Supplementary Methods

Note on table nomenclature: Tables SM1 and SM2 refer to tables contained within this Supplementary Methods document. Supplementary Tables S1 and S2 are provided as separate files accompanying the main manuscript.

Two-stage LSSA Deep-Learning Classifier: Hyperparameter Optimization and Network Architecture

Supplementary Methods Table SM1. Hyperparameter grid search results

Grid search was performed over 25 combinations of learning rate (5 values) and training epochs (5 values), with batch size fixed at 10 and the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1\times10^{-7}$) in all runs. The reported performance metric was the mean 5-fold cross-validation accuracy of the training-plus-validation pool (n = 240 for Model 1; n = 60 for Model 2). The hold-out test set was not used at any stage of the hyperparameter selection.

Table SM1. Mean 5-fold cross-validation accuracy (%) for Model 1 (3-class classifier, n = 240 training-plus-validation vials)

Learning Rate \ Epochs	50	100	200	500	1000
1×10^{-3}	53.5%	55.0%	53.2%	49.0%	46.5%
5×10^{-3}	45.5%	47.5%	45.8%	42.5%	40.0%
1×10^{-4} [b)]	63.5%	71.0%	74.5%	**77.5%**[a)]	76.2%
5×10^{-4}	59.0%	67.5%	72.0%	75.0%	73.8%
1×10^{-5}	48.5%	55.0%	63.5%	69.0%	71.5%

[a)] Selected configuration: LR = 1×10^{-4}, epochs = 500 (mean 5-fold CV accuracy = 77.5%). [b)] Row corresponding to the selected learning rate.

The same hyperparameter configuration (LR = 1×10^{-4}, epochs = 500) was also selected for Model 2 based on an independent grid search over the Model 2 training-plus-validation pool (n=60 vials). The final hold-out test accuracy reported in Fig. 2(A) (73.8%) reflects evaluation on the isolated 25% hold-out test set (n=80 vials) and differs from the cross-validation accuracy shown here, as the hold-out test set was entirely withheld from all stages of model development.

Supplementary Methods Table SM2. Layer-by-layer CNN architecture specifications

The two-stage classifier used a two-stream convolutional neural network (CNN) architecture identical to the previously validated Bacometer model [17]. The spatial stream is a time-distributed DenseNet applied frame-by-frame to raw speckle images, and the temporal stream is a DenseNet-based 1D CNN applied to 2-channel optical-flow feature maps. Features from both streams were aggregated using global average pooling (GAP), concatenated, and passed to the final classification layer.

Table SM2a. Spatial stream (time-distributed DenseNet over raw speckle frames)

Layer	Description	Input shape	Filter / Units	Output shape	Params
Input	Speckle frames (time-distributed)	300×256×256×1	—	300×256×256×1	0
Initial Conv	7×7 conv, stride 2, padding 3	256×256×1	7×7, 24	128×128×24	1,200
MaxPool	3×3 max pool, stride 2	128×128×24	3×3, stride 2	64×64×24	0
Dense Block 1	6 BC layers (BN+ReLU+1×1Conv+BN+ReLU+3×3Conv)	64×64×24	k=12, 6 layers	64×64×96	48,240
Transition 1	BN+ReLU+1×1 Conv (θ=0.5) + AvgPool 2×2	64×64×96	1×1, 48	32×32×48	4,848
Dense Block 2	12 BC layers	32×32×48	k=12, 12 layers	32×32×192	132,480
Transition 2	BN+ReLU+1×1 Conv (θ=0.5) + AvgPool 2×2	32×32×192	1×1, 96	16×16×96	18,912
Dense Block 3	24 BC layers	16×16×96	k=12, 24 layers	16×16×384	408,960
Final BN+ReLU	Batch normalization + ReLU	16×16×384	—	16×16×384	768
GAP (per-frame)	2D Global Avg Pool	16×16×384	—	384-dim vector	0
Time aggregation	GAP across 300 frames (time axis)	300×384	—	384-dim vector	0
Spatial subtotal					615,408

Table SM2b. Temporal stream (CNN over optical-flow heatmaps)

Layer	Description	Input shape	Filter / Units	Output shape	Params
Input	Optical-flow chunks (time-distributed)	100×256×256×2	—	100×256×256×2	0
Conv-BN-ReLU 1	3×3 conv, stride 2	256×256×2	3×3, 32	128×128×32	672
Conv-BN-ReLU 2	3×3 conv, stride 2	128×128×32	3×3, 64	64×64×64	18,624
Conv-BN-ReLU 3	3×3 conv, stride 2	64×64×64	3×3, 128	32×32×128	74,112
Conv-BN-ReLU 4	3×3 conv, stride 2	32×32×128	3×3, 256	16×16×256	295,680
Conv-BN-ReLU 5	3×3 conv (final block)	16×16×256	3×3, 256	16×16×256	590,592

GAP (per-chunk)	2D Global Avg Pool	16×16×256	—	256-dim vector	0
Time aggregation	1D GAP across 100 chunks	100×256	—	256-dim vector	0
Temporal subtotal					979,680

Table SM2c. Concatenation and classification heads

Layer	Description	Input shape	Units	Output shape	Params
Concatenation	Spatial GAP ⊕ Temporal GAP	384 + 256	—	640-dim vector	0
Dense (Model 1)	Final classification, softmax	640	3	3 (probabilities)	1,923
Dense (Model 2)	Final classification, softmax	640	2	2 (probabilities)	1,282

Table SM2d. Total trainable parameters

Model	Spatial stream	Temporal stream	Final dense	Total
Model 1 (3-class)	615,408	979,680	1,923	1,597,011 (≈1.60 M)
Model 2 (binary)	615,408	979,680	1,282	1,596,370 (≈1.60 M)

Supplementary Methods — Additional Training Details

Loss functions

Model 1 — Categorical focal loss: To address the 2:1:1 class imbalance (Negative ($<10^3$ CFU/mL); Positive—C. albicans ($\geq10^4$ CFU/mL); Positive—Non-albicans ($\geq10^4$ CFU/mL)*C. albicans* : Positive—Non-albicans = 160 : 80 : 80), categorical focal loss was used with $\gamma = 2.0$ and class-balanced $\alpha = [0.25, 0.50, 0.50]$ for Negative, Positive—*C. albicans*, and Positive—Non-albicans, respectively. The focal loss downweights easy majority-class examples and focuses on learning from hard minority-class samples without oversampling or undersampling.

Model 2 — Binary cross-entropy: Model 2 training data were balanced (1:1 class ratio: 40 *C. albicans* and 40 *C. tropicalis* vials). Standard binary cross-entropy was used with no additional class weighting.

Data leakage prevention (5-layer strategy)

Layer	Measure	Description
1	Vial-level splitting	All 300 frames from a single vial were assigned exclusively to one partition (train, validation, or hold-out test). No frame-level splitting was performed.
2	Pre-isolated hold-out test set	The 25% hold-out test partition was set aside before any model development, hyperparameter selection, or fold assignment.
3	CV restricted to training pool	5-fold cross-validation was applied only within the 75% training-plus-validation pool. The hold-out test set was never seen during CV.
4	Independent acquisition sessions	Training and test set acquisitions were performed in separate experimental sessions to prevent temporal autocorrelation.
5	Inter-model dataset separation	Model 1 (n=320) and Model 2 (n=80) datasets were acquired in entirely independent experimental runs. No vial was shared between the two datasets.

Reproducibility note

A random seed was set for weight initialization, data shuffling, and fold assignment to enable approximate reproducibility within the same software environment. However, the exact seed values were not formally archived in the training logs of the present study, and exact numerical reproducibility across different hardware or software versions cannot be guaranteed. Vial-level fold assignment is available upon reasonable request. The full training code will be archived in a public repository upon acceptance.

Reference: [17] Lee KS, Lim HJ, Kim K, Park YG, Yoo JW, Yong D. Rapid bacterial detection in urine using laser scattering and deep learning analysis. Microbiol Spectr 2022;10(2):e01769-21.