

Quality-Guided Focal Loss: Enhancing Minority Class Detection in Haematological Imaging

Thabang Fenge Isaka^{1[0009-0000-5216-112X]}, Claire Wynne^{2[0000-0002-0481-8366]},
and Jane Courtney^{1[2222-3333-4444-5555]}

¹ School of Electrical and Electronic Engineering, Technological University Dublin,
Dublin, Ireland

² School of Biological, Health and Sports Sciences, Technological University Dublin,
Dublin, Ireland

Abstract. In the critical race against malaria, the most dangerous parasites often hide in plain sight. When parasitaemia falls below 1%, precisely when early detection matters most, conventional AI detection systems falter despite impressive aggregate metrics. This paradox of "seeing everything except what matters most" stems from a fundamental detection dilemma: infected cells comprise a vanishingly small minority that conventional approaches systematically overlook. We propose a methodical Quality-Guided Focal Loss (QGFL), a framework that reconceptualizes how detection systems learn from imbalanced data. By integrating class-specific focusing parameters, quality-guided weighting, and spatial awareness through UIoU, QGFL achieves a remarkable improvement in detecting infected cells in the clinically vital 1–3% parasitaemia range. Our cross-dataset validation confirms QGFL's generalizability across diverse imaging conditions without requiring dataset-specific tuning. This work advances the approach to minority class detection in medical imaging, demonstrating how prediction quality can guide model optimization, ensuring that what matters clinically also matters computationally.

Keywords: Medical Image Analysis · Object Detection · Class Imbalance · Adaptive Focal Loss · Malaria Detection.

1 Introduction

In the fight against malaria, early detection can mean the difference between life and death. Yet in resource-constrained settings where malaria remains endemic, diagnostic capabilities are severely limited by the availability of trained microscopists. Automated detection systems offer transformative potential, but face a fundamental challenge: infected red blood cells often constitute merely 1–3% of all visible cells, creating extreme class imbalance that causes conventional AI systems to miss critical infections. Microscopy faces similar challenges, with average microscopists detecting only 50–100 parasites/ μl versus expert limits of 5 parasites/ μl [13], consistent with broader medical image imbalance issues [3]. This "needle-in-haystack" problem is particularly acute in low-density infections,

precisely when early intervention is most critical. Mosquera et al. [7] demonstrated that imbalanced datasets can lead to 20% drops in F1-score for malaria detection systems, while Ramos-Briceño et al. [9] showed that state-of-the-art neural networks struggle with species-level *Plasmodium* identification in sparse distributions. As illustrated in Figure 1, morphological similarity between infected and uninfected cells creates scenarios where subtle infections are easily missed, representing the most dangerous failure mode in automated diagnosis.

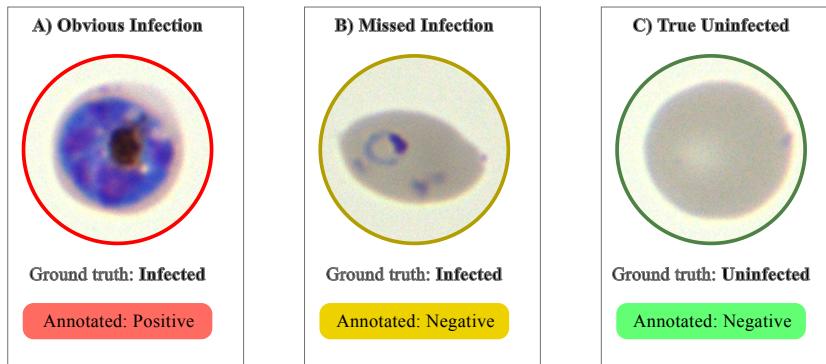


Fig. 1. Critical detection scenarios: (A) obvious infection correctly identified, (B) missed infection incorrectly classified as negative, (C) true uninfected cell correctly identified.

While focal loss approaches have shown promise for addressing class imbalance in object detection [6], their application to medical imaging with extreme intra-image imbalance remains underexplored. Current approaches lack systematic frameworks addressing minority class detection challenges where false negatives carry severe clinical consequences. Additionally, difficulty thresholding emerges as critical, requiring mechanisms to focus on challenging examples while avoiding noise from trivial cases [2]. This paper proposes a Quality-Guided Focal Loss (QGFL), a progressive adaptation framework for haematological imaging that addresses intra-image class imbalance through quality-guided components and difficulty thresholding. Building upon recent advances in quality-guided approaches [17], our framework demonstrates substantial improvements in detecting infected cells in low-density scenarios, showing promise for enhanced automated malaria detection applications.

The key contributions of this work are:

1. A progressive focal loss adaptation framework building from basic focal loss through quality-guided components to difficulty thresholding.
2. A quality-guided component that dynamically adjusts loss contribution based on detection quality, improving performance without sacrificing majority class accuracy.
3. Comprehensive validation across three datasets with varying imbalance characteristics, demonstrating consistent improvements in low-density scenarios.

4. A difficulty thresholding strategy that focuses training on genuinely challenging examples while avoiding noise from trivial cases.

2 Related Work

2.1 Object Detection Architectures for Haematological Imaging

Object detection in haematological imaging has evolved from traditional image processing to sophisticated deep learning architectures. Davidson et al. [1] implemented a two-stage approach achieving 92% accuracy on balanced test sets but showing significant performance degradation on images with natural parasite distributions. This limitation stems from treating detection as sequential segmentation and classification tasks, failing to leverage contextual information across the entire field of view. Hung and Carpenter [5] pioneered Faster R-CNN application to malaria microscopy, improving localization accuracy but relying heavily on mean Average Precision (mAP), which obscured performance disparities between infected and uninfected cell detection. Recent approaches focus on computational efficiency while maintaining accuracy. Guemas et al. [4] employed Real-Time Detection Transformer (RT-DETR) for species-level recognition, demonstrating improved species classification but continued struggles with intra-image class imbalance. Their evaluation showed detection recall for *P. falciparum* dropped from 87% to 63% when infection ratio fell below 3%, highlighting persistent minority class detection challenges. Current architectures face critical limitations: optimization for balanced datasets, evaluation methodologies masking performance disparities, and lack of specialized components for haematological imaging challenges such as morphological similarity and variable staining conditions.

2.2 Class Imbalance and Focal Loss Adaptations in Medical Imaging

Class imbalance in medical imaging manifests at both dataset and intra-image levels with distinct performance implications. Gao et al. [3] distinguished between these imbalance types, achieving AUC of 0.91 on LUNA16 and 0.87 on DDSM using deep autoencoder architectures, though primarily effective for dataset-level imbalance. Mosquera et al. [7] demonstrated F1-scores dropping up to 20% when minority classes constitute less than 5% of datasets, with melanoma detection showing 17.3% lower F1-score compared to majority classes. Focal Loss, introduced by Lin et al. [6], represented significant advancement in addressing class imbalance by dynamically adjusting loss contribution of well-classified examples, achieving 3.2% Average Precision improvement in RetinaNet. Medical imaging adaptations include Yeung et al. [15] Unified Focal Loss framework, achieving 7.3% Dice score improvement for brain tumor segmentation, and Su et al. [11] Adaptive Focal Loss with dynamic focusing parameter adjustment, achieving 4.7% improvement in medical X-ray keypoint detection. Zhao et al. [17] proposed cross-domain Focal Loss adaptation for low-contrast medical image segmentation, demonstrating significant improvements through adversarial learning integration.

Current approaches face limitations: few address extreme intra-image imbalance in haematological imaging, most focus on either classification or segmentation tasks, and quality-guided component integration with Focal Loss requires exploration.

2.3 Threshold Calibration and Quality-Guided Components

Threshold calibration represents a critical aspect of object detection in imbalanced scenarios, where minority classes require substantially lower confidence thresholds for optimal recall. Esposito et al. [2] demonstrated that class-specific calibration significantly improves minority class detection without compromising overall performance. Rajaraman et al. [8] revealed Expected Calibration Error (ECE) 2.3 times higher for minority classes, highlighting fundamental trade-offs between calibration and discrimination in imbalanced contexts. Quality assessment in object detection extends beyond simple IoU metrics to incorporate localization precision, classification confidence, and contextual relevance. Zhao et al. [17] demonstrated that quality-guided components improve performance by dynamically shifting attention between low-quality and high-quality prediction regions. Zhan et al. [16] showed detection performance drops by 27.3% when image quality falls below PSNR threshold of 25dB, particularly affecting small, low-contrast objects analogous to haematological imaging challenges. Current approaches inadequately address extreme threshold disparities in haematological imaging, focus primarily on post-hoc adjustment rather than integrated training considerations, and insufficiently explore quality assessment’s relationship with detection reliability in low-density scenarios. This analysis of current approaches across object detection architectures, class imbalance methods, focal loss adaptations, threshold calibration, and quality-guided components identifies specific limitations that the Quality-Guided Focal Loss framework addresses, providing a comprehensive solution to minority class detection challenges in haematological imaging.

3 Materials and Methods

3.1 Dataset Characteristics and Preprocessing

This study utilized three publicly available malaria microscopy datasets: D1 [1], D2 [5], and D3 [4], all characterized by pronounced intra-image sparsity. Table 1 reveals that 87–98% of images contain ≤ 2 infected cells, creating a clinically realistic “needle-in-haystack” detection scenario.

D1 comprises 398 microscopy images from laboratory *P. falciparum* cultures with standardized Giemsa staining. D2 consists of 1,328 images from ex vivo *P. vivax* patient samples, presenting severe class imbalance with 97% uninfected cells. D3 represents the largest collection with 29,228 images from 475 patient samples across 6 French hospitals, spanning multiple *Plasmodium* species. All datasets were standardized to binary infected/uninfected COCO-annotated classifications and partitioned using 70/20/10% splits to prevent data leakage [14]. D3 required

Table 1. Intra-Image Class Imbalance Across Datasets

Dataset	Split	Img.	Inf.	Uninf.	Mean Inf.%	Std. Dev	≤ 2 Inf.
D1	Train	279	1,614	23,372	2.8%	$\pm 1.1\%$	243 (87%)
	Valid	79	599	6,799	3.2%	$\pm 1.4\%$	62 (78%)
	Test	40	164	3,516	1.9%	$\pm 1.0\%$	38 (95%)
D2	Train	930	1,791	57,912	0.7%	$\pm 0.6\%$	915 (98%)
	Valid	265	446	16,825	0.8%	$\pm 0.5\%$	259 (98%)
	Test	133	215	8,400	0.6%	$\pm 0.4\%$	131 (98%)
D3	Train	865	3,600	65,678	1.4%	$\pm 1.2\%$	812 (94%)
	Valid	247	997	18,221	1.7%	$\pm 1.3\%$	226 (91%)
	Test	123	467	9,565	1.3%	$\pm 1.0\%$	117 (95%)

strategic subsampling: retaining 100% of images containing infected cells and 5% of uninfected-only images, preserving natural intra-image sparsity essential for clinical relevance. No synthetic augmentation or artificial class balancing was employed, aligning with findings that preserving natural distributions improves generalization in medical imaging [10].

3.2 Base Architecture

The Quality-Guided Focal Loss framework builds upon RetinaNet with ResNet50 backbone pre-trained on ImageNet, selected for established performance in detecting small objects under class imbalance [6]. The architecture integrates Feature Pyramid Network for multi-scale feature representation, critical for detecting parasites across developmental stages. Domain-specific adaptations included re-configuring the classification head for binary classification and disabling in-place ReLU operations for stable gradient flow. Training employed SGD with momentum 0.9, learning rate 0.003, and weight decay 0.0005. For our experiments, the Baseline (B) model in Table 2 refers to the RetinaNet architecture that uses default library parameters for its standard focal loss. In contrast, our Standard Focal Loss (FL) model, which represents the first level of our progressive framework, utilizes the standard implementation from Lin et al. [6] with our own tuned hyperparameters:

$$\text{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

Where p_t represents predicted probability for the true class, α_t is class-balancing weight (0.9 for infected, 0.1 for uninfected), and γ is the focusing parameter (2.0) that down-weights easy examples.

3.3 Progressive Loss Adaptations

The Quality-Guided Focal Loss framework in Figure 2 employs progressive adaptation, systematically building upon standard focal loss [6] to address minority class detection challenges. This methodical progression, inspired by adaptive focal loss methodologies [11] and calibration-aware approaches [17], enables precise identification of effective components while maintaining experimental control.

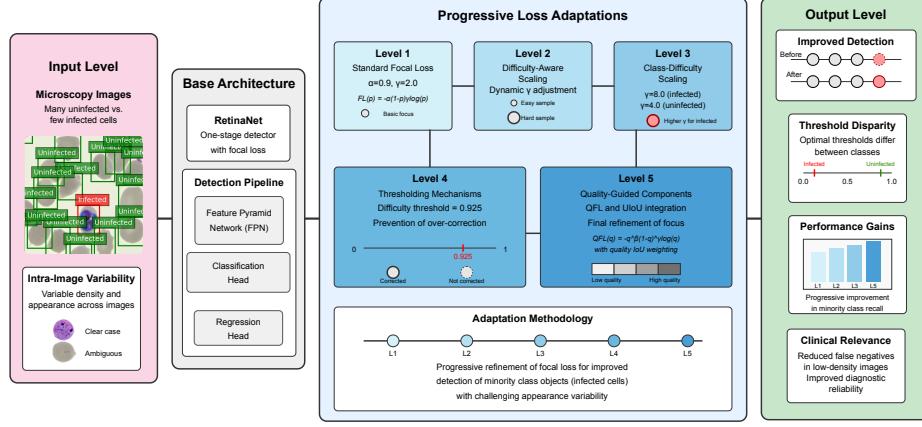


Fig. 2. The five-level progressive adaptation approach, from standard focal loss through difficulty-aware and class-specific modifications, to the complete quality-guided framework with UIoU integration.

Level 2: Difficulty-Aware Scaling Dynamic adjustment of focusing parameter based on sample difficulty, building upon adaptive focal loss approaches [11]:

$$\text{FL}_{\text{diff}}(p_t) = -\alpha_t(1 - p_t)^{\gamma_{\text{eff}}} \log(p_t) \quad (2)$$

Where $\gamma_{\text{eff}} = \gamma + (\max_{\gamma} - \gamma) \times \text{difficulty}$ with difficulty $= (1 - p_t)$ and optimal $\max_{\gamma} = 4.0$. This adaptation addresses the inherent challenge where difficult examples often receive insufficient attention under standard focusing mechanisms.

Level 3: Class-Difficulty Scaling Class-specific maximum focusing parameters, extending class-aware focal loss methodologies [12]:

$$\text{FL}_{\text{class-diff}}(p_t) = -\alpha_t(1 - p_t)^{\gamma_{\text{eff}}} \log(p_t) \quad (3)$$

This is the case where γ_{eff} applies class-specific parameters: infected_max $_{\gamma} = 8.0$ and uninfected_max $_{\gamma} = 4.0$, recognizing the inherent asymmetry in detection tasks.

Level 4: Thresholding Mechanisms Difficulty thresholding to focus on challenging examples, adapting threshold calibration principles [2]:

$$\text{FL}_{\text{thresh}}(p_t) = -\alpha_t(1 - p_t)^{\gamma_{\text{eff}}} \log(p_t) \quad (4)$$

Where difficulty is calculated as:

$$\text{difficulty} = \frac{\max(\text{raw_difficulty} - \text{threshold}, 0)}{1 - \text{threshold}} \quad (5)$$

With raw_difficulty $= (1 - p_t)$ and optimal threshold $= 0.925$.

Level 5: Quality-Guided Components Integration of quality measurement and Unified IoU components, following calibration error principles and synthesizing quality-guided approaches [17]:

$$\text{QGFL}(p_t) = -\alpha_t(1 - p_t)^{\gamma_{\text{eff}}}(1 + \text{quality_weight}) \times \text{uiou_ratio} \times \log(p_t) \quad (6)$$

Where quality is measured as the absolute difference between prediction and target:

$$\text{quality} = |p - \text{target}| \quad (7)$$

Adjusted using:

$$\text{quality_adjusted} = \max(\text{quality} - \text{quality_margin}, 0) \quad (8)$$

And transformed to:

$$\text{quality_weight} = \min(\text{quality_adjusted}^{\text{quality_factor}}, 10.0) \quad (9)$$

With $\text{quality_margin} = 0.5$ and $\text{quality_factor} = 2.0$. The UIoU component varies from 2.0 to 0.5 using linear decay during training.

3.4 Adaptation Methodology

The progressive refinement methodology systematically isolates challenges of minority class detection, as visualized in Figure 2. Parameter calibration followed rigorous grid search on Dataset D1: alpha (0.25–0.9), gamma (1.0–4.0), max gamma (4.0–20.0), and thresholds (0.05–0.96). Error analysis employed TIDE methodology categorizing detection failures. Cross-dataset validation employed direct parameter transfer from D1 to D2 and D3 without dataset-specific tuning, providing stringent generalizability testing across varying class imbalance ratios, parasite species, and imaging protocols.

4 Results

4.1 Cross-Dataset Performance

Table 2 reports the performance of the Quality-Guided Focal Loss (QGFL) framework across three haematological imaging datasets. Relative to the baseline RetinaNet, QGFL improves mean average precision (mAP) by 5.7% on D1, 11.2% on D2, and 10.4% on D3. Infected-class F1 scores show substantial gains in D1 (from 0.562 to 0.655) and D2 (from 0.648 to 0.743), corresponding to 16.5 and 14.7 percent relative improvement, respectively. On D3, recall increases from 0.675 to 0.807, indicating improved coverage of sparse positives. The Class-Diff+Thresh adaptation yields mAP improvements of 10.3% on D1 and 12.6% on D2, suggesting that explicit modulation of class difficulty contributes significantly to overall performance.

Table 2. Performance comparison of QGFL variants and baseline RetinaNet across three haematological imaging datasets. See legend below for full adaptation level names.

Adapt.	D1				D2				D3			
	mAP	Prec.	Recall	F1	mAP	Prec.	Recall	F1	mAP	Prec.	Recall	F1
B	0.672	0.824	0.427	0.562	0.681	0.718	0.591	0.648	0.772	0.903	0.675	0.772
FL	0.741	0.683	0.604	0.641	0.763	0.682	0.767	0.722	0.815	0.700	0.794	0.744
DA	0.704	0.674	0.579	0.623	0.752	0.652	0.740	0.693	0.845	0.661	0.812	0.729
CD	0.690	0.674	0.543	0.601	0.720	0.667	0.698	0.682	0.808	0.746	0.717	0.731
CD+T	0.741	0.650	0.634	0.642	0.767	0.731	0.735	0.733	0.815	0.701	0.797	0.745
FL+Q	0.713	0.754	0.598	0.667	0.726	0.747	0.688	0.717	0.812	0.717	0.760	0.738
FL+Q+U	0.711	0.797	0.573	0.667	0.759	0.728	0.758	0.743	0.816	0.732	0.784	0.757
QGFL	0.710	0.786	0.561	0.655	0.757	0.728	0.758	0.743	0.852	0.735	0.807	0.769

Adaptation Legend:

B = Baseline RetinaNet, FL = Standard Focal Loss, DA = Difficulty-Aware, CD = Class-Difficulty, CD+T = Class-Difficulty + Threshold, FL+Q = Focal Loss + Quality, FL+Q+U = Focal Loss + Quality + UIoU, QGFL = Complete Quality-Guided Loss Framework.

Figure 3 supports these findings. Average precision for the infected class increases across all datasets: from 0.557 to 0.627 in D1, from 0.554 to 0.666 in D2, and from 0.764 to 0.795 in D3. The QGFL curves consistently span a broader recall range, particularly in D1 and D2 where intra-image imbalance is more severe.

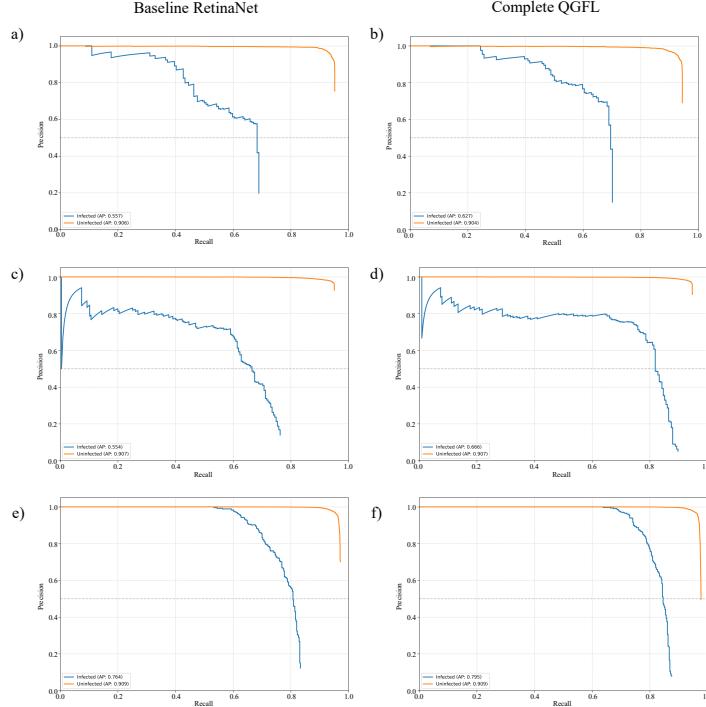


Fig. 3. Precision-recall curves comparing the baseline RetinaNet against the complete QGFL framework for infected-class detection across datasets D1, D2, and D3

4.2 Confidence Calibration and Error Analysis

Figure 4 presents optimal confidence thresholds across adaptation strategies and datasets. Infected-class thresholds consistently fall below the standard 0.5 detection threshold, ranging from 0.14–0.87, while uninfected cells achieve optimal performance at 0.56–0.93. This pattern persists across all three datasets regardless of adaptation complexity. Complete QGFL demonstrates more consistent infected-class thresholds across datasets while preserving uninfected performance above 0.8.

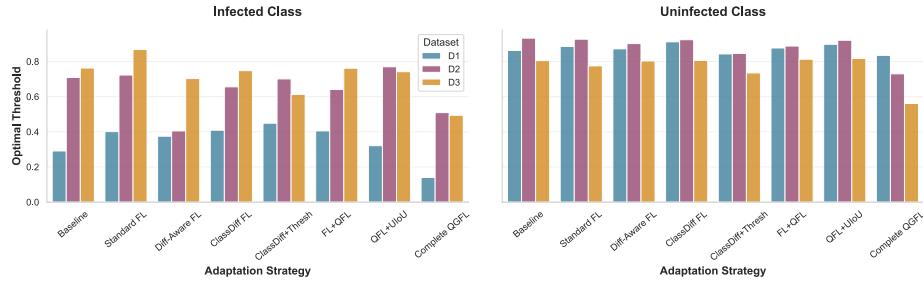


Fig. 4. Optimal confidence thresholds for infected and uninfected classes. A systematic disparity is observed, with infected cells consistently requiring lower thresholds across all datasets.

Figure 5 presents missed detection rates for infected cells across adaptation strategies. Complete QGFL achieves the lowest error rates across all datasets (38.1%, 18.8%, and 14.9% on D1, D2, and D3), with D1 showing the most substantial improvement from baseline (52.5% to 38.1%).

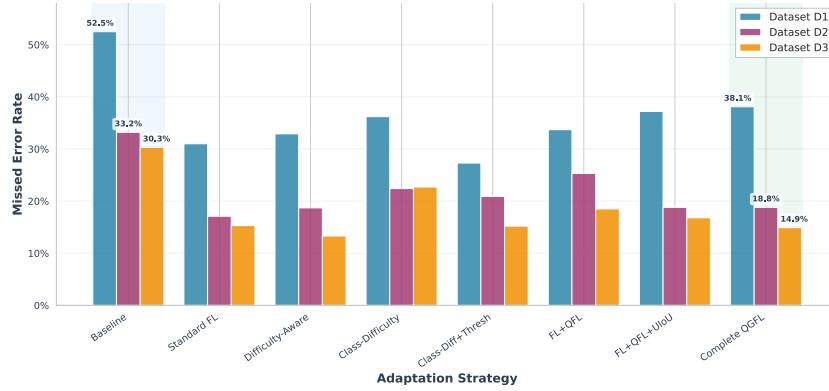


Fig. 5. Missed detection rates for infected cells across all datasets, where the complete QGFL model consistently achieves the lowest error rates.

4.3 Low-Density Performance

Figure 6 examines recall performance across infection density bins, focusing on the clinically critical 1–3% range where Complete QGFL demonstrates substantial improvements: 46% enhancement on D1 (0.42 to 0.61 recall), 93% on D2 (0.28 to 0.54), and 8% on D3 (0.71 to 0.76). These density-stratified results validate that QGFL’s systematic approach addresses minority class detection precisely where clinical impact is greatest.

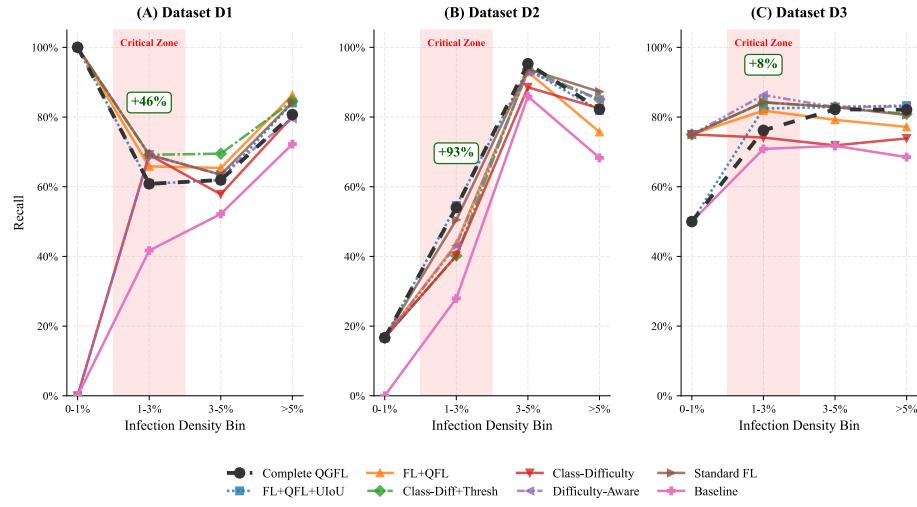


Fig. 6. Density-stratified recall performance, highlighting the substantial improvements by QGFL in the clinically critical 1–3% infection density range across all three datasets.

4.4 Qualitative Results

Figure 7 presents representative detection outputs across adaptation strategies, illustrating progressive improvement in minority class detection. Complete QGFL demonstrates the most comprehensive detection performance, consistently identifying infected cells across density scenarios while maintaining precision.

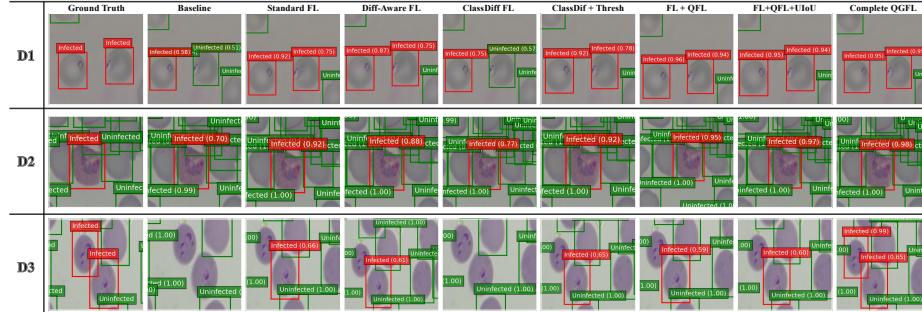


Fig. 7. Qualitative results for datasets D1-D3, where the complete QGFL framework demonstrates superior detection coverage of infected cells.

5 Discussion and Limitations

This study introduced Quality-Guided Focal Loss (QGFL), a progressive adaptation framework addressing minority class detection in haematological imaging. QGFL substantially improves infected cell detection across three diverse datasets, with notable gains in clinically relevant low-density scenarios. The progressive adaptation methodology revealed critical insights about focal loss modifications for medical object detection. Class-specific threshold calibration yielded substantial performance improvements (10.3% mAP improvement on D1, 12.6% on D2), confirming that explicit modulation of class difficulty contributes significantly to overall performance. This aligns with Rajaraman et al. [8], who found that deep learning models are poorly calibrated for minority classes in medical imaging, with Expected Calibration Error 2.3 times higher for minority classes. QGFL addresses this calibration challenge through targeted adjustments to gradient weighting and loss formulation.

Density-stratified recall analysis in Figure 6 provides compelling evidence for QGFL’s clinical utility. In the critical 1–3% infection density range where early diagnosis is crucial, QGFL demonstrated substantial improvements: 46% enhancement on D1, 93% on D2, and 8% on D3. This pattern suggests that QGFL’s effectiveness scales inversely with baseline performance, offering the most substantial improvements in the most challenging detection scenarios. The 93% improvement on D2, which features the most severe class imbalance (0.7% infected cells), demonstrates that QGFL effectively addresses the fundamental challenge of extreme intra-image sparsity. Confidence calibration properties (Figure 4) reveal a systematic pattern with direct clinical implications. The consistent disparity between infected-class thresholds (0.14–0.87) and uninfected thresholds (0.56–0.93) across all datasets explains why conventional detection systems fail in low-prevalence scenarios. This threshold disparity creates a fundamental detection dilemma: using a single global threshold inevitably sacrifices either sensitivity or specificity for the minority class. QGFL resolves this dilemma by dynamically adjusting loss contributions based on class membership and

prediction quality, enabling the model to maintain high sensitivity for infected cells without compromising overall precision. The stability of this threshold pattern across adaptation levels confirms it represents an intrinsic property of the detection task rather than an artifact of model design. It is important to note that this focus on improving recall involves a deliberate trade-off with precision. For instance, on dataset D1, QGFL improves recall from a baseline of 0.427 to 0.561, while precision moves from 0.824 to 0.786. Our framework is designed to manage this balance, prioritizing the reduction of clinically critical false negatives in alignment with the goals of early-stage screening. The integration of quality-guided components with focal loss represents a significant contribution that addresses a specific gap in medical object detection. While previous approaches like Zhao et al. [17] demonstrated quality-guided components for segmentation tasks, and Lin et al. [6] introduced focal loss for general object detection, neither approach adequately addressed the unique challenges of haematological imaging where objects of interest are morphologically similar but functionally distinct. QGFL bridges this gap by combining three critical elements: (1) class-specific focusing parameters that acknowledge the asymmetric detection challenges, (2) quality-guided weighting that dynamically adjusts loss contribution based on prediction quality, and (3) UIoU integration that addresses the spatial aspects of detection. This comprehensive framework specifically targets the needle-in-haystack detection scenario encountered in early-stage infection. Our evaluation is centered on the RetinaNet architecture, a choice made deliberately as it was the pioneering architecture for focal loss, thus providing the most direct and logical platform to isolate and rigorously evaluate the impact of our progressive loss adaptations.

While these results are promising, three primary limitations affect clinical applicability. First, while the QGFL framework is theoretically more complex, our empirical analysis found its practical computational overhead to be modest. For example, on a representative dataset (D3), it required only 5.1% more training time, demonstrating that its performance benefits are accessible without a prohibitive resource burden. Second, our analysis is constrained in two main ways: its focus on binary classification and its evaluation on a single architecture. The binary approach, while effective for proving the core concept, does not address the clinical need for species-level *Plasmodium* identification. Similarly, while our choice of RetinaNet was justified for isolating the effects of our loss function, the performance of QGFL on other modern architectures remains unevaluated. Third, while datasets preserved natural class distributions, they represent controlled laboratory conditions rather than the full variability encountered in clinical practice, including staining quality variations and microscope calibration differences.

Future work will address these limitations through optimized implementations that maintain performance while reducing computational requirements, extension to multi-class scenarios with hierarchical class relationships, and prospective validation in clinical environments. Additionally, systematic evaluation of established class balancing techniques alongside the progressive QGFL framework will identify synergistic combinations and optimize performance across diverse

medical imaging applications. Crucially, future validation will involve implementing QGFL across state-of-the-art architectures, including YOLO variants and transformer-based models, to rigorously establish its generalizability.

6 Conclusion

This work demonstrates that by reconceptualizing loss functions through a quality-guided approach, we can achieve improved detection capabilities in minority class scenarios. The Quality-Guided Focal Loss framework achieves consistent improvements in detecting infected cells, with up to 93% improvement in recall for the clinically critical 1–3% parasitaemia range where early intervention matters most. Beyond the quantitative gains, QGFL represents a systematic approach that moves from treating all predictions equally to acknowledging their inherent quality differences. This principle extends beyond haematological imaging, offering a methodical pathway toward more reliable automated detection across medical domains where subtle abnormalities demand heightened sensitivity, ultimately contributing to earlier diagnosis and improved patient outcomes.

Acknowledgments. Funded by Taighde Éireann – Research Ireland through the Research Ireland Centre for Research Training in Machine Learning (18/CRT/6183).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Davidson, M.S., Andradi-Brown, C., Yahiya, S., Chmielewski, J., O'Donnell, A.J., Gurung, P., Jeninga, M.D., Prommana, P., Andrew, D.W., Petter, M., Uthaipibull, C., Boyle, M.J., Ashdown, G.W., Dvorin, J.D., Reece, S.E., Wilson, D.W., Cunningham, K.A., Ando, D.M., Dimon, M., Baum, J.: Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks. *Biological Imaging* **1**, e2 (2021). <https://doi.org/10.1017/s2633903x21000015>
2. Esposito, C., Landrum, G.A., Schneider, N., Stiefl, N., Riniker, S.: GHOST: Adjusting the decision threshold to handle imbalanced data in machine learning. *Journal of Chemical Information and Modeling* **61**(6), 2623–2640 (2021). <https://doi.org/10.1021/acs.jcim.1c00160>
3. Gao, L., Zhang, L., Liu, C., Wu, S.: Handling imbalanced medical image data: A deep-learning-based one-class classification approach. *Artificial Intelligence in Medicine* **108**, 101935 (2020). <https://doi.org/10.1016/j.artmed.2020.101935>
4. Guemas, E., Routier, B., Ghelfenstein-Ferreira, T., Cordier, C., Hartuis, S., Marion, B., Bertout, S., Varlet-Marie, E., Costa, D., Pasquier, G.: Automatic patient-level recognition of four plasmodium species on thin blood smear by a real-time detection transformer (RT-DETR) object detection algorithm: a proof-of-concept and evaluation. *Microbiology Spectrum* **12**(2), e01440–23 (2024). <https://doi.org/10.1128/spectrum.01440-23>

5. Hung, J., Lopes, S.C.P., Nery, O.A., Nosten, F., Ferreira, M.U., Duraisingh, M.T., Marti, M., Ravel, D., Rangel, G., Malleret, B., Lacerda, M.V.G., Rénia, L., Costa, F.T.M., Carpenter, A.E.: Applying faster r-CNN for object detection on malaria images. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) **2017**, 808–813 (2017). <https://doi.org/10.1109/cvprw.2017.112>
6. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(2), 318–327 (2018). <https://doi.org/10.1109/tpami.2018.2858826>
7. Mosquera, C., Ferrer, L., Milone, D.H., Luna, D., Ferrante, E.: Class imbalance on medical image classification: towards better evaluation practices for discrimination and calibration performance. European Radiology **34**(12), 7895–7903 (2024). <https://doi.org/10.1007/s00330-024-10834-0>
8. Rajaraman, S., Ganesan, P., Antani, S.: Deep learning model calibration for improving performance in class-imbalanced medical image classification tasks. PLoS ONE **17**(1), e0262838 (2022). <https://doi.org/10.1371/journal.pone.0262838>
9. Ramos-Briceño, D.A., Flammia-D'Aleo, A., Fernández-López, G., Carrión-Nessi, F.S., Forero-Peña, D.A.: Deep learning-based malaria parasite detection: convolutional neural networks model for accurate species identification of plasmodium falciparum and plasmodium vivax. Scientific Reports **15**(1), 3746 (2025). <https://doi.org/10.1038/s41598-025-87979-5>
10. Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R., Sieh, W.: Deep learning to improve breast cancer detection on screening mammography. Scientific Reports **9**(1), 12495 (2019). <https://doi.org/10.1038/s41598-019-48995-4>
11. Su, Z., Adam, A., Nasrudin, M.F.: Adaptive focal loss for keypoint-based deep learning detectors addressing class imbalance. IEEE Access **13**, 31842–31856 (2025). <https://doi.org/10.1109/access.2025.3538917>
12. Tan, J., Wang, C., Li, B., Li, Q., Ouyang, W., Yin, C., Yan, J.: Equalization loss for long-tailed object recognition. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) **00**, 11659–11668 (2020). <https://doi.org/10.1109/cvpr42600.2020.01168>
13. Tangpukdee, N., Duangdee, C., Wilairatana, P., Krudsood, S.: Malaria diagnosis: A brief review. Korean Journal of Parasitology **47**(2), 93–102 (2009). <https://doi.org/10.3347/kjp.2009.47.2.93>, <https://doi.org/10.3347/kjp.2009.47.2.93>
14. Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L.R., Summers, R.M., Rubin, D.L., Lungren, M.P.: Preparing medical imaging data for machine learning. Radiology **295**(1), 4–15 (2020). <https://doi.org/10.1148/radiol.2020192224>
15. Yeung, M., Sala, E., Schönlieb, C.B., Rundo, L.: Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics **95**, 102026 (2022). <https://doi.org/10.1016/j.compmedimag.2021.102026>
16. Zhan, B., Song, E., Liu, H.: FSA-net: Rethinking the attention mechanisms in medical image segmentation from releasing global suppressed information. Computers in Biology and Medicine **161**, 106932 (2023). <https://doi.org/10.1016/j.combiomed.2023.106932>
17. Zhao, Y., Shen, X., Chen, J., Qian, W., Ma, H., Sang, L.: ARD loss for low-contrast medical image segmentation. Machine Learning: Science and Technology **5**(1), 015013 (2024). <https://doi.org/10.1088/2632-2153/ad1d06>