Scientific Reports

https://doi.org/10.1038/s41598-026-48015-2

Article in Press

Explainable AI to interpret advanced computer vision fungal pathogen prediction

Received: 14 January 2026

Accepted: 6 April 2026

Published online: 18 April 2026

Cite this article as: Guthrie J.D., Shankarnarayan S.A. & Charlebois D.A. Explainable AI to interpret advanced computer vision fungal pathogen prediction. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-48015-2

Joshua D. Guthrie, Shamanth A. Shankarnarayan & Daniel A. Charlebois

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Explainable AI to Interpret Advanced Computer Vision Fungal Pathogen Prediction

Joshua D. Guthrie[1,+], **Shamanth A. Shankarnarayan**[1,+], **and Daniel A. Charlebois**[1,2,*]

[1]Department of Physics, University of Alberta, Edmonton, Alberta, Canada
[2]Department of Biomedical Engineering, University of Alberta, Edmonton, Alberta, Canada
*Corresponding author: dcharleb@ualberta.ca
+These authors contributed equally to this work

ABSTRACT

Antimicrobial resistance is a growing concern, with pathogenic fungi making a substantial contribution to untreatable life-threatening infections across the globe. Artificial intelligence is increasingly used in microbiology and antimicrobial resistance research, with promise to improve clinical diagnostics and infectious disease treatments. However, how AI models make predictions remains largely unknown, which is a major hurdle for human trust and regulatory approval. We train vision transformers (Swin Transformer-Tiny and Vision Transformer-Base 16) and convolutional neural networks (DenseNet121 and InceptionV3) to quickly and accurately identify pathogenic yeast species from microscopy images. Using explainable AI (Occlusion Sensitivity and Grad-CAM), we identify biologically relevant features (organelle, cell interior, cell wall, budding patterns/scars, and optical patterns) and irrelevant image features (background artifacts) these high-performance computer vision models may be using for pathogenic yeast classification. These findings improve AI-based microscopy classification of pathogenic yeast species and advance our understanding of the visual features these models rely on to make predictions.

ARTICLE IN PRESS

1 Introduction

Globally, 6.5 million fungal infections are estimated to occur annually and are associated with 3.8 million deaths, of which 2.5 million deaths are directly attributed to these infections[1]. *Candida* bloodstream infection (candidemia) and invasive candidiasis result in nearly 1 million deaths annually[1]. Many fungal infections are underreported and underdiagnosed, and treatment is restricted to a few classes of antifungal drugs[2]. Candidemia is the fourth most common infectious disease in intensive care units and occurs predominantly in hospitalized patients with underlying risk factors; however, candidemia is not limited to patients who are immune-compromised and/or critically ill[3]. The emergence of multiple new human fungal pathogens each year[2] and increasing resistance to antifungal drugs has escalated the demands on healthcare facilities, leading to higher economic burden and mortality[4]. *Candida albicans* is the leading cause of candidemia in many geographical locations around the world. Despite an increase in species diversity and the emergence of *Candidozyma auris* (previously *Candida auris*), approximately 90-95% of candidemia cases worldwide continue to be caused by five major species: *C. albicans*, *Nakaseomyces glabratus* (previously *Candida glabrata*), *Candida parapsilosis*, *Candida tropicalis* and *Pichia kudriavzevii* (previously *Candida krusei*)[3,5]. The multidrug-resistant pathogenic yeast *C. auris* has appeared in over 60 countries during the last decade, with hospital-acquired infections and outbreaks being a major source of concern[6].

There are multiple challenges in identifying pathogenic yeast species. The blood culture method is the gold standard for identifying yeasts in candidemia cases[7]. Alternatively, non-culture-based methods, such as identifying serum biomarkers (β-(1,3)-D-glucan) and nucleic acids (nucleic acid amplification tests)[8], are also employed. However, these methods are limited due to low sensitivity, long turnaround times, and sophisticated instrumentation, specialized laboratory space, and highly-trained personnel requirements[7]. To overcome these challenges, several new detection methods are being developed for faster detection of yeast species directly from blood samples, as well as yeasts grown on agar plates[7]. These detection methods include the use of chrome agar[9], matrix-assisted laser desorption/ionization-time of flight mass spectroscopy (MALDI TOF MS)[10], DNA sequencing[11], T2 magnetic resonance[12], volatile organic compounds[13], Fourier transform infrared with attenuated total reflectance[14], imperfect match probes[15], recombinase polymerase amplification[16], high resolution melting curves[17], Raman spectroscopy[18], and laser-induced breakdown spectroscopy[19]. However, there are many obstacles, such as the integration of these methods into routine diagnostic workflows, species-level misidentification, and limited access to advanced equipment and highly trained personnel in low- and middle-income countries[1].

The use of artificial intelligence (AI) and machine learning (ML) has seen remarkable growth in recent years, driven by advances in model architectures, training algorithms, computational hardware, open-source AI frameworks (e.g., PyTorch[20]),

ARTICLE IN PRESS

and the availability of vast amounts of training data[21]. Artificial intelligence and ML approaches are increasingly being applied in scientific research to model complex data[21]. For example, AlphaFold can determine the 3D structures of proteins and their bimolecular interactions with accuracies comparable to experiments and is accelerating drug discovery[22]. Recent studies have investigated ML methods in the diagnosis and treatment of infections[23], and to discover antibiotics[24]. The field of computer vision, which applies AI to analyze image data, has made substantial progress from advances in artificial neural networks and attention-based transformer models[25–27]. Computer vision models based on convolutional neural network (CNN) and vision transformer architectures have achieved state-of-the-art performance on gold-standard image classification tasks[25–28], such as the ImageNet competition[26]. These models have been applied in biological and medical image analysis[29], achieving performance comparable to current diagnostic tools and human professionals[30]. They have also been applied in recent studies for the classification of yeast species in microscopy images[31–33]. Furthermore, advances in computer vision segmentation models have enabled advanced automated image processing[34], finding various applications in cell microscopy image analysis[35–38].

Deep learning methods have been applied in the context of yeast pathogens, including segmenting microscopy images[37,39] and inner organelles[40], tracking cellular growth[41], and measuring cell dimensions[42]. InceptionV3[43], a deep CNN architecture that applies convolutions to capture features at multiple scales, was previously combined with transfer learning to classify *Candida* species[33]. Specifically, the InceptionV3-based model was able to classify four pathogenic yeast species (*C. albicans*, *C. auris*, *N. glabratus*, and *C. haemulonii*) from brightfield microscopy images with an overall accuracy of 78%; a subsequent study reproduced these results using similar models and *Candida* species[44]. Most recently, pathogenic yeast species were successfully classified by CNNs using light scattering data with comparable accuracy to microscopy images[45]. DenseNet121, ViT, and Swin (Shifted Window) Transformer models have not been used for yeast species classification and provide improvements over InceptionV3 and other computer vision architectures[25,46–48]. Vision transformers produce comparable or better results than CNNs in image classification tasks, including the identification of mold species[49]. Despite this, only a limited number of studies have applied vision transformers to identify pathogens from microscopy images[50,51]. This can be attributed to their novelty and the fact that vision transformers require more data than CNNs to achieve comparable performance[25]. Swin transformers are a modification of the vision transformer architecture that address this limitation through a hierarchical representation of feature maps and the use of shifting window attention, improving local feature extraction capabilities while reducing computational complexity[48].

The “black box" nature of AI models often renders their predictions difficult, or even impossible, to understand. This limits the application of AI in fields where model interpretation and human trust are of particular importance, such as scientific research and medical diagnostics[52]. The emerging field of explainable AI (XAI) aims to elucidate how AI/ML models make predictions[53], with an emphasis on computer vision[54]. Although complete explainability of computer vision models has yet to be achieved, XAI methods can indicate image features that are most important for model predictions[54]. Understanding the decision-making process underlying AI models will enable scientific and medical fields to adopt AI-driven tools with greater confidence. Explainable AI methods, such as Grad-CAM (Gradient Class Activation Mapping)[55] and Occlusion Sensitivity[56], can be used to investigate predictions made by computer vision models. Grad-CAM uses information flowing into the final convolutional layer of a CNN to produce a coarse-grained visualization of areas in an image that were most important for a given prediction[55]. This makes Grad-CAM suitable for indicating if the model places high prediction importance on a high-level object within an image, but not for determining specific predictive features. Occlusion Sensitivity is a model-agnostic perturbation method that measures the change in prediction output of a model when occluding specific regions of the input image, with large changes in model output indicating prediction importance of the occluded region[56]. Occlusion Sensitivity is capable of producing fine-grained importance visualizations to suggest specific features that may lead to model predictions[56].

In this study, we develop high-performance computer vision models (InceptionV3[43], DenseNet121[46], Vision Transformer (ViT)-Base 16[47], and Swin Transformer-Tiny[48]) to quickly and accurately identify yeast pathogens from microscopy images. We address some current limitations in computer vision fungal pathogen identification by applying XAI to help interpret the predictions made by these advanced computer vision classification models. We use segmentation models to automate data processing, enabling the generation of large datasets for model training and evaluation, while increasing data processing efficiency, eliminating a major bottleneck of previous work[31–33]. This increase in training data and improvements in model development, particularly the use of state-of-the-art model architectures and full model fine-tuning, leads to substantial improvements in model performance that scales with classification complexity through the inclusion of additional pathogenic yeast species (e.g., *P. kudriavzevii*, *C. parapsilosis*, and *C. tropicalis*). The use of XAI methods to interpret computer vision predictions supports model verification, indicates biologically relevant and irrelevant features used for species predictions, identifies areas for model improvement, and enhances trust in the application of AI models for pathogen identification.

ARTICLE IN PRESS

ARTICLE IN PRESS

2 Materials and Methods

2.1 Sample Preparation, Imaging, and Data Processing

Clinical isolates of *C. albicans*, *C. auris*, and *N. glabratus* were obtained from the Alberta Precision Laboratory - Public Health Laboratory. the The American Type Culture Collection (ATCC) provided the strains for *C. parapsilosis* ATCC-22019, *P. kudriavzevii* ATCC-6258, *C. haemulonii* ATCC-22991, and *C. tropicalis* ATCC-750. All strains were preserved in 25% glycerol at -80°C until further use. The clinical and standard strains were revived by culturing from frozen stock in Sabouraud glucose agar (SDA) plates (Millipore, Darmstadt, Germany, 1.05438.0500) and incubated at 35°C for 48 h. Fresh subcultures were grown on SDA agar plates and incubated at 35°C for 24 h. Isolated colonies of different pathogenic yeast species grown in SDA medium for 24 h were used to prepare wet mount slides (Fisherbrand, Pittsburg, USA, 22-034486) using sterile distilled water, overlayed with 22 mm coverslips (Fisherbrand, Pittsburg, USA). The raw images for each species were captured using the EVOS M7000 imaging system (Invitrogen, Thermo Fisher Scientific, Massachusetts, USA, AMF7000) with an 100x oil immersion objective l. Coarse and fine adjustments were made to focus the cell on a single plane prior to capturing images. The brightness of the trans-illumination was adjusted to 0.033 before image capture.

Images of individual cells were cropped from the raw microscope images to build training, validation, and test datasets[33]. This process was automated using a CNN-based yeast cell segmentation model (yeaZ) for brightfield images[37]. Applying yeaZ produced segmentation masks for each raw microscope image, which contained the boundary coordinates of the identified yeast cells. To ensure each cell was fully contained within its segmentation border, the boundary coordinates were expanded by a factor of 0.15 using morphological dilation (implemented with scikit-image[57]). The choice of a morphological dilation factor was determined empirically to ensure that the segmented cell masks fully encompass the cell body and immediate peri-cellular context, optimizing for robust feature extraction without introducing excessive background noise. The expanded boundary coordinates were used to calculate square (1:1 aspect ratio) bounding boxes around each identified cell, which were then used to crop individual cell images with a padding of 10 pixels. For each crop, pixels that were outside of the expanded cell segmentation boundary coordinates were set to 0 (black) to remove the background around each cell, reducing irrelevant image features. Due to intrinsic errors of the segmentation model and image data, a large number of low-quality images (e.g., slide artifacts and poorly segmented cells) were cropped using this process. To automate the removal of these images, crops that were smaller than 50x50 pixels were excluded (as these were generally found to be slide artifacts), and feature extraction with ResNet18[58] and k-means clustering[59] with k=2 were used to separate the remaining high-quality crops from low-quality crops[60], which we found to produce consistent results for our datasets. This process was applied to the raw microscope images collected for training, validation, and testing. To ensure dataset balance and avoid biases during the training and validation process[61], the training and validation sets were randomly downsampled to match the species with the lowest number of crops in each set (each species had a different number of individual cell images due to the random distribution of cells across the raw microscopy images). The test set was randomly downsampled to match the size of the validation set to maintain consistency when comparing model validation and test results.

ARTICLE IN PRESS

2.2 Image Dataset Collection

A total of 900-1,200 raw microscope images for model training, 100-300 for validation, and 100-250 for testing were collected for each species from different subcultures. The data processing system described in Section 2.1 was applied to the raw images, resulting in 32,809 training, 4,214 validation, and 4,214 test images after k-means to remove low-quality crops and randomly downsampling to maintain balanced datasets. Examples of discarded k-means cluster images from the *C. albicans* test set are shown in Supplementary Figure 1. Although k-means removed most low-quality crops, some still appeared in the final datasets, particularly for *C. auris* and *P. kudriavzeviii*. This was likely due to the smaller cell size and other morphological differences compared to the yeast species that the segmentation model was trained on (*S. cerevisiae*)[37], which may have caused challenges during segmentation and clustering. The outlier rates in the test set were quantified by visually inspecting the images and counting the number of outliers for each species, showing that 2.4% of the overall test set consisted of low-quality crops that slipped through detection (705 out of $7 \times 4,214 = 29,498$ images). Most species had an outlier rate of 1% or less, while *C. auris* and *P. kudriavzevii* had relatively high rates of 9.1% (387 outliers out of 4,214 images) and 4.8% (204 outliers out of 4,214 images), respectively. To ensure that these outliers (artifacts and poor crops) did not bias our test results, we manually removed them from the test datasets. Model test performance without manual outlier removal was comparable to test performance with manual outlier removal (Supplementary Section 3, Supplementary Figure 6, and Supplementary Table 4). The number of images for each species was then randomly down-sampled to maintain a balanced test set, resulting in 3,827 individual cell images per species. The training, validation, and test raw microscopy images and individual cell crops after each data processing step for each species are summarized in Supplementary Table 1. Pixel and image size distributions for the final training, validation, and test datasets are shown in Supplementary Figures 2, 3, and 4, respectively.

2.3 Computer Vision Model Training and Evaluation

Transfer learning and fine-tuning were used to train the computer vision models[62] to classify pathogenic yeasts. The pre-trained base network acts as a generalized model of image classification tasks, while the added network learns features specific to the new training data to enable pathogen classification. To adapt the models to microscopy images of pathogenic yeast species, fine-tuning was applied to the full networks (the base model architecture with a randomly initialized 7-class classification head) by initializing the base model with ImageNetV1 weights and then training the entire network using the microscopy image training dataset. This process adjusted the pre-trained weights to enable the network to learn the visual characteristics specific to the pathogenic yeast species, leveraging the advanced network architectures of the base models and their pre-trained weights to improve the classification performance of the models.

Four computer vision architectures were considered for the base model, implemented using PyTorch[20]: InceptionV3[43], DenseNet121[46], a base 16 Vision Transformer (ViT)[47], and a Swin Transformer-Tiny[48]. Image transformations specific to each pre-trained model (using PyTorch's default transformations[20] for the ImageNetV1 dataset[63]) were applied to all images before they were used for training, validation, and testing. These transformations consisted of resizing (bilinear interpolation to [342], [256], and [256] for InceptionV3, DenseNet121, and ViT, respectively, and bicubic interpolation to [232] for Swin Transformer-Tiny), followed by a central crop (of [299], [224], [224], [224] for InceptionV3, DenseNet121, ViT, and Swin Transformer-Tiny, respectively), value rescaling to [0.0, 1.0], and then normalization using mean=[0.485, 0.456, 0.406] and std=[0.229, 0.224, 0.225][20]. Each model was trained on the training set using an AdamW optimizer[64] to minimize a cross-entropy loss function[65]. To ensure consistent model comparisons, the same hyperparameters were used to train each model, with a learning rate of 3e-5, an AdamW weight decay of 0.1, and a batch size of 32. As the training, validation, and test sets were constructed to be balanced across all seven *Candida* species, we trained the models without the use of class weighting. Given the large size and diversity of the dataset, along with initial testing showing minimal impact on validation performance while slowing the training process (data not shown), artificial data augmentation was not used to train the models. The validation set was used to evaluate model performance (overall accuracy and loss) after each iteration (epoch) of the training process. Early stopping[66] (with a patience of 5) was used to stop the training process once the validation loss no longer decreased, with the model weights from the epoch with the best validation performance (lowest validation loss) saved. After the training and validation process was completed, the models were applied to the test set to evaluate their unbiased generalization performance. Standard classification metrics were used to quantify model performance, including the overall model accuracy and class-specific precision, recall, and F_1-score, and their averages over all classes. Precision is defined as the ratio between the true positive (TP) and all positive predictions (TP plus false positives) for a given class. Recall and class accuracy are used interchangeably in this study and defined as the correct predictions (TP) divided by the total number of samples (TP plus false negatives) for a given class. The F_1-score, defined as the harmonic mean between precision and recall, was used to evaluate the balance between these metrics. Confusion matrices were used to visualize model performance and identify classification trends. The classification metrics, confusion matrices, and principal component analysis (PCA) were determined using scikit-learn[67].

To investigate the feature space of the images for each species, we performed a PCA[68] on the test set images after using our trained model as a feature extractor. PCA reduced the dimensionality of the feature representations learned by our trained model to visualize their principal 2D and 3D components. This enabled us to identify differences and similarities between the representations of each species and to better understand common misidentification patterns that occurred during model evaluation by showing species that overlap in the feature space.

2.4 Explainable AI for Model Verification and Prediction Interpretation

We applied XAI methods to investigate if the high-performance computer vision models were learning relevant features (e.g., cellular characteristics such as cell wall and organelles) or irrelevant features (e.g., background artifacts). Additionally, we used XAI to provide additional verification to model predictions and identify features that were most important for species predictions. We chose to use Grad-CAM[55] and Occlusion Sensitivity[56] for our image classification models. Grad-CAM was chosen because it has been extensively tested[69] and used for computer vision model interpretation[70]. Occlusion Sensitivity was selected because it does not rely on specific model architectures[56] and is capable of producing more fine-grained feature importance visualizations than Grad-CAM. Grad-CAM and Occlusion Sensitivity, implemented using the PyTorch-based library Captum[71], produce normalized attribution maps (values ranging from 0 to 1) that highlight regions of the input image with high importance for a given model prediction. Different sets of test predictions (correct and incorrect) made by the model were investigated by generating Grad-CAM and Occlusion Sensitivity attribution maps for a large number of input images and visually inspecting them to identify image features with the highest attribution (equal to 1). The results of the visual inspection were then consolidated by calculating the proportion of images in the set that contained high attribution on each of the features.

First, we used Grad-CAM to distinguish whether there was high attribution inside the cell, outside the cell, or partly inside and outside the cell in each image. This resulted in three measurable binary variables (high attribution yes/1 or no/0 for each case) for each attribution map. Next, Occlusion Sensitivity (with a small non-overlapping 2x2 pixel sliding window) was used to produce fine-grained attribution maps to identify specific image features contributing to model predictions. We identified nine visually distinguishable high-level features that could be present in the input images: internal organelles (nucleus and/or vacuoles), the cell interior (not on a specific organelle), the inner optical pattern, the cell wall, the outer optical pattern, budding patterns/scars, partially cropped cells outside of the main cell, the slide background (between the cell and the background removal border), and the black background. This resulted in nine measurable binary variables for each Occlusion Sensitivity attribution map (high attribution yes/1 or no/0 for each feature), with six of the features being biologically relevant (organelle, cell interior, cell wall, budding pattern/scar, inner and outer optical patterns) and three of these features being irrelevant (other cells in the image, slide background, and black background). As more than one feature had high attribution in the majority of cases, each attribution map generally had more than one high-importance feature.

The data collection/processing, model training and evaluation, and XAI workflow are shown in Figure 1.

3 Results and Discussion

3.1 Advanced Computer Vision to Classify Pathogenic Yeasts

We developed high-performance computer vision models to classify seven pathogenic yeast species (Section 2.3). All our computer vision models produced comparable overall test results (Figure 2). DenseNet121 was the best-performing CNN model, with an optimal performance (lowest validation loss) obtained after epoch 5 with a training loss, training accuracy, validation loss, and validation accuracy of 0.148, 95%, 0.367, and 88%, respectively (Figure 3a). Vision Transformer-Base 16 was the best-performing vision transformer model, with an optimal performance obtained after epoch 7 with a training loss, training accuracy, validation loss, and validation accuracy of 0.107, 96%, 0.392, 88%, respectively (Figure 3c). The corresponding validation confusion matrices (Figure 3b and Figure 3d) and the validation performance metrics (Supplementary Table 2a and Supplementary Table 2c) were also determined.

DenseNet121 had an overall test accuracy of 87% with strong classification performance across all species, correctly classifying *C. albicans*, *C. auris*, *N. glabratus*, *C. haemulonii*, *P. kudriavzevii*, *C. parapsilosis*, and *C. tropicalis* in 89%, 90%, 85%, 76%, 90%, 97%, and 82% of cases, respectively (Figure 4a and Supplementary Table 2b). The test performance was generally balanced between precision and recall across the species, having F_1 scores ranging from 0.83 to 0.92, with an average value of 0.87 (Supplementary Table 2b). The most substantial differences between precision and recall were for *C. parapsilosis* (0.73 precision and 0.97 recall) and *C. haemulonii* (0.90 precision and 0.76 recall), with specific predictions contributing to these imbalances (Figure 4a). The relatively low precision of *C. parapsilosis* arose due to other species, primarily *N. glabratus*, *C. haemulonii*, and *C. auris*, being misidentified as *C. parapsilosis*, while the low recall of *C. haemulonii* was due to it being misidentified as other species, primarily *C. parapsilosis*, *C. albicans*, and *N. glabratus*.

Vision Transformer-Base 16 had an overall test accuracy of 86% with performance across species similar to DenseNet121, correctly classifying *C. albicans*, *C. auris*, *N. glabratus*, *C. haemulonii*, *P. kudriavzevii*, *C. parapsilosis*, and *C. tropicalis* in 90%, 93%, 78%, 75%, 90%, 96%, and 77% of cases, respectively (Figure 4b and Supplementary Table 2d). Vision Transformer-Base 16 performed slightly better than DenseNet121 on *C. albicans* and *C. auris* (1% and 3%, respectively), but worse on *N. glabratus*, *C. haemulonii*, *C. parapsilosis*, and *C. tropicalis* (7%, 1%, 1%, and 5%, respectively). The species misidentification patterns of Vision Transformer-Base 16 (Figure 4b) and balance between precision and recall were similar to DenseNet121, having F_1 scores ranging from 0.82 to 0.93 with an average value of 0.86 (Supplementary Table 2d). The most substantial differences between precision and recall were for *C. parapsilosis* (0.73 precision and 0.96 recall), *C. tropicalis* (0.87 precision and 0.77 recall), and *C. haemulonii* (0.90 precision and 0.75 recall). The specific predictions contributing to the imbalances for *C. parapsilosis* and *C. haemulonii* were similar to DenseNet121 as described above, while the larger imbalance for *C. tropicalis* was mainly caused by additional misidentifications of *C. tropicalis* as *P. krudiavzevii* (Figure 4b).

The test set confusion matrices and performance metrics for InceptionV3 and Swin Transformer-Tiny were also obtained (Supplementary Section 2, Supplementary Figure 5, and Supplementary Table 3). All models produced similar test results and classification patterns (Figure 4 and Supplementary Figure 5), with DenseNet121 producing the best results for *N. glabratus* and *C. haemulonii* (Figure 2). To investigate how outliers affect these results when using a fully automated test set without manual outlier removal, we applied the DenseNet121 model to a test set containing outliers (Supplementary Section 3, Supplementary Figure 6, and Supplementary Table 4). The inclusion of outliers resulted in a 1% decrease in overall test accuracy, with a minor effect on performance for all species except *C. auris* and *P. krudiavzevii*, which respectively resulted in a 4% decrease in recall and a 4% decrease in precision.

We trained, validated, and tested the high-performance computer vision models using our new approach on four pathogenic yeast species previously considered to compare with an existing machine learning-based pathogenic yeast classifier[33]. Performance was similar across each of our model architectures, with the Vision Transformer producing the best test results, with an overall accuracy of 92% and correctly classifying *C. albicans*, *C. auris*, *N. glabratus*, and *C. haemulonii* in 95%, 97%, 96%, and 82% of cases, respectively (Supplementary Section 4 and Supplementary Figure 7). This is a substantial improvement over the previously published best-performing CNN (InceptionV3) test results[33], which had an overall accuracy of 78% and

ARTICLE IN PRESS

correctly classified *C. albicans*, *C. auris*, *N. glabratus*, and *C. haemulonii* in 97%, 73%, 69%, and 73% of cases, respectively. These results highlight the improvements we have made in data collection, processing, and model development. The model architecture test set comparisons, the Vision Transformer confusion matrix, and a species accuracy comparison to previous work[33] are shown in Supplementary Figure 7.

3.2 Principal Component Analysis to Investigate Predictive Feature Characteristics and Model Confusion

As our DenseNet121 model performed marginally better than our other computer vision models (Figure 2), we used DenseNet121 to investigate the feature characteristics of the test set for each species used to make predictions and to identify reasons for model confusion. Specifically, we performed a 2D and 3D principal component analysis (PCA) on the features extracted from the penultimate layer (before the classification head) of the trained DenseNet121 model. For the 2D PCA analysis, the variance explained by PC1 and PC2 were 20.3% and 13.1%, respectively, for a total explained variance of 33.4%. For the 3D PCA analysis, the variance explained by PC1, PC2, and PC3 were 20.3%, 13.1%, and 10.4%, respectively, for a total explained variance of 43.8%. This analysis implied that the pathogenic yeast species that were confused by the DenseNet121 model had feature spaces with similar shapes and extensive overlap (e.g., *C. albicans* and *C. tropicalis*; Figure 5 and Supplementary Video 1). Furthermore, species that the model could easily distinguish between had minimal overlap in the PCA (including *C. albicans* and *C. auris*; Figure 5). In some cases, DenseNet121 confused one species for another without the reverse being true. For instance, *N. glabratus* was often misidentified as *C. parapsilosis*, but not vice versa (Figure 4). The PCA results indicate that these scenarios occur when a species with a spread-out feature space such as *N. glabratus*, overlaps with a species that has a more concentrated feature space, such as *C. parapsilosis* (Figure 5). This suggests that our DenseNet121 model is biased toward species with more concentrated feature spaces because a lower variance in visible features is easier for the model to learn. This likely occurs due to model confusion when an image of a species with more variability in visible features contains features similar to another species, resulting in inaccurate species predictions. We note that these PCA visualizations of the learned feature space are suggestive, rather than definitive, in terms of species separability.

3.3 Explainable AI to Interpret Computer Vision Model Predictions

The size and budding patterns of pathogenic yeast cells were previously hypothesized to contribute correct species classification by computer vision models[33]. However, this hypothesis could not be verified due to the black-box nature of AI models[52]. Here, we used XAI, namely Grad-CAM[55] and Occlusion Sensitivity[56], to interpret and provide supporting evidence for the predictions of our computer vision models. Specifically, we used these XAI methods to identify image features that our models may rely on to accurately identify the species of pathogenic yeast. Grad-CAM[55] and Occlusion Sensitivity[56] were used to generate attribution maps for three sets of one hundred randomly selected TP test set predictions for each species made by our best-performing DenseNet121 model. Additionally, we generated attribution maps for the eight most frequent DenseNet121 misidentification cases (Figure 4), specifically *N. glabratus* misidentified as *C. parapsilosis*, *C. haemulonii* as *C. parapsilosis*, *C. tropicalis* as *P. kudriavzevii*, *C. albicans* as *C. tropicalis*, *C. auris* as *C. parapsilosis*, *C. haemulonii* as *C. albicans*, *C. haemulonii* as *N. glabratus*, and *C. tropicalis* as *C. albicans*, which respectively contributed to 12%, 11%, 8%, 6%, 6%, 6%, 4%, and 4% of the test predictions for the corresponding true species. Representative Grad-CAM and Occlusion Sensitivity attribution maps are shown in Figure 6, which demonstrate the level of detail and features that each of these methods identify.

We calculated the proportion of TP predictions made by DenseNet121 with high Grad-CAM attribution inside the cell, outside of the cell, or both for each pathogenic yeast species (Figure 7a). These results indicate that there was high prediction importance fully inside the cell in nearly all cases for each species, with no attribution maps indicating high importance fully outside of the cell. For the top eight misidentification cases, the proportion of images with high Grad-CAM attribution was 100% for all species (Figure 7b). These results indicate that the model places high prediction importance on the cell in each image in all correct and incorrect predictions investigated. However, the low resolution of the Grad-CAM attribution maps did not allow specific cell characteristics that contribute to model predictions to be identified. Therefore, we applied Occlusion Sensitivity to produce fine-grained attribution maps to identify specific high-importance image features (Section 2.4).

Next, we calculated the proportion of TP predictions with high Occlusion Sensitivity attribution on each high-level feature for all species (Figure 7c). These results suggest that the DenseNet121 model places high importance on biologically relevant features in the majority of TP predictions, while placing little importance on non-biologically relevant features. The most frequent high-importance feature was found to be the cell wall for TP predictions of *C. albicans* and *C. tropicalis*, the cell interior (not specific to an organelle) for *C. auris*, *C. haemulonii*, *P. kudriavzevii*, and *C. parapsilosis*, and internal organelles (nuclei and vacuoles) for *N. glabratus*. Although these were the features with the most frequent high attribution, the predictions generally had more than one feature with high attribution, and a combination of feature characteristics likely contributed to each model prediction. As for model misidentifications, we again found that the features with high Occlusion Sensitivity attribution were biologically relevant in the vast majority of cases (Figure 7d). The cell wall was the most frequent high-importance feature for *P. kudriavezevii* misidentified as *C. tropicalis*, *C. albicans* as *C. tropicalis*, and *C. haemulonii* as *C. albicans*, while the cell interior was the most frequent for the other five misidentification cases.

ARTICLE IN PRESS

The Grad-CAM and Occlusion Sensitivity XAI methods indicated that the cell in each image had the highest importance for species classification predictions made by the DenseNet121 model, suggesting that the model relied primarily on biologically relevant cell features to make predictions (Figures 7). The detailed Occlusion Sensitivity attribution maps identified specific biologically relevant cell characteristics and irrelevant image features that had high importance for given DenseNet121 predictions (Figure 7c and 7d), indicating that the interior cell structure, cell wall, and organelles were of the most important for distinguishing between the species in the dataset, followed by the optical patterns around the cells produced by diffraction of the microscope light source and the optical properties of the cells[72].

4 Conclusion

Driven by advances in data collection and processing, model architectures, and training methods, we developed high-performance computer vision models with strong generalization performance on a large dataset consisting of seven pathogenic yeast species. Based on bright-field microscopy images, our approach has the potential to be integrated into current diagnostic processes[7]. The increase in performance of our advanced computer vision models to more accurately classify a wider range of pathogenic yeasts compared to previous work[31–33] was achieved through technical improvements in microscopy image collection, data processing, and model development. Specifically, we obtained a substantially larger training dataset of individual cell images (32,809 per species) with lower time and resource costs compared to previous studies by implementing the yeaZ yeast segmentation model[37] and the k-means algorithm[59,60] to automate raw microscope image processing. Along with full model fine-tuning, this increase in training data enabled our advanced vision models to learn nuanced morphological differences between pathogenic yeast species.

The application of XAI methods indicated that biologically relevant features are of high importance in correct and incorrect species predictions made by our advanced computer vision models. Grad-CAM[55] was used for high-level attribution maps, whereas Occlusion Sensitivity[56] was used to elucidate finer details. Grad-CAM indicated that the computer vision models place high importance on the cell in each image to make predictions, while Occlusion Sensitivity identified important features unique to each pathogenic yeast species, including the cell wall, internal organelles, cell interior structure, and optical properties. The predictions made by our best-performing model, DenseNet121, were suggested by the XAI methods to be primarily based on cell morphology, enhancing transparency and trust in model predictions[52]. The XAI methods used in this study provided supporting evidence for species-specific features (e.g., optical patterns generated by cellular organelles) that are most important to distinguish yeast species and identified potential reasons (e.g., morphological similarities between species) for incorrect model predictions.

The similarity in performance between the computer vision models developed in our study, together with the XAI model interpretation, suggests that the current approach may be reaching a species classification performance plateau on the current microscopy image dataset. Further improvements in classification performance may be achieved through advances in experimental data collection, data processing (e.g., using the trained models together with ResNet18 clustering to detect and remove additional low-quality images), species-distinguishing feature enhancement, model architectures, training methods, and evaluation frameworks. These improvements, especially those using robust data augmentation and an ensemble prediction approach using multiple models[73], will be important when expanding these models to multi-species classification tasks, as observed in other image-based deep learning studies[74]. This may be achieved using high-resolution single-cell imaging, staining methods, or computational feature enhancements. Additional XAI methods could be used to interpret predictions on mixed-species cultures[75], such as Grad-CAM++[76], which has better object localization and explanation of occurrences of multiple objects in a class compared to Grad-CAM[55].

Each pathogenic yeast species considered in this study has a distinct size, shape, cell wall thickness and curvature, internal structure/organelle distribution, budding pattern, and optical properties[77]. This unique distribution of visual features is captured through bright-field microscopy, which our advanced computer vision models learned to make accurate species classification predictions. However, there are overlaps in the visual feature distributions (Figure 5, Supplementary Video 1, and Figure 7), which led to incorrect model predictions when the cell images do not contain sufficient distinguishing information. Other methods are capable of generating unique fingerprints of each species for accurate classification, including MALDI TOF MS by generating mass/charge ratio spectra[10], Raman spectroscopic methods that detect vibrational signatures of molecules within the yeast cells[18], using unique volatile organic compounds that are produced by each species[13], and DNA sequencing[8]. However, we demonstrated that our advanced vision models can successfully classify pathogenic yeast species based on their unique visual fingerprints captured by bright-field microscopy, while having a substantially lower resource cost compared to other methods.

Although our advanced computer vision models demonstrated robust classification performance on a large dataset, there are limitations that must be overcome to successfully translate this research into clinical applications. All training and testing data were acquired using a single protocol and microscopy imaging system for seven clinically isolated pathogenic yeast species. Consequently, the performance of our advanced vision models may vary when applied to microscope images acquired from

ARTICLE IN PRESS

ARTICLE IN PRESS

different laboratories using different equipment (e.g., objective magnification, camera sensor, and lighting calibration), different sample preparation techniques, or clinical samples. Furthermore, exploring pathways to integrate our computer vision models into clinical settings will require assessing their compatibility with existing diagnostic workflows, performing a cost-benefit analysis, and obtaining regulatory approval.

Our study highlights the power of advanced computer vision models to identify pathogenic yeast species and XAI to support the interpretability of these models, indicating the biological and non-biological features responsible for model predictions. We anticipate that this approach of combining computer vision models with XAI will advance antimicrobial resistance research and lead to transparent AI-driven diagnostic methods and treatments for infectious disease patients.

References

1. Denning, D. W. Global incidence and mortality of severe fungal disease. *The Lancet Infect. Dis.* **24**, e428–e438 (2024).
2. Corrêa-Junior, D., Andrade, I. B. d., Alves, V., Araújo, G. R. d. S. & Frases, S. Clinical challenges of emerging and re-emerging yeast infections in the context of the covid-19 pandemic. *Microorganisms* **10**, 2223 (2022).
3. Pappas, P. G., Lionakis, M. S., Arendrup, M. C., Ostrosky-Zeichner, L. & Kullberg, B. J. Invasive candidiasis. *Nat. Rev. Dis. Primers* **4**, 1–20 (2018).
4. Cornely, O. A. *et al.* Global guideline for the diagnosis and management of candidiasis: an initiative of the ecmm in cooperation with isham and asm. *The Lancet Infect. Dis.* (2025).
5. CDC. Data and statistics on candidemia. https://www.cdc.gov/candidiasis/data-research/facts-stats/index.html (2024). Accessed: (6 April 2025).
6. CDC. Increasing threat of spread of antimicrobial-resistant fungus in healthcare facilities. https://www.cdc.gov/media/releases/2023/p0320-cauris.html (2023). Accessed: (6 April 2025).
7. Mendonca, A., Santos, H., Franco-Duarte, R. & Sampaio, P. Fungal infections diagnosis–past, present and future. *Res. microbiology* **173**, 103915 (2022).
8. Iyer, V. *et al.* Culture-independent identification of bloodstream infections from whole blood: Prospective evaluation in specimens of known infection status. *J. Clin. Microbiol.* **62**, e01498–23 (2024).
9. Birinci, A., Akkurt, L., Acuner, C., Unlu, M. & Durupinar, B. Rapid identification of the candida species from direct blood cultures by chromagar™ candida. *J. international medical research* **32**, 484–487 (2004).
10. Panda, A. *et al.* Maldi-tof mass spectrometry for rapid identification of clinical fungal isolates based on ribosomal protein biomarkers. *J. Microbiol. Methods* **109**, 93 (2015).
11. Gudisa, R. & Rudramurthy, S. M. Role of gene sequencing for the diagnosis, tracking and prevention of fungal infections. *J. The Acad. Clin. Microbiol.* **24**, S15–S24 (2022).
12. Wilson, N. M. *et al.* T2 magnetic resonance assay improves timely management of candidemia. *JAMS* **1**, 12–18 (2017).
13. Bastos, M. L. *et al.* Breaking barriers in candida spp. detection with electronic noses and artificial intelligence. *Sci. Reports* **14**, 956 (2024).
14. Franconi, I. *et al.* Rapid identification of clinically relevant candida spp. by i-done software using attenuated total reflectance fourier transform infrared (atr-ftir) spectroscopy. *J. Fungi* **11**, 40 (2025).
15. Higashi, Y., Niimi, H., Sakamaki, I., Yamamoto, Y. & Kitajima, I. Rapid identification of candida species in candidemia directly from blood samples using imperfect match probes. *Sci. Reports* **10**, 5828 (2020).
16. Wang, L. *et al.* Rapid detection of candida tropicalis in clinical samples from different sources using rpa-lfs. *Front. Cell. Infect. Microbiol.* **12**, 898186 (2022).
17. Nemcova, E. *et al.* Rapid identification of medically important candida isolates using high resolution melting analysis. *PLOS One* **10**, e0116940 (2015).
18. Koya, S. K. *et al.* Rapid identification of candida auris by raman spectroscopy combined with deep learning. *J. Raman Spectrosc.* **56**, 218–227 (2025).
19. Rehse, S. J. A review of the use of laser-induced breakdown spectroscopy for bacterial classification, quantification, and identification. *Spectrochimica Acta Part B: At. Spectrosc.* **154**, 50–69 (2019).
20. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Adv. neural information processing systems* **32** (2019).

ARTICLE IN PRESS

ARTICLE IN PRESS

21. Xu, Y. *et al.* Artificial intelligence: A powerful paradigm for scientific research. *The Innov.* **2** (2021).

22. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* **630**, 493–500 (2024).

23. Zhang, X., Zhang, D., Zhang, X. & Zhang, X. Artificial intelligence applications in the diagnosis and treatment of bacterial infections. *Front. Microbiol.* **15**, 1449844 (2024).

24. Zheng, E. J. *et al.* Discovery of antibiotics that selectively kill metabolically dormant bacteria. *Cell Chem. Biol.* **31**, 712–728 (2024).

25. Guo, M.-H. *et al.* Attention mechanisms in computer vision: A survey. *Comput. visual media* **8**, 331–368 (2022).

26. Byerly, A., Kalganova, T. & Ott, R. The current state of the art in deep learning for image classification: A review. In Arai, K. (ed.) *Intelligent Computing*, 88–105 (Springer International Publishing, Cham, 2022).

27. Zhao, X. *et al.* A review of convolutional neural networks in computer vision. *Artif. Intell. Rev.* **57**, 99 (2024).

28. Chai, J., Zeng, H., Li, A. & Ngai, E. W. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Mach. Learn. with Appl.* **6**, 100134 (2021).

29. Ali, M. *et al.* Applications of artificial intelligence, deep learning, and machine learning to support the analysis of microscopic images of cells and tissues. *J. Imaging* **11**, 59 (2025).

30. Esteva, A. *et al.* Deep learning-enabled medical computer vision. *NPJ digital medicine* **4**, 5 (2021).

31. Zieliński, B., Sroka-Oleksiak, A., Rymarczyk, D., Piekarczyk, A. & Brzychczy-Włoch, M. Deep learning approach to describe and classify fungi microscopic images. *PLOS One* **15**, e0234806 (2020).

32. Rahman, M. A. *et al.* Classification of fungal genera from microscopic images using artificial intelligence. *J. Pathol. Informatics* **14**, 100314 (2023).

33. Shankarnarayan, S. A. & Charlebois, D. A. Machine learning to identify clinically relevant candida yeast species. *Med. Mycol.* **62**, myad134 (2024).

34. Guo, Y., Liu, Y., Georgiou, T. & Lew, M. S. A review of semantic segmentation using deep neural networks. *Int. journal multimedia information retrieval* **7**, 87–93 (2018).

35. Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. *Nat. Methods* **22**, 592–599 (2025).

36. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. In *Medical image computing and computer assisted intervention–MICCAI 2018: 21st international conference, Granada, Spain, September 16-20, 2018, proceedings, part II 11*, 265–273 (Springer, 2018).

37. Dietler, N. *et al.* A convolutional neural network segments yeast microscopy images with high accuracy. *Nat. communications* **11**, 5723 (2020).

38. Bunk, D. *et al.* Yeastmate: neural network-assisted segmentation of mating and budding events in saccharomyces cerevisiae. *Bioinformatics* **38**, 2667–2669 (2022).

39. Salem, D. *et al.* Yeastnet: Deep-learning-enabled accurate segmentation of budding yeast cells in bright-field microscopy. *Appl. Sci.* **11**, 2692 (2021).

40. Lee, M., Kunzi, M., Neurohr, G., Lee, S. S. & Park, Y. Hybrid machine-learning framework for volumetric segmentation and quantification of vacuoles in individual yeast cells using holotomography. *Biomed. Opt. Express* **14**, 4567–4578 (2023).

41. Szabó, G., Bonaiuti, P., Ciliberto, A. & Horváth, A. Enhancing yeast cell tracking with a time-symmetric deep learning approach. *npj Syst. Biol. Appl.* **11**, 25 (2025).

42. Lawson, K., Skrtic, S., Vo, M. & Escorcia, W. *Measuring Cell Dimensions in Fission Yeast Using Machine Learning*, 33–46 (Springer US, New York, NY, 2025).

43. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).

44. Parker, R. A. *et al.* A complete transfer learning-based pipeline for discriminating between select pathogenic yeasts from microscopy photographs. *Pathogens* **14**, 504 (2025).

45. Liu, X. *et al.* Machine learning-assisted classification of pathogenic yeasts using laser light scattering and conventional microscopy. *J. Imaging* **12**, 136 (2026).

ARTICLE IN PRESS

ARTICLE IN PRESS

46. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708 (2017).

47. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

48. Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022 (2021).

49. Mansourvar, M., Charylo, K. R., Frandsen, R. J. N., Brewer, S. S. & Hoof, J. B. Automated fungal identification with deep learning on time-lapse images. *Information* **116**, 109 (2025).

50. Prasad, K., Mukhopadhyay, C., Banerjee, B. *et al.* Vision transformer based bacteria classification model for gram-stained direct smear images. *Multimed. Tools Appl.* 1–21 (2025).

51. Sengar, N., Burget, R. & Dutta, M. K. A vision transformer based approach for analysis of plasmodium vivax life cycle for malaria prediction using thin blood smear microscopic images. *Comput. Methods Programs Biomed.* **224**, 106996 (2022).

52. Durán, J. M. & Jongsma, K. R. Who is afraid of black box algorithms? on the epistemological and ethical basis of trust in medical ai. *J. medical ethics* **47**, 329–335 (2021).

53. Dwivedi, R. *et al.* Explainable ai (xai): Core ideas, techniques, and solutions. *ACM Comput. Surv.* **55**, 1–33 (2023).

54. Kamakshi, V. & Krishnan, N. C. Explainable image classification: The journey so far and the road ahead. *AI* **4**, 620–651 (2023).

55. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626 (2017).

56. Zeiler, M. Visualizing and understanding convolutional networks. In *European conference on computer vision/arXiv*, vol. 1311 (2014).

57. Van der Walt, S. *et al.* scikit-image: image processing in python. *PeerJ* **2**, e453 (2014).

58. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).

59. Steinley, D. K-means clustering: a half-century synthesis. *Br. J. Math. Stat. Psychol.* **59**, 1–34 (2006).

60. Guérin, J., Thiery, S., Nyiri, E., Gibaru, O. & Boots, B. Combining pretrained cnn feature extractors to enhance clustering of complex natural images. *Neurocomputing* **423**, 551–571 (2021).

61. Kumar, P., Bhatnagar, R., Gaur, K. & Bhatnagar, A. Classification of imbalanced data: review of methods and applications. In *IOP conference series: materials science and engineering*, vol. 1099, 012077 (IOP Publishing, 2021).

62. Shaha, M. & Pawar, M. Transfer learning for image classification. In *2018 second international conference on electronics, communication and aerospace technology (ICECA)*, 656–660 (IEEE, 2018).

63. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255 (Ieee, 2009).

64. Loshchilov, I. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).

65. Mao, A., Mohri, M. & Zhong, Y. Cross-entropy loss functions: Theoretical analysis and applications. In *International conference on Machine learning*, 23803–23828 (PMLR, 2023).

66. Prechelt, L. Early stopping-but when? In *Neural Networks: Tricks of the trade*, 55–69 (Springer, 2002).

67. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. machine Learn. research* **12**, 2825–2830 (2011).

68. Greenacre, M. *et al.* Principal component analysis. *Nat. Rev. Methods Primers* **2**, 100 (2022).

69. Adebayo, J. *et al.* Sanity checks for saliency maps. *Adv. neural information processing systems* **31** (2018).

70. Van der Velden, B. H., Kuijf, H. J., Gilhuijs, K. G. & Viergever, M. A. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Med. Image Analysis* **79**, 102470 (2022).

71. Kokhlikyan, N. *et al.* Captum: A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv:2009.07896* (2020).

72. Yu Wan, W. *et al.* Integration of light scattering with machine learning for label free cell detection. *Biomed. Opt. Express* **12**, 3512–3529 (2021).

ARTICLE IN PRESS

73. Mohammed, A. & Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ. - Comput. Inf. Sci.* **35**, 757–774 (2023).

74. Bai, B. *et al.* Deep learning-enabled virtual histological staining of biological samples. *Light. Sci. & Appl.* **12**, 57 (2023).

75. Charlebois, D. A. Quantitative systems-based prediction of antimicrobial resistance. *npj Syst. Biol. Appl.* **9**, 40 (2023).

76. Chattopadhay, A., Sarkar, A., Howlader, P. & Balasubramanian, V. N. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, 839–847 (IEEE, 2018).

77. Walsh, T. J., Hayden, R. T. & Larone, D. H. *Larone's medically important fungi: A guide to identification*, 121–146 (John Wiley & Sons, 2018).

5 Acknowledgments

This research was funded by grants to D.A.C. from the Human Frontier Science Program (RGEC30/2024) and the Alberta Innovates Accelerating Innovations into Care - Concepts program (AICE-Concepts 597389). We thank Dr. Tanis Dingle at the Alberta Precision Laboratories - Public Health Laboratory and Dr. Amrita Bharat at the Public Health Agency of Canada - National Microbiology Laboratory for providing the pathogenic yeast species isolates, Asst. Prof. Michael Manhart for suggesting the use of principal component analysis, and Prof. Randy Goebel for helpful discussions on explainable AI.

6 Author Contributions Statement

D.A.C., S.A.S., and J.D.G. conceptualized and designed the study. J.D.G. carried out the computational work, data visualization, and analytics. S.A.S. carried out the experimental work. S.A.S. and J.D.G. interpreted the results. J.D.G., S.A.S., and D.A.C. wrote the manuscript. D.A.C. acquired the funding and supervised the research. All authors reviewed the manuscript.

7 Data Availability

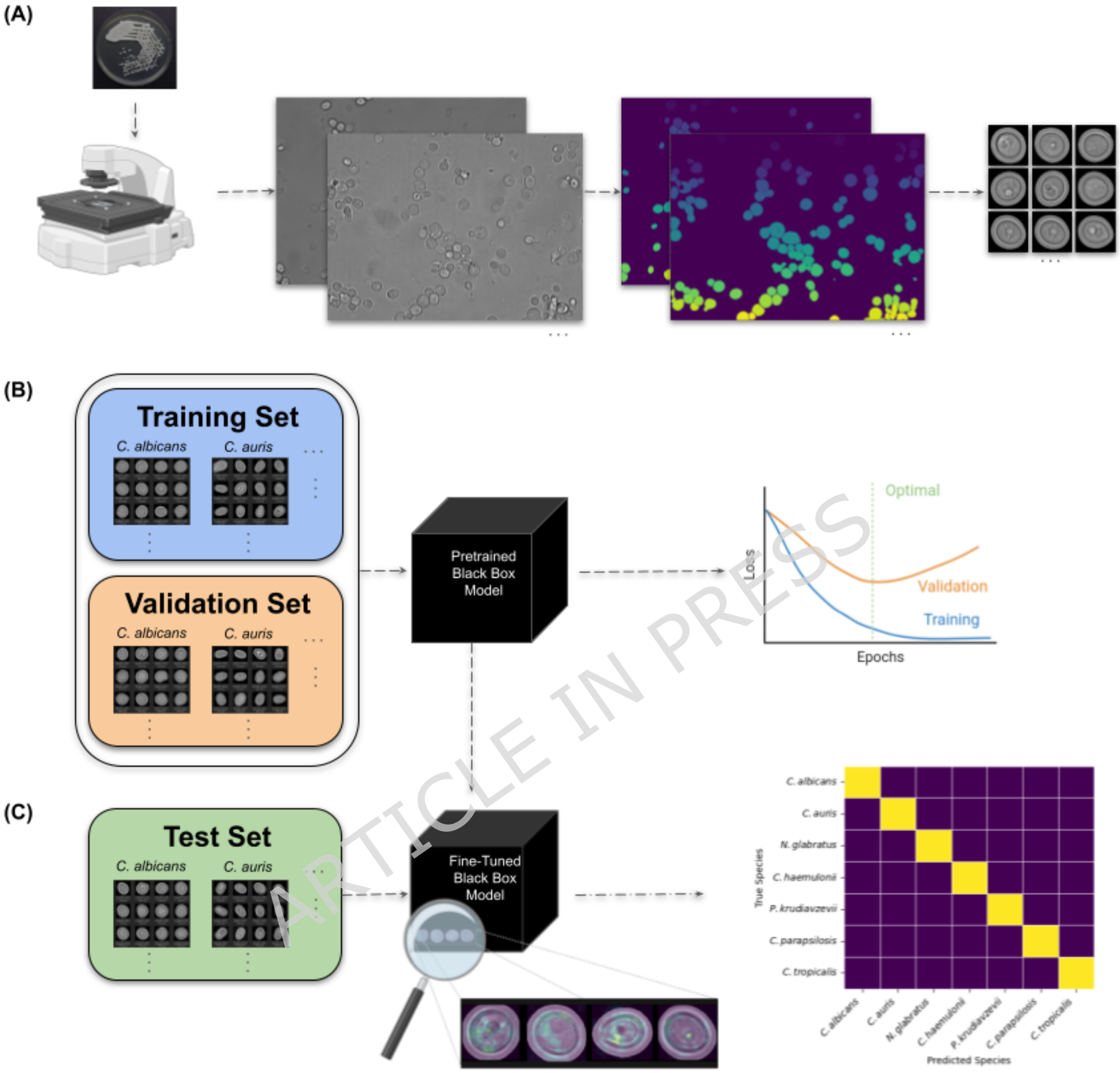
The details required to reproduce the results presented in this study are contained within the manuscript and its Supporting Information. Due to the proprietary nature of the trained machine learning models and microscopy datasets, these assets are not publicly available. However, they are available from the corresponding author, D.A.C., upon reasonable, non-commercial request and the completion of a Data Transfer Agreement. The code used for data processing, model training and evaluation, XAI, PCA, and analysis code is available for non-commercial use at: https://doi.org/10.5281/zenodo.19137095.

8 Competing Interests

D.A.C. is funded by an AICE-Concepts grant to develop a commercial point-of-care medical diagnostic device based on artificial intelligence.

ARTICLE IN PRESS

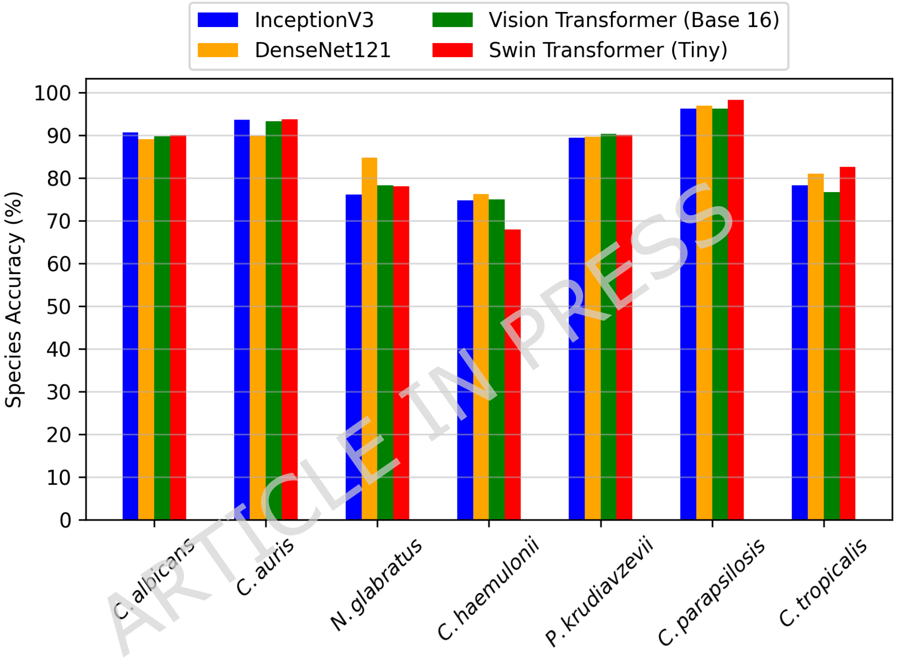
ARTICLE IN PRESS



ARTICLE IN PRESS

Figure 1. Workflow for training and evaluating the advanced computer vision pathogenic yeast species classification models. (a) Raw microscope images for each pathogenic species were collected in three independent groups to create training, validation, and test datasets. Automated data processing was used to crop images of individual cells from the raw microscopy images. (b) The training data set was used to train "black-box" computer vision models, with model performance being evaluated on the validation set after each epoch to optimize hyperparameters. (c) The test data set was used to evaluate model generalization performance, visualized using confusion matrices. Explainable AI was used to interpret model predictions and to identify biologically relevant and irrelevant image features with high prediction importance. Parts of this figure were created using Bio-Render.com.

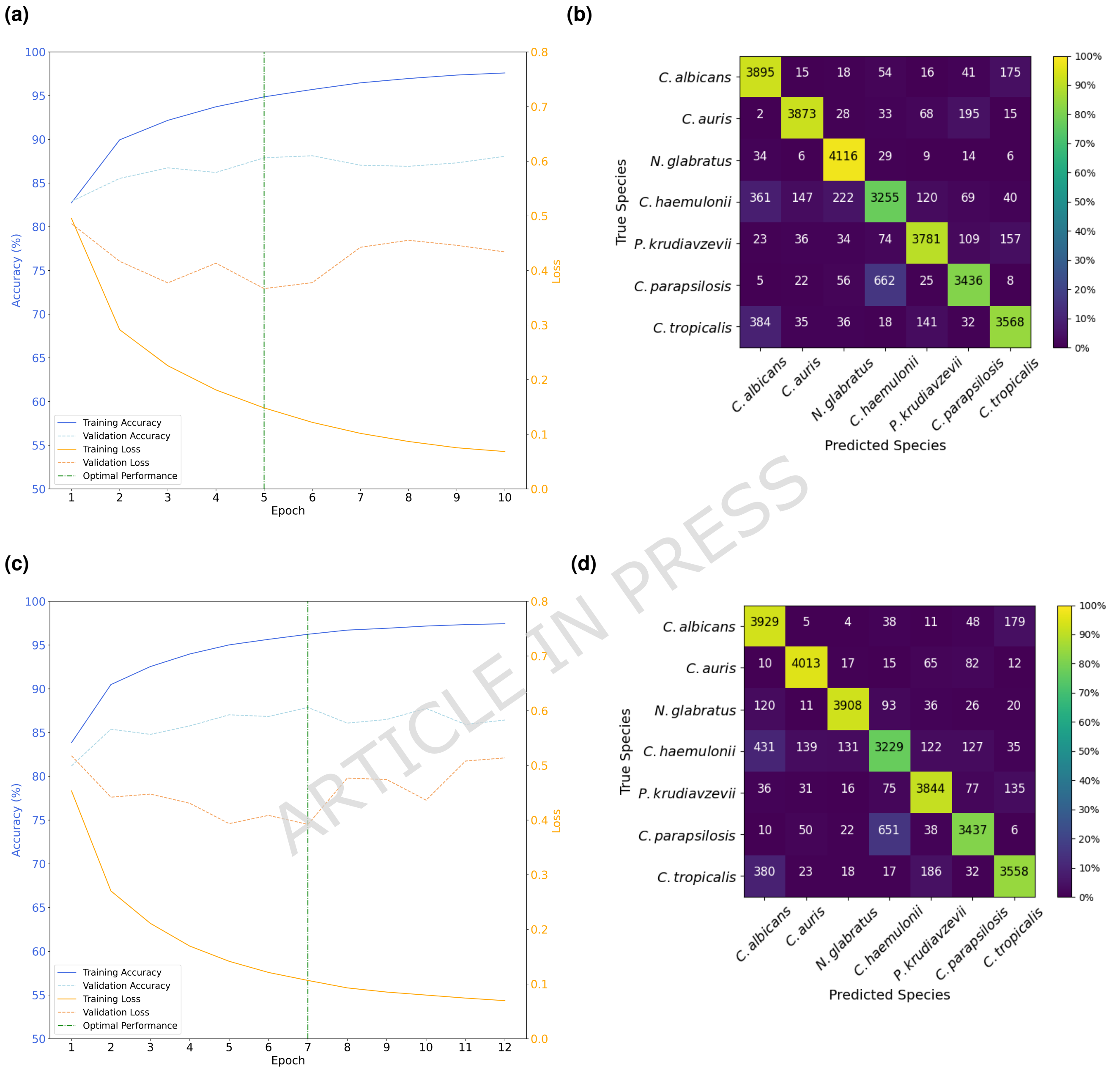
ARTICLE IN PRESS



ARTICLE IN PRESS

Figure 2. Pathogenic yeast species classification accuracies for advanced computer vision models. DenseNet121, Swin Transformer (Tiny), Vision Transformer (Base 16), and InceptionV3 models had overall test accuracies of 87%, 86%, 86%, and 86%, respectively.

ARTICLE IN PRESS



(b)

	C. albicans	*C. auris*	*N. glabratus*	*C. haemulonii*	*P. krudiavzevii*	*C. parapsilosis*	*C. tropicalis*
C. albicans	3895	15	18	54	16	41	175
C. auris	2	3873	28	33	68	195	15
N. glabratus	34	6	4116	29	9	14	6
C. haemulonii	361	147	222	3255	120	69	40
P. krudiavzevii	23	36	34	74	3781	109	157
C. parapsilosis	5	22	56	662	25	3436	8
C. tropicalis	384	35	36	18	141	32	3568

(d)

	C. albicans	*C. auris*	*N. glabratus*	*C. haemulonii*	*P. krudiavzevii*	*C. parapsilosis*	*C. tropicalis*
C. albicans	3929	5	4	38	11	48	179
C. auris	10	4013	17	15	65	82	12
N. glabratus	120	11	3908	93	36	26	20
C. haemulonii	431	139	131	3229	122	127	35
P. krudiavzevii	36	31	16	75	3844	77	135
C. parapsilosis	10	50	22	651	38	3437	6
C. tropicalis	380	23	18	17	186	32	3558

ARTICLE IN PRESS

Figure 3. Training and validation performance of the DenseNet121 and Vision Transformer-Base 16 pathogen classification models. The heatmaps show the percentage of samples in each cell relative to the validation dataset size of each class (4214 images). **(a)** Training and validation accuracy and loss curves for DenseNet121. **(b)** Confusion matrix for the DenseNet121 model applied to the validation set. **(c)** Training and validation accuracy and loss curves for Vision Transformer-Base 16. **(d)** Confusion matrix for the Vision Transformer-Base 16 model applied to the validation set.

ARTICLE IN PRESS

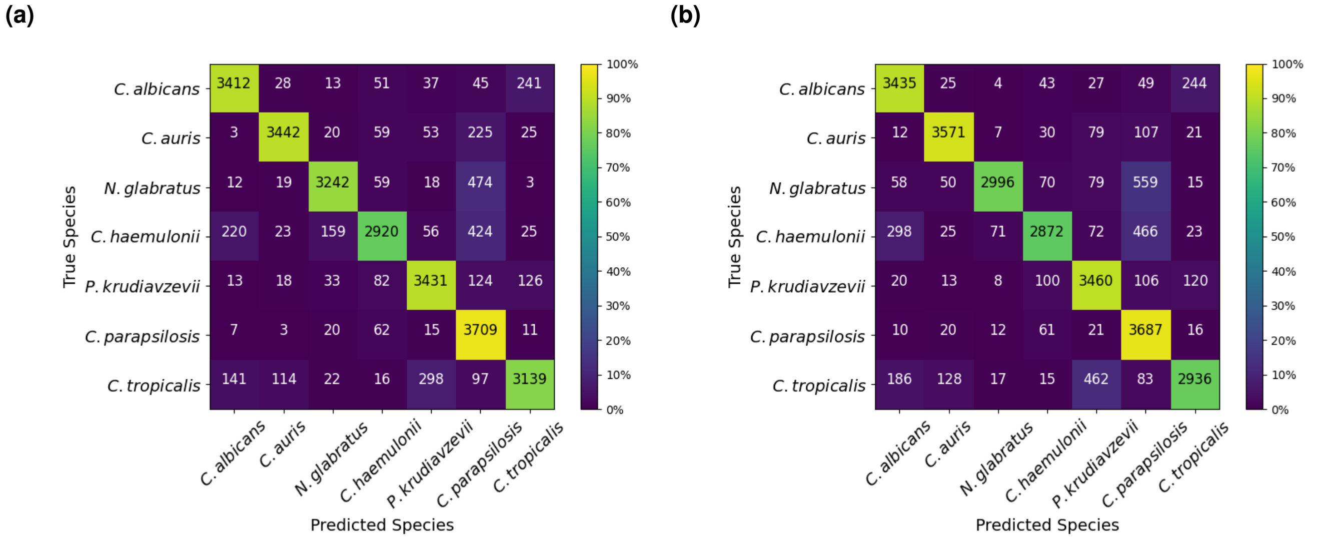


Figure 4. Confusion matrices of test set predictions made by the best-performing CNN and vision transformer pathogen classification models. The heatmaps show the percentage of samples in each cell relative to the test dataset size of each class (3827 images). **(a)** Test set confusion matrix for DenseNet121. **(b)** Test set confusion matrix for Vision Transformer-Base 16.

ARTICLE IN PRESS

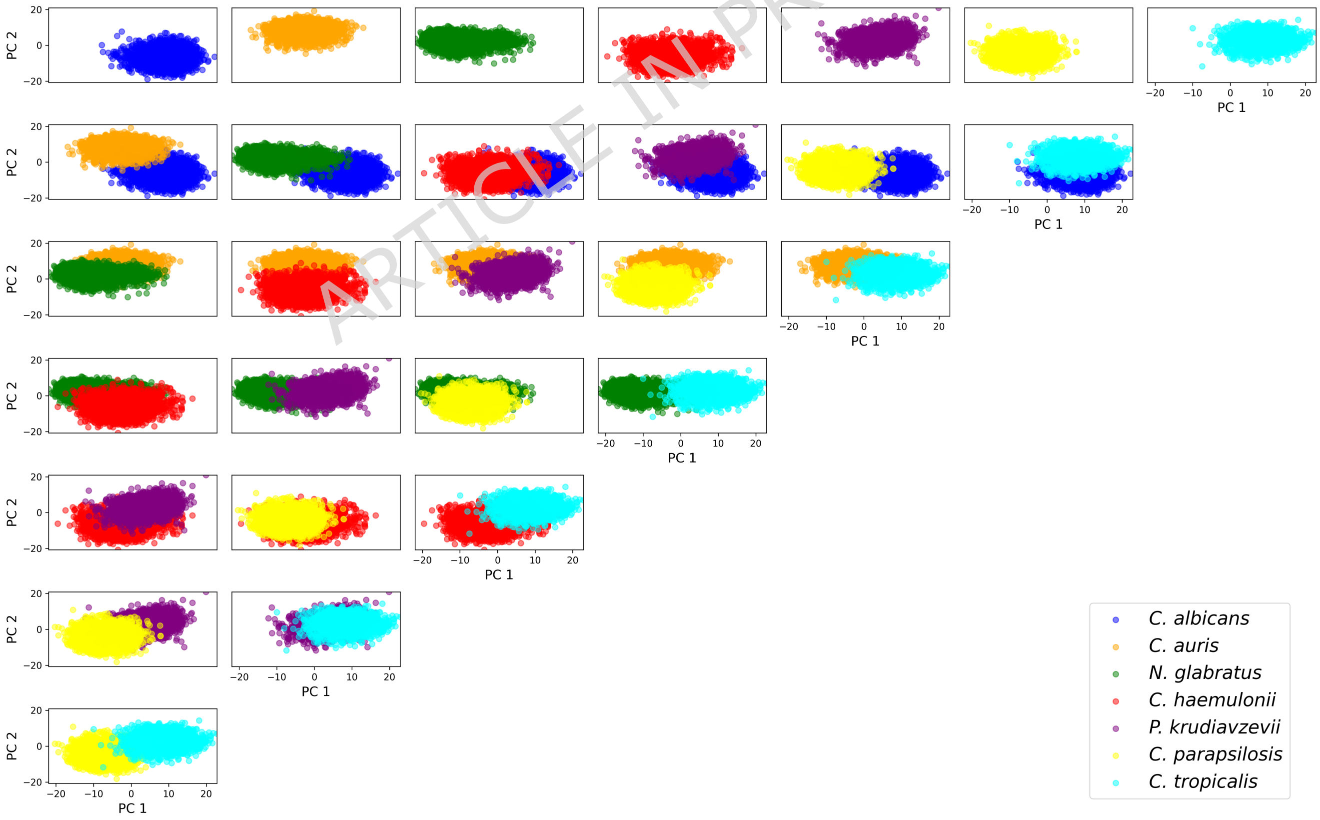
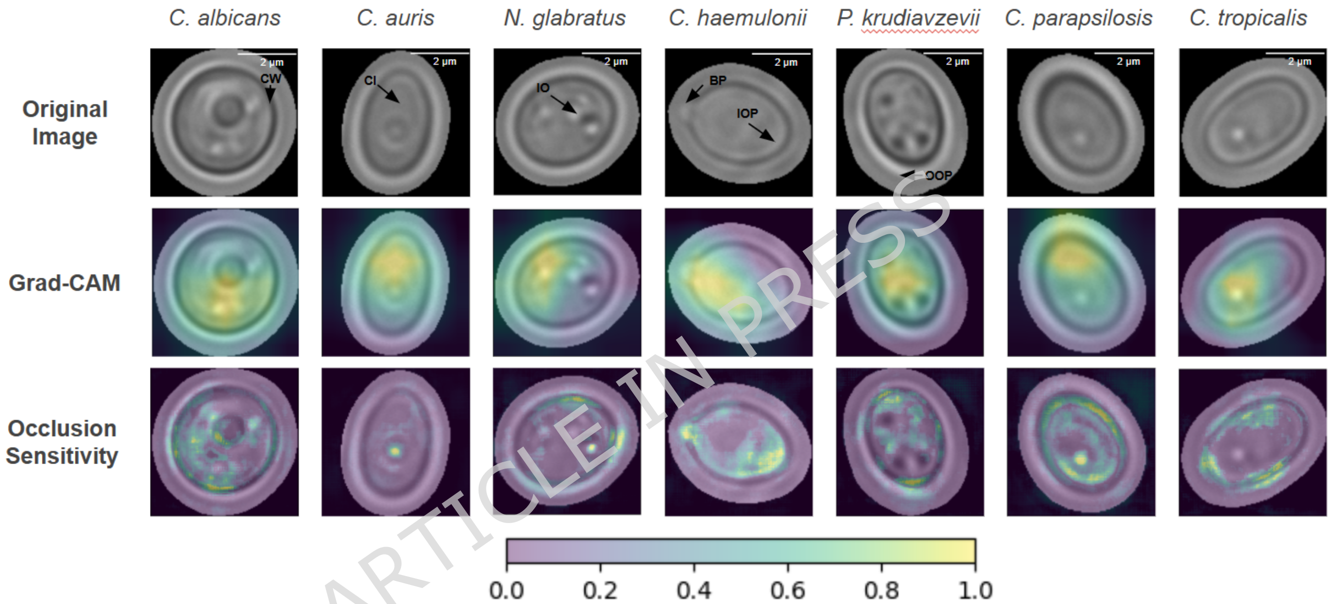


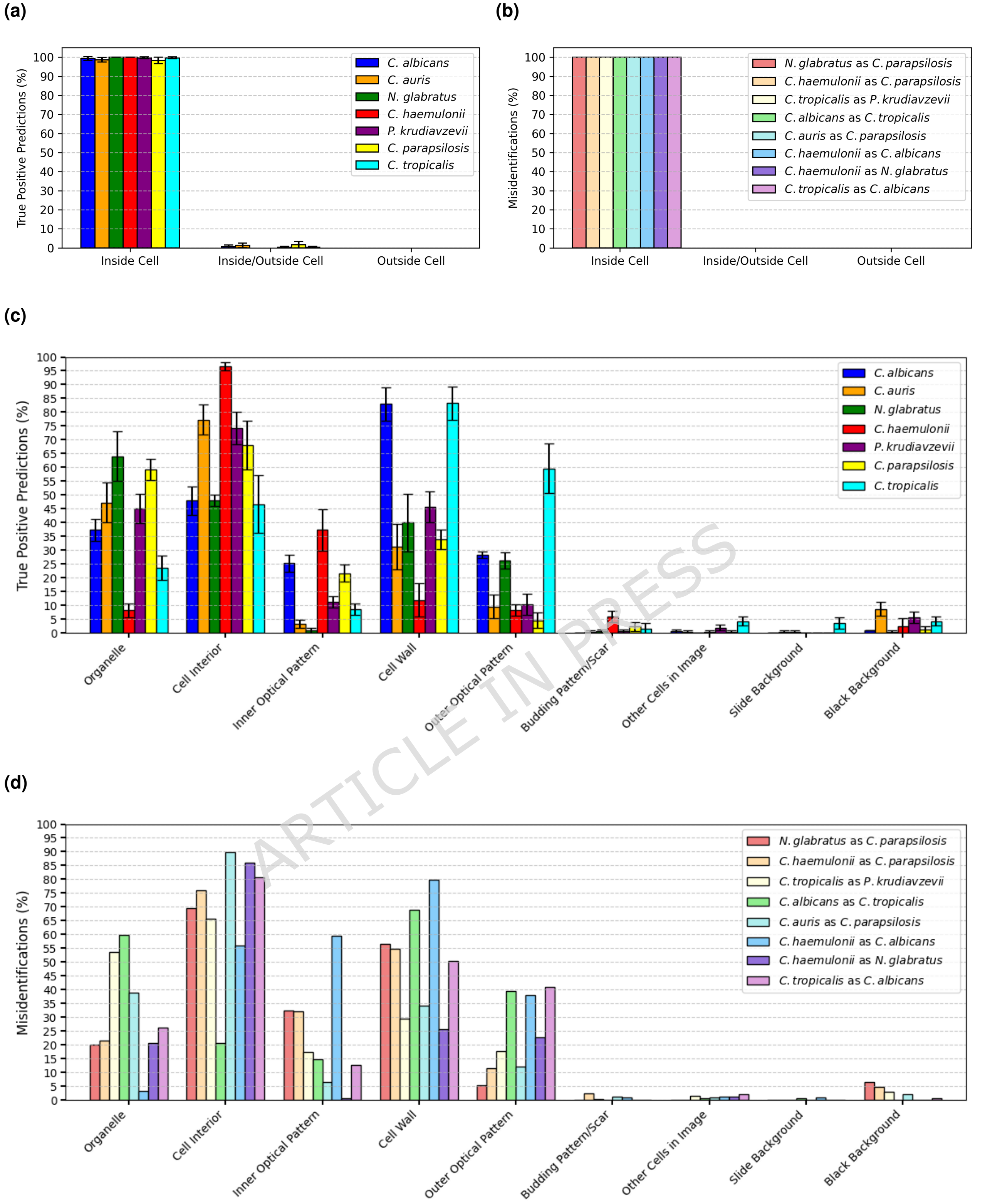
Figure 5. A 2-dimensional principal component analysis (PCA) of test set images for each pathogenic yeast species using the trained DenseNet121 model as a feature extractor. The first row shows the PCA for individual species, followed by rows showing comparisons of species pairs. A 3-dimensional PCA is provided in Supplementary Video 1.



ARTICLE IN PRESS

Figure 6. Representative cell images with overlayed Grad-CAM[55] and Occlusion Sensitivity[56] attribution maps for true positive test set predictions made by the DenseNet121 pathogen classification model. The colour bar shows the normalized attribution values. CW: cell wall; CI: cell interior; IO: internal organelles (nucleus, vacuole, etc); BP: budding pattern; IOP: inner optical pattern; and OOP: outer optical pattern. The scale bar provides an approximate length.

ARTICLE IN PRESS



ARTICLE IN PRESS

Figure 7. Proportions of identified features with high explainable AI attribution for test set predictions made by the DenseNet121 pathogen classification model. Grad-CAM[55] results are shown for true positive (TP) predictions in **(a)** and for misidentification predictions in **(b)**. Occlusion Sensitivity[56] results are shown for TP predictions in **(c)** and for misidentification predictions in **(d)**. TP results were averaged over 3 sets of 100 randomly sampled predictions from the test set for each pathogenic yeast species, with the error bars showing the standard deviation. No error bars are shown in the misidentification results, as all images for these cases were analyzed. The proportions of each case do not sum to 100%, as many images had high attribution on more than one feature.