Supplementary Table S1. Dataset composition and class-stratified split assignment for the two-stage LSSA deep-learning classifier

This Supplementary Table reports the dataset composition and data-splitting strategy of the two convolutional neural network (CNN) classifiers developed in this study. Tables S1a and S1c summarize the per-species × per-concentration vial counts for Model 1 and Model 2 reference strain datasets, respectively. Tables S1b and S1d summarize the class-level stratified split assignments (training, 60%; validation, 15%; holdout test, 25%) used to develop and evaluate each model. Cross-validation was performed within the 75% train-plus-validation pool using a stratified 5-fold cross-validation. Vial-level fold assignments were not formally archived in the training logs of the present study; this limitation has been added to the reporting checklist for future work.

Table S1a. Model 1 dataset composition (n = 320 vials, per-species × per-concentration vial counts)

Species	**10^2 CFU/mL**	**10^3 CFU/mL**	**10^4 CFU/mL**	**10^5 CFU/mL**	**Total**
C. albicans ATCC 15043	40	40	40	40	**160**
N. glabratus KCTC 7219[a)]	10	10	10	10	**40**
C. tropicalis KCTC 7212	10	10	10	10	**40**
C. parapsilosis ATCC 22019	10	10	10	10	**40**
P. kudriavzevii ATCC 6258[b)]	10	10	10	10	**40**
Total	**80**	**80**	**80**	**80**	**320**

[a)] Formerly *Candida glabrata*; reclassified as *Nakaseomyces glabratus*.

[b)] Formerly *Candida krusei*; reclassified as *Pichia kudriavzevii*.

Note on in-vial concentrations: Concentration values (10^2, 10^3, 10^4, 10^5 CFU/mL) refer to the working suspension concentrations used to prepare the inocula, as confirmed by concurrent SDA culture. A Candida suspension of 0.5 McFarland standard (approximately 2×10^6 CFU/mL) was prepared, and 25 µL of a 10-fold serial dilution was added to 5 mL of LB broth to prepare working solutions with final concentrations of 10^2, 10^3, 10^4, and 10^5 CFU/mL. Two milliliters of each of these solutions were measured using the Bacometer.

Table S1b. Model 1 class-stratified split assignment (3-class classifier)

Class	**Composition**	**Total**	**Train (60%)**	**Val (15%)**	**Hold-out test (25%)**
Negative ($<10^3$ CFU/mL)	*C. albicans* 80 + non-albicans 80	160	96	24	40
Positive — *C. albicans* ($\geq 10^4$)	*C. albicans* only	80	48	12	20
Positive — Non-albicans ($\geq 10^4$)	4 non-albicans species	80	48	12	20

Total	—	**320**	**192**	**48**	**80**

Stratification was performed at the class level. Stratified five-fold cross-validation was applied to the 75% train-plus-validation pool (n=240). The holdout test partition (n=80) was isolated before fold assignment and used exclusively for the final performance estimation reported in Fig. 2(A).

Table S1c. Model 2 dataset composition (n = 80 vials, per-species × per-concentration vial counts)

Species	10^4 CFU/mL	10^5 CFU/mL	Total
C. albicans ATCC 15043	20	20	**40**
C. tropicalis KCTC 7212	20	20	**40**
Total	**40**	**40**	**80**

Model 2 was trained on a fully independent dataset acquired in separate experimental sessions from the Model 1 dataset. No vial used for Model 1 training, validation, or testing was used for Model 2 development.

Table S1d. Model 2 class-stratified split assignment (binary classifier)

Class	Total	Train (60%)	Val (15%)	Hold-out test (25%)
C. albicans	40	24	6	10
C. tropicalis	40	24	6	10
Total	**80**	**48**	**12**	**20**

Stratification was performed at the class level. Stratified five-fold cross-validation was applied to the 75% train-plus-validation pool (n=60). The hold-out test partition (n=20) was isolated before fold assignment and used exclusively for the final performance estimation. The cascade evaluation results reported in Fig. 2(B) (n=36) used reference strain samples independent of both the Model 1 and Model 2 training sets.

Notes on data splitting and reproducibility

1. Stratification unit: Data splitting was performed with class-level stratification (three classes for Model 1 and two classes for Model 2) and not species- or concentration-level stratification. Because the class labels are defined by concentration thresholds ($<10^3$ vs. $\geq 10^4$ CFU/mL), class-level stratification inherently preserves the concentration distribution across partitions.

2. Cross-validation: Stratified 5-fold cross-validation was applied within the train-plus-validation pool, preserving the class proportions in each fold.

3. Hold-out test isolation: The 25% hold-out test partition was set aside before any cross-validation fold assignment and used exclusively for the final performance evaluation.

4. Inter-model dataset independence: The Model 1 (n = 320) and Model 2 (n = 80) datasets were acquired in independent experimental runs and were entirely non-overlapping.

5. Reproducibility limitation: Vial-level fold assignment was not formally archived in the training logs of the present study. The full training code and per-vial fold assignment will be archived in a public repository upon acceptance.