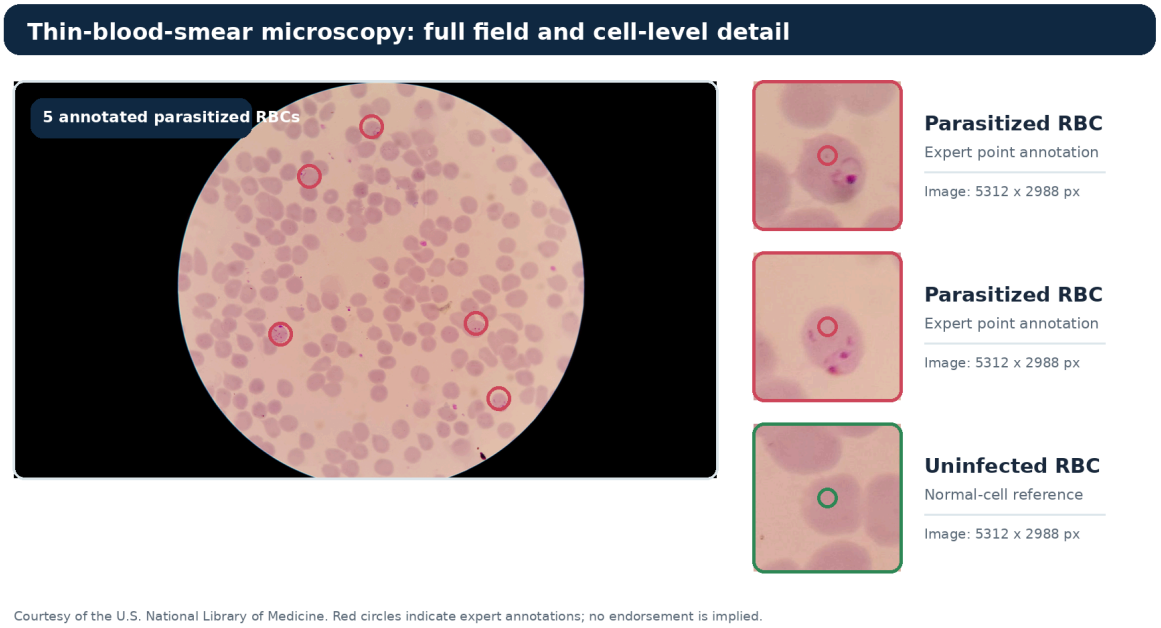
IDNEO | EDGE/DM Medtech DataSpace

Self-Supervised Hierarchical Cell Detection

and Malaria Evidence Scoring in Thin Blood Smear Microscopy

Public Technical Note and Research Model Card



Courtesy of the U.S. National Library of Medicine. Red circles indicate expert annotations; no endorsement is implied.

Self-supervised vision	Dense cell localisation	Hierarchical classification	Governed inference

EDGE/DM Data Product: Malaria Microscopy Inference Service

https://idneo.edgedm.eu/malaria-inference.html

Research and technical validation only. Not intended for diagnosis, patient management or clinical decision support during the pilot phase.

EXECUTIVE OVERVIEW

A high-resolution microscopy demonstrator for advanced machine-learning workflows

This Data Product demonstrates IDNEO's capability to design, train, validate and expose an end-to-end computer-vision pipeline for biomedical microscopy. The system analyses complete thin-blood-smear fields rather than pre-cropped cells, localises individual cells, classifies them through a biologically informed hierarchy, reconstructs image-level evidence and returns structured results through a controlled inference interface.

The demonstrator is intentionally positioned as a research asset. It is designed to support technical evaluation, collaboration and reuse of machine-learning patterns across MedTech imaging problems.

965	196,934	0.964	0.982
microscopy images	**cell annotations**	**image AUROC**	**image AUPRC**
full-resolution RGB fields	RBC, infected RBC and WBC	internal fold-0 ranking	internal fold-0 ranking

What the project demonstrates

Self-supervised representation transfer, dense small-object localisation, hierarchical classification, full-image tiling and reconstruction, quantitative image analysis, patient-grouped validation, structured API outputs and governed Data Product publication.

Public positioning

- **A reusable** computer-vision architecture for cellular and microscopic imaging research - not merely a binary image classifier.
- **A transparent** validation package that separates cell-level, image-level and patient-group aggregation metrics.
- **A governed** service concept that permits controlled inference without public release of the source dataset or model weights.
- **A starting point** for future multimodal research combining image evidence with acquisition metadata, structured knowledge or other contextual variables.

Scope boundary

The current operational model is image-centric. Multimodal fusion, detailed morphology, production-grade deployment and external clinical validation are research extensions rather than validated capabilities of this version.

DATA PRODUCT PROFILE

Governed inference service in the EDGE/DM catalogue

EDGE/DM publishes this resource as a sandbox-validated inference service for controlled technical validation. Public catalogue information describes a request-based access model, with operational availability subject to runtime/connector integration, eligibility review, applicable DUA/DUP conditions and a traceable access decision [1].

Data Product	Malaria Microscopy Inference Service
Provider	IDNEO
Resource type	Controlled inference service / API
EDGE/DM status	Validated in sandbox; runtime/connector integration pending
Purpose	Controlled inference, technical validation and research collaboration
Input	RGB thin-blood-smear microscopy image plus research identifiers
Output	Cell counts, parasitemia proxy, continuous malaria evidence score and cell-level evidence
Permitted use	Research inference, technical validation, service evaluation
Prohibited use	Clinical diagnosis, medical decision-making, patient management, unauthorised data or model extraction

Why a governed service

The value of the Data Product lies not only in the model, but in the combination of a documented input/output contract, versioned evidence, controlled execution and explicit usage conditions. This supports collaboration while preserving provider control over model artefacts and execution context.

- Public discovery of the capability and its evidence package.
- Request-based, purpose-bound access to an inference runtime.
- Traceable model version, thresholds and execution metadata.
- No public distribution of model weights or the source dataset through the service.

Data minimisation

The inference request requires an image and research identifiers only. The demonstrator should use pseudonymous or synthetic identifiers and must not receive directly identifying patient information.

DATASET AND EVIDENCE BASE

High-resolution, expert-annotated thin-blood-smear microscopy

Development uses the public NIH-NLM Thin Blood Smears Pf collection. According to the dataset ReadMe, it contains images from 193 patient groups, with five images per group, acquired at Chittagong Medical College Hospital in Bangladesh. Giemsa-stained thin smears from Plasmodium falciparum-infected cases and healthy controls were photographed using a smartphone camera through a microscope eyepiece. The images were manually annotated by an expert, de-identified and released with point and polygon annotations [2,4].

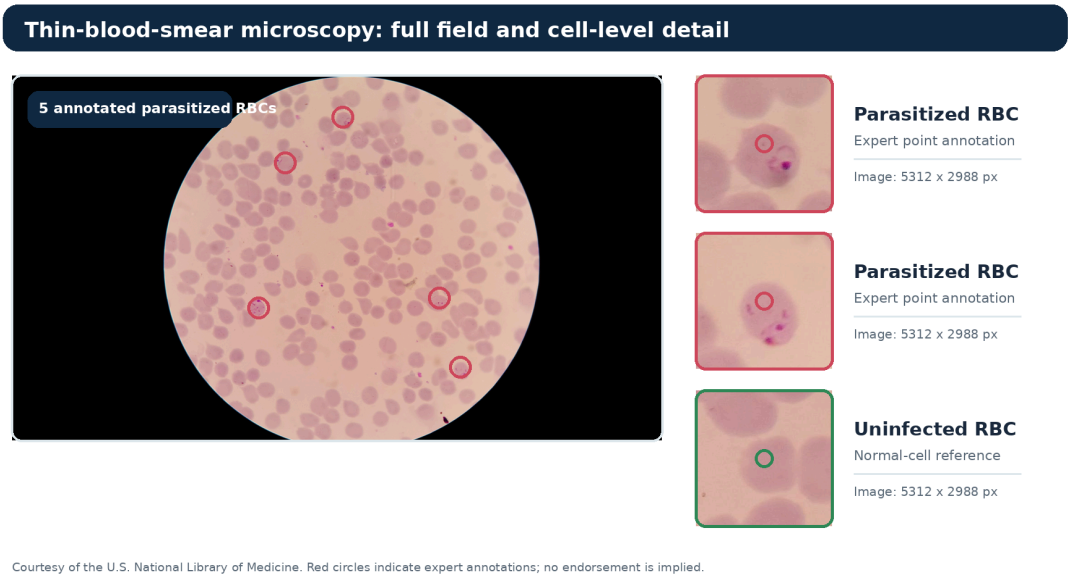


Figure 1. Full microscopy field and annotated cell-level detail. Courtesy of the U.S. National Library of Medicine; no endorsement is implied. Dataset and requested scientific citation: [2,4].

Dataset profile used by the pipeline

Property	Value
Images	965 full-resolution RGB microscopy fields
Patient groups	193; five images per patient
Image dimensions	5312 x 2988 pixels
Cell annotations	196,934 total
Class counts	188,711 uninfected RBC; 7,952 parasitized RBC; 271 WBC

Property	Value
Annotation formats	Expert point and polygon annotations

Image-use attribution

Dataset images in this note are reproduced only as limited technical illustrations. Courtesy of the U.S. National Library of Medicine; no endorsement is implied. The Data Product does not redistribute the dataset. Any reuse must follow the current dataset agreement and NLM terms, retain the requested attribution and cite the source publication [2-4].

MACHINE-LEARNING ARCHITECTURE

Self-supervised visual representations combined with task-specific dense prediction

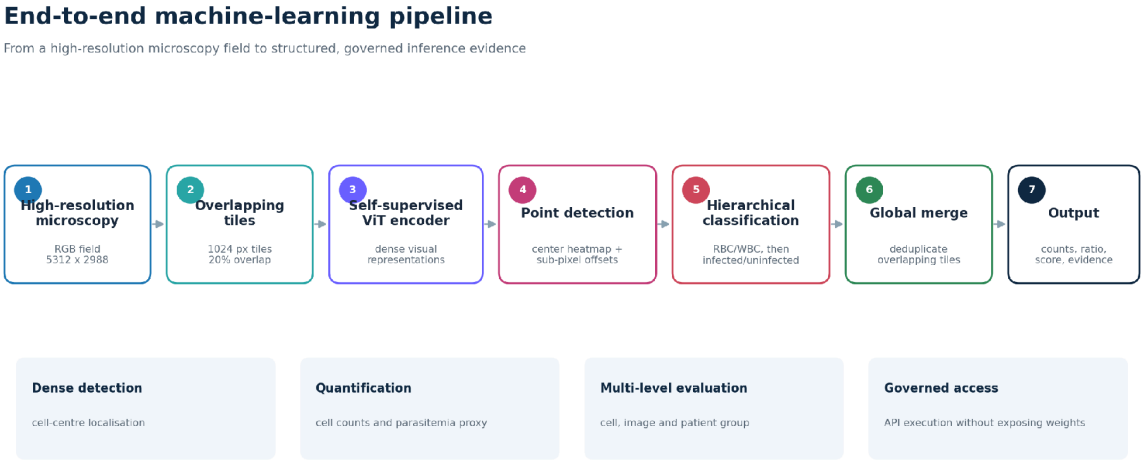


Figure 2. End-to-end inference pipeline from a full microscopy field to structured evidence.

Core architecture

Component	Implementation	Role
Input handling	1024 x 1024 tiles, 20% overlap; tiles resized to 768 x 768	Scales high-resolution fields to GPU memory while preserving local detail
Visual encoder	Small-patch Vision Transformer initialised through self-supervised visual pretraining; 384-dimensional patch embeddings	Transfers generic visual representations to microscopy while supporting task-specific adaptation [5,6]
Feature pyramid	Single-scale ViT map projected to 256 channels and upsampled	Provides a denser feature grid for small-cell localisation
Point detector	One-channel centre heatmap plus two sub-pixel offsets	Localises cell centres without requiring bounding boxes
Cell classifier	Two binary heads: RBC/WBC and infected/uninfected for RBC	Encodes domain hierarchy instead of a flat three-class softmax
Global reconstruction	Coordinate conversion and 20-pixel duplicate merge	Combines overlapping tile predictions into one image result
Image evidence score	Maximum centre x RBC x infection confidence among infected predictions	Ranks image-level evidence; not a calibrated clinical probability

ENGINEERING AND TRAINING

A modular research implementation designed for controlled experimentation

The codebase separates dataset preparation, tiling, target generation, model components, training, evaluation, calibration, deployment and API layers. This modular structure supports ablation studies and reuse in adjacent microscopic-imaging use cases.

Training configuration

Optimiser	AdamW
Epoch budget	60
Batching	Batch size 8; gradient accumulation 2
Learning rates	Heads 3e-4; backbone 1e-5
Backbone adaptation	Last four transformer blocks unfrozen
Schedule	5-epoch warm-up followed by cosine decay
Numerics	BF16 automatic mixed precision on CUDA
Stability controls	Gradient clipping, finite-loss checks and deterministic seeds

Multi-task supervision

- Centre heatmap focal loss for dense, class-agnostic localisation [7].
- Offset regression loss for sub-pixel cell-centre refinement.
- RBC/WBC binary classification loss.
- Infected/uninfected binary classification loss restricted to RBC semantics.
- Optional count, global-image and mask heads retained for future ablation, but disabled in the validated configuration.

Capabilities demonstrated beyond model training

Reproducible data splits Patient-grouped folds and versioned manifests	**Full-image inference** Tiling, black-region rejection and coordinate reconstruction	**Metric hierarchy** Cell, image and patient-group evaluation
Structured evidence	**Service packaging**	**Research extensibility**

Counts, score, ratio and coordinates in JSON	FastAPI, Docker and health/metadata endpoints	Optional KG, morphology and edge-optimisation pathways

VALIDATION METHODOLOGY

Patient-grouped validation with full-image reconstruction

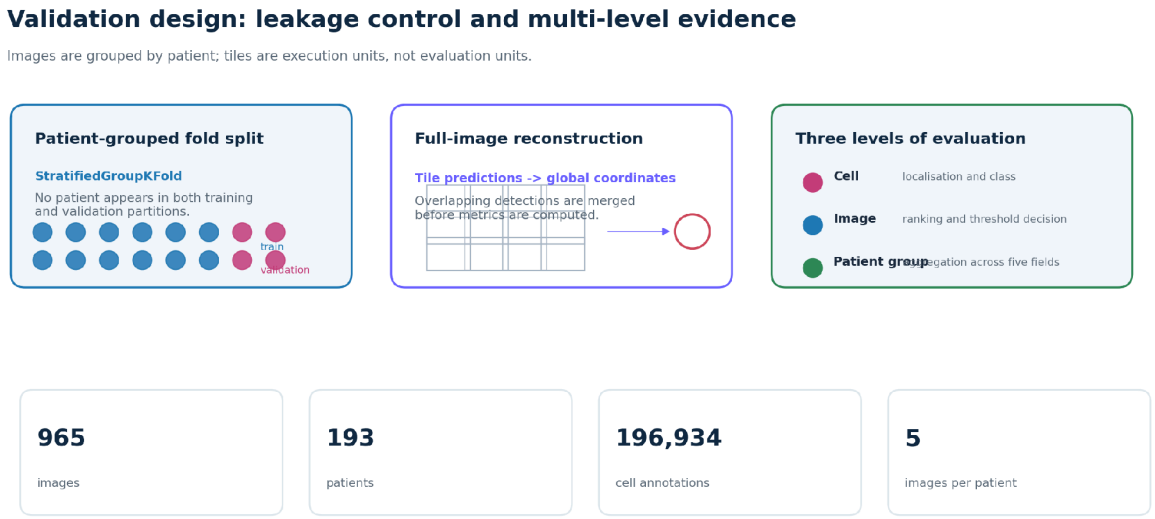


Figure 3. Validation design and evidence hierarchy.

Fold-0 evaluation protocol

Validation subset	190 images from 38 held-out patient groups
Group composition	29 positive and 9 negative patient groups
Cell annotations in fold 0	39,565 total: 1,665 parasitized, 37,855 uninfected and 45 WBC
Centre threshold	0.35
Local NMS kernel	11 feature-map pixels
Duplicate merge radius	20 image pixels
Image decision threshold	0.50 on continuous evidence score
Evaluation levels	Cell localisation/classification, image ranking/decision, patient-group aggregation

Interpretation of the validation

Fold 0 is an internal research validation partition, not an external test cohort. The same fold was used for checkpoint/threshold-related development decisions, so the results should be interpreted as preliminary technical evidence rather than a final unbiased benchmark.

RESULTS

Strong image-level ranking with moderate cell-level localisation and classification

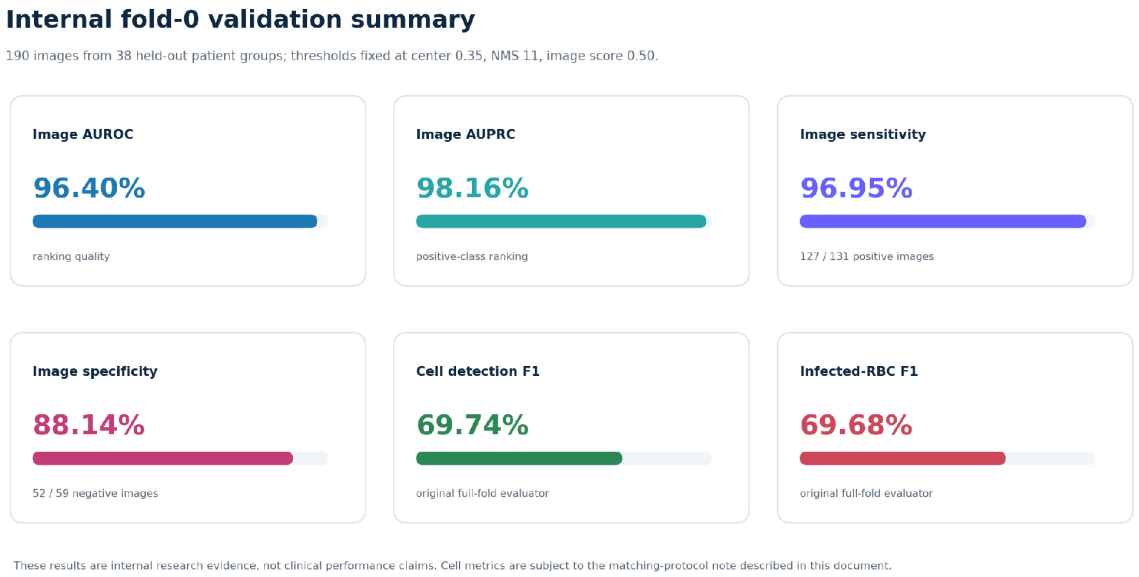


Figure 4. Internal fold-0 performance summary. Cell metrics shown are from the original full-fold evaluator.

Image and patient-group results

Level	Metric	Result	Observed counts	95% exact CI
Image	Sensitivity	96.95%	127 TP / 4 FN	92.37-99.16%
Image	Specificity	88.14%	52 TN / 7 FP	77.07-95.09%
Image	AUROC	0.9640	190 images	-
Image	AUPRC	0.9816	190 images	-
Patient group	Sensitivity	100.00%	29 TP / 0 FN	88.06-100.00%
Patient group	Specificity	55.56%	5 TN / 4 FP	21.20-86.30%

Cell and quantitative results

Metric	Result	Interpretation
Cell-centre detection F1	69.74%	Original full-fold evaluator; localisation within 20 px
Infected-RBC F1	69.68%	Original full-fold evaluator
Mean localisation distance	10.96 px	Mean distance for matched detections
Parasitized count MAE	0.7947 cells/image	Absolute error in infected-cell count
Parasitemia MAE	0.0043 ratio units	Absolute error of the detected-cell ratio proxy

Threshold trade-off

At an image threshold of 0.60, patient-group sensitivity decreased to 96.55% while specificity increased to 88.89%. This illustrates that the service threshold is a research operating point, not a clinically established decision boundary.

METRIC QUALITY ASSURANCE

Transparent audit of the cell-matching protocol

Review of the original cell evaluator identified a matching edge case: fine-class compatibility was heavily weighted before the spatial-radius constraint was applied. A prediction could therefore be paired with a distant same-class cell and later discarded, even when a spatially valid cell of another class was nearby. This affects the interpretation of localisation and hierarchical classification metrics.

Partial corrected recalculation

The supplied checkpoint was re-run on the five curated fold-0 images included in the technical package. A corrected one-to-one spatial matching procedure first restricts candidate pairs to 20 pixels, maximises the number of valid matches and then evaluates class labels. The remaining 185 fold-0 images were not available in the package, so this is a quality-assurance audit rather than a replacement for the full-fold metrics.

Score, not calibrated probability

The current score is computed from centre, RBC and infection confidences and aggregated by the maximum infected-cell evidence. It is useful for ranking and threshold-based research experiments, but it has not been calibrated as a probability of disease. Public-facing schemas should use the term malaria evidence score; any legacy probability field must be interpreted accordingly.

Metric	Original matching	Corrected matching	Change
Detection true positives	774	782	+8
Detection false positives	322	314	-8
Detection false negatives	313	305	-8
Detection precision	70.62%	71.35%	+0.73 pp
Detection recall	71.21%	71.94%	+0.74 pp
Detection F1	70.91%	71.64%	+0.73 pp

Illustrative audit examples

Figure 5. Curated inference examples used in the partial audit. All five image-level decisions were correct, but the subset was selected for demonstration and is not an unbiased performance estimate.

WBC metric caveat

The five audited images contain no WBC ground-truth annotations. Therefore WBC sensitivity, precision and F1 are undefined on this subset. The historical 100% RBC-positive score must not be interpreted as validated WBC discrimination.

INFERENCE INTERFACE

Structured outputs designed for governed evaluation

The service exposes health, metadata and prediction endpoints. Prediction requests use multipart form data and return a deterministic image-level structure containing the model decision, continuous evidence score, cell counts, parasitemia proxy, selected cell-level evidence and inference metadata.

Request example

```
curl -X POST 'https://<governed-runtime>/predict' \
  -H 'Authorization: Bearer <access-token>' \
  -F 'file=@thin_smear_field.jpg' \
  -F 'patient_id=research-subject-001' \
  -F 'image_id=field-001'
```

Output contract

Output block	Representative fields	Meaning
Identification	schema_version, patient_id, image_id	Versioning and request traceability
Assessment	decision, indications_present, score, threshold	Research image-level result
Quantification	cell counts; parasitemia ratio and percentage	Detected cellular composition and ratio proxy
Evidence	class, centre coordinates, confidence components	Inspectable cell-level evidence
Inference metadata	latency, device, thresholds, number of detections	Reproducibility and audit context
Operations and errors	/health, /metadata; structured error codes	Runtime discovery and controlled failure modes

Recommended evidence metadata for each governed execution

- Model version and checkpoint SHA-256.
- Input hash, execution timestamp and trace identifier.
- Thresholds, device, latency and output schema version.
- Access-policy reference and purpose-bound authorisation record.

RESPONSIBLE USE AND LIMITATIONS

What the current evidence supports - and what it does not

Supported technical claim	Boundary / limitation
The model performs full-field cellular localisation and hierarchical classification.	Cell-level F1 is moderate; localisation of every infected RBC is not yet reliable.
Image-level scores show strong ranking on the internal fold-0 partition.	The partition is internal and was involved in model/threshold development; no external cohort has been evaluated.
Patient-group aggregation can combine five fields.	The OR/max strategy improves sensitivity but can reduce specificity; the negative group is small.
The pipeline estimates infected-cell counts and a parasitemia proxy.	The estimate depends on detection/classification balance and is not a validated laboratory measurement.
The architecture uses a modern self-supervised visual encoder.	The operational input is currently image-only; this version is not a complete multimodal system.
Cell centres and types enable quantitative cellular analysis.	Detailed morphology, segmentation, parasite stage and shape descriptors are not validated outputs.
The code contains deployment pathways.	The validated runtime is PyTorch-based; complete end-to-end TensorRT parity is not demonstrated.

Mandatory public-use statement

This demonstrator is provided for research purposes, software evaluation and controlled technical validation. It must not be used to diagnose malaria, guide treatment, manage patients or replace qualified microscopy review.

Research roadmap

- Recalculate all fold-0 cell metrics with the corrected spatial matcher and retain per-image prediction artefacts.
- Generate five-fold out-of-fold evidence and a strictly separated final evaluation protocol.
- Add image-quality and out-of-distribution checks with an inconclusive outcome.
- Calibrate image and patient-group scores; assess Brier score and calibration error.
- Evaluate WBC performance with adequate positive examples and macro-averaged metrics.
- Extend to cell segmentation and validated morphology descriptors where data support them.
- Explore multimodal fusion with acquisition metadata and structured domain knowledge.

EVIDENCE PACKAGE AND REFERENCES

Recommended public citation

Suggested citation

IDNEO. Self-Supervised Hierarchical Cell Detection and Malaria Evidence Scoring in Thin Blood Smear Microscopy: Public Technical Note and Research Model Card. EDGE/DM Data Product, version 1.1, 2026.

References

[1] EDGE/DM. Malaria Microscopy Inference Service. Public Data Product page. Accessed 29 July 2026. https://idneo.edgedm.eu/malaria-inference.html

[2] U.S. National Library of Medicine, National Institutes of Health. NIH-NLM Thin Blood Smears Pf dataset, ReadMe and current dataset files. Accessed 29 July 2026. https://data.lhncbc.nlm.nih.gov/public/Malaria/NIH-NLM-ThinBloodSmearsPf/

[3] U.S. National Library of Medicine. Terms and Conditions for downloaded NLM data. Accessed 29 July 2026. https://www.nlm.nih.gov/databases/download/terms_and_conditions.html

[4] Kassim YM, Palaniappan K, Yang F, et al. Clustering-Based Dual Deep Learning Architecture for Detecting Red Blood Cells in Malaria Diagnostic Smears. IEEE J Biomed Health Inform. 2021;25(5):1735-1746. doi:10.1109/JBHI.2020.3034863. https://doi.org/10.1109/JBHI.2020.3034863

[5] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. ICLR 2021. https://openreview.net/forum?id=YicbFdNTTy

[6] Caron M, Touvron H, Misra I, et al. Emerging Properties in Self-Supervised Vision Transformers. ICCV 2021:9650-9660. https://openaccess.thecvf.com/content/ICCV2021/html/Caron_Emerging_Properties_in_Self-Supervised_Vision_Transformers_ICCV_2021_paper.html

[7] Lin T-Y, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. ICCV 2017:2980-2988. https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

Contact and access

Access is requested through the EDGE/DM participation and Data Product workflow: https://idneo.edgedm.eu/participation.html