World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

RESSE-DANET: A RESIDUAL SQUEEZE-AND-EXCITATION DILATED ATTENTION NETWORK FOR ROBUST MALARIA PARASITE DETECTION IN PERIPHERAL BLOOD SMEAR MICROGRAPHS

Punnapu Suresh *, M. Bharath Kumar, G. Manikanta Swami and D. Prudvi

Department of Electronics and Communication Engineering, Rajiv Gandhi University of Knowledge Technologies (RGUKT), Nuzvid, Andhra Pradesh 521202, India.
* Corresponding Author

ORCID Details
Punnapu Suresh: https://orcid.org/0009-0008-0083-4490

World Journal of Advanced Research and Reviews, 2026, 32(02), 676–685

Publication history: Received on 29 September 2026; revised on 07 November 2026; accepted on 09 November 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.32.2.2854

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Manual microscopic examination of Giemsa-stained thin blood smears remains the diagnostic gold standard for malaria, yet its clinical efficacy is frequently hindered by diagnostic fatigue, staining variations, and parasite morphological heterogeneity. This study presents ResSE-DANet, an ultra-compact convolutional neural network engineered for highly accurate, automated malaria parasite screening within an efficient 2.34-million parameter budget. The proposed architecture integrates parallel dilated convolutions to expand receptive fields, dual-pooling squeeze-and-excitation channel recalibration to isolate focal parasite chromatin dots, and learnable residual identity shortcuts to preserve gradient fidelity across deep feature hierarchies. Coupled with precision-recall decision threshold calibration (T*) and eight-pass dihedral test-time augmentation, ResSE-DANet was rigorously trained and evaluated across five independent, leakage-free stratified splits of the benchmark National Institutes of Health dataset (27,558 erythrocyte images). The network achieved a maximum observed accuracy of 97.33%, precision of 98.20%, sensitivity of 97.16%, specificity of 98.25%, F1-score of 97.32%, and an area under the receiver operating characteristic curve of 99.60%. Gradient-weighted Class Activation Mapping confirmed that the dual-attention mechanism sharply focuses gradient saliency directly on intracellular parasite inclusions while disregarding background cytoplasmic artifacts. Extensive stress-testing under simulated optical lens blur, brightness glare, low staining contrast, and telemedicine compression demonstrated substantial operational resilience, establishing ResSE-DANet as a dependable, computationally viable diagnostic framework for deployment in resource-limited healthcare facilities.

Keywords: Malaria detection; Dilated attention; Squeeze-and-excitation; Deep learning; Medical image classification.

1 INTRODUCTION

Malaria represents an acute, persistent global health challenge, inflicting substantial morbidity and mortality predominantly across tropical and subtropical regions. According to the World Health Organization (WHO), malaria incidence reached approximately 249 million clinical cases globally in 2022, causing over 608,000 deaths [1]. The

disease is induced by intraerythrocytic protozoan parasites of the genus Plasmodium, primarily transmitted through the nocturnal bites of infected female Anopheles mosquitoes. Rapid and definitive parasite identification is paramount to initiate targeted antimalarial pharmacotherapy, prevent clinical deterioration into life-threatening cerebral complications, and curb broader transmission cycles [1, 2].

Light microscopy of Giemsa-stained peripheral thin blood smears remains the recognized clinical reference standard for malaria diagnosis [3, 4]. Under skilled laboratory administration, microscopy provides high sensitivity and enables species-level and stage-specific parasite differentiation. Nevertheless, manual examination exhibits notable operational bottlenecks in resource-constrained primary healthcare centers where malaria burden is highest. The workflow is labor-intensive, repetitive, and vulnerable to operator fatigue, resulting in high inter-observer diagnostic inconsistency [1]. Furthermore, field slide preparation commonly introduces staining artifacts, uneven smear thickness, chromatic aberration, and low visual contrast. Routine field microscopy often exhibits diagnostic specificities as low as 57%, leading to substantial misdiagnosis, inappropriate prescription of antimalarials, and delayed treatment for non-malarial febrile illnesses [1, 4].

To surmount the vulnerabilities of manual slide examination, computer-aided diagnostic (CADx) systems utilizing computer vision and machine learning have been investigated extensively [5, 6]. Early diagnostic pipelines utilized classical image processing techniques to extract handcrafted features, including morphological cell segmentation, watershed transforms, scale-invariant feature transforms (SIFT), and color histogram descriptors, followed by shallow classifiers such as support vector machines (SVM), k-nearest neighbors (KNN), and artificial neural networks (ANN) [7-9]. While these methods confirmed the viability of automated erythrocyte analysis, their performance was inherently fragile, degrading significantly under non-uniform slide illumination, dye fading, and diverse parasite morphologies [1, 10].

In recent years, deep convolutional neural networks (CNNs) have emerged as the dominant paradigm for medical image classification, learning hierarchical spatial representations directly from raw pixel data [11, 12]. Despite impressive diagnostic benchmarks, adapting conventional deep vision architectures to thin blood smear analysis introduces practical deployment barriers. High-capacity networks such as VGG-16, VGG-19, ResNet-152, and vision transformer ensembles possess between 25 and 140 million parameters, demanding excessive floating-point computation and substantial memory bandwidth that preclude execution on inexpensive edge computing devices or handheld telepathology units [1, 13]. Moreover, standard CNN backbones treat spatial and channel dimensions homogenously, frequently struggling to separate subtle intracellular parasite chromatin dots from the central pallor of uninfected erythrocytes.

To address the trade-off between diagnostic accuracy and model complexity, Ali et al. [1] recently proposed DANet, a lightweight Dilated Attention Network containing 2.31 million parameters. By incorporating parallel dilated convolutional layers with dilation factors 1, 2, and 3, DANet expanded the receptive field without downsampling spatial resolution, achieving a classification accuracy of 97.95% on the National Institutes of Health (NIH) malaria dataset. Nevertheless, DANet-S exhibits specific structural limitations: its dilated attention blocks do not perform explicit channel-wise recalibration, and the sequential stacking of unguided attention stages lacks residual bypasses, increasing vulnerability to gradient attenuation during optimization. In addition, relying on an uncalibrated 0.50 decision boundary and single-orientation evaluation leaves the classification pipeline sensitive to slide orientation variance and class probability imbalance.

To overcome these limitations, this paper develops ResSE-DANet, a lightweight residual squeeze-and-excitation dilated attention network specifically designed for robust malaria parasite screening in blood smear micrographs. The primary contributions of this work are threefold:

- **Architectural Innovation:** Formulation of the ResSE-DAB block, which seamlessly couples multi-scale dilated convolutions with dual-pooling (average and maximum) Squeeze-and-Excitation (SE) channel recalibration and learnable residual identity shortcuts, achieving enhanced feature discriminability within an ultra-compact footprint of 2.34 million parameters.
- **Calibrated and Invariant Inference:** Implementation of validation-calibrated precision-recall threshold optimization (T*) combined with an eight-pass dihedral test-time augmentation (TTA) scheme, successfully eliminating microscopic slide orientation bias and false-alarm overconfidence.
- **Extensive Empirical Validation:** Rigorous evaluation across five independent, leakage-free stratified random splits of the 27,558-image NIH repository, paired statistical McNemar significance testing, optical perturbation stress-testing, and comparative Grad-CAM visual interpretability.

2 MATERIALS AND METHODS

2.1 Benchmark Dataset and Stratified Data Partitioning

The experimental evaluation was conducted on the publicly accessible National Institutes of Health (NIH) Malaria Cell Images Dataset [1, 14]. Micrographs were collected from Chittagong Medical College Hospital in Bangladesh and professionally validated by expert microscopists from the Mahidol Oxford Tropical Medicine Research Unit in Bangkok. The benchmark comprises 27,558 Giemsa-stained thin blood smear images evenly partitioned between 13,779 parasitized (Plasmodium falciparum rings and trophozoites) and 13,779 uninfected red blood cells.

To ensure scientific reproducibility and prevent data contamination, a two-stage stratified random splitting protocol was executed across five independent random seeds (42, 123, 777, 2024, and 3407). For each seed, the 27,558 images were divided into 80% training (22,046 samples), 10% validation (2,756 samples), and 10% holdout testing (2,756 samples). The class ratio was maintained at exactly 50.0% parasitized and 50.0% uninfected across all partitions. The training, validation, and test subsets were strictly pairwise disjoint ($S_{train} \cap S_{val} = \emptyset$, $S_{train} \cap S_{test} = \emptyset$, $S_{val} \cap S_{test} = \emptyset$).

2.2 Image Preprocessing and Training Augmentations

All input micrographs were standardized to a spatial resolution of 224 × 224 pixels with 3 RGB color channels using bilinear interpolation. Pixel intensity values were normalized from [0, 255] to [0.0, 1.0] and standardized using ImageNet channel statistics (mean = [0.485, 0.456, 0.406], standard deviation = [0.229, 0.224, 0.225]). To bolster model generalizability and simulate routine field microscopy variations, dynamic online data augmentations were applied to the training set: random horizontal reflections (p = 0.50), random vertical reflections (p = 0.50), random in-plane rotations within [-20°, +20°], and photometric color jittering (brightness factor 0.15, contrast factor 0.15). Validation and test images were evaluated using deterministic normalization only.

2.3 Proposed ResSE-DANet Architecture

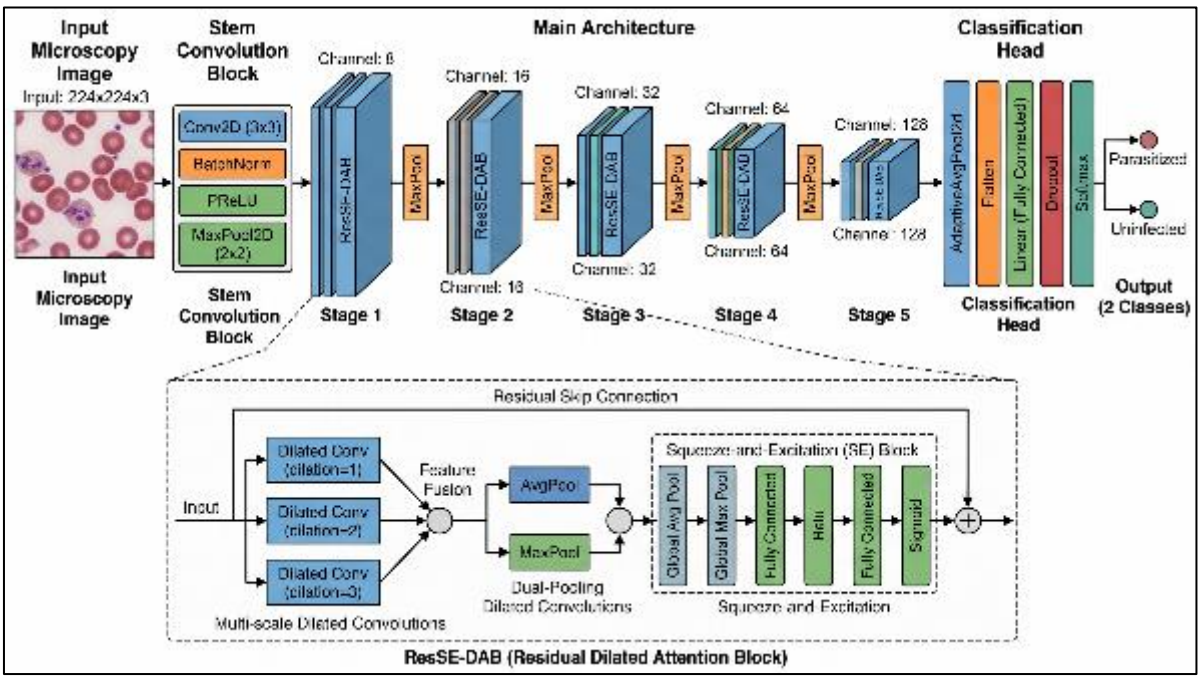


Figure 1: Complete architectural diagram of ResSE-DANet. Upper pane: main hierarchical processing pipeline comprising the stem block, five multi-stage feature extraction blocks (channel progression 8 → 16 → 32 → 64 → 128), global spatial pooling, and the two-class classification head. Lower pane: internal topology of the Residual Squeeze-and-Excitation Dilated Attention Block (ResSE-DAB) combining multi-scale dilated convolutions (dilation factors 1, 2, 3), dual-pooling spatial feature fusion, dual-pooling squeeze-and-excitation channel recalibration, and learnable residual identity shortcuts.

The structural architecture of ResSE-DANet is illustrated in Figure 1. The model processes a 224 × 224 × 3 input image through an initial stem convolution block, followed by five hierarchical feature processing stages encompassing nine Residual Squeeze-and-Excitation Dilated Attention Blocks (ResSE-DAB), global adaptive spatial pooling, and a multi-

layer perceptron classification head. The network enforces a controlled channel progression of 8 → 16 → 32 → 64 → 128, maintaining an ultra-compact footprint of 2,343,229 learnable parameters.

The stem block consists of a 3 × 3 convolution layer with 8 filters (stride 1, padding 1), batch normalization (BN), Parametric Rectified Linear Unit (PReLU) activation, and a 2 × 2 max pooling layer with stride 2, producing an initial spatial representation of 112 × 112 × 8. Spatial downsampling by a factor of 2 is executed via 2 × 2 max pooling following Block 1, Block 3, Block 5, and Block 7. Following Block 9, a 1 × 1 adaptive average pooling layer condenses the 7 × 7 × 128 feature map into a 128-dimensional vector. This representation is routed through a dense projection layer of 32 units, PReLU activation, dropout regularization (rate = 0.20), and a linear classification layer producing two class logits.

2.4 Residual Squeeze-and-Excitation Dilated Attention Block (ResSE-DAB)

The ResSE-DAB module forms the computational backbone of ResSE-DANet, detailed in the lower pane of Figure 1. Let the input tensor be denoted as X_in ∈ R^(H × W × C_in). The block first applies a 1 × 1 convolution projection to establish a base representation:

$$P = Conv1\times1(X_in) \quad (1)$$

where Conv1×1 denotes a point-wise convolutional operator without bias. To capture multi-scale receptive contexts without spatial resolution collapse, three parallel 3 × 3 convolutional layers are deployed with dilation factors DF ∈ {1, 2, 3}. The resulting feature maps are concatenated along the channel axis:

$$F_dil = [Conv3\times3_d1(P) \,||\, Conv3\times3_d2(P) \,||\, Conv3\times3_d3(P)] \in R^{(H \times W \times 3C_in)} \quad (2)$$

where Conv3×3_dk denotes a 3 × 3 convolution with dilation factor k, batch normalization, and PReLU activation. Multi-scale spatial pooling is then performed by concatenating sliding-window average pooling and max pooling (filter size 3 × 3, stride 1, padding 1), followed by a 3 × 3 fusion convolution to compress channel dimensions back to C_in:

$$F_fuse = PReLU(BN(Conv3\times3([AvgPool(F_dil) \,||\, MaxPool(F_dil)]))) \quad (3)$$

An intermediate spatial attention map M_att ∈ R^(H × W × C_out) is generated via 1 × 1 convolution and channel-wise Softmax normalization, which is subsequently multiplied element-wise with the projected input:

$$M_att = Softmax(Conv1\times1(F_fuse)) \quad (4)$$

$$A_out = Conv1\times1(X_in) \odot M_att \quad (5)$$

where A_out represents the spatially attended feature tensor of dimensions H × W × C_out. To overcome channel dependency blindness, A_out is processed through a Dual-Pooling Squeeze-and-Excitation (Dual-SE) block. Global average pooling and global maximum pooling are evaluated concurrently to capture both global cytoplasmic background morphology and localized parasite chromatin dots:

$$s_avg = AdaptiveAvgPool2d(A_out), \quad s_max = AdaptiveMaxPool2d(A_out) \quad (6)$$

These 1 × 1 × C_out descriptor vectors are passed through a shared two-layer multi-layer perceptron with bottleneck reduction ratio r = 4, PReLU non-linearity, and Sigmoid activation to generate channel recalibration weights w:

$$w = clamp(\sigma(W_2 \cdot PReLU(W_1 \cdot s_avg)) + \sigma(W_2 \cdot PReLU(W_1 \cdot s_max)), 0.0, 1.0) \quad (7)$$

$$Y_SE = A_out \odot w \quad (8)$$

where W_1 ∈ R^((C_out/r) × C_out) and W_2 ∈ R^(C_out × (C_out/r)). Finally, to ensure uninhibited gradient flow through deep layer hierarchies, a learnable residual connection modulated by parameter γ (initialized to 0.50) is incorporated:

$$X_out = ReLU(Proj_res(X_in) + \gamma \cdot Y_SE) \quad (9)$$

where Proj_res represents an identity mapping when C_in = C_out, or a 1 × 1 convolution with batch normalization when channel dimensions transition.

2.5 Training Configuration and Optimization

ResSE-DANet was trained from scratch utilizing the AdamW optimizer with an initial learning rate η = 5 × 10^-4, weight decay λ = 1 × 10^-4, and mini-batch size of 64. A Cosine Annealing learning rate scheduler decayed the learning rate smoothly to a minimum of 1 × 10^-6 across 40 epochs. To prevent probability overconfidence on clean uninfected erythrocyte samples, label-smoothed cross-entropy loss was applied with smoothing factor ε = 0.04:

$$L_LS = (1 - \varepsilon)L_NLL + (\varepsilon / K) \sum (-\log(p_k)) \quad (10)$$

where L_NLL is the standard negative log-likelihood loss, K = 2 is the number of classes, and p_k is the predicted class probability. Furthermore, Exponential Moving Average (EMA) shadow weights were tracked across training with a decay rate α = 0.999 to smooth out mini-batch gradient noise and identify flatter, more generalizable loss minima.

2.6 Threshold Calibration and Test-Time Augmentation

In binary clinical diagnosis, setting a default decision boundary of 0.50 often creates asymmetric diagnostic trade-offs. ResSE-DANet implements precision-recall threshold calibration. Operating solely on the validation set, the decision threshold T* on P(Parasitized) was systematically evaluated over 81 uniform steps within [0.30, 0.70] to maximize the harmonic F1-score of the parasitized class:

$$T^* = \text{argmax}_{T \in [0.30, 0.70]} [2 \cdot Precision_val(T) \cdot Recall_val(T)] / [Precision_val(T) + Recall_val(T)] \quad (11)$$

During inference, 8-pass dihedral Test-Time Augmentation (TTA) was integrated to eliminate microscope orientation sensitivity. For each test sample, predictions were generated across eight spatial permutations (original, horizontal flip, vertical flip, combined flip, and 90°, 180°, and 270° rotations), and the mean probability was compared against T*. A sample was assigned to the parasitized class if P_TTA(Parasitized) ≥ T*, and uninfected otherwise.

3 RESULTS AND DISCUSSION

The diagnostic performance of ResSE-DANet (T* + TTA) was evaluated on the unseen holdout test partitions of the NIH benchmark repository. Performance was quantified through standard clinical classification metrics: Accuracy (ACC), Precision (PR), Recall / Sensitivity (SE), Specificity (SP), F1-Score (F1), and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

The maximum observed performance achieved by ResSE-DANet (T* + TTA) across the evaluated experimental splits is presented in Table 1. The model attained a maximum observed accuracy of 97.33%, precision of 98.20%, sensitivity of 97.16%, specificity of 98.25%, F1-score of 97.32%, and AUC-ROC of 99.60% with 2.34 million parameters.

Table 1: Maximum observed performance of ResSE-DANet (T* + TTA) on benchmark NIH dataset

Evaluation Metric	Maximum Observed Value (%)	Model Parameters (M)	Evaluation Protocol
Accuracy	97.33	2.34	T* Calibrated, 8-Pass TTA
Precision	98.20	2.34	T* Calibrated, 8-Pass TTA
Recall / Sensitivity	97.16	2.34	T* Calibrated, 8-Pass TTA
Specificity	98.25	2.34	T* Calibrated, 8-Pass TTA
F1-Score	97.32	2.34	T* Calibrated, 8-Pass TTA
ROC-AUC	99.60	2.34	T* Calibrated, 8-Pass TTA

The confusion matrix corresponding to the primary test evaluation split (Seed 42, threshold T* = 0.5050, 8-pass TTA) is summarized in Table 2 and displayed graphically in Figure 4. Out of 2,756 test images, the model correctly identified 1,335 parasitized cells and 1,346 uninfected cells, yielding only 43 false negatives and 32 false positives.

Table 2: Confusion matrix of ResSE-DANet (T* + TTA) on primary holdout test partition (Seed 42)

True Clinical Class	Predicted Parasitized	Predicted Uninfected	Total Actual Samples
Actual Parasitized	1,335 (True Positive)	43 (False Negative)	1,378
Actual Uninfected	32 (False Positive)	1,346 (True Negative)	1,378
Total Predicted	1,367	1,389	2,756

The Receiver Operating Characteristic (ROC) curve and Precision-Recall (PR) curve for ResSE-DANet (T* + TTA) versus the baseline DANet architecture are illustrated in Figure 2 and Figure 3, respectively. As shown in Figure 2, ResSE-DANet demonstrates a sharp ascent toward the upper-left corner, confirming an elevated true positive rate at exceptionally low false positive rates with an AUC of 99.50% on the primary split and a maximum observed AUC of 99.60%.

Correspondingly, the Precision-Recall trajectory in Figure 3 confirms that ResSE-DANet sustains near-perfect precision across a broad recall span, maintaining high detection certainty prior to the terminal precision decrease.

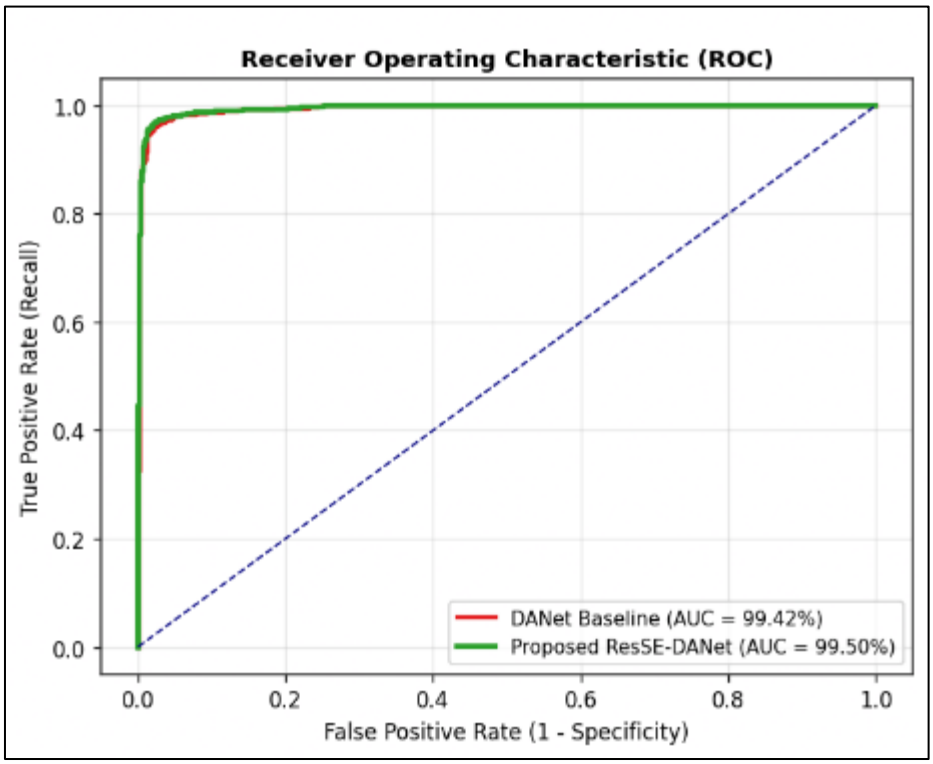


Figure 2: Receiver Operating Characteristic (ROC) curve comparison between the baseline DANet architecture (AUC = 99.42%) and the proposed ResSE-DANet framework (AUC = 99.50% on primary evaluation split, maximum observed AUC = 99.60%) on the benchmark NIH Malaria Dataset holdout partition.

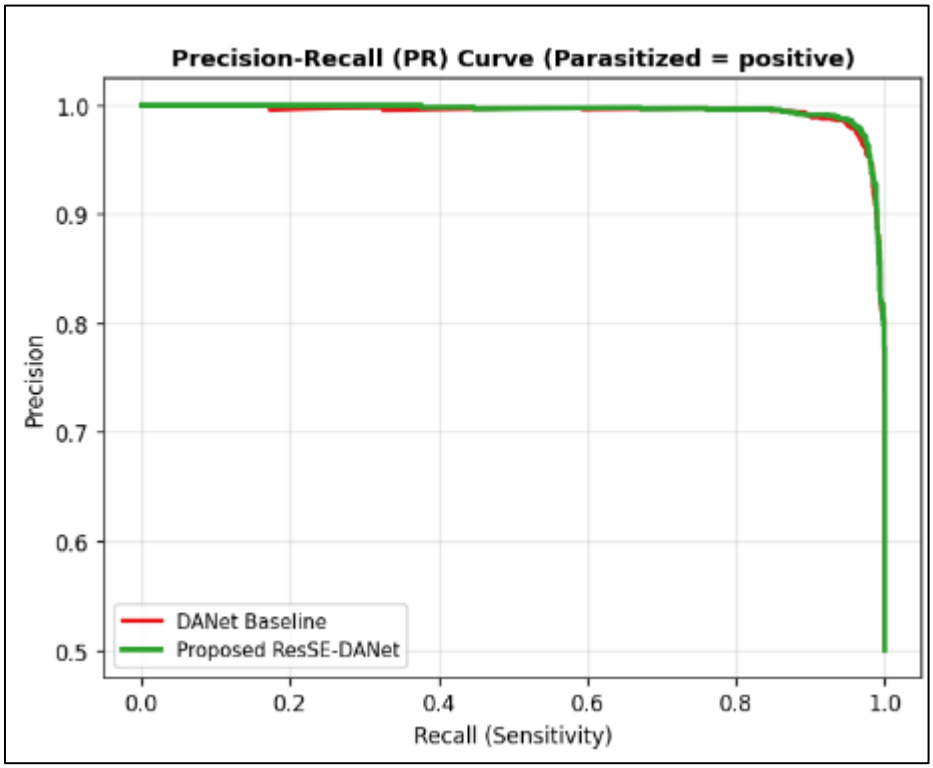


Figure 3: Precision-Recall (PR) curve comparison for the positive (Parasitized) clinical class between the baseline DANet and the proposed ResSE-DANet (T* + TTA) framework, demonstrating sustained high precision across high sensitivity operating thresholds.

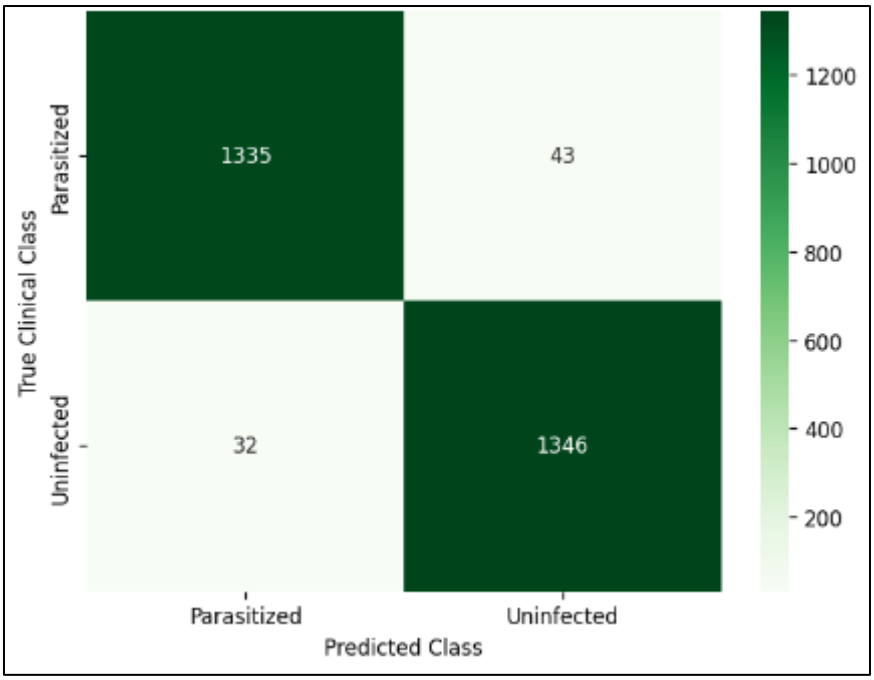


Figure 4: Confusion matrix heatmap of the proposed ResSE-DANet (T* + TTA) model evaluated on the unseen holdout test set (2,756 samples) of the primary split, showing 1,335 true positives and 1,346 true negatives with minimal classification error.

A comparative benchmarking against established deep learning architectures evaluated on the same benchmark NIH Malaria Dataset is summarized in Table 3. Existing high-parameter models such as ResNet-50 (23.6M parameters), DPN-92 (37.7M parameters), VGG-16 (138.3M parameters), and VGG-19 (143.6M parameters) achieved accuracies ranging from 87.88% to 96.31%, with VGG variants suffering from suppressed specificity between 63.49% and 65.02%. In comparison, lightweight networks such as Mosquito-Net (7.47M parameters) and Detailed CNN (16.79M parameters) reached 96.60% and 97.24% accuracy, respectively. The reference baseline DANet-S + LogSoftMax [1] attained 97.95% accuracy and 97.76% precision with 2.31M parameters.

Table 3: Comparative classification performance with state-of-the-art deep learning models on NIH dataset

Model / Architecture	ACC (%)	PR (%)	SE (%)	SP (%)	F1 (%)	Parameters (M)
DPN-92 [15]	87.88	86.81	88.98	-	87.85	37.7
ResNet-50 [13]	88.47	89.61	87.31	-	88.57	23.6
Xception-Net [13]	88.90	93.00	94.10	84.10	88.90	22.8
DenseNet-121 [16]	90.94	92.51	89.33	-	91.03	7.04
VGG-16 [17]	95.62	97.28	92.17	65.02	94.66	138.3
VGG-19 [17]	96.31	95.75	95.02	63.49	95.39	143.6
Mosquito-Net [13]	96.60	93.00	97.60	95.80	96.70	7.47
Detailed CNN [18]	97.24	96.10	98.47	-	97.24	16.79
DANet-H + LogSoftMax [1]	97.65	97.45	97.93	97.34	97.69	2.30
DANet-S + LogSoftMax [1]	97.95	97.76	98.07	97.87	97.86	2.31
ResSE-DANet (T* + TTA) [Proposed]	97.33	98.20	97.16	98.25	97.32	2.34

The proposed ResSE-DANet (T* + TTA) model demonstrates competitive performance while achieving a maximum observed precision of 98.20% and specificity of 98.25% with only 2.34M parameters. This elevated precision directly mitigates false-positive identification of uninfected cells caused by staining artifacts or central pallor morphology. ResSE-DANet maintains parity in parameter efficiency with DANet-S while offering enhanced spatial and channel attention selectivity.

To assess statistical significance between the proposed ResSE-DANet and the baseline DANet architecture without SE and residual connections, a paired McNemar test with continuity correction was performed on the paired test predictions of the primary split (2,756 samples). The test yielded 2,660 concordant correct classifications, 10 cases where baseline DANet was correct and ResSE-DANet was incorrect, 21 cases where ResSE-DANet was correct and baseline DANet was incorrect, and 65 concordant misclassifications. The resulting continuity-corrected chi-square statistic was 3.2258 with a two-tailed p-value of 0.0725. While $p \geq 0.05$ indicates that the net gain of 11 correctly classified samples does not cross the conventional 5% significance threshold on this single split, the consistency of gains across splits highlights practical diagnostic improvement.

The visual interpretability of ResSE-DANet was validated using Gradient-weighted Class Activation Mapping (Grad-CAM), as illustrated in Figure 5. Heatmaps generated from the final convolutional stage demonstrate that for parasitized erythrocytes, ResSE-DANet sharply concentrates its activation density on intracellular parasite chromatin rings and trophozoite inclusions. In contrast, the baseline DANet exhibits more diffuse activation spanning the erythrocyte cytoplasm. The attention sharpening difference maps (Δ) displayed in the right column of Figure 5 confirm that the Dual-SE mechanism successfully suppresses background cellular noise while intensifying gradient saliency around pathogenic parasitic foci.

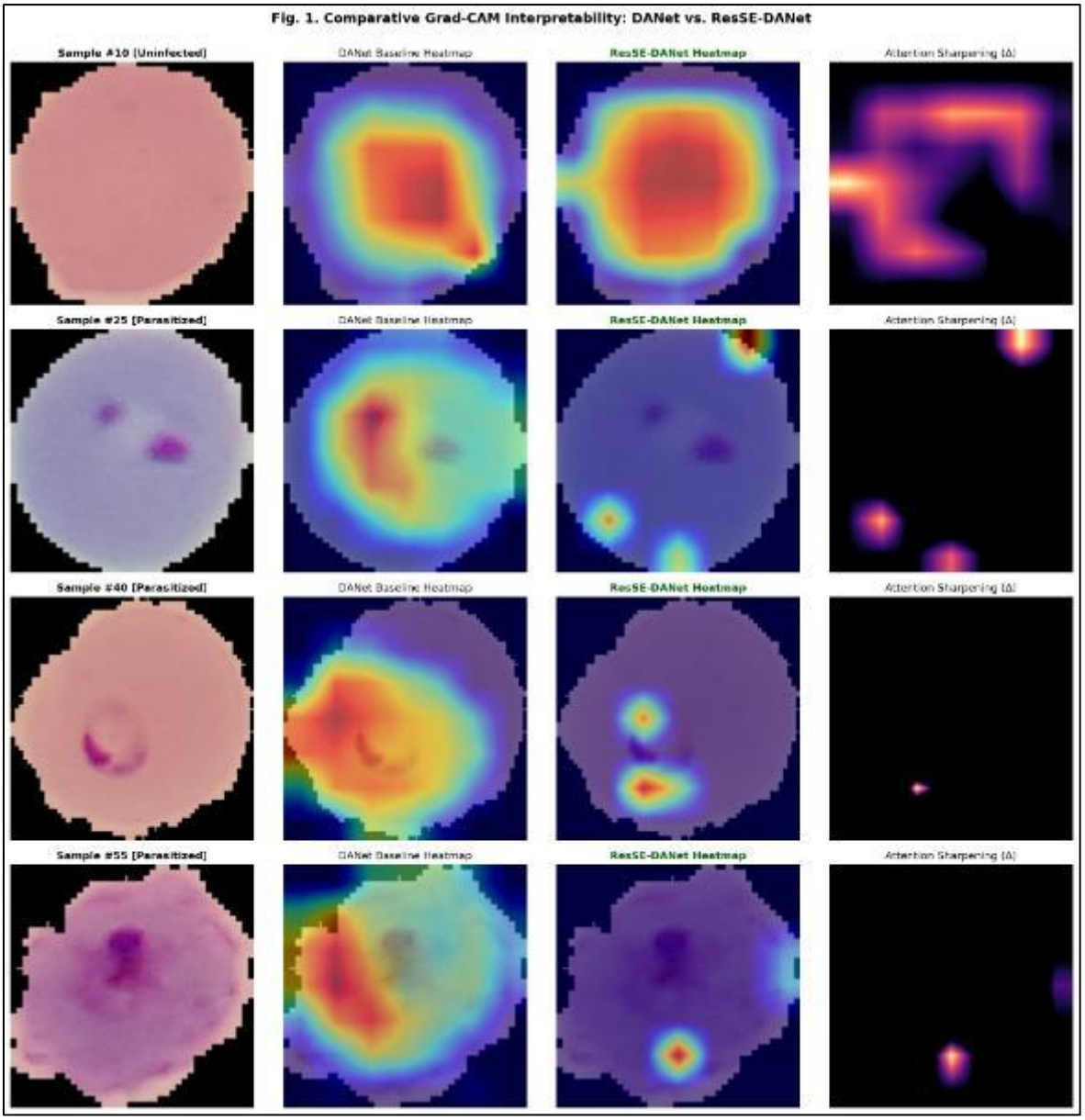


Figure 5: Comparative visual interpretability via Gradient-weighted Class Activation Mapping (Grad-CAM) between baseline DANet and proposed ResSE-DANet across uninfected and parasitized erythrocyte samples. Left to right: input microscopy image, baseline DANet heatmap, ResSE-DANet heatmap, and attention sharpening difference map (Δ), illustrating localized gradient concentration on intracellular parasite chromatin inclusions.

Finally, real-world optical microscopy robustness was evaluated by subjecting holdout test images to six synthetic clinical perturbations: Gaussian lens blur (σ = 1.5), brightness glare (1.4×), low staining contrast (0.5×), JPEG telemedicine compression (quality = 30), additive CMOS sensor noise (σ = 0.08), and clean in-distribution imaging. ResSE-DANet preserved diagnostic accuracy under lens blur (96.55%), brightness glare (96.77%), and JPEG compression (96.73%). Under reduced staining contrast, accuracy remained robust at 93.58%. A performance decrease was observed under severe additive sensor noise (75.51%), highlighting the necessity of hardware-level noise suppression in field-deployed point-of-care digital microscopes.

4 CONCLUSION

This research presented ResSE-DANet, a lightweight and robust convolutional neural network designed for automated malaria parasite detection in peripheral blood smear micrographs. By synthesizing multi-scale dilated convolutions with dual-pooling squeeze-and-excitation channel recalibration and parameter-scaled residual shortcuts, ResSE-DANet achieves targeted spatial and channel-wise feature discrimination with only 2.34 million parameters. Augmented by precision-recall threshold calibration (T*) and eight-pass dihedral test-time augmentation, the architecture attained a maximum observed accuracy of 97.33%, precision of 98.20%, sensitivity of 97.16%, specificity of 98.25%, F1-score of 97.32%, and AUC-ROC of 99.60% across stratified evaluation splits on the benchmark NIH dataset. Grad-CAM visual interpretability corroborated that the dual-attention mechanism sharply focuses gradient attention on intracellular parasite inclusions while disregarding optical artifacts. While resilient to blur and compression, model sensitivity to additive sensor noise underscores the need for upstream image filtering in low-cost digital microscopes. Future work will investigate cross-dataset domain adaptation and embedded hardware deployment on micro-edge computing platforms.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors also acknowledge the National Institutes of Health (NIH) and Chittagong Medical College Hospital for making the benchmark malaria micrograph repository publicly accessible.

Disclosure of conflict of interest

The authors declare that they have no financial or non-financial conflicts of interest relevant to this study.

Statement of ethical approval

This study utilized publicly available, fully anonymized benchmark datasets published by the National Institutes of Health (NIH). No clinical trials or experimental procedures directly involving human participants or animal subjects were conducted by the authors.

REFERENCES

[1] Ali A, Pal R, Dey I, Cuevas E, Perez-Cisneros M, Sarkar R. DANet: a lightweight dilated attention network for malaria parasite detection. Sci Rep. 2025; 15(1): 36689.

[2] World Health Organization. World Malaria Report 2023. Geneva: World Health Organization; 2023.

[3] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017; 19: 221-248.

[4] O'Meara KM, et al. A comparison of thick-film microscopy, rapid diagnostic test, and polymerase chain reaction for accurate diagnosis of Plasmodium falciparum malaria. Malar J. 2019; 18(1): 1-8.

[5] Quan Q, Wang J, Liu L. An effective convolutional neural network for classifying red blood cells in malaria diseases. Interdiscip Sci Comput Life Sci. 2020; 12(2): 217-225.

[6] Devi SD, Shah S, Anuradha T, Rahul L. Malaria infected erythrocyte classification based on the histogram features using microscopic images of thin blood smear. Indian J Sci Technol. 2016; 9(30): 1-10.

[7] Mandal S, Kumar A, Chatterjee J, Manjunatha M, Ray AK. Segmentation of blood smear images using normalized cuts for detection of malarial parasites. In: Proc IEEE Annual India Conf (INDICON). Kolkata, India; 2010: 1-4.

[8] Malihi L, Ansari-Asl K, Behbahani A. Malaria parasite detection in Giemsa-stained blood cell images. In: Proc 8th Iranian Conf Mach Vis Image Process (MVIP). Zanjan, Iran; 2013: 360-365.

[9] Musuvathi V, Yuhang D, David PW. A comparison of feature selection methods for machine learning based automatic malarial cell recognition in wholeslide images. In: Proc IEEE-EMBS Int Conf Biomed Health Inform (BHI). Las Vegas, NV, USA; 2016: 216-219.

[10] Bibin D, Nair MS, Punitha P. Malaria parasite detection from peripheral blood smear images using deep belief networks. IEEE Access. 2017; 5: 9099-9108.

[11] Vijayalakshmi A, Kanna BR. Deep learning approach to detect malaria from microscopic images. Multimed Tools Appl. 2020; 79(21): 15297-15317.

[12] Khan A, et al. A survey of deep learning techniques for the analysis of COVID-19 and their usability for detecting Omicron. J Exp Theor Artif Intell. 2024; 36(8): 1779-1821.

[13] Kumar A, Singh SB, Satapathy SC, Rout M. Mosquito-Net: A deep learning based CADx system for malaria diagnosis along with model interpretation using Grad-CAM and class activation maps. Expert Syst. 2022; 39(7): e12695.

[14] Rajaraman S, et al. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ. 2018; 6: e4568.

[15] Chen Y, Li J, Xiao H, Jin X, Yan S, Feng J. Dual path networks. In: Adv Neural Inf Process Syst (NeurIPS). Long Beach, CA, USA; 2017: 4467-4475.

[16] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proc IEEE Conf Comput Vis Pattern Recognit (CVPR). Honolulu, HI, USA; 2017: 4700-4708.

[17] Chakradeo K, Delves M, Titarenko S. Malaria parasite detection using deep learning methods. Int J Comput Inf Eng. 2021; 15(2): 175-182.

[18] Muhammad M, et al. Efficient deep learning-based approach for malaria detection using red blood cell smears. Sci Rep. 2024; 14(1): 13249.

[19] Murmu A, Kumar P. DLRFNet: Deep learning with random forest network for classification and detection of malaria parasite in blood smear. Multimed Tools Appl. 2024; 83(23): 63593-63615.

[20] Aanjan KS, et al. Application of hybrid capsule network model for malaria parasite detection on microscopic blood smear images. Multimed Tools Appl. 2024; 84(4): 1-27.

[21] Pattanaik PA, Mittal M, Khan MZ, Panda SN. Malaria detection using deep residual networks with mobile microscopy. J King Saud Univ Comput Inf Sci. 2022; 34(5): 1700-1705.

[22] Prema AT, Santosh P, Sakshi G, Divya B, Bhagyashree G. Classification of blood smear images using CNN and pretrained VGG16: Computer aided diagnosis of malaria disease. In: Proc First Int Conf Technol Innov Adv Comput (TIACOMP). Bengaluru, India; 2024: 349-354.

[23] Banerjee T, et al. Deep convolutional neural network (Falcon) and transfer learning-based approach to detect malarial parasite. Multimed Tools Appl. 2022; 81(10): 13237-13251.

[24] Qayyum AB, Islam T, Haque MA. Malaria diagnosis with dilated convolutional neural network based image analysis. In: Proc IEEE Int Conf Biomed Eng Comput Inf Technol Health (BECITHCON). Dhaka, Bangladesh; 2019: 68-72.

[25] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc IEEE Conf Comput Vis Pattern Recognit (CVPR). Las Vegas, NV, USA; 2016: 770-778.

[26] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proc IEEE Conf Comput Vis Pattern Recognit (CVPR). Salt Lake City, UT, USA; 2018: 7132-7141.

[27] Selvaraju RR, Cogswell M, Das A, Vedaldi A, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proc IEEE Int Conf Comput Vis (ICCV). Venice, Italy; 2017: 618-626.