World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3119

DESIGN OF A MULTIMODAL AI-BASED INTELLIGENT BLOOD CELL DIAGNOSTIC SYSTEM WITH EDGE-ENHANCED YOLOV8

WeiPeng Yu[1*], YongQiang He[1], Tian He[1], MeiXin Lu[2], WeiDong Ma[3]
[1]*College of Computer Science and Engineering, Guilin University of Technology, Guilin 541006, Guangxi, China.*
[2]*School of Artificial Intelligence, Guilin University of Technology, Guilin 541006, Guangxi, China.*
[3]*College of Earth sciences, Guilin University of Technology, Guilin 541006, Guangxi, China.*
[*]*Corresponding Author: WeiPeng Yu*

Abstract: This study presents a practical design for an intelligent blood cell diagnostic system that integrates microscopic images, clinical text, and an edge-enhanced YOLOv8 detector. The scheme addresses the low efficiency, subjective variation, and limited specialist resources associated with manual blood-smear review. A standardized multimodal dataset is organized around normal and abnormal red blood cells, white blood cells, and platelets. The detection backbone is strengthened with a cell morphology feature enhancement module, weighted bidirectional feature fusion, and a combined localization and classification loss to improve recognition under cell overlap, uneven staining, and blurred boundaries. A PySide6 interface, FastAPI service, SQLite database, and cloud-edge collaborative workflow support image, batch, video, and camera-based analysis. The available training curves, bounding-box distribution, detection visualization, and recall-confidence results are used as feasibility evidence without adding unsupported experimental claims. The resulting design provides a modular route for efficient blood cell screening and can be transferred to bacterial examination, urinary sediment analysis, and sperm quality assessment through dataset and output-layer adaptation.
Keywords: Blood cell detection; YOLOv8; Multimodal AI; CMEM; Cloud-edge collaboration; System design

1 INTRODUCTION

Blood cell morphology is a fundamental basis for screening and diagnosing hematological disorders. Manual slide review, however, depends heavily on professional experience, requires considerable time for each specimen, and may produce inconsistent judgments when cell boundaries are blurred, cells overlap, or abnormal cells are rare. These limitations are particularly restrictive in primary medical institutions, where specialist personnel and high-end equipment are limited. Artificial intelligence is increasingly used to support clinical interpretation and workflow automation, but its value depends on whether the model, software, data governance, and clinical deployment are designed as one system rather than as isolated technical components [1].

Deep learning has demonstrated strong diagnostic potential in medical imaging, while recent evaluations also emphasize the need for representative data, rigorous validation, and transparent reporting [2]. In blood morphology analysis, convolutional networks can learn differences in nuclei, cytoplasm, staining, shape, and texture that are difficult to encode manually. Large-scale morphology studies show that deep models can distinguish fine-grained bone marrow cell classes when high-quality annotations and clinically meaningful categories are available [3]. Public white blood cell resources further demonstrate the importance of cell-level labels, nucleus and cytoplasm information, and acquisition diversity for reproducible development [4].

The objective is to construct an intelligent diagnostic scheme that automatically detects cells, quantifies their distribution, provides graded abnormality prompts, and supports primary-care deployment. The design uses YOLOv8 as the detection core, introduces a cell morphology feature enhancement module (CMEM) for blurred boundaries, combines image and clinical-text information, and integrates the model into an accessible software platform. This article focuses on design logic, mathematical formulation, implementation workflow, and deployment feasibility while avoiding unsupported experimental claims beyond the available figures and performance evidence.

2 DESIGN REQUIREMENTS AND TECHNICAL BASIS

2.1 Clinical and Functional Requirements

The system is intended to improve efficiency, diagnostic consistency, and accessibility. It should identify red blood cells, white blood cells, and platelets in microscopic images; support abnormal morphology prompts; process individual images, image batches, video, and real-time camera streams; and export structured results. For clinical use, detection results must be visualized with bounding boxes and labels so that medical personnel can inspect the evidence rather than receive only a final class. Account management, record storage, report generation, and interfaces with laboratory information systems are also required.

The intended dataset combines images acquired from optical microscopes and automated blood analyzers with clinical text such as patient history, routine blood indicators, and previous diagnostic results. Images cover venous and peripheral blood and include normal cells and abnormal forms. The proposed dataset construction plan requires standardized resolution, noise removal, color correction, double-blind annotation by multiple experienced laboratory physicians, and consistency checking with a target Kappa coefficient of at least 0.85. These requirements align with the broader observation that blood-cell models require consistent labeling and explicit morphology definitions rather than image quantity alone [3-4].

2.2 Detection and Trustworthiness Basis

YOLO-family detectors are suitable for the proposed workflow because they perform localization and classification in one forward process and offer a practical balance between accuracy and speed. Recent reviews describe the progression from early YOLO versions to anchor-free and decoupled-head designs represented by YOLOv8 [5]. Accordingly, this study selects YOLOv8 and improves its feature fusion, loss function, and morphology sensitivity instead of replacing the detection pipeline with a larger architecture.

Clinical use also requires interpretable outputs. Explainable medical-image analysis research distinguishes visual explanations, concept-level interpretation, and example-based reasoning, while warning that explanation quality must be evaluated rather than assumed [6]. The present scheme uses direct bounding-box visualization, confidence values, abnormal-cell prompts, and clinician correction feedback. These elements do not replace medical judgment; they make the detector output inspectable and create a route for controlled model refinement.

2.3 Development Context and Design Direction

Domestic development has moved from isolated algorithm experiments toward integrated analysis platforms. Universities, research institutes, in-vitro diagnostic enterprises, and medical AI companies increasingly combine detection models with image acquisition, data management, and user-facing software. Existing related work identifies YOLO and Faster R-CNN as common technical foundations and emphasizes targeted optimization for staining differences, weak cell details, and limited annotated samples. Data augmentation and transfer learning are therefore important when high-quality medical annotations are expensive. Platform functions are also expanding from single-image inference to batch images, video, and real-time microscope-camera detection.

International work places greater emphasis on large multi-center datasets, cell instance segmentation, fine-grained leukocyte subtyping, robustness across devices and staining protocols, and formal clinical integration. Vision Transformer architectures, multimodal analysis, continual adaptation, and explainable AI are explored alongside convolutional detectors. System development is often connected with medical institutions and instrument manufacturers so that automated analysis can enter laboratory workflows and satisfy medical-device approval requirements.

The common direction is not merely higher benchmark accuracy. Detection must remain stable for clustered cells, rare abnormalities, and uneven illumination; outputs must provide clinically useful prompts; and the software must connect with cloud services, mobile devices, and existing information systems. Low-cost portable analysis is particularly valuable for resource-limited regions. These trends support the proposed system-level design, which combines a detection core, edge enhancement, multimodal information, human-readable visualization, and a deployment architecture suitable for both tertiary and primary medical institutions.

3 OVERALL SYSTEM SCHEME

3.1 Layered Architecture

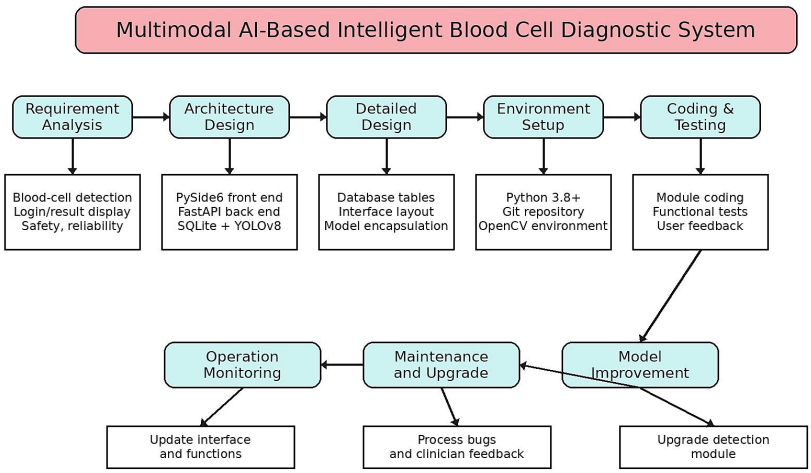

Figure 1 System Development and Implementation Route

The overall scheme is organized into data, model, service, application, and deployment layers. The data layer stores standardized microscopic images, labels, clinical text, user information, and detection records. The model layer contains preprocessing, the YOLOv8 detector, CMEM boundary enhancement, multimodal feature alignment, and model-management functions. FastAPI exposes inference and record services, SQLite supports lightweight local data management, and PySide6 provides the user interface. The deployment layer connects a lightweight edge model at the primary institution with a higher-precision cloud model for complex samples.

Figure 1 summarizes the proposed system development and implementation route. Requirement analysis defines blood cell detection, login, result display, safety, performance, reliability, and extensibility. System architecture design then selects PySide6, FastAPI, SQLite, and YOLOv8. Detailed design covers database tables, interface layout, and model encapsulation. Environment construction, modular coding, testing, operation monitoring, maintenance, and model upgrading form a continuous engineering cycle.

3.2 Core Modules and Data Flow

A user imports a microscope image or activates a camera stream. The preprocessing stage normalizes size and color and reduces background noise. The detector outputs cell locations, classes, and confidence values. Image-derived features may then be aligned with clinical-text features through a cross-modal attention mechanism, allowing cell morphology, patient age, symptoms, and routine blood indices to contribute to an integrated suggestion. Detection records are stored locally, while difficult or suspected abnormal specimens can be transmitted to the cloud for deeper analysis. The result is returned to the edge terminal and can be exported as a report.

Table 1 Core Modules of the Intelligent Blood Cell Diagnostic Scheme

Module	Main Design Content
Data layer	Microscopic images, clinical text, labels, user accounts, and detection records
Model layer	YOLOv8, WB-FPN, CMEM, multimodal fusion, pruning, distillation, and quantization
Service layer	FastAPI inference service, model management, report generation, and LIS interface
Application layer	PySide6 login, image/batch/video/camera detection, visualization, and export
Deployment layer	Lightweight edge screening and cloud-assisted analysis of difficult specimens

Table 1 organizes the core modules and their responsibilities within the proposed diagnostic scheme. Privacy protection is treated as part of the architecture. The proposed scheme uses institution-local storage of original records, federated exchange of model parameters, and differential-privacy noise when cross-institution collaboration is required. Recent federated-learning research provides a practical basis for collaborative model evaluation and distributed learning without centralizing all medical data [7-8]. The design nevertheless requires local access control, de-identification, encrypted communication, and explicit clinical governance before real deployment.

3.3 Engineering and Validation Targets

Table 2 lists the engineering and validation targets for the proposed system; these values are planned criteria rather than completed clinical results. They provide measurable criteria for later implementation. The detector is expected to complete a single-image analysis within 0.5 s after inference acceleration. The lightweight model is planned to reduce the initial 12 GB model scale to less than 1 GB and lower memory use by 80%, allowing operation on ordinary primary-care computers. Multi-center validation is intended to examine whether cell classification accuracy reaches 98%, lesion recognition reaches 92%, missed detection remains within 3%, and false detection remains within 2%.

Table 2 Engineering and Clinical Validation Targets

Target Item	Planned Criterion
Single-image inference time	≤ 0.5 s after inference acceleration
Lightweight model size	Reduced from 12 GB to < 1 GB
Memory use	Reduced by 80%
Cell classification accuracy	≥ 98%
Lesion recognition accuracy	≥ 92%
Missed / false detection rate	≤ 3% / ≤ 2%
Tertiary-hospital review efficiency	Improvement > 50%
Primary-care screening capability	Improvement target: 60%
Continuous-operation failure rate	≤ 1%

Application evaluation also includes workflow outcomes. In tertiary hospitals, AI-assisted slide review is expected to improve efficiency by more than 50% when compared with the existing manual process. In primary institutions, the intended goal is a 60% improvement in initial hematological screening capability, assessed together with installation convenience, learning cost, cloud consultation frequency, and user trust. Continuous operation should maintain a failure rate no greater than 1%. These targets must be verified using the locked multi-center protocol described later and should not be interpreted as achieved values in the present design study.

4 EDGE-ENHANCED YOLOV8 MODEL DESIGN

4.1 Detection Backbone and Feature Fusion

Blood cells present three major detection difficulties: severe overlap, large scale differences between platelets and larger cells, and weak boundaries caused by staining or illumination variation. The proposed scheme improves the YOLOv8 neck with a weighted bidirectional feature pyramid so that shallow spatial detail and deep semantic information can be combined. A CIoU and Focal Loss combination is used to improve localization while reducing the influence of class imbalance when abnormal cells are scarce. TensorRT quantization, pruning, and knowledge distillation are planned for efficient inference. Knowledge distillation is a mature model-compression strategy in which a compact student learns structured information from a stronger teacher [9].

Figure 2 presents the training and validation curves used for feasibility analysis. The training box, classification, and distribution-focal losses decrease during optimization, while precision, recall, mAP50, and mAP50-95 rise and then stabilize. Validation losses fluctuate more strongly than training losses, which supports this study's emphasis on continued data expansion, multi-center validation, and feedback-based fine-tuning rather than relying on a single training run.

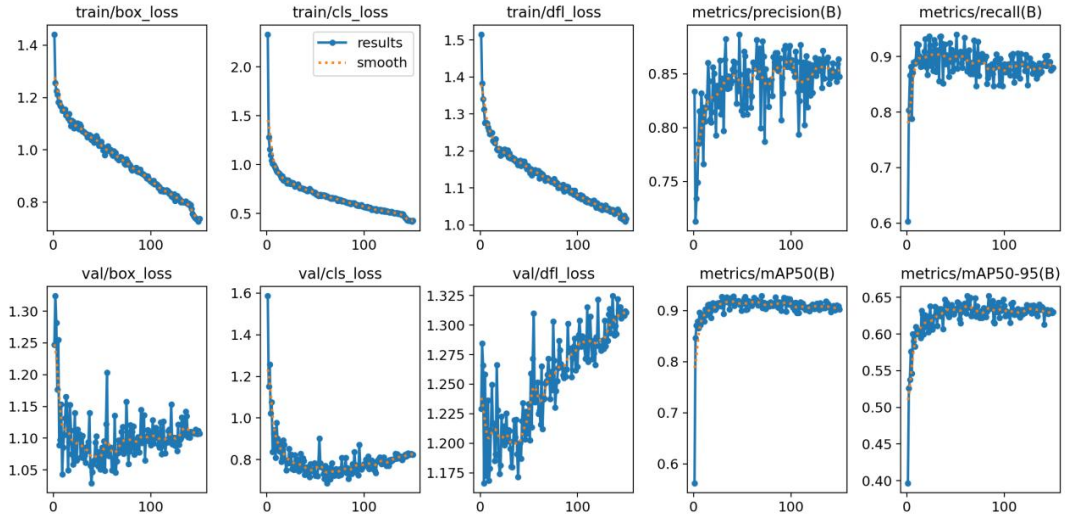

Figure 2 Training and Validation Losses and Evaluation Metrics

4.2 CMEM Mathematical Formulation

The CMEM component is designed to strengthen blurred cell boundaries by combining a deep semantic feature map with an explicit edge response. For an input image I(x,y), the local intensity gradient is defined as

$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right) \tag{1}$$

The gradient magnitude represents edge strength and is written as

$$G(x,y) = |\nabla I| = \sqrt{\frac{\partial I^2}{\partial x} + \frac{\partial I^2}{\partial y}} \tag{2}$$

The gradient direction represents the local boundary normal:

$$\theta(x,y) = \operatorname{atan2}\left(\frac{\partial I}{\partial y}, \frac{\partial I}{\partial x}\right) \tag{3}$$

An ideal edge produces a clear magnitude peak. A blurred boundary has a weaker response and less consistent direction in its neighborhood. Let Fbackbone denote the deep feature map extracted by the YOLOv8 backbone and Fedge denote the edge-strength feature. CMEM converts the edge response into a spatial attention map:

$$A = \sigma(F_{edge}) \tag{4}$$

The attention map modulates boundary-related regions through element-wise multiplication:

$$F_{enhanced} = A \odot F_{backbone} \tag{5}$$

The enhanced feature is fused with the original semantic feature so that boundary emphasis does not discard high-level cell information:

$$F_{fused} = \alpha F_{backbone} + \beta F_{enhanced} \tag{6}$$

Here, α and β are learnable fusion weights or may be initialized as residual weights. The complete information flow can be expressed as

$$F_{backbone} = \text{Backbone}(I) \tag{7}$$

$$F_{edge} = \text{EdgeDetector}(I) \quad (8)$$

$$F_{enhanced} = \text{CMEM}(F_{backbone}, F_{edge}) \quad (9)$$

$$Y = \text{DetectionHead}(F_{fused}) \quad (10)$$

The final fused representation is delivered to the YOLOv8 detection head for bounding-box regression and cell classification. This formulation follows the source document's edge-guided feature optimization concept and preserves its intended relationship between the backbone, edge detector, CMEM, and detection head.

5 SOFTWARE IMPLEMENTATION AND INTERACTION DESIGN

5.1 Data and Interface Design

The software is designed around a user table, a detection-record table, and model-management metadata. The login window controls account access, while the main interface supports one-click image upload, automatic analysis, visualized abnormal-cell marking, record retrieval, and report export. The detection engine is encapsulated so that different YOLO versions or heterogeneous models can be loaded without rewriting the interface. This separation improves maintainability and leaves space for later clinical functions.

Figure 3 shows the bounding-box parameter distribution. The coordinate and size distributions describe the spatial coverage and scale characteristics of annotated objects. They are useful for examining whether training samples are concentrated in particular regions or dominated by one object size. In the proposed workflow, this analysis is performed before training and after data expansion to identify sampling bias and insufficient coverage.

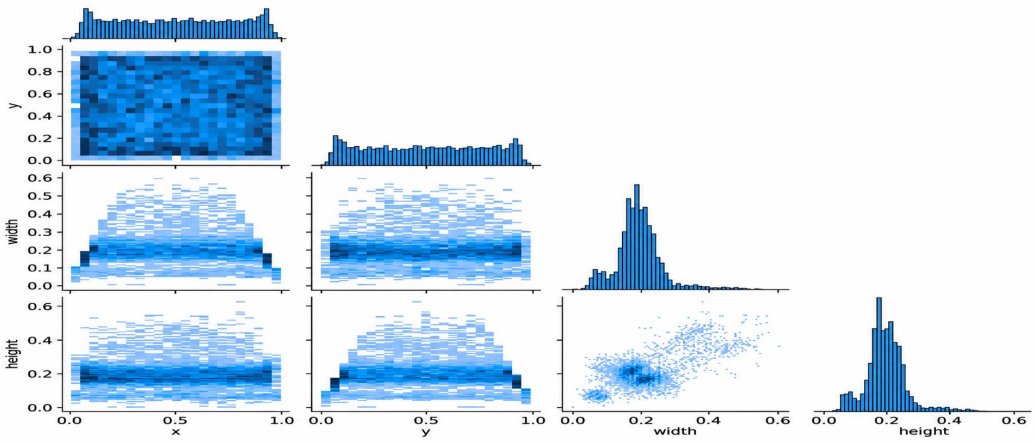

Figure 3 Distribution of Target Bounding-Box Parameters

5.2 Detection Workflow and Visual Output

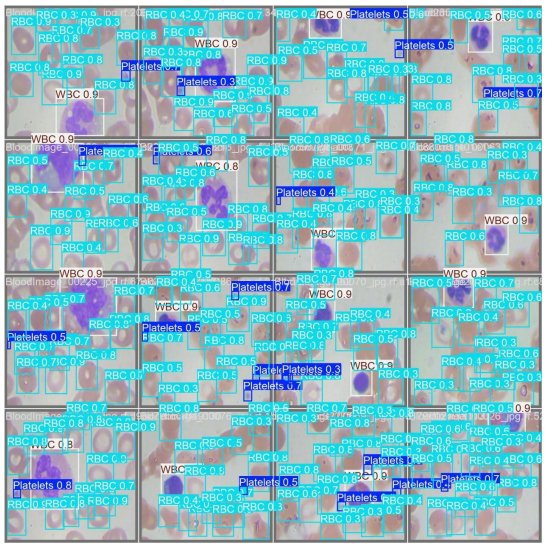

Figure 4 Visualization of Blood Cell Detection Results

Coding and testing are divided by module. Image, video, and camera inputs share the same preprocessing and inference interface. Unit tests cover login, image upload, model loading, result storage, and report generation; integrated tests examine interface responsiveness and continuous inference. Operational monitoring records upload failures, report-format errors, latency, and long-running stability so that maintenance can respond to actual usage.

Figure 4 presents the blood-cell detection visualization generated by the proposed workflow. Red blood cells, white blood cells, and platelets are marked with class labels and confidence values across multiple microscope fields. The visualization demonstrates the intended interaction form: the system does not hide the detector output but exposes candidate cells for professional review. Dense overlap and small platelets remain important design targets for CMEM enhancement and multiscale fusion.

6 PERFORMANCE EVIDENCE AND VALIDATION DESIGN

6.1 Evidence Retained from the Source

The available experimental materials provide training curves, bounding-box distributions, detection examples, and a recall-confidence curve rather than a complete independent clinical trial. Accordingly, this paper interprets these materials as feasibility evidence and does not convert them into unreported accuracy claims. Figure 5 shows the class-specific recall-confidence relationship for platelets, red blood cells, white blood cells, and all classes. Recall decreases as the confidence threshold increases, with distinct decline patterns among cell categories. The curve therefore supports class-aware threshold selection rather than a single rigid threshold for every cell type.

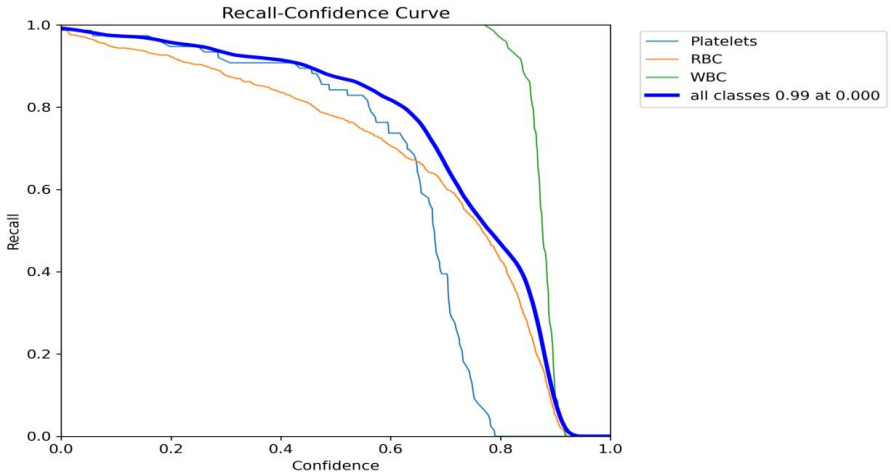

Figure 5 Recall-Confidence Curves for Blood Cell Categories

The validation plan proposes multi-center testing in tertiary and primary hospitals, with comparison against manual slide review and evaluation of classification accuracy, lesion recognition, missed detection, false detection, efficiency, deployment convenience, and user dependence. A rigorous implementation should lock the test set before model adjustment, report patient- or specimen-level partitioning, and document acquisition devices and staining protocols. Methodological reviews of medical imaging warn that leakage, weak external validation, and nonrepresentative samples can overstate model performance [10].

6.2 Clinical Feedback and Iterative Improvement

Clinician feedback forms a controlled update loop. Corrections to cell labels and abnormality prompts are reviewed before being added to the training set. Multi-center validation is used to examine differences among devices, specimens, and institutions. Difficult samples, including hemolytic and diluted specimens, are analyzed separately. This feedback mechanism is intended to improve consistency while preventing uncontrolled online learning from changing a deployed diagnostic model without verification.

7 DEPLOYMENT AND EXTENSIBILITY

Primary-care deployment uses a lightweight edge model for routine screening and a complete cloud model for complex specimens. The edge terminal performs rapid cell counting and common abnormality detection on ordinary computers. Suspected leukemia or other difficult samples can be uploaded through a controlled interface for deeper analysis, after which the result is returned to the local terminal. Model pruning, quantization, and teacher-student distillation reduce storage and memory requirements, while modular service interfaces permit connection with existing LIS workflows.

The design follows a one-core, multi-domain strategy. The common detection engine, preprocessing pipeline, data-management layer, and model-loading interface can be adapted to bacterial morphology, urinary sediment components, and sperm analysis. Transfer requires new domain-specific datasets and labels, adjustment of output categories and feature layers, and, for sperm quality analysis, an additional motion-trajectory module. This path avoids rebuilding the complete software platform for every microscopic examination task while keeping the clinical model and validation requirements domain specific.

8 CONCLUSION

This paper presents a design study for an intelligent blood cell diagnostic system. The scheme combines standardized multimodal data, an edge-enhanced YOLOv8 detector, image-text fusion, interpretable result visualization, lightweight deployment, and cloud-edge collaboration. CMEM uses explicit gradient and edge information to modulate deep features while retaining the original semantic representation. PySide6, FastAPI, SQLite, model encapsulation, record management, and report export provide an implementable software route. The retained figures indicate feasible optimization and detection behavior, but clinical effectiveness must be established through specimen-level, multi-center validation. The modular architecture also provides a reusable foundation for bacterial examination, urinary sediment detection, and sperm quality analysis after domain-specific data construction and model adaptation.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nature Medicine, 2022, 28(1): 31-38.
[2] Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. NPJ Digital Medicine, 2021, 4: 65.
[3] Matek C, Kerkhoffs T, Marr C, et al. Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. Blood, 2021, 138(20): 1917-1927.
[4] Kouzehkanan Z M, Sadeghian A, Hosseinzadeh M, et al. Raabin-WBC: A large free access dataset of white blood cells images. Medical & Biological Engineering & Computing, 2022, 60: 1427-1443.
[5] Terven J, Córdova-esparza D M, Romero-gonzález J A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
[6] Van Der Velden B H M, Kuijf H J, Gilhuijs K G A, et al. Explainable artificial intelligence in deep learning-based medical image analysis. Medical Image Analysis, 2022, 79: 102470.
[7] Karargyris A, Umeton R, Sheller M J, et al. Federated benchmarking of medical artificial intelligence with MedPerf. Nature Machine Intelligence, 2023, 5: 799-810.
[8] Wen J, Zhang Z, Lan Y, et al. A survey on federated learning: Challenges and applications. International Journal of Machine Learning and Cybernetics, 2023, 14(2): 513-535.
[9] Gou J, Yu B, Maybank S J, et al. Knowledge distillation: A survey. International Journal of Computer Vision, 2021, 129(6): 1789-1819.
[10] Varoquaux G, Cheplygina V. Machine learning for medical imaging: Methodological failures and recommendations for the future. NPJ Digital Medicine, 2022, 5: 48.