higher average performance. A p-value of less than 0.05 means that the difference is very unlikely to have happened by chance. This shows that using a hybrid kernel and CSA-based parameter optimization provides a significant improvement over standard SVM. Compared to SVM-GA, HKSVM-CSA is still superior, even though both use metaheuristic optimization. CSA provides more stable exploitation-exploration than GA, enabling it to determine more optimal parameters. On the other hand, compared to KNN, the large t-value indicates a significant performance gap between HKSVM-CSA and KNN. KNN is far behind in malaria classification using color and GLCM features. Table 3 shows that HKSVM-CSA is the best-performing model, with performance significantly better than that of all other models. Using CSA is also more effective than GA for optimizing SVM hyperparameters.

Meanwhile, based on the training and testing time tests (Table 4), KNN has the shortest training time (0.01 seconds) because it does not involve a complex training process. SVM standard shows the fastest test time (0.19 seconds) because it does not include an optimization process, which makes the computation run efficiently. Meanwhile, HKSVM-CSA had the longest time. This is natural because the proposed method must extract features using two kernels, which makes it computationally more complex. Additionally, the CSA optimization process requires more iterations to determine the optimal SVM parameters.

Table 4. Comparison of computational time for training and testing

Model	Training time (s)	Testing time (s)
SVM standard	2.97	0.19
KNN	0.01	0.43
SVM-GA	5.72	0.39
The proposed method (HKSVM-CSA)	238.19	5.31

3.4. Discussion

The experimental results presented demonstrate the superiority of the HKSVM-CSA model in classifying malaria-infected blood cells. HKSVM-CSA integrates a hybrid kernel (RBF and histogram intersection) and optimizes hyperparameters using the CSA. The RBF kernel's ability to handle nonlinear spatial features, combined with the histogram intersection kernel's sensitivity to color distribution, results in a more discriminative feature space. The HKSVM-CSA model achieved 94% accuracy, 0.93 sensitivity, 0.94 specificity, and an AUC of 0.98, significantly outperforming the baseline methods. These results confirm that a multi-kernel function is sufficient to fully represent the complex visual characteristics of malaria-infected and non-infected cells. Moreover, hyperparameter optimization using CSA significantly improved performance. CSA capability enabled the model to effectively fine-tune its decision boundary to the specific feature distribution in the malaria cell images dataset.

From a clinical perspective, a high-sensitivity automated malaria detection system reduces the likelihood of false negatives, thereby ensuring that infected individuals receive timely treatment. The balanced specificity also minimizes false positives, reducing unnecessary treatments and associated costs. Despite its advantages, the current study has several limitations. The experiments were conducted using a single dataset under controlled imaging conditions. Performance may vary when applied to images captured from different sources or under varying lighting and staining conditions. Furthermore, future research should investigate domain adaptation techniques to enhance model generalizability across datasets from various medical centers. The integration of deep feature representations—such as those derived from pre-trained CNNs—with HKSVMs may further enhance performance. Additionally, developing an end-to-end system that includes segmentation, feature extraction, classification, and report generation would offer greater practical value.

4. CONCLUSION

This study presents a novel, computationally efficient framework for the automatic detection of malaria-infected red blood cells, which enhances the SVM method with a hybrid kernel and hyperparameter optimization via the CSA. The integration of the RBF and histogram intersection kernels enabled the model to capture both spatial and chromatic image characteristics. At the same time, CSA effectively tuned the classifier's parameters to maximize performance. Experimental results on the malaria cell images dataset demonstrated significant improvements over a standard SVM, with the proposed model achieving 94% accuracy, 0.93 sensitivity, 0.94 specificity, and an AUC of 0.98. These findings suggest that the proposed HKSVM-CSA is a robust and scalable solution for malaria detection from microscopic images. Further

development could include integrating deep learning-based features and evaluating the model on a diverse dataset. Implementation on mobile devices or embedded systems can also increase clinical benefits in resource-constrained regions.

ACKNOWLEDGMENTS

We are very grateful to Universitas Sahid Surakarta and Telkom University for providing exceptional support in enhancing our research competence.

FUNDING INFORMATION

This research is collaborative, with funding shared between Universitas Sahid Surakarta and Telkom University Bandung.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sri Huning Anwariningsih	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
Indrarini Dyah Irawati	✓	✓				✓				✓	✓	✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper. The authors state no conflict of interest.

DATA AVAILABILITY

The data supporting this study are publicly available in Kaggle at https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria.

REFERENCES

[1] D. Đumić, D. Kečo, and Z. Mašetić, "Automatization of microscopy malaria diagnosis using computer vision and random forest method," *IFAC-PapersOnLine*, vol. 55, no. 4, pp. 80–84, 2022, doi: 10.1016/j.ifacol.2022.06.013.
[2] H.-I. Shin *et al.*, "2023 world malaria report (status of world malaria in 2022)," *Public Health Weekly Report*, vol. 17, no. 32, pp. 1351–1377, 2024, doi: 10.56786/PHWR.2024.17.32.1.
[3] P. Venkatesan, "WHO world malaria report 2024," *The Lancet Microbe*, vol. 6, no. 4, 2025, doi: 10.1016/j.lanmic.2025.101073.
[4] I. M. D. Maysanjaya, "Comparative study of classification method on diagnosis of plasmodium phase," *Journal of Physics: Conference Series*, vol. 1516, no. 1, 2020, doi: 10.1088/1742-6596/1516/1/012021.
[5] B. M. Cooke, N. Mohandas, and R. L. Coppel, "The malaria-infected red blood cell: structural and functional changes," in *Advances in Parasitology*, vol. 50, 2001, pp. 1–86, doi: 10.1016/S0065-308X(01)50029-9.
[6] A. Loddo, C. Di Ruberto, and M. Kocher, "Recent advances of malaria parasites detection systems based on mathematical morphology," *Sensors*, vol. 18, no. 2, 2018, doi: 10.3390/s18020513.
[7] M. L. McMorrow, M. Aidoo, and S. P. Kachur, "Malaria rapid diagnostic tests in elimination settings—can they find the last parasite?," *Clinical Microbiology and Infection*, vol. 17, no. 11, pp. 1624–1631, 2011, doi: 10.1111/j.1469-0691.2011.03639.x.
[8] S. Momčilović, C. Cantacessi, V. A. -Arsenijević, D. Otranto, and S. T. -Otašević, "Rapid diagnosis of parasitic diseases: current scenario and future needs," *Clinical Microbiology and Infection*, vol. 25, no. 3, pp. 290–309, 2019, doi: 10.1016/j.cmi.2018.04.028.
[9] K. Yahuza, A. M. Umar, S. D. Baha'uddeen, E. T. Atalabi, G. L. Mukhtar, and B. Abdulkadir, "Recent advancements in detection and quantification of malaria using artificial intelligence," *UMYU Journal of Microbiology Research*, vol. 9, no. 2, pp. 1–21, 2024, doi: 10.47430/ujmr.2492.001.
[10] C. R. Maturana *et al.*, "Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: a review," *Frontiers in Microbiology*, vol. 13, 2022, doi: 10.3389/fmicb.2022.1006659.
[11] Y. Zhong *et al.*, "Efficient malaria parasite detection from diverse images of thick blood smears for cross-regional model accuracy," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 4, pp. 226–233, 2023, doi: 10.1109/OJEMB.2023.3328435.
[12] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Analysis of infected blood cell images using morphological operators," *Image and Vision Computing*, vol. 20, no. 2, pp. 133–146, 2002, doi: 10.1016/S0262-8856(01)00092-0.

[13] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of k-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, 2022, doi: 10.1016/j.dajour.2022.100071.
[14] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An overview on the advancements of support vector machine models in healthcare applications: a review," *Information*, vol. 15, no. 4, 2024, doi: 10.3390/info15040235.
[15] R. Anugrah, K. Usman, and L. Novamizanti, "Classification of malaria in red blood cell microscopic images using deep learning with EfficientNet architecture and SVM," in *2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, 2023, pp. 1–6, doi: 10.1109/ICRAIE59459.2023.10468300.
[16] D. Li and Z. Ma, "Residual attention learning network and SVM for malaria parasite detection," *Multimedia Tools and Applications*, vol. 81, no. 8, pp. 10935–10960, 2022, doi: 10.1007/s11042-022-12373-6.
[17] S. Shashikiran and H. D. Sunitha, "Malaria cell detection using advanced SVM techniques," in *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*, 2024, pp. 1–6, doi: 10.1109/NMITCON62075.2024.10699300.
[18] K. Ohdar and A. Nigam, "A robust approach for malaria parasite identification with CNN based feature extraction and classification using SVM," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–6, doi: 10.1109/ICCCNT56998.2023.10306840.
[19] M. Mujahid *et al.*, "Efficient deep learning-based approach for malaria detection using red blood cell smears," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-63831-0.
[20] A. Dev, M. M. Fouda, L. Kerby, and Z. Md Fadlullah, "Advancing malaria identification from microscopic blood smears using hybrid deep learning frameworks," *IEEE Access*, vol. 12, pp. 71705–71715, 2024, doi: 10.1109/ACCESS.2024.3402442.
[21] T. V. Saradhi, "A study on hyperparameter tuning in support vector machines and its impact on model accuracy," *Global Journal of Engineering Innovations and Interdisciplinary Research*, vol. 5, no. 1, 2025, doi: 10.33425/3066-1226.1063.
[22] H. R. Goudarzi and Z. Farzinrad, "A combined kernel function for dynamic support vector machines: exploiting hybrid similarity in data clustering," *International Journal of Data Science and Analytics*, vol. 20, no. 5, pp. 4511–4526, 2025, doi: 10.1007/s41060-025-00732-z.
[23] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World Congress on Nature & Biologically Inspired Computing*, 2009, pp. 210–214, doi: 10.1109/NABIC.2009.5393690.
[24] M. Shehab, A. T. Khader, and M. A. Al-Betar, "A survey on applications and variants of the cuckoo search algorithm," *Applied Soft Computing*, vol. 61, pp. 1041–1059, 2017, doi: 10.1016/j.asoc.2017.02.034.
[25] A. Chakraborty, "Malaria cell images dataset," *Kaggle*. 2019. Accessed: Jan. 16, 2025. [Online]. Available: https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria
[26] I. Pérez, I. Y. T. -Vicenti, C. G. G. -Xuriguera, C. Carvajal, J. V. A. Franco, and C. Loézar, "How to interpret diagnostic tests," *Medwave*, vol. 21, no. 7, pp. e8432–e8432, 2021, doi: 10.5867/medwave.2021.07.8432.
[27] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.
[28] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-56706-x.

BIOGRAPHIES OF AUTHORS

Sri Huning Anwariningsih holds a Doctor of Computer Science degree from Universitas Gadjah Mada, Yogyakarta, Indonesia, in 2025. She received her bachelor's degree (informatics) from Sekolah Tinggi Teknologi Telkom, Bandung, Indonesia, in 2002 and her master's degree (informatics) from Institut Teknologi Sepuluh November (ITS), Surabaya, Indonesia, in 2010. She is currently an associate professor at the Department of Informatics in Universitas Sahid Surakarta, Surakarta, Indonesia. Her research includes computer vision, global optimization, machine learning, data mining, bioinformatics, robotics, and artificial intelligence. She has published over 30 papers in journals and conferences. She is also actively writing informatics textbooks for elementary school students and middle-grade students. She can be contacted at email: huning1706@gmail.com or srihuning@usahidsolo.ac.id.

Indrarini Dyah Irawati holds a doctorate degree in Telecommunication Engineering from Institut Teknologi Bandung, Indonesia, in 2019. She received her B.Eng. in Telecommunication Engineering from Sekolah Tinggi Teknologi Telkom and M.Eng. from Institut Teknologi Telkom. She is currently a full professor at the School of Applied Science, Telkom University, Indonesia, with expertise in telecommunications networks and multimedia engineering, compressive learning, signal processing, the internet of things (IoT), and artificial intelligence. Her research portfolio includes over 50 funded projects that emphasize innovation and real-world impact through collaborations within the pentahelix framework, involving academia, industry, government, community, and media. She has published more than 50 Scopus and Web of Science-indexed journal articles, authored numerous academic books, and hold over 30 intellectual property rights in areas such as smart agriculture, secure communications, and environmental monitoring. She is also highly active in community empowerment, leading various technology-based initiatives that bridge research and public benefit, and frequently serves as a keynote and invited speaker at international scientific forums. She can be contacted at email: indrarini@telkomuniversity.ac.id.