Malaria Cell Image Classification via Majority Voting Ensembling aided Label Contrastive Learning

Shreyan Kundu[1], Rahul Talukdar[2], Nirban Roy[1], Semanti Das[3], Soumili Basu[4], Souradeep Mukhopadhyay[5*], Biswadip Basu Mallik[6]

[1*]Dept. of CSE (AI), Institute of Engineering & Management, University of Engineering & Management, Kolkata, India.
[2]Dept. of CSE (core), Institute of Engineering & Management, Kolkata, India.
[3]Centre for Ecological Sciences, Indian Institute of Science, Bangalore, India.
[4]Department of Computer Science & Engineering, Academy of Technology, Adisaptagram, India.
[5]Computer Science and Automation, Indian Institute of Science, Bangalore, India.
[6]Dept. of BSH, Institute of Engineering & Management, University of Engineering & Management, Kolkata, India.

*Corresponding author(s). E-mail(s): souradeepm@iisc.ac.in;
Contributing authors: shreyankundu@gmail.com; rahultalukdar33714@gmail.com; nirbanroy04@gmail.com; semantidas@iisc.ac.in; soumili.sjc@gmail.com; biswadip.basumallik@iem.edu.in;

Abstract

The plasmodium group single-celled parasite that causes malaria is contagious. A female Anopheles mosquito with the infection is most frequently the one spreading it. Over 40of the world's population is at danger; 219 million illnesses and around 435,000 deaths were recorded in 2017 alone. Even with many sophisticated evaluation tools, microscopists in resourceconstrained settings still struggle to improve the accuracy of diagnosis. Cell picture classification using

deep learning avoids making incorrect diagnosis conclusions. This research aims to increase diagnostic accuracy by classifying malarial-infected cells using majority voting ensembling in conjunction with triplet loss aided label contrastive learning. The outcomes of the experiment demonstrate how well our method works with microscopic cell images in terms of accuracy, precision, recall, and other parameters.

Keywords: Malaria, Ensemble, Label contrastive learning, Triplet loss, Majority voting

1 Introduction

The Plasmodium group of single-celled protozoan parasites is the infectious agent that causes malaria. The disease is mostly transmitted via the bite of a female Anopheles mosquito carrying the infection. With around 240 million cases reported annually, the disease puts approximately 40% of the world's population at risk. Malaria primarily afflicted countries in Africa. Two-thirds of malaria-related deaths worldwide occur in children under the age of five, accounting for 92% of malaria-related deaths in Africa. Typical malaria symptoms include fever, headaches, nausea, and, in more severe cases, convulsions, yellow skin, and a coma that results in death.

Every year, qualified professionals analyze millions of blood samples to look for signs of malaria infection. Manually counting the parasites and infected red blood cells is the method used to detect malaria. But the microscopist's experience and expertise are the only factors that matter. Working in an environment with few resources and without a system that would aid in skill maintenance will impact diagnostic quality. This results in incorrect diagnostic choices.

2 Related Work

The categorization of cell pictures using deep learning can help avoid making incorrect diagnosis conclusions. One kind of machine learning called deep learning has shown remarkable results in a variety of non-medical domains. Deep learning has had little use in the medical industry because of privacy issues and a lack of specialised understanding in that area. However, deep learning has been applied in numerous medical domains recently. Liang et al.'s paper from[1] is the first to apply deep learning convolutional neural networks (CNNs) for the diagnosis of malaria. A study by Odu et al.[2] built a machine learning model using XGBoost to predict malaria outbreaks in six African nations based on climate data. While successful, the model's dependence on potentially limited climate data hinders real-time application.

Mahdieh Poostchi et al.[3] extensively reviewed image analysis and machine learning methods for malaria detection from blood slides, covering imaging, preprocessing, parasite detection, cell segmentation, and classification. However, the study lacks a consolidated framework, requiring substantial customization for practical implementation. Narayanan [4] assessed various machine learning and deep learning methods

for malaria detection from cell images. They introduced a new Convolutional Neural Network (CNN) and contrasted it with established models like AlexNet, ResNet, VGG-16, and DenseNet. Despite achieving 96.7% accuracy and 0.994 AUC, computational demands restrict deployment in resource-constrained environments. Reddy et al.[5] implemented transfer learning using the ResNet-50 model to classify malaria-infected cells. The model showed high diagnostic accuracy, addressing challenges faced by microscopists in resource-constrained regions. However, one disadvantage is that transfer learning models like ResNet-50 require significant computational resources and may not be suitable for real-time applications in low-resource environments. Manning et al.[6] developed a malaria assessment system using image analysis and a Feed-forward Neural Network (FNN) for classifying cells. The system achieved 92% accuracy on a benchmark dataset, offering a promising tool for automated malaria diagnosis. Nonetheless, the primary limitation is its relatively lower accuracy compared to other advanced machine learning models, potentially reducing its reliability for clinical use. Ma et al.[7] study leverages active learning to improve malaria cell classification, achieving 93% accuracy using only 26% of the dataset. The research demonstrates that active learning, particularly uncertainty sampling, outperforms random sampling. However, a notable disadvantage is the increased computational time required for training. Das et al.[8] employ machine learning to automate malaria parasite detection from microscopic images, achieving 84% accuracy. Their approach integrates feature selection using F-statistic, Bayesian learning, and SVM, highlighting complexities in optimizing classification performance due to the diverse nature of malaria stages and image variability.

The remainder of the paper is organized in this manner. Section III demonstrates our motivation and our contribution. Section IV contains our suggested technique on the classification of malaria cell images. Section V goes into depth on the experiments we ran and the outcomes that followed. Section VI describes this conclusion and the work that has to be done.

3 Motivation and Contribution

The reviewed studies underscore the significant advancements in using machine learning and deep learning techniques for malaria detection and prediction. While these methods demonstrate high accuracy and potential for improving malaria diagnosis, several limitations persist, such as reliance on high-quality data, computational intensity, and challenges in real-world applicability. Addressing these drawbacks is crucial for developing robust and scalable solutions for malaria control and prevention.

The main contributions of this study are as follows:

1. To the best of our knowledge, the malaria cell classification problem is the first to use the combination of triplet loss (a contrastive environment) and cross-entropy loss (a supervised setting). By combining contrastive and supervised learning techniques for improved feature representation, this innovative method improves the model's ability to categorize pests with more accuracy.

2. In contrastive learning, using a label memory bank reduces the number of erroneous negative examples. Through the retention and use of label data from prior

batches, the model enhances its ability to discriminate between instances that are similar and those that are not, hence enhancing representation learning and overall performance by mitigating the influence of false negative samples.

3. Finally, the application of a majority voting-based ensembling of 3 top-performing label contrastive learning backbones (Resnet, Densenet and Efficientnet) proves the efficiency of our method in terms of accuracy.

4 Our Method

4.1 Label Contrastive Learning with triplet loss

Yang et al.[10] first proposed Label Contrastive learning for image classification with InfoNCE loss as contrastive loss. But We have slightly change their method by incorporating Triplet loss in place of InfoNCE loss. Its two basic components are contrastive learning and supervised learning. The model distinguishes between various image classes through the use of supervised learning. While the contrastive learning component pushes examples of different classes as far away as feasible, supervised learning promotes the model to place examples of the same class near together in the representation space. Two identical image encoders, IE_q and IE_k, are part of the contrastive learning component. They are in charge of encoding the representations of the input query picture and key image, respectively. While momentum is used to update encoder IE_k, back-propagation is used to update encoder IE_q. Both the image representations and the associated class labels are kept in the label memory bank. The sum of the weighted classification loss and contrastive loss determines the overall loss of the model.

We treat the contrastive learning part of our design like a dictionary look-up problem. We query Q (a set of sample representations) for an encoding query, and the dictionary's associated positive key, $Z+$. In order for there to be a contrastive loss, there needs to be a significant loss between the query Q and the negative key $Z-$ and a little loss between the positive key $Z+$ and the query Q. The momentum update technique proposed in MoCo [11] is used to update the parameters of the E_k encoder, and the label memory bank method proposed in [12] is put into practice. Below, we shall delve deeply into each element of the Contrastive loss formulation.

4.1.1 Explanation of Image encoders

Two encoders, IE_q and IE_k, are used to extract features from a cell image to create its feature representation. While IE_k pulls features from the key image, IE_q pulls features from the input query image. The model uses these extracted features for subsequent processing or comparison.

4.1.2 memory bank with label (class)

Greater batch sizes in contrastive learning lead to higher learned representation quality. However, the batch size is constrained by the memory capacity and cannot be increased indefinitely. The label memory bank that Ren et al.[12] suggested for our

model is implemented using queues. Unlike the memory bank, the label(Class) memory bank retains both the associated label and its image representation.

4.1.3 Update on momentum

Since there are a lot of picture representations and labels in the label memory bank, back-propagation updating of the encoder IE_k's parameters can result in a significant overhead. The momentum updating method reported in MoCo [11] can maintain the consistency of image representations in the memory bank in addition to reducing the computational cost. Moreover, our method leverages the notion of momentum update, which performs gradient update operations for encoder IE_q only via back-propagation. For updates to the encoder IE_k parameter, the momentum update technique is employed; the precise update computation is as follows:

$$G_k \leftarrow \alpha G_k + (1-\alpha) G_q \tag{1}$$

where G_k and G_q are the parameters of IE_k and IE_q, respectively, and α is the momentum parameter, which can have values between [0, 1]. MoCo notes that a fairly big momentum parameter α will yield the best outcomes.

4.1.4 Classification loss

The fully connected layers E_q encoder output is sent to a classifier, which computes the classification loss. Here is the definition of a cross-entropy loss function, which is the classification loss function.

$$L_{\mathrm{CLA}} = -\frac{1}{|Q|} \sum_{q_i \in Q} \sum_{k \in C} 1_{q_{i,k}} \log(P_{q_{i,k}}) \tag{2}$$

where Q is the set of portrayals for the query picture, P is the estimated probability, $1_{q_{i,k}}$ is the value of the indicator function, which is 1 when the query image's classification is accurate, and $q_{i,k}$ is the prediction that the query image q_i will be k.

4.1.5 Triplet Loss as Contrastive loss

Let the i-th sample of our dataset J be denoted by (I_i, C_i). The feature representation of the picture $R_i = IE(I_i)$ is obtained by passing it through any image encoder IE. During the neural network training process, training samples are selected and generated into triplets. Each triplet $(Trip^i = (I^i_{anc}, I^i_{pos}, I^i_{neg}))$ consists of an anchor sample (I_{anc}), a positive sample (I_{pos}), and a negative sample (I_{neg}). The tiplet labels are as follows: $C_{anc} = C_{pos} \neq C_{neg}$.Triplet loss aims to bring samples in the same class into neighboring locations on a manifold surface and push samples with different labels apart. The optimisation objective of the triplet $Trip^i$ is,

$$L^i_{CON} = 0.5[||R^i_{pos} - R^i_{anc}||^2_2 - ||R^i_{neg} - R^i_{anc}||^2_2 + \delta]_+ \tag{3}$$

where,

$$[U]_+ = \max(0, U)$$

and δ is a hyperparameter that helps us to maintain $||R^i_{neg} - R^i_{anc}||^2_2$ s to be greater than the distance $||R^i_{pos} - R^i_{anc}||^2_2$. For all triplet $Trip^i$ in the training dataset, the final objective to minimize is

$$L_{CON} = \frac{\Sigma_{|Trip|} L^i_{CON}}{|Trip|} \tag{4}$$

where $|Trip|$ is the number of triplets in the training set. The visualization of triplet loss in our method is shown in Figure 1.

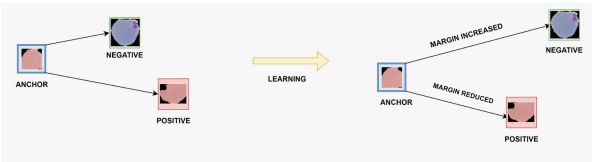

Fig. 1 Visualization of triplet loss in our method

4.1.6 Overall Loss

Our loss is the weighted sum of the classification loss and the contrastive loss. This is how the total loss is calculated:

$$L_{Overall} = \gamma L_{CON} + (1 - \gamma) L_{CLA} \tag{5}$$

where the hyper-parameter γ modifies the relative weight of contrastive loss and classification loss. This method seeks to minimize $L_{Overall}$, the contrastive loss that determines where in the representation space similar images are located as close together and other classes of images as far apart. The classification loss encourages the model to distinguish between many image classes. Once the complete model has been trained, pictures can be classified directly using the encoder IE_q and classifier. The triplet loss aided LCL method is shown in Figure 2.

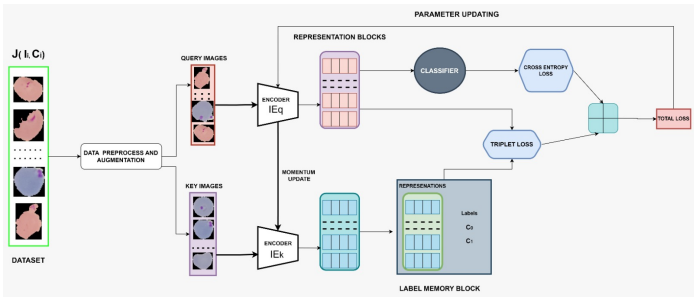

Fig. 2 Triplet Loss aided LCL

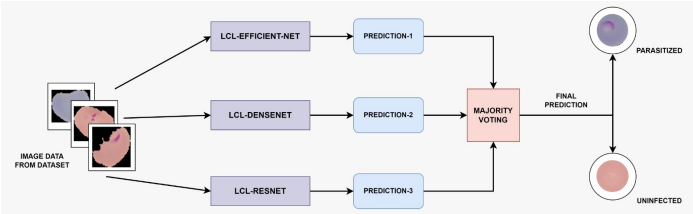

Fig. 3 Majority Voting Based LCL-Ensembling

4.2 Majority voting based Ensembling

In our LCL-based architecture, we have shown that EfficientNet, Densenet, and Resnet are the top 3 performing backbones. So we apply majority voting to ensemble on the predictions of the 3 LCL-backbones to get more efficiency in the malaria cell image classification problem. Our overall method is shown in Figure 3.

5 Results

5.1 Dataset Description

This Collection is extracted from the NIH website[9]. A professional slide reader at the Mahidol-Oxford Tropical Medicine Research Unit in Bangkok, Thailand, painstakingly tagged the photos. The annotations and de-identified photos are stored at NLM. 27,558 cell pictures total—equal numbers of parasitized and uninfected cells—make up the dataset. Plasmodium was present in positive samples, but other sorts of items, such as staining artifacts or contaminants, were absent from negative samples. Our dataset is split into a train set (70%), a validation set (20%), and a test set (10%).

5.2 Experimental setup

The computer used for all of the trials has a 12-GB NVIDIA Tesla T4 GPU, and Python 3.6 is the programming language used. The deep learning model is created in the Tensorflow environment using the Keras library.

5.3 Parameters and hyperparameters

The values of the parameters and hyperparameters that we employed in our method will be covered in this section. The values are shown in Table 1.

5.4 Evaluation parameters

A test gets results in 4 ways: correctly identified positive (True Positive-TP), correctly identified negative (True Negative-TN), mistakenly called negative (False Negative-FN) and mistakenly called positive (False Positive-FP). We assess our model using established measures such as F measure, Accuracy, Precision, and Recall.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \tag{6}$$

Table 1 Parameters and Hyperparameters Used in Our Method

Parameter/Hyperparameter	Value
Batch size	32
Optimizer	SGD
Weight decay	10^{-4}
Temperature hyperparameter	0.5
α	0.9
Learning rate	0.001
Training epochs	10
δ	0.5
γ	0.3

$$\text{Precision(P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall(R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{F measure} = \frac{2 * \text{R} * \text{P}}{\text{R} + \text{P}} \tag{9}$$

5.5 Our Result and visualization

We have got 97.53% as overall accuracy in the test dataset. Similarly, we got 97.4% as precision, recall, and F-score. Here, class 0 means malaria infected cell and class 1 means uninfected cell. For a better understanding of the readers, we have computed the confusion matrix as well as the ROC curve in Figure 4 and 5 respectively. The AUC score is 0.99.

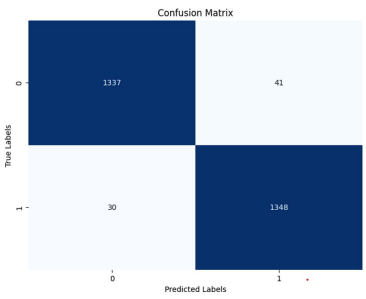

Fig. 4 Confusion Matrix

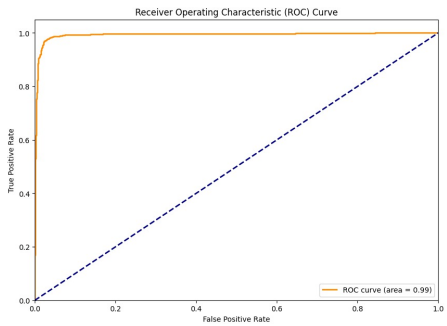

Fig. 5 ROC Curve

Now, we have shown TSNE to prove the efficacy of our classification algorithm in Figure 6 and 7.

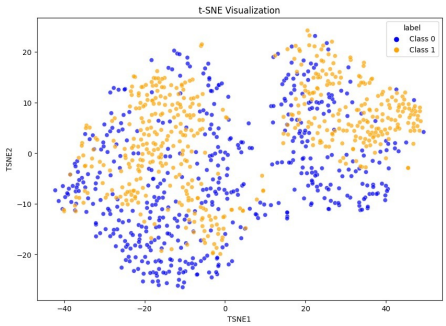

Fig. 6 t-SNE Visualization Before Applying Our Method

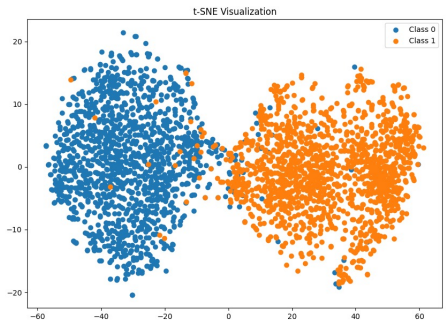

Fig. 7 t-SNE Visualization After Applying Our Method

5.6 Comparison with different backbones

First, we have compared our methods with different CNN architectures without LCL and with LCL setting in terms of accuracy, precision, recall, and F-score in table 2 and 3 respectively.

Table 2 Performance Comparison Without LCL Setting

Backbone	Accuracy	Precision	Recall	F1 Score
ResNet50	86.00%	82.00%	83.00%	85.00%
MobileNetV3	96.02%	97.00%	97.00%	97.00%
DenseNet121	96.23%	96.00%	96.00%	96.00%
VGG	91.80%	92.00%	91.00%	91.00%
EfficientNetB3	97.00%	97.00%	97.00%	97.00%
Our Method	**97.53%**	**97.40%**	**97.40%**	**97.40%**

Table 3 Performance Comparison With LCL Setting

Backbone	Accuracy	Precision	Recall	F1 Score
ResNet50	97.02%	97.00%	97.00%	97.00%
MobileNet	96.70%	96.75%	96.70%	96.70%
DenseNet	97.31%	97.30%	97.30%	97.30%
VGG	96.22%	96.20%	96.15%	96.22%
EfficientNet	97.06%	96.10%	96.10%	96.10%
Our Method	**97.53%**	**97.40%**	**97.40%**	**97.40%**

5.7 Comparison with SOTA

We have compared our method with various state of the art methods. Table 4 proves how our method outperforms other works in terms of accuracy.

Table 4 Comparison with State-of-the-Art Methods

Method	Accuracy	Precision	Recall	F1 Score
Liang et al.[1]	91.99%	95.12%	89.00%	90.24%
Reddy et al.[5]	95.40%	-	-	-
Das et al.[8]	83.50%	88.51%	96.62%	92.38%
Ma et al.[7]	75.00%	-	-	-
Manning et al.[6]	92.20%	-	-	-
Our Method	**97.53%**	**97.40%**	**97.40%**	**97.40%**

6 Takeaways

Our research created a new label contrastive learning architecture with triplet loss and majority voting ensembling method, testing several backbones and selecting Resnet, Densenet and Efficientnet for its high accuracy (97.53%). This label contrastive approach, combining triplet loss and cross-entropy, achieved excellent precision, recall, and F-score. We believe this approach, along with careful hyperparameter tuning, paves the way for future advancements in label contrastive learning. Future work exploring dual backbones, and various ensembling techniques in broader applications in medical imaging and other areas is planned. This research lays a strong foundation for further exploration in this rapidly evolving field.

References

[1] Z. Liang, A. Powell, I. Ersoy, et al., "CNN-based approach for malaria diagnosis," in *Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, pp. 493-496, 2016.

[2] N. Odu, R. Prasad, C. Onime, "Predicting malaria incidence using machine learning and climate variability," *Informatics in Medicine Unlocked*, vol. 22, 2021, Art. no. 100508. doi: 10.1016/j.imu.2020.100508.

[3] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, G. Thoma, "An overview of image analysis and machine learning techniques for malaria detection," *Translational Research*, vol. 194, pp. 36-55, 2018. doi: 10.1016/j.trsl.2017.12.004.

[4] B. Narayanan, R. Ali, R. Hardie, "Comparative performance analysis of machine learning architectures for malaria detection in cell images," presented at *SPIE Defense + Commercial Sensing*, 2019. doi: 10.1117/12.2524681.

[5] A. S. B. Reddy, D. S. Juliet, "Malaria cell-image classification using transfer learning with ResNet-50," in *Proc. IEEE Int. Conf. on Communication and Signal Processing (ICCSP)*, pp. 945-949, 2019. doi: 10.1109/ICCSP.2019.8697909.

[6] K. Manning, X. Zhai, W. Yu, "A machine learning-based system for malaria diagnosis using image analysis," *Digital Communications and Networks*, vol. 8, no. 2, pp. 132-142, 2022. doi: 10.1016/j.dcan.2021.07.011.

[7] M. Ma, "Experiments on malaria cell classification using active machine learning," in *Proc. IEEE 3rd Int. Conf. on Frontiers Technology of Information and Computer (ICFTIC)*, pp. 133-137, 2021. doi: 10.1109/ICFTIC54370.2021.9647411.

[8] D. K. Das, M. Ghosh, M. Pal, A. K. Maiti, C. Chakraborty, "Automated screening of malaria parasite using machine learning techniques and light microscopy images," *Micron*, vol. 45, pp. 97-106, 2013. doi: 10.1016/j.micron.2012.11.002.

[9] S. Rajaraman, S. K. Antani, M. Poostchi, K. Silamut, M. A. Hossain, R. J. Maude, S. Jaeger, G. R. Thoma, "Using pre-trained convolutional neural networks for enhanced malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, Art. no. e4568, 2018. doi: 10.7717/peerj.4568.

[10] H. Yang, J. Li, "Label contrastive learning for image classification," *Soft Computing*, vol. 27, pp. 13477–13486, 2023. doi: 10.1007/s00500-022-07808-z.

[11] K. He, H. Fan, Y. Wu, et al., "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp. 9729–9738, 2020.

[12] Y. Ren, J. Bai, J. Zhang, "Label contrastive coding for graph neural network-based graph classification," presented at *Database Systems for Advanced Applications*, 2021.