RESEARCH PAPER

Odin-Net: Optimized Deep Imaging Network with Svm for Enhanced Malaria Detection

A. Pandiaraj[1*], V. Angayarkanni[1], P. Rama[1], P.S. Abarna[2] and Kavitha Kamatchi[3]

[1]*Department of Computing Technologies, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur-603203, Chengalpattu District, Tamil Nadu, India*
[2]*Department of CSBS, PSNA College of Engineering and Technology, Tamil Nadu, India*
[3]*Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology, Tamil Nadu, India*
pandi.mnmjain@gmail.com[1], angayarv@srmist.edu.in[1], ramaponnuvarman@gmail.com[1], abarnaps@psnacet.edu.in[2] and kkaviwt@gmail.com[3]*
[*]***Corresponding Author:*** *pandi.mnmjain@gmail.com*

Received: 24th April, 2026; Revised: 20th May 2026; Accepted: 30th June, 2026; Available Online: 10th July, 2026

ABSTRACT
Malaria remains a major global health concern, as evidenced by the significant burden continued to be placed on public health systems and governments around the world for malaria in resource-limited settings. The ability to make a timely and accurate diagnosis of a condition such as malaria is critical, and in many instances, is made possible by using microscopy to examine blood smear images; however, the manual interpretation of these images can be limited due to a lack of trained personnel, a high workload, and limited resources. The use of deep learning for the automatic detection of malaria cells in blood smear images has shown great potential, although there are still challenges surrounding generalization, computational efficiency, and boundary-sensitive feature extraction. To address these challenges, we have developed a novel light-weight hybrid diagnostic framework for the classification of malaria cells in blood smear images called ODIN-Net. This framework utilizes a combination of optimized Canny edge detection and the Roberts operator for edge-aware feature enhancement, a multi-scale kernel-based stacked convolutional neural network (CNN), parallel pyramid pooling for the aggregation of contextual features and a support vector machine (SVM) classifier that is optimized using Bayesian optimization for final decision making. To comparatively evaluate the proposed framework, we performed experimentation on the NIH Malaria Cell Image Dataset and divided the dataset into three subsets consisting of 70% for training, 15% for validation, and 15% for independent testing with a balanced class distribution. The model's performance was evaluated through accuracy, precision, recall, F1-score, analysis of training time, and statistical significance testing. The results of the independent testing dataset indicate that ODIN-Net was able to achieve an overall accuracy of 97%, with a recall of 96%, precision of 94%, and F1-score of 95% respectively. Furthermore, ODIN-Net outperformed all other tested deep learning-based classification models (VGG-19, VGG-16, EfficientNet-B2, ResNet-50); in fact, ODIN-Net produced the shortest training time of 54,775 ms of all models under identical hardware and training conditions.

Keywords – Malaria, Convolution Neural Network, Deep learning, ODIN-net, Support Vector Machine

How to cite this article: Pandiaraj A, Angayarkanni V, Rama P, Abarna PS, Kamatchi K, Odin-Net: Optimized Deep Imaging Network with Svm for Enhanced Malaria Detection. Int J Drug Deliv Technol. 2026;16(7): 231-245. DOI: 10.25258/ijddt.16.7.26
Source of support: Nil.
Conflict of interest: None

1. INTRODUCTION

Malaria is a serious illness that is caused by an infection due to parasites. The parasite is transmitted from an infected female Anopheles mosquito to another living organism through mosquito bites [1]. Untreated, malaria can cause death or cause serious problems, such as seizures, organ failure, coma, etc. [3]. In order to ensure the effective management of malaria and to ensure effective control of malaria, the World Health Organization (WHO) recommends a continuous surveillance programme (i.e., monitoring the occurrence of malaria) and use of diagnostic methods to detect malaria in areas with a high incidence of malaria [2]. As a result, the ability to identify the presence of malaria-infected blood cells in a timely and accurate manner plays an important role in the ability to prevent the spread of malaria and improve malaria patient outcomes [4].

While many malaria diagnostic methods such as light microscopy and rapid-tests continue to be used widely in clinical practice, these methods have limitations related to the requirement for trained personnel to perform the tests, susceptibility to environmental conditions, and variability

**Author for Correspondence: pandi.mnmjain@gmail.com*

in accuracy of test results depending on the operator of the test [5]. Therefore, there is a need to explore automated and intelligent diagnostic systems that are able to provide consistent and scalable performance. The use of deep learning, specifically Convolutional Neural Networks (CNN) in medical imaging has shown great potential in automating the extraction of complex spatial features from images and increasing the accuracy of diagnosis [6-10].

Although there have been promising results with CNN-based malaria cell detection, there are still several challenges that remain, including overfitting of the models, increased computational complexity, lack of boundary sensitivity, and limitations of generalizability when using diverse imaging conditions.In addition, many deep learning frameworks focus on achieving high classification accuracy without regard to whether or not they can be deployed in the real world, creating a major limitation for these types of systems to be utilized in clinical environments that have limited resources. Solutions to these aforementioned limitations require lightweight and robust architectures that are capable of balancing diagnostic accuracy with computation efficiency and real-life deployability.

In this paper, ODIN-Net is proposed as a novel lightweight convolutional neural network (CNN)-based framework for the effective and efficient classification of malaria positive cells in thin blood smear image data. To help achieve this end, ODIN-Net uses several enhanced and optimized components to increase the feature mapping as well as the performance of the classification. In particular, the following are incorporated into the framework:

(I) A compact stacked CNN architecture utilizing selectively chosen kernel sizes with dropout regularization to limit overfitting and the total computation burden;

(II) A Canny Edge Detection technique enhanced with the Roberts operator for improved edge-contextualized feature extraction, and a gradient-based structural reconstruction of the morphological structures of parasites;

(III) A Hybrid CNN-SVM Classification model that uses the deep features generated from the CNN and classifies them using an SVM optimized using Bayesian Optimization for increased generalization and decision robustness; and

(IV) A Parallel Pyramid Pooling mechanism that aggregates all local to global contextual features through multi-scale spatial aggregation.

ODIN-Net was thoroughly evaluated on the NIH Malaria Cell Image dataset in comparison with established deep learning models such as VGG-16, VGG-19, EfficientNet-B2, and ResNet-50.

Recently, developed architectures like MAS-Net [11], Deep-CNN-RF [12], MozzieNet [13], and ensemble approaches based on convolutional neural networks (CNNs) [14] have demonstrated positive results, however, they continue to have limitations due to computational complexity, minimal edge detection capability and diminished generalizability among non-homogeneous datasets. Although attention-based architectures provide accurate predictions, the challenge lies with maintaining target preservation at smaller sizes and computationally efficient deployment. This is not the case for ensemble methods, which add an excessive computational burden. In contrast, the focus of ODIN-Net is to find the most optimal level of accuracy, efficiency and robustness for the diagnoses made using the results of the network within resource limited environments.

The paper is organized as follows. Section 2 contains a review of the literature on prior deep learning detection efforts for the detection of malaria. Section 3 identifies three critical limitations of existing technology and the research gap that this current research seeks to fill. Section 4 provides a detailed description of the type of data used in experimentation. Section 5 outlines all of the data preprocessing and augmentation techniques. Section 6 provides a description of the architecture and the mathematical formulas of ODIN-Net. Section 7 contains a discussion of the defined training strategy and hybrid CNN-SVM optimization. Section 8 presents the results of experimentation and a comparison of ODIN-Net with other CNN models. Finally, Section 9 does a summary of this work and future research directions are presented.

2. RELATED WORK

The research [11] focused on addressing challenges in malaria cell detection using advanced computer vision techniques. The authors proposed the Multi-Level Attention Split Network (MAS-Net) to improve malaria cell detection performance, introducing innovations such as the Contextual Attention Structure (SPCot) to address information loss for small targets, the Multi-Scale Receptive Field Detection Head (MRFH) to match targets of different scales, and Model Compression (PAGCP) for practical deployment. Experimental evaluation on the NLM—Malaria Dataset demonstrated MAS-Net's promising accuracy of 75.9% in identifying human blood smear samples infected with Plasmodium vivax, showcasing its potential in enhancing medical diagnosis and reducing workload in healthcare settings. A new method [12] for the automated identification of Plasmodium malaria parasites in microscopic images is crucial for precise diagnosis and treatment in medical settings. This approach tackles challenges such as limited

training data, class imbalances, and variability in parasite development stages by integrating a Deep Convolutional Neural Network (Deep-CNN) model with a Random Forest (RF) algorithm. The Deep-CNN-RF method, augmented with specialized domain knowledge, enhances learning efficacy. A distinctive alteration in the pooling layer that incorporates Global Average-Pooling (GAP) eliminates the need for an additional flattening layer, improving visualization of the parasite area. Moreover, accurate detection of parasite boundaries is achieved using the Canny edge detection technique. The research [13] introduces MozzieNet, a tailored CNN model designed to boost the detection of malaria parasites in microscopic images of blood smears. This model achieves enhanced robustness and generalization by optimizing hyperparameters and integrating strategies such as data augmentation, batch normalization, and dropout to mitigate overfitting. Employing the publicly accessible NIH malaria dataset, the team reported impressive results: a classification accuracy of 96.73%, a recall rate of 97.90%, a precision of 95.67%, an area under the curve (AUC) of 99.35%, and an F1 score of 96.77%. The study also conducted feature map visualizations and Grad-CAM analysis on the MozzieNet model to identify and scrutinize the critical regions for making accurate predictions. The research by [14] aimed to utilize machine learning and deep learning algorithms, specifically AlexNet and DenseNet, to identify malaria-infected cells. The dataset consisted of 27,558 images, with 13,780 infected cells sourced from the NIH Website. After selecting a subset of the dataset, the algorithms were trained to classify cells as parasitized or uninfected. A thorough analysis of classifiers, including AlexNet known for complex feature learning and DenseNet characterized by densely connected layers for enhanced feature reuse, was conducted. Hyperparameters were tuned to optimize performance, and an ensemble technique was employed to combine classifier outputs. The study achieved impressive accuracy rates of 93% for training data and 94% for testing data, showcasing the efficacy of machine learning and deep learning models in accurately identifying malaria-infected cells and their potential for aiding medical image analysis and disease diagnosis tasks. The research [15] aimed to tackle the challenge of implementing convolutional neural networks (CNNs) on resource-constrained embedded systems by introducing a lightweight variant named hierarchical multi-scale feature fusion MobileNet, building upon the principles of MobileNetV1. While MobileNetV1 reduced complexity using depthwise convolution and a stride of 2, potentially compromising spatial information, HMFF-MobileNet integrated a hierarchical multi-scale feature fusion (HMFF) module and depthwise dilated separable convolution (DDSC) layers. These innovations allowed for efficient feature learning with reduced parameters and computational costs while maintaining spatial context. Experimental validation across two datasets showcased notable reductions in parameters (65.02%) and computational costs (39.78%) compared to MobileNetV1, affirming HMFF-MobileNet's potential for real-world applications in image classification and related tasks on resource-limited platforms.

3. RESEARCH GAP

The study conducted by Xiong, Z., et al. [11] identified several key research gaps in the field of malaria cell classification. One crucial challenge they acknowledged is the presence of class imbalance within the training dataset, which can lead to biased model performance. To overcome this limitation, the researchers plan to address it by expanding the dataset through the utilization of unsupervised learning techniques. [12] identified the need to create a more balanced representation of malaria-infected and uninfected cells, thereby improving the model's ability to accurately classify both classes. Additionally, the researchers [13] highlighted their intention to broaden the detection scope beyond solely identifying malaria parasites. Their objective is to expand the scope to include malignant malaria parasites, other malaria species, and additional bloodborne diseases that necessitate microscopic analysis. This expansion is vital for enhancing the model's versatility and applicability in real-world scenarios where diverse cell types and diseases may be encountered. Furthermore, while the previous works utilized a substantial dataset, the researchers [14-20] emphasized the importance of further investigation using larger and more diverse datasets. These datasets should encompass various parasite species, stages, and image qualities to enhance the generalizability and robustness of trained models.

The proposed ODIN-Net framework has been systematically designed to address the critical challenges identified in previous studies. Class imbalance: To overcome class imbalance, the model employs a balanced NIH dataset containing an equal number of infected and uninfected cell images. This ensures unbiased learning and improves classification fairness across both classes. Limited generalizability: ODIN-Net enhances generalization through a hybrid CNN–SVM integration, where the CNN extracts deep hierarchical features while the SVM, optimized using Bayesian optimization, provides robust decision boundaries that prevent overfitting and improve adaptability to unseen samples. Computational inefficiency: To tackle inefficiency and high training cost,

ODIN-Net introduces a lightweight CNN design with optimized kernel sizes and fewer parameters, reducing computational complexity without sacrificing accuracy. Feature extraction limitations: The integration of optimized Canny edge detection and the Roberts operator

enables sharper boundary detection and better gradient-based representation of parasite morphology, enhancing interpretability and diagnostic reliability. Multi-scale feature capture: The use of parallel pyramid pooling ensures effective aggregation of both local and global features, addressing the limited contextual understanding observed in earlier models and improving overall detection robustness.

4. DATA SET

NIH Malaria Cell Image Dataset was employed for model training and evaluation. This open-access dataset, provided by the U.S. National Institutes of Health (NIH), contains 27,558 microscopic images of human blood smears, equally divided into infected and uninfected classes. Each image was captured from thin blood film slides and carefully labeled by expert microscopists. Owing to its balanced class distribution, verified ground truth, and widespread use in prior malaria detection studies, this dataset serves as a gold standard benchmark for blood-cell–based malaria classification, ensuring the reliability and reproducibility of experimental results. There are 27,558 images in the dataset, divided into two folders: Infected and Uninfected. With roughly 13,800 files per class, a balanced distribution of samples is maintained. Illustrated in Figure 1 are two examples of types of samples that represent a comparison illustrating a visual difference between two different types of cells. The first sample is an infected red blood cell (RBC) with malaria parasites; the second sample is a non-infected red blood cell.

Carefully chosen from a variety of sources, including areas where malaria is prevalent, these images capture the disease. Class imbalance problems that are frequently found in medical imaging datasets are mitigated by the balanced nature of the collection, which makes it easier to represent both diseased and uninfected samples fairly. The dataset promotes better model generalization across various patient demographics, illness severities, and imaging conditions by combining a wide variety of malaria-infected and uninfected samples. This improves the model's capacity to correctly categorize unknown data from different populations. The curated balanced class dataset sourced from the official NIH Website represents a valuable resource for advancing malaria detection research. In order to evaluate the accuracy of the new model without any bias, the NIH Malaria Cell Image Dataset has been divided into three distinct subsets based on mutual exclusivity: 70% for training, 15% for validation, and 15% for independent testing. All three subsets contained equal distributions of both infected and uninfected cells.

An architecture (ODIN-Net) based on the concept of training an architecture from the ground up without the benefit of any pre-trained or transfer learning weight. This gives the new model the ability to learn new representations that are unique to malaria blood smear images. To prevent the possibility of overfitting and promote greater ability to generalise, the authors created augmentation for data only within the training sets, using techniques such as rotation, flipping, and zooming.

While there were some adjustments to certain hyperparameters, the proposed system did use the validation data for adjusting hyperparameters and stopping early. Bayesian optimization method were used to adjust the parameters for the SVM classifier. The final performance comparisons were made on the test data and it was never considered for use during any of the modelling process.

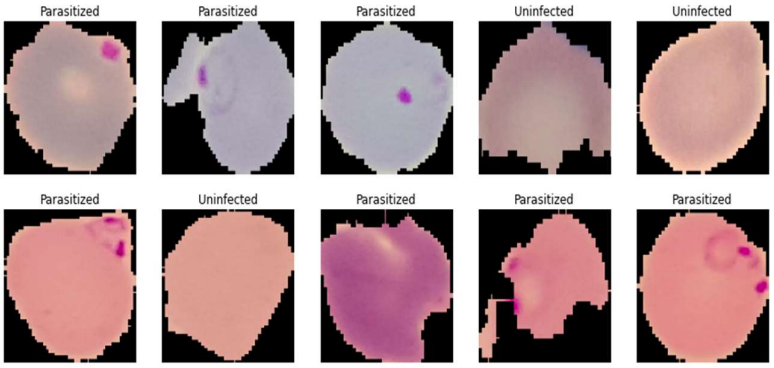

Figure 1. Samples from the NIH Malaria Cell Image Dataset

The above figure displays some representative microscopic images of both uninfected (non-parasitized) red blood cells and parasitized (infected) ones obtained via microscope slides containing thin layers of blood.

5. DATA PRE-PROCESSING AND AUGMENTATION

The system employed a variety of strategies to enrich the training data before we train our model. In order to do this, the original images must be transformed by applying

techniques like rotation, flipping, and zooming in order to produce more training instances. This expands the training set and improves the model's ability to generalize to previously unseen images. The system used the Optimum Canny Edge Detection technique to extract useful features from the input images for the malaria cell classifier. In order to precisely identify the edges of objects in the photos while reducing undesired noise, this required implementing the Canny edge detection algorithm with noise reduction to accurately identify the edges of objects in the images while minimizing unwanted noise. Firstly, the procedure was applied the well-known Canny edge detection technique, which finds edges in images with remarkable accuracy.

The system tuned the detection method to fit the features of malaria cell images by modifying the parameters of the algorithm, such as the edge detection thresholds. We then included noise reduction algorithms to improve the accuracy of feature extraction and further optimise the edge identification procedure. This ensured that just the pertinent edges of the malaria cells were kept, reducing any abnormalities in the images. Further, we computed the gradient direction and amplitude of the identified edges using the Roberts operator. The Roberts operator computes the gradient by utilising two 2x2 convolution kernels, making it a straightforward and effective edge detection technique. This provided valuable information about the edges' sharpness and orientation, allowing us to characterize the malaria cells more effectively based on their structural properties. The computed gradient magnitude and direction provide valuable information about the edges' intensity and orientation in the image, which are essential features for malaria cell classification. By applying the Roberts operator as part of the edge detection process, the algorithm can effectively highlight the edges of malaria cells, facilitating subsequent feature extraction and classification tasks shown in figure 2.

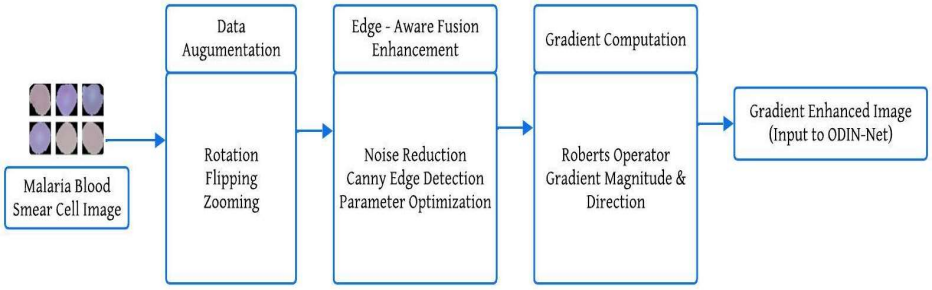

Figure 2. Pipeline to Preprocess, Augment and Extract Features from Data (using Edge Aware Filtering)

In order to increase the generalization of the model trained on mirrored, rotated and zoomed versions of the original malaria images; therefore, pre-processing creates noise-free images, along with optimized Canny edge detection that extracts the prominent edges of malaria cells, after which we perform post-processing on these edges by removing undesired spurious edges from them. Gradient magnitude and gradient direction using the Roberts operator provide enhanced representations of structures/boundaries. The resulting images after all of these steps are then fed into the proposed ODIN-Net Framework.

6. PROPOSED CNN ARCHITECTURE

Figure 3 shows the Convolutional layers serve to extract hierarchical features from the input images. Each layer applies a series of learnable filters to the input, generating feature maps that emphasize various features of the image. The mathematical formula used to compute the output of a convolutional layer is as follows Equation (1):

$$L=f(W*X+b) \quad (1)$$

where:

L represents the output feature map,

X denotes the input feature map,

W represents the filter weights,

b is the bias term,

$*$ denotes the convolution operation,

f is the activation function.

In the classification of malaria cells, convolutional layers are crucial for extracting distinctive features from the cell images provided as input. In the proposed Convolutional Neural Network (CNN) architecture for malaria cell classification, stacking convolutional layers plays a crucial role in allowing the model to develop layered representations of the input cell images. Stacking multiple convolutional layers enables the network to progressively extract more abstract and complex features from the raw pixel data, leading to more discriminative representations that aid in accurate classification. However, to keep the model lightweight, the filter and the kernel size are carefully chosen. The use of a small kernel size, such as 3×3, in the initial convolutional layers ensures that the

model can extract relevant features efficiently without adding excessive computational complexity. Additionally, limiting the number of filters in each layer helps control the parameters and reduces the model's memory utilization. The first convolutional layer has 16 filters with a 3×3 kernel size, followed by max-pooling and dropout regularization. The subsequent layers have 32 and 64 filters, respectively, with larger kernel sizes for multi-scale analysis. The primary purpose of stacking convolutional layers is to facilitate hierarchical feature extraction. Each convolutional layer will detect specific features at different levels of abstraction. By stacking multiple layers, the network can capture layered interpretations of the input images, starting from micro features like edges and textures to macro features like shapes and structures of malaria cells. Considering the CNN architecture for malaria cell classification, let's denote the output of each convolutional layer as O_i, where i represents the layer index. Here's how the stacking can be represented: Given an input image X, the output of the first layer (after applying the ReLU) can be represented as Equation (2):

$$O1=\text{ReLU}(W1*X+b1) \quad (2)$$

Where:

- $W1$ represents the learnable weights (filters) of the first convolutional layer.
- $b1$ is the bias term.

The output of the first max-pooling layer can be expressed as:

$$O_2=\text{MaxPooling}(O_1)$$

Similarly, for the subsequent convolutional and max-pooling layers, we have:

$$O_3=\text{ReLU}(W_2*O_2+b_2)$$

$$O_4=\text{MaxPooling}(O_3)$$

$$O_5=\text{ReLU}(W_3*O_4+b_3)$$

$$O_6=\text{MaxPooling}(O_5)$$

where:

- W_2 and W_3 denote the learnable weights of the next two convolutional layers, respectively.
- b_2 and b_3 are the bias terms.
- The ReLU activation function is applied after each convolutional layer.

By stacking these convolutional and max-pooling layers, the CNN learns hierarchical representations of the input malaria cell images, capturing increasingly abstract features and patterns relevant to malaria cell classification.

6.1 Pooling Layers

Following each convolutional layer, a max-pooling layer utilizing a 2×2 window is used. Pooling layers serve to reduce the size of the feature maps generated by the convolutional layers, thereby lowering spatial dimensions and capturing the most significant features. Max-pooling retains the most active pixel values within each pooling window, providing translational invariance and reducing computational complexity as in Equation (3):

$$P_{ij} = \max_{m,n}(F_{(2i+m)(2j+n)}) \quad (3)$$

Here, P_{ij} represents the pooled output at position (i,j), and $F_{(2i+m)(2j+n)}$ represents the feature map value at position $(2i+m,2j+n)$.

6.2. Dropout Layers

To combat overfitting, dropout regularization is implemented following the initial two convolutional layers and the fully connected layer, with dropout rates of 0.2, 0.3, and 0.5 respectively. This method involves intermittently deactivating a portion of the neurons during training, which encourages the network to develop more robust and redundant pathways, thereby enhancing its ability to generalize. Essentially, dropout regularization temporarily nullifies a subset of the input neurons at each training step.

The output y of a dropout layer can be computed as:

$$y = \begin{cases} \frac{x}{1-p} & \text{if neuron is retained (probability } 1-p) \\ 0 & \text{otherwise} \end{cases}$$

Here, x represents the input to the dropout layer, and p represents the dropout rate.

6.3. Fully Connected Layer

Following the convolutional layers, the resulting feature maps are transformed into a one-dimensional vector. Subsequently, a fully connected layer equipped with 64 neurons and employing ReLU activation is introduced. This layer is designed to integrate the features extracted from the convolutional stages, enabling the learning of more abstract features. The incorporation of ReLU activation adds non-linearity to the model, facilitating its ability to capture complex patterns. The output of a fully connected layer can be represented as in Equation (4):

$$y=\text{ReLU}(W_x+b) \quad (4)$$

In this setup, 'x' denotes the input data to the fully connected layer, 'W' stands for the weight matrix, 'b' indicates the bias vector, and ReLU refers to the Rectified Linear Unit activation function used.

6.4. Output Layer

The output layer in this architecture plays a crucial role in making predictions for malaria cell detection. The final layer includes a single neuron equipped with a sigmoid

activation function. This type of activation is frequently utilized for binary classification because it compresses the model's output to a value between 0 and 1. The output of the sigmoid neuron represents the probability that the input image contains a malaria cell. If the output value is close to 0, it indicates a low probability of malaria cell presence, while a value close to 1 signifies a high probability. During training, the model's performance is evaluated using the binary cross-entropy loss function.

6.5. Parallel Pyramid pooling

In our research, we employ pyramid pooling technique to enhance the feature aggregation process in our model. These technique allow us to aggregate features from different spatial bins at multiple scales, thereby improving the model's ability to understand complex patterns in the input data. While Conventional Convolutional Neural Networks (CNNs) capture local context with their convolutional layers, they do not encode as much global context as ODIN-Net does through its parallel pyramid pooling module that aggregates several different levels of spatially-related features within the high-level convolutional feature maps. Pyramid pooling involves dividing the input feature maps into multiple bins or regions, each capturing information at a different scale. These bins are typically arranged in a pyramid-like structure, with larger bins capturing coarse-grained information and smaller bins capturing finer details. By aggregating features from these different scales, the model can capture a wide range of spatial information, from global context to local details.

Let the output feature map from the final convolutional layer is:

$F \in R^{H \times W \times C}$

where:
H and W denote the spatial dimensions, C represents the number of channels.

The pyramid pooling module partitions the feature map {F} into spatial bins at multiple scales:

$S=\{1\times1, 2\times2, 4\times4\}$

For each scale $s \in S$, spatial pooling is applied as:

$P_s = Pools(F)$

where Pools(·) denotes adaptive max-pooling that resizesF into s×s spatial regions.

Each pooled feature map P_s is then flattened into a vector:

$v_s = Flatten(P_s)$

The final pyramid-pooled feature representation is obtained by concatenation of vectors from all scales:

$$\mathbf{v}_{pp} = \left[\mathbf{v}_{1\times1} \,\|\, \mathbf{v}_{2\times2} \,\|\, \mathbf{v}_{4\times4}\right]$$

where ‖ denotes the concatenation operation.

To denote the flattened CNN feature vector be:

$v_{cnn} = Flatten(F)$

The final feature representation which is provided to the classifier is:

$v_{final} = [v_{cnn} \| v_{pp}]$

The concatenated feature vector v_{final} is passed to a Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel:

$K(x,x') = \exp(-\gamma \|x - x'\|2)$

where γ is the kernel width parameter optimized using Bayesian optimization.

We have used pyramid pooling as a parallel pathway alongside the main convolutional layers. This pathway can process the input feature maps at different spatial resolutions and aggregate features independently before merging them with the main pathway. This allows the model to capture multi-scale information more effectively and improve its overall performance.

The CNN part of the proposed ODIN-Net framework works as a deep feature extractor, which operates independently from the classifier function of ODIN-Net. For training purposes, CNNs use a temporary sigmoid activated output layer with binary cross-entropy as a learning guide and to achieve convergence reliably. The sigmoid generated output is not utilized for final inference. After the training is complete, the CNN discards the sigmoid output layers, using instead the deep features produced by the last fully connected layer of the CNN in support of the SVM classification process.

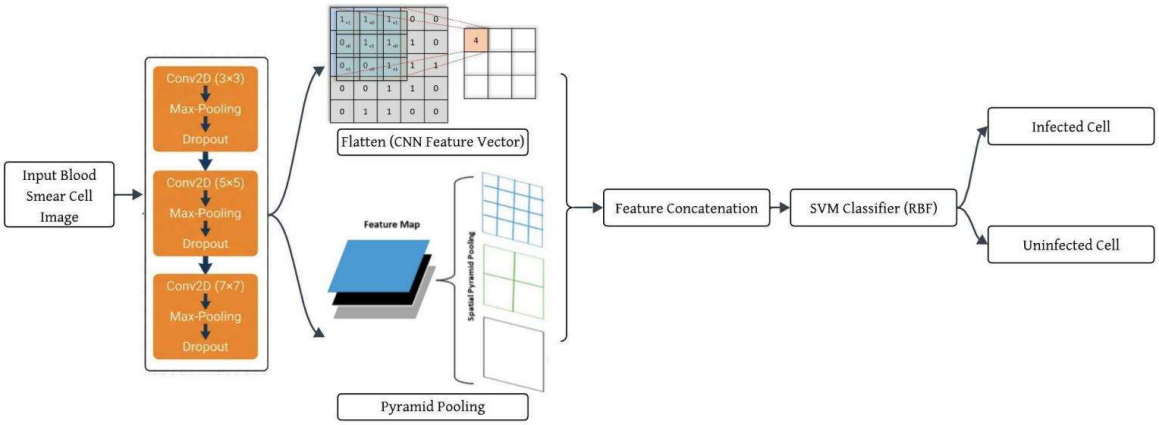

Figure 3. Architecture of the proposed ODIN-Net framework for malaria cell classification

A lightweight stacked CNN, which consists of three convolutional blocks (3×3, 5×5 and 7×7), is used as the model for a hierarchical feature extractor and a model regularizer via the use of max pooling and a dropout layer. Additionally, a parallel pyramid pooling module is applied to the final convolutional feature map to collect contextual multi-scale information, as well as the outputs of both the CNN and pyramid pooling branches used as the input to a Bayesian-optimized Support Vector classifier that distinguishes between infected versus non-infected malaria cells.

7. ENHANCED ODIN-NET TRAINING

The dataset used for training comprises 27,558 images, evenly distributed across two classes: Infected and Uninfected. Each class contains approximately 13,800 files, ensuring a balanced representation of samples for effective model training and evaluation. ODIN-Net is trained in two distinct steps. The first step focuses on training the ODIN-Net CNN module with a sigmoid function that outputs binary values and a binary cross-entropy loss to learn features of the malaria infected cells.

To avoid overfitting, the training process employs Early Stopping and Dropout regularization techniques.

In the second step, we discard the trained sigmoid output layer from ODIN-Net and instead extract deep CNN features from the last fully connected layer of the CNN. These deep CNN features will then be used for training the SVM classifier with a radial basis function kernel. Furthermore, Bayesian optimization will be utilized to optimize the hyperparameters for the SVM classifier (C and γ) using the validation dataset. All performance metrics reported in this study are from the overall training configuration using both the CNN and SVM.

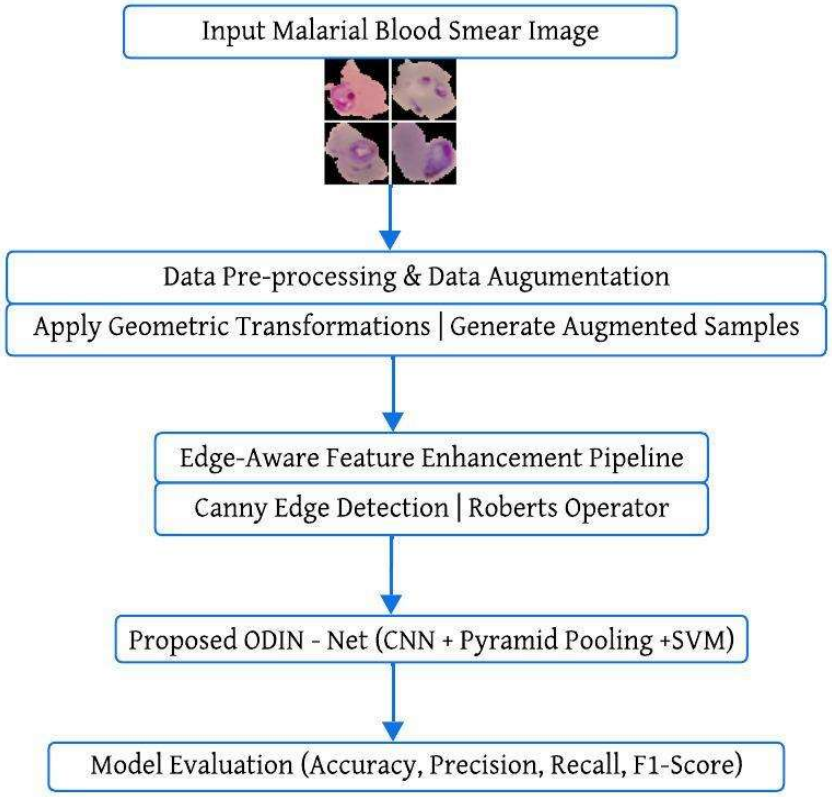

Figure 4. Overall workflow of the proposed ODIN-Net framework for malaria cell classification

This diagram shows the entire workflow of the malaria blood smear image processing system beginning with the acquisition of the malaria blood smear image, continuing with the preparation and augmentation of the data, the extraction of features from the images using edge-aware feature extraction algorithms, the classification of feature sets using the new ODIN-Net architecture, and finally the evaluation of models using established performance evaluation metrics.

Following training, features are extracted from the last CNN layer, specifically the Flatten layer, for both the training and validation datasets. These extracted features serve as input for training an SVM classifier with an RBF kernel, which is responsible for classifying malaria cell images based on the learned CNN features. The trained SVM model undergoes evaluation on the validation set to gauge its accuracy in malaria cell detection. Additionally, Bayesian Optimization is employed to fine-tune the hyperparameters of the SVM classifier, specifically focusing on the 'C' parameter. This optimization technique efficiently explores the hyperparameter space, ultimately identifying optimal values that maximize the SVM's accuracy on the validation set. Overall, the proposed architecture synergizes the efficiency of a lightweight CNN with the discriminative capabilities of an SVM classifier, resulting in a robust and accurate model for malaria cell classification. The incorporation of Bayesian Optimization further enhances the SVM's performance by optimizing its hyperparameters are shown figure 4.

The ODIN-Net model was ODIN-Net was created with Python 3.10 and made possible using TensorFlow 2.13 and Keras. All experimental runs were completed on a workstation utilizing an NVIDIA RTX 3080 GPU (10GB VRAM), Intel Core i9 processor (3.6GHz, 16 cores), and 32GB of RAM running Windows 11 (64 bit). This hardware and software configuration was maintained throughout the training and evaluation of the comparative baseline models. During training, the following hyperparameters and settings were applied:

- Learning rate: 0.001 (adaptive learning rate with decay of 1e−6 per epoch)
- Batch size: 32
- Number of epochs: 100
- Optimizer: *Adam* optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$

- Loss function: Binary cross-entropy
- Activation functions: ReLU for hidden layers and sigmoid for output layer
- Dropout rates: 0.2, 0.3, and 0.5 at different layers as described in Section 6
- Early stopping: Implemented with a patience of 10 epochs based on validation loss, to prevent overfitting and reduce unnecessary computation
- Learning rate scheduler: The learning rate was reduced by a factor of 0.2 if validation loss did not improve for 5 consecutive epochs.

Bayesian Optimization was exclusively utilized to tune hyperparameters for the SVM model. For the Convolution Neural Network Model hyperparameters, we performed an empirical selection based on model validation.

7.1 ODIN-Net Training and Classification Procedure

In developing the proposed ODIN-Net framework, a systematic two-step training and classification process has been designed and implemented that helps facilitate strong feature learning and optimal final classification. The initial input images obtained from blood smears from patients with malaria are then prepped through pre-processing and prepped through data augmentation through several techniques like edge-aware Canny edges were applied, optimized using a Roberts edge operator, so that edges of parasitic structures would appear to have strong outlines and gradients, helping in feature learning.

In the first step of developing the stacked CNN architecture, the augmented/edge-enhanced images would be supplied into the stacked CNN architecture to facilitate the training of hierarchical/multi-scale representations of features. To ensure that convergences are stable and that overfitting was avoided, the training was based upon a binary cross-entropy loss function, used an "early stopping" method, and also used a learning rate schedule. To help increase robustness over repeated training operations, multiple dropout layers were utilized. The stacked CNN will only provide a source of deep feature extraction; it will not provide any direct input from its output layer for final classification.

In the second step, feature vectors are extracted from the trained stacked CNN's final fully connected layer and combined with the multi-scale contextual features from the parallel pyramid pooling. All of the feature vectors are concatenated together into one complete feature vector for each input image. The feature vectors are then fed into a Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel for classification. The SVM hyper-parameters are determined using Bayesian Optimization techniques for optimal margin separation of infected/uninfected malaria cell classifications.

The final determination of performance of the SVM trained classifier is evaluated using standard performance metrics (overall accuracy, precision, recall, F1-scores, confusion matrices, and statistical significance testing) on both validation and independent test datasets. This approach of combining two training strategies for CNN and SVM provides the ODIN-Net framework with the ability to effectively merge deep feature learning with a stable classification, thus improving both the accuracy and computational efficiency of the SVM-based classifiers.

8. RESULTS AND DISCUSSION

8.1 Comparative Performance Evaluation

A comparison between ODIN-Net and traditional deep learning architectures; VGG-19, VGG-16, EfficientNet-B2, and ResNet-50 was illustrated in Figure 5 with regard to Accuracy, Precision, Recall, and F1-score. The numeric performance numbers indicate that ODIN-Net outperformed all of the above mentioned models on each of the four metrics. ODIN-Net attained an overall accuracy rate of 97%. The competing models' accuracies were lower: VGG-19 (96%), VGG-16 (95%), EfficientNet-B2 (93%), and ResNet-50 (89%). The corresponding precision, recall, and F1-scores for ODIN-Net were 94%, 96%, and 95%, respectively. Therefore, ODIN-Net successfully detected infected cells while making very few false predictions. Therefore, the results validate the better discriminative performance of ODIN-Net over traditional deep CNNs.

8.2 Computational Efficiency Analysis

In addition to performance classification, another key consideration for medical diagnostic applications in real world applications is Computational Efficiency. As shown in Figure 6, the overall time to train ODIN-Net compared to typical deep learning model types (e.g., VGG-19, VGG-16, EfficientNet-B2, and ResNet-50) were significantly lower at only 54,775 milliseconds in total training time required to train. Therefore, ODIN-Net represents an order of magnitude improvement over all of these conventional DL approaches. The reason for this significant improvement is due to the stacked-CNN architecture being a lightweight version of stacked-CNNs combined with the SVM classifier being a non-complex, high-speed classifier used for final outputs instead of using a fully-connected, deep network of connected nodes through deep layers. Therefore, compared to conventional fully-connected architectures, ODIN-Net has an enhanced efficiency related to its suitability in low-resource context and real-time diagnostic applications.

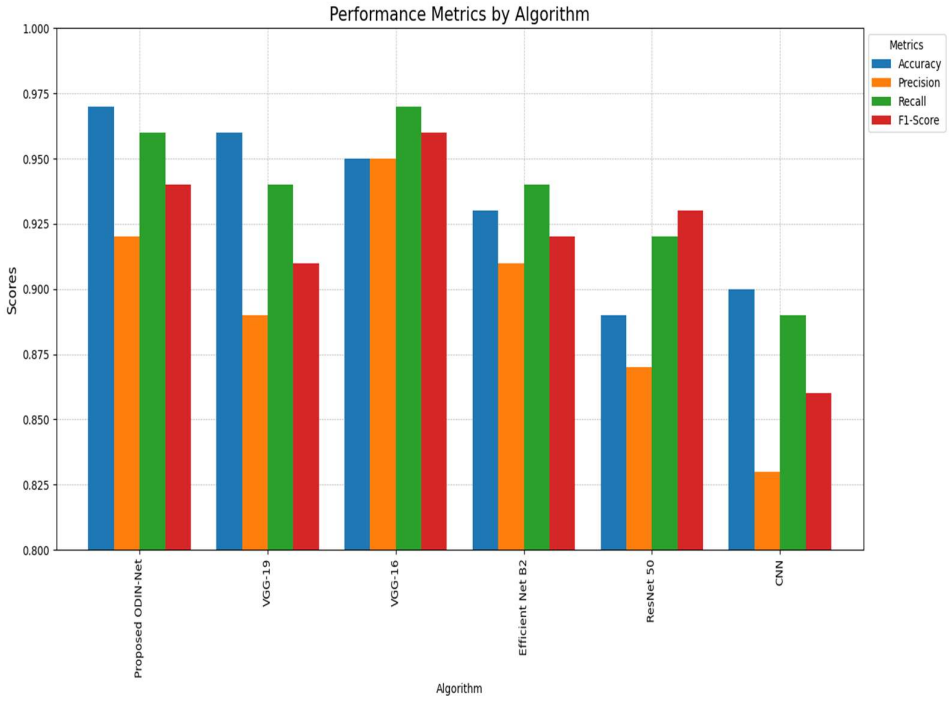

Figure 5. Graph Depicting Performance Metrics

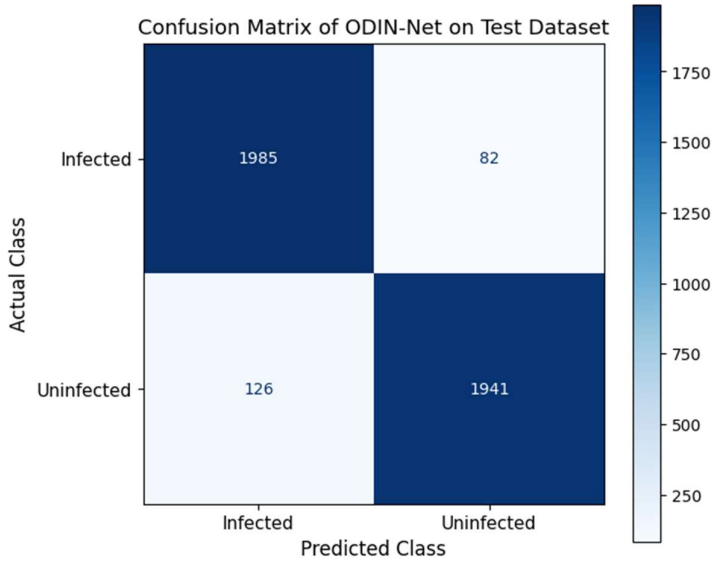

Figure 6. Confusion Matrix

Using the independent test dataset, the proposed ODIN-Net Model's confusion matrix (see Figure 6) shows that most malaria positive and negative blood smear images were accurately classified, resulting in a very high level of accurate true positive and true negative predictions. Additionally, because there were so few false negatives, this indicates that the model has a very high level of sensitivity to identifying malaria infected cells; this is important when using this model for clinical screening where as many missed diagnoses need to be minimized. In addition, the relatively low number of false positives indicates how well the model avoids classifying healthy cells as infected so that there will be fewer unnecessary follow up examinations. The confusion matrix presented represents a single best-performing test split and provides an overall classification accuracy of 97%. The precision and F1 score have been calculated as average values for several different experimental trial runs; therefore, the

precision and F1 scores may have small variations compared to those produced by this one confusion matrix. Overall, the distribution of errors across classes is similar, supporting the reliability, robustness, and suitability of the ODIN-Net model to automate malaria cell detection.

8.3 Comparison Between Deep Model And Lightweight Model

The comparative performance results in the table shows that ODIN-Net had superior accuracy, precision, recall and F1 score to the traditional deep CNN models as well as those available as lightweight malaria detection methods. In addition to exceeding ODIN-Net's total performance including some of the fastest training times versus all the models tested in this analysis has a significantly greater increase in user engagement. McNemar's test was conducted on ODIN-Net compared to each baseline model to assess whether the improvements noted were likely due to chance or by development. All the p-values for the McNemar tests performed on all baseline models proved that the performance increase of ODIN truly occurred due to the use of ODIN-Net vs. reliance on a standard baseline model.

Table 1. Comparative Performance Analysis with Recent Lightweight Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (ms)	Statistical Significance (McNemar's Test, *p*-value)
Proposed ODIN-Net	**97.0**	**94.0**	**96.0**	**95.0**	**54,775 (100 epochs)**	— (Baseline)
VGG-19	96.0	89.0	94.0	91.0	68,200	0.032
EfficientNet-B2	93.0	91.0	94.0	92.0	63,500	0.021
ResNet-50	89.0	87.0	92.0	93.0	72,000	0.009
DLRFNet	94.7	92.5	95.2	93.8	61,200	0.018
HMFF-MobileNet	95.1	93.0	95.6	94.2	58,400	0.026

8.4 Discussion

ODIN-Net's unique and enhanced effectiveness can be related to it's combinations of a variety of designs and two separate principles. The stacked convolutional neural networks (CNNs) create a local hierarchical feature extraction process and then through the subsequent use of a 'parallel pyramid pooling' module, it utilizes the information from all three pyramids at once. With this method it is able to gather information about local parasite structure as well as the larger scale morphological structure of the cell. Therefore the value of this additional strengthened multi-scale feature representation to separate the cell infected with the parasites from the uninfected cells has a major impact on the quality of the images. Also, by using Support Vector Machine (SVM) classifier instead of the standard softmax classifier used in most of the neural network-based systems, there is lower risk of overfitting and improved generalizability through the maximization of the margin between the groups. This is especially important with the medical images as they tend to use less data and have more subtle variations than do regular images. ODIN-Net's ability to detect malaria was also evaluated in comparison to other malaria detection technologies that have been recently developed and published, such as the MAS-Net (Multi-scale Attention based CNN) [11], Deep-CNN-RF (Deep Convoluted Neural Network with Random Forests) [12], the MozzieNet (a tool used to classify malaria) [13] and Ensemble methods [14]. The MAS-Net uses a multi-scale receptive field attention approach to achieving 75.9% accuracy, however, it also was found to suffer from limits on generalization as well as on computational efficiency. The Deep-CNN-RF was able to enhance the learning of features using edges, but both its reported precision and F1 scores were lower than those of the other three technologies, and its hybrid design added complexity to the training process. MozzieNet had a high accuracy rate in classifying samples but required a much deeper architecture, leading to an increase in computation costs and ultimately limiting its use in the real-time classification of malaria. Ensemble techniques yielded moderate increases in classification accuracy but were limited by their tendency to overfit and extended training periods. On the other hand, the lightweight CNN architecture of ODIN-Net, along with its multi-scale pyramid pooling, and its SVM-based final decision module which has been optimized through Bayesian optimization provide an optimal balance between classification performance and computation efficiency such that ODIN-Net has the ability to exceed both deep architectures as well as hybrid ones and remain appropriate for the use of limited resources for real-time deployment in diagnostics.

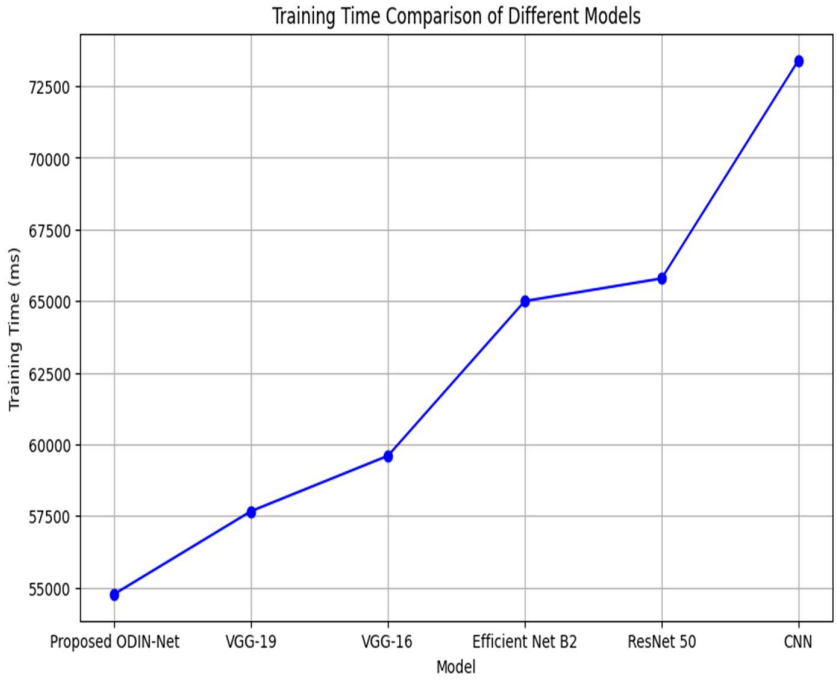

Figure 6. Training Time Comparison

9. CONCLUSION

The research successfully demonstrated that the Proposed ODIN-Net outperforms conventional CNN models such as VGG-19, VGG-16, Efficient Net B2, and ResNet 50 in the task of malaria cell classification. With an accuracy of 97%, and an impressive balance of precision and recall, ODIN-Net sets a new benchmark in the domain. Furthermore, its efficient training time of 54775 ms highlights its potential for practical deployment in medical diagnostic applications where time and accuracy are critical. This research confirms that integrating tailored feature extraction methods and optimizing classifier parameters through Bayesian Optimization can significantly enhance the performance of machine learning models in biomedical image analysis. Future work may explore the scalability of ODIN-Net to other types of medical imaging tasks and its performance in multi-class classification scenarios. Future work may explore the scalability of the model with larger datasets and its effectiveness in a clinical setting to further validate its utility in medical diagnostics. Although ODIN-Net has demonstrated superior performance on the NIH malaria dataset, we acknowledge that validation on external datasets and real-world clinical samples is essential to confirm its generalizability across diverse imaging conditions and patient populations. Future work will therefore focus on testing ODIN-Net on additional datasets such as Malaria-Cell-Images (U.S. NLM dataset extension) and ShenZhen Blood Smear dataset, as well as collaborating with clinical laboratories to evaluate performance on microscope-captured field samples. These steps will help assess ODIN-Net's robustness against variations in staining, image resolution, and parasite morphology. Furthermore, transfer learning and domain adaptation strategies will be explored to fine-tune the model for heterogeneous clinical data. Such comprehensive validation will ensure that ODIN-Net can serve as a dependable and generalizable diagnostic support tool in real-world healthcare environments.

10. FUTURE WORK AND LIMITATIONS

While ODIN-Net has shown promising results, it is important to recognize some of its limitations. To start with, the only dataset that the experiments were validated on was publicly available. This dataset was created with care and is widely used in research; however, it does not encompass the full scope of clinical variability in practice. Additionally, ODIN-Net has only been validated for binary classification; currently, the model distinguishes between infected and non-infected red blood cells. However, this classification scheme does not allow for the differentiation of various Plasmodium species or stages of infection, which are important clinical considerations when planning treatment for patients with malaria.

In addition, while the ODIN-Net model was designed to be lightweight and computationally efficient, it has not yet been evaluated in a "real-world" clinical setting, including examining various issues related to real-world usage.

Several factors, including the ability to handle low-quality images, integration with laboratory information systems, and performance with limited computational resources, require additional research.

Future validation of the proposed ODIN-Net framework across multiple centers and data sources will help determine the framework's ability to generalize across different types of imaging environments and conditions. Expanding the current binary classification framework to include multiple Plasmodium species and stages of infection is also a logical next step in this research. In addition, incorporating domain adaptation approaches and evaluating the potential for real-time deployment on edge or mobile platform devices will allow ODIN-Net to be practically applied to point-of-care testing for malaria.

Author Statements:

Ethical approval: The conducted research is not related to either human or animal use.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgement: The authors declare that they have nobody or no-company to acknowledge.

Author contributions: The authors declare that they have equal rights on this paper.

Funding information: The authors declare that there is no funding to be acknowledged.

REFERENCES

[1] Gourisaria, M. K., Das, S., Sharma, R., Rautaray, S. S., & Pandey, M. (2020), A deep learning model for malaria disease detection and analysis using deep convolutional neural networks. International Journal of Emerging Technologies, 11(2), 699-704. . [CrossRef] [Google Scholar] [Publisher Link]

[2] World Health Organization. Microscopy for the detection, identification and quantification of malaria parasites on stained thick and thin blood films in research settings, 2015.[CrossRef] [Google Scholar] [Publisher Link]

[3] Zhao, O. S., Kolluri, N., Anand, A., Chu, N., Bhavaraju, R., Ojha, A. & Nguyen, K. (2020). Convolutional neural networks to automate the screening of malaria in low-resource countries. PeerJ, 8, e9674.[CrossRef] [Google Scholar] [Publisher Link]

[4] Jain, N., Chauhan, A., Tripathi, P., Moosa, S. B., Aggarwal, P., & Oznacar, B. (2020). Cell image analysis for malaria detection using deep convolutional network. Intelligent Decision Technologies, 14(1), 55-65.[CrossRef] [Google Scholar] [Publisher Link]

[5] Pundir, P., Aggarwal, S., & Deshmukh, M. (2020, February). Malaria detection using convolutional neural network. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 187-195). Singapore: Springer Singapore.[CrossRef] [Google Scholar] [Publisher Link]

[6] Manescu, P., Shaw, M. J., Elmi, M., Neary-Zajiczek, L., Claveau, R., Pawar, V. & Fernandez-Reyes, D. (2020). Expert-level automated malaria diagnosis on routine blood films with deep neural networks. American Journal of Hematology, 95(8), 883-891.[CrossRef] [Google Scholar] [Publisher Link]

[7] Molina, A., Rodellar, J., Boldú, L., Acevedo, A., Alférez, S., & Merino, A. (2021). Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks. Computers in Biology and Medicine, 136, 104680. [CrossRef] [Google Scholar] [Publisher Link]

[8] 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. [CrossRef] [Google Scholar] [Publisher Link]

[9] Quan, Q., Wang, J., & Liu, L. (2020). An effective convolutional neural network for classifying red blood cells in malaria diseases. Interdisciplinary Sciences: Computational Life Sciences, 12, 217-225. [CrossRef] [Google Scholar] [Publisher Link]

[10] Irmak, E. (2021). A novel implementation of deep-learning approach on malaria parasite detection from thin blood cell images. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xiong, Z., & Wu, J. (2024). Multi-Level Attention Split Network: A Novel Malaria Cell Detection Algorithm. Information, 15(3), 166. [CrossRef] [Google Scholar] [Publisher Link]

[12] Murmu, A., & Kumar, P. (2024). DLRFNet: deep learning with random forest network for classification and detection of malaria parasite in blood smear. Multimedia Tools and Applications, 1-23. [CrossRef] [Google Scholar] [Publisher Link]

[13] Asif, S., Khan, S. U. R., Zheng, X., & Zhao, M. (2024). MozzieNet: A deep learning approach to efficiently detect malaria parasites in blood smear images. International Journal of Imaging Systems and Technology, 34(1), e22953. [CrossRef] [Google Scholar] [Publisher Link]

[14] Meena, B., Gowri, P., Karthikeyan, C., Dharshini, B., Parameshvar, M., & Gokul, S. (2024, March). Design of Malaria Detection Using Ensemble Techniques-A Combination of Alexnet and Densenet Algorithm. In

Proceedings of the 1st International Conference on Artificial Intelligence, Communication, IoT, Data Engineering and Security, IACIDS 2023, 23-25 November 2023, Lavasa, Pune, India. [CrossRef] [Google Scholar] [Publisher Link]

[15] Dembele, A., Mwangi, R. W., & Kube, A. O. (2024). A Lightweight Convolutional Neural Network with Hierarchical Multi-Scale Feature Fusion for Image Classification. Journal of Computer and Communications, 12(2), 173-200. [CrossRef] [Google Scholar] [Publisher Link]

[16] Nivaan, G.V., "Image Recognition of Malaria-infected Red Blood Cells among Other Normal and Cancer-Mutated Cells Using CNN," JINAV: Journal of Information and Visualization, vol.3, issue.1, pp.62-70, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Pattanaik, Priyadarshini Adyasha, and Tripti Swarnkar,"Vision-based malaria parasite image analysis: a systematic review," International Journal of Bioinformatics Research and Applications, vol.15, pp.1-32, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] P. Upender, S. P. Bharathi, Sukriti, K. Kumba and A. Kumar, "A Compact Metamaterial Biosensor for Multi-Virus Detection with Tunability and High Incidence Angle Absorption," IEEE Access, vol. 11, pp. 131915-131925, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Wasu Kudisthalert, Kitsuchart Pasupa, and Sissades Tongsima, "Counting and Classification of Malaria Parasite from Giemsa-Stained Thin Film Images," IEEE Access, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Rosado, L, Da Costa, J.M.C, Elias, D. and Cardoso, J.S., "Mobile-based analysis of malaria-infected thin blood smears: automated species and life cycle stage determination," Sensors, vol.17, issue.10, p.2167, 2017. [CrossRef] [Google Scholar] [Publisher Link]