

# OD-Abs: A One-Class Outlier Detection-Based Approach for Classifying Images of Bacteria and Fungi from Abscesses

Gabriela CZIBULA<sup>1</sup>, Teodora-Alexandra TOADER<sup>1,\*</sup>,  
Cristina MIRCEA<sup>2</sup>

<sup>1</sup> *Department of Computer Science, Babeş-Bolyai University,  
Cluj-Napoca, 1, M. Kogălniceanu street, Romania*

<sup>2</sup> *Department of Molecular Biology and Biotechnology, Babeş-Bolyai University,  
Cluj-Napoca, 44, Gheorghe Bîlaşcu street, Romania*

*e-mail: gabriela.czibula@ubbcluj.ro, teodora.toader@ubbcluj.ro, cristina.mircea@ubbcluj.ro*

Received: January 2026; accepted: July 2026

**Abstract.** Accurately differentiating between fungal and bacterial infections is critical for treatment but remains challenging due to slow manual diagnostics and class imbalance in medical datasets. This paper introduces OD-Abs, a one-class outlier detection classifier based on an autoencoder architecture to distinguish infection types in microscopic images. By training solely on the reference fungal class, the model learns to identify bacterial instances as outliers. Our approach yields statistically significant improvements compared to other one-class approaches, achieving an F1-score of 0.967, an AUC of 0.980, and a specificity of 0.992. Additionally, we explore the explainability of our method by analysing reconstruction error maps from both a technical and biological perspective and conclude that the model's decision-making aligns with biologically relevant morphological features. This proof of concept study establishes the OD-Abs autoencoder model as a highly robust tool for handling the task of classifying images of bacteria and fungi from abscesses.

**Key words:** one-class classification, outlier detection, encoder-decoder architectures, bacterial versus fungal abscesses classification.

## 1. Introduction

Accurate classification between fungal and bacterial infections represents a critical challenge in clinical diagnostics, with significant implications for patient outcomes and treatment decisions (Rawson *et al.*, 2024). Abscesses, defined as localized collections of pus, can have a wide distribution in the human body and can virtually localize in any organ or tissue (Mehta *et al.*, 2025). Usually, all of the scanning techniques, Computed Tomography (CT), Ultrasound (US), and Magnetic Resonance Imaging (MRI), can provide image data of abscesses. Generally, CT serves as a rapid initial screening tool, while MRI provides more detailed and specific information. Most frequently, bacterial and fungal abscesses

---

\*Corresponding author.

have very similar appearances on standard CT and MRI, but a particular MRI technique, Diffusion-Weighted Imaging (DWI), can provide images that can differentiate between the two, especially when localized in the Central Nervous System (CNS) (Mueller-Mang *et al.*, 2007). The type of infection, bacterial or fungal, is confirmed with traditional microbiological methods, including culture-based microscopy and biochemical assays, which are often slow and labour-intensive, delaying targeted therapy and contributing to elevated morbidity and mortality (Rawson *et al.*, 2024).

The integration of Machine Learning (ML) (Filatovas *et al.*, 2024) into microscopic image analysis is critical for overcoming the limitations of manual diagnosis. Since examining microscopic slides can be very time-consuming, ML models provide a powerful alternative that can analyse complex visual data with great speed and consistency. Currently, the most common approaches in the literature rely on standard supervised learning, particularly Convolutional Neural Networks (CNNs) and Vision Transformers. These architectures have demonstrated strong performance in species-level classification within a single domain, effectively differentiating between various fungal or bacterial genera. While research comparing fungi and bacteria exists, it is largely concentrated in ophthalmology for diagnosing keratitis, rather than microscopic analysis. In the domain of digital pathology, studies have addressed the identification of mixed microorganisms, classifying multiple bacterial (Dey *et al.*, 2022) and fungal species simultaneously in a multi-class setting. However, to the best of our knowledge, the specific problem of binary classification distinguishing fungi from bacteria from abscesses in microscopic images (Satvati *et al.*, 2024) remains unexplored. Solving this task could be particularly useful for scenarios where the clinical priority is to rapidly distinguish between these two broad categories rather than identifying the specific species.

The datasets for classifying bacteria and fungi from abscess images are generally imbalanced, as one class (bacterial or fungal) usually has more available images than the other. In such cases, supervised binary classifiers trained on imbalanced data will be biased towards predicting the majority class. One way to address class imbalance is through *one-class classification* (OCC), also known as one-class outlier detection. OCC refers to a group of outlier detection methods that are trained using data from a single class, known as the reference class. By learning the characteristics of the reference class, the OCC models can identify instances that conform to it. During inference, any new instance whose feature-based representation deviates significantly from the learned representation is considered an outlier and lies outside the decision boundary defined by the OCC method (Alam *et al.*, 2020).

The problem of differentiating between images of fungi and bacteria from abscesses is formulated in the paper as a one-class outlier detection problem, with the majority class representing the reference class. Although this task is fundamentally a binary classification problem, we hypothesize that formulating it as a one-class outlier detection provides a more robust solution than standard binary classifiers. While this formulation effectively addresses the class imbalance issue, relevant given the smaller number of images of bacteria from abscesses from our case study, its primary value lies in shifting the focus toward representation learning by aiming to capture the fundamental structural characteristics of

the reference class rather than just separating it from a different class. This offers a flexible framework for feature extraction and potential future applications in generalized anomaly detection, regardless of the specific morphology of the anomaly.

The main contribution of the paper is to introduce, as a proof of concept, a one-class outlier detection-based classifier OD-Abs to distinguish between images of bacteria or fungi from abscesses. The main goal is to encode in images of abscesses patterns and structural information that would be relevant to differentiate between fungal and bacterial cell structures. The proposed OD-Abs classifier uses an encoder-decoder architecture that can accurately reconstruct the data sampled from the same distribution as the training data. Thus, the encoder-decoder models trained on images of abscesses of the reference (majority) class will learn to preserve the structure of these images and uncover the underlying structural relationships within them. After training, an image of an abscess will be classified as belonging to the non-reference (minority) class if it is dissimilar to the representations encoded by the model. We hypothesize that the OD-Abs model, through the embedded encoder-decoder architectures, can learn through self-supervision features relevant to distinguishing structural relationships in images and thus differentiate between bacteria and fungi. With this goal, the proof of concept contains only one case study. If our hypothesis holds, the evaluation can be extended to multiple real datasets. Beyond classification metrics, clinical trust also requires interpretability. Our approach leverages the properties of the encoder-decoder architecture. We utilize reconstruction error maps to visualize the pixel-wise differences between the original input image and the model's reconstruction. These maps highlight the specific morphological regions that the model finds difficult to reconstruct, effectively pointing out the structural features that deviate from the learned fungi class. This allows us to qualitatively verify if the model is focusing on relevant biological structures. To the best of our knowledge, the research conducted in the paper is new in the literature, as we did not find similar approaches addressing the image classification of fungi versus bacteria from abscesses from a one-class classification perspective.

To summarize, the research questions addressed in this paper are the following:

- RQ1** To what extent can a one-class outlier detection model accurately uncover hidden patterns and structural information useful for classifying images of abscesses as having a fungal or bacterial origin?
- RQ2** Does the approach proposed for answering RQ1 bring a statistically significant performance improvement in classifying images of bacteria and fungi from abscesses compared to existing solutions?
- RQ3** What insights does the visualization of reconstruction residuals provide about the model's decision-making process, and how well do these interpretations align with those from a biological perspective?

The paper is organized as follows. Section 2 reviews some background knowledge related to the encoder-decoder architectures used in the paper and presents a related work on the classification of fungi versus bacteria from abscesses. In Section 3 we introduce the methodology for developing our OD-Abs one-class outlier detection-based classifier,

while Section 4 presents the experimental evaluation of the proposed approach. Section 5 discusses the explainability of the OD-ABS model and how these interpretations are correlated with the biological perspective. Section 6 outlines the conclusions of the paper and presents directions for future improvements.

## 2. Background

This section presents the encoder-decoder architectures used in the paper (Section 2.1) and reviews several recent approaches from the literature addressing the task of fungi versus bacteria classification (Section 2.2).

### 2.1. Deep Learning Models Used

The architectures that we employed belong to the autoencoder (AE) family, a family of neural networks composed of an encoder and a decoder that learns to map inputs to outputs with minimal information loss through a low-dimensional latent space. For anomaly detection, this architecture works by reconstructing normal samples while failing to reconstruct outliers.

#### 2.1.1. Variational Autoencoders

Unlike a standard autoencoder that maps an input to a fixed vector in the latent space, the Variational Autoencoder (VAE), introduced by Kingma and Welling (2014), maps the input to a probability distribution. The encoder predicts the parameters of this distribution: usually a mean vector ( $\mu$ ) and a variance vector ( $\sigma^2$ ), from which a latent vector  $z$  is sampled ( $z \sim \mathcal{N}(\mu, \sigma^2)$ ). The decoder then attempts to reconstruct the input from this sampled  $z$ . To train this network, the negative Evidence Lower Bound (negative ELBO) is minimized (equation (1)). This objective consists of two terms: a negative expected log-likelihood term, which encourages accurate reconstruction of the input, and a Kullback–Leibler (KL) divergence regularization term, which encourages the approximate posterior to remain close to a standard Gaussian prior.

$$\mathcal{L}_{VAE} = -\mathbb{E}_{q(z|x)}[\log p(x|z)] + D_{KL}(q(z|x) \| p(z)). \quad (1)$$

#### 2.1.2. $\beta$ -Variational Autoencoders

A common extension to the VAE is the  $\beta$ -Variational Autoencoder ( $\beta$ -VAE), proposed by Higgins *et al.* (2017). This architecture modifies the VAE objective by introducing a hyperparameter,  $\beta$ , to weight the KL divergence term:

$$\mathcal{L}_{\beta\text{-VAE}} = -\mathbb{E}_{q(z|x)}[\log p(x|z)] + \beta \cdot D_{KL}(q(z|x) \| p(z)).$$

By setting  $\beta > 1$ , the model increases the weight of the KL regularization term, encouraging disentangled representations.  $\beta$ -VAEs are valuable in anomaly detection for their robustness. The disentangled latent space separates the underlying generative factors of

the “normal” data more effectively. This ensures that the reconstruction error is strictly driven by the presence of abnormal features rather than entangled noise or correlations in the latent space.

### 2.1.3. Convolutional Autoencoders

While the standard autoencoder relies on fully connected layers, which necessitate flattening input data and discarding spatial structure, Convolutional Autoencoders (CAEs) are designed specifically to process high-dimensional, structured inputs such as images.

In a CAE, the encoder consists of convolutional layers that perform feature extraction, capturing hierarchical patterns while progressively reducing spatial dimensionality. The decoder employs transposed convolutional layers to upsample the latent representation back to the original input dimensions.

## 2.2. Related Work on Image Classification of Bacteria and Fungi from Abscesses

Many approaches in the literature focus on either fungi or bacteria classification, most commonly performing species-level classification within these two groups. Classification among different fungal genera, species, or higher-level classes that encapsulate multiple similar species (depending on the dataset) is most often addressed using CNN-based architectures, as demonstrated in works such as Bhimavarapu *et al.* (2024), Mohamed *et al.* (2024), Nawarathne and Kumari (2023), Rahman *et al.* (2023), Prommakhot and Srinonchat (2024). These studies show that CNN-based models can achieve strong performance on fungal image classification tasks. Transformer-based architectures have also been employed for this purpose and have achieved competitive results (Rahman *et al.*, 2025; Gümüş, 2024; Ahmed and Haque, 2023). Similarly, CNN-based methods (Visitsatapong *et al.*, 2024; Poomrittigul *et al.*, 2022), as well as transformer-based approaches (Prasad *et al.*, 2025), have been proposed for bacterial classification tasks.

Research explicitly addressing bacterial versus fungal classification is less extensively explored in the literature; nevertheless, several relevant studies exist. One notable application is the differentiation between bacterial keratitis and fungal keratitis. In Redd *et al.* (2022), the authors evaluate five different CNN architectures, as well as a CNN ensemble, for classifying anterior segment eye photographs according to the etiological category of keratitis. Another approach is presented in Won *et al.* (2023), where a CNN with a ResNet (He *et al.*, 2016) backbone is augmented with a Lesion Guiding Module and a Masked Adjusting Module, and qualitative analysis is performed using Grad-CAM visualizations. A multimodal approach is explored in Prajna *et al.* (2025), where both clinical data and images are used. The proposed model employs separate feature extractors for each modality, followed by a fusion mechanism prior to classification. The study reports that the vision-based model outperforms the clinical-data-only model and that the multimodal approach does not yield a significant improvement over the vision-only approach.

Studies focusing on bacteria and fungi classification from microscopic images can also be found in the literature. In Sroka-Oleksiak *et al.* (2025), the authors propose a method for identifying 14 bacterial species and 3 fungal species from Gram-stained microscopic images. Their pipeline includes cell segmentation using the Cellpose model,

patch extraction, patch encoding, and aggregation prior to final classification. Separate classifiers are trained for bacterial and fungal samples. Other works, such as Treebupachatsakul and Poomrittigul (2020) and Zawadzki (2020), investigate multiclass classification of mixed bacterial and fungal microscopic images. In Treebupachatsakul and Poomrittigul (2020), a LeNet CNN is trained to distinguish between three bacterial species and one yeast species. In Zawadzki (2020), the authors evaluate multiple CNN architectures (Xception, ResNet, Inception-V3, MobileNetV2, and DenseNet201) on microscopic images stained using Congo Red and modified Gram stain across three datasets. They report accuracies of up to 100% on DS1 and 96% on DS2 using the Xception architecture.

### 3. Methodology

To answer RQ1, this section introduces OD-Abs, a one-class outlier detection-based image classifier to distinguish between bacteria and fungi from abscesses. Deep encoder-decoder architectures are used to encode patterns and structural information in images of fungal abscesses, patterns that would be relevant to distinguish between fungal and bacterial cell structures.

#### 3.1. Problem Definition and Formalisation

The problem of classifying abscess images as bacterial or fungal is a binary classification problem. Denoting by  $\mathcal{C} = \{bacterial, fungal\}$  the set of two targeted classes, the learning task is to estimate for an abscess image  $I$  the most probable class of  $\mathcal{C}$  to which it belongs.

We formulate the problem as a one-class outlier detection problem in which the learning model is trained on only one class. The goal is to learn the main characteristics of the reference class and to detect as outliers those instances whose features differ significantly from the learned representation.

Let us consider a dataset of images of fungi and bacteria from abscesses. We consider the *reference* class (and denote it by  $c_{ref}$ ) as the majority class (i.e. the one that has more available images), while the minority (less represented) class will be considered as the *other* class and will be denoted by  $c_{other}$ . For instance, if more images of fungi from abscesses are available than those of bacteria from abscesses, then the reference class is the “fungal” one ( $c_{ref} = fungal$ ) and the *other* class is the “bacterial” one ( $c_{other} = bacterial$ ). Otherwise, if the fungal class is underrepresented in the dataset, then  $c_{ref} = bacterial$  and  $c_{other} = fungal$ .

OD-Abs model is trained only on the reference class  $c_{ref}$  and learns the characteristics of this class by compressing and then recreating the images from the reference class. Theoretically, the model learns a decision boundary surrounding the images from the reference class. Thus, the model will be able to identify images that resemble the learned representation. During inference, new images that fall outside the decision boundary are classified as outliers (i.e. belonging to the  $c_{other}$  class), as they exhibit significantly different characteristics than the representation of the reference class. Since OD-Abs is an

encoder-decoder model, images from the  $c_{other}$  class typically result in high reconstruction errors because the model has not learned to represent them.

More specifically, we would like to detect whether an abscess image  $I$  resembles images belonging to the reference class or is dissimilar to it and thus represents an anomaly. To determine whether the image  $I$  does not belong to the class  $c_{ref}$  we compute the likelihood that the image is an anomaly of the *reference* class. This likelihood is expressed as the dissimilarity degree of the given image with respect to the structure of the class  $c_{ref}$  encoded in a *encoder-decoder* architecture. Through their latent representations, the encoder-decoder architectures are used to uncover, through self-supervision, features and structural information from abscess images, which are relevant for distinguishing between fungi and bacteria.

Denoting by  $\mathcal{Img}$  the input space of abscess images, the classification of images of bacteria versus fungi from abscesses can be formalized as an approximation of a target function  $f : \mathcal{Img} \rightarrow \mathcal{C}$  that maps an abscess image to the “fungal” or “bacterial” class. Thus, the OD-Abs classification model will approximate an additional function  $p_{ref} : \mathcal{Img} \rightarrow [0, 1]$  expressing the probability that an image  $I \in \mathcal{Img}$  belongs to the class  $c_{ref}$ . Certainly, the probability  $p_{other}$  that an image  $I \in \mathcal{Img}$  belongs to the class  $c_{other}$  (i.e. it is an outlier) is obtained as  $p_{other}(I) = 1 - p_{ref}(I)$ . Thus, after the approximation  $\tilde{p}_{ref} \approx p_{ref}$  is learned from a training data consisting of images belonging to the majority (reference) class, an approximation of the target function  $\tilde{f} \approx f$  will be simply computed as shown in equation (2).

$$\tilde{f}(I) = \begin{cases} c_{ref}, & \text{if } p_{ref}(I) \geq 0.5, \\ c_{other}, & \text{otherwise,} \end{cases} \quad \forall I \in \mathcal{Img}. \quad (2)$$

### 3.2. Overview of OD-Abs

The overall OD-Abs methodology consists of three main stages. Figure 1 illustrates the two operational steps of the final classifier, namely training and thresholding, followed by classification of unseen images.

1. **Encoder-decoder model selection.** To decide the best-performing encoder-decoder model to be integrated in the OD-Abs one-class anomaly detector, the performance of the three encoder-decoder models will be comparatively analysed.
2. **Building the OD-Abs classification model.** During this stage, the one-class outlier detection model OD-Abs will be built. A set of images of the majority class (i.e. the class – fungal or bacterial – that has the most available images) is used for self-supervised training of the encoder-decoder model that was selected at the previous stage to learn to represent relevant structural relationships in images of fungi from abscesses. The trained OD-Abs will be further used to discriminate between images representing fungal and bacterial infections.
3. **Testing the OD-Abs classifier.** The predictive model OD-Abs built during the previous stage is tested on abscess images (both fungal and bacterial) that were not seen

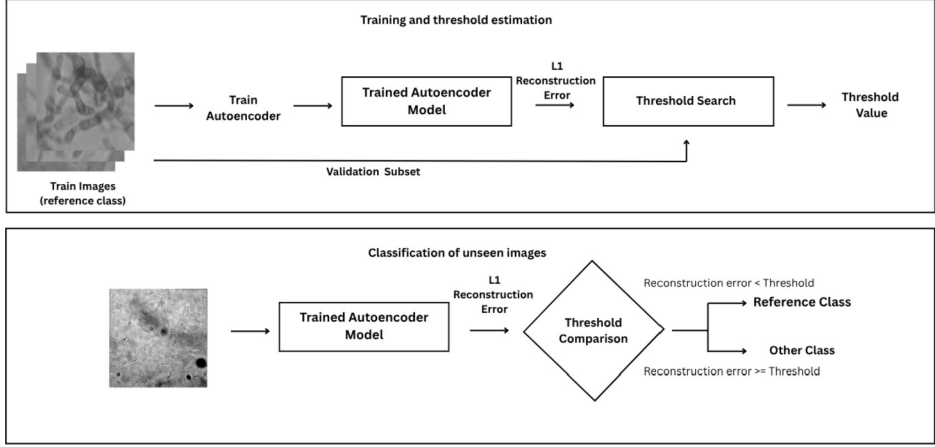


Fig. 1. Overview of the OD-Abs training, threshold estimation, and classification workflow.

during training. The performance of OD-Abs will be assessed using supervised classification metrics.

### 3.3. Encoder-Decoder Model Selection

As illustrated in Fig. 1, our proposed OD-Abs approach uses an encoder-decoder model  $M$  that is trained through self-supervision on the images from the majority class (representing fungi or bacteria) and will learn to encode the structure of these images.

For the encoder-decoder model  $M$ , three variants are considered. The architectures of the three encoder-decoder models employed in our experiments, together with details regarding the training process, are described below.

#### 3.3.1. Variational Autoencoders

The first model is a Variational Autoencoder (VAE) designed to process grayscale images of size  $256 \times 256$ . The encoder network utilizes a ResNet-18 architecture (He *et al.*, 2016), initialized with weights pretrained on ImageNet. To accommodate the single-channel grayscale input, the first convolutional layer is modified by summing the original RGB weights along the channel dimension, resulting in a Conv2d layer with a kernel size of  $7 \times 7$ , stride of 2, and padding of 3. The encoder retains the layers up to the fourth residual block, extracting a deep feature map with dimensions  $512 \times 8 \times 8$ . This feature map is flattened into a vector of size 32768 and passed through two parallel fully connected layers to produce the latent mean ( $\mu$ ) and log-variance ( $\log \sigma^2$ ) vectors, with a latent dimensionality of  $d = 32$ .

The decoder network generally mirrors the encoder's depth but utilizes a different up-sampling strategy to mitigate checkerboard artifacts commonly associated with transposed convolutions. It begins with a fully connected layer that maps the sampled latent vector  $z$  back to the feature map shape of  $512 \times 8 \times 8$ . This is followed by five up-sampling blocks to

reach the target resolution. Each block employs a nearest-neighbour upsampling operation (scale factor 2) to mitigate checkerboard artifacts, followed by a standard 2D convolution ( $3 \times 3$  kernel, stride 1, padding 1), Batch Normalization, and a LeakyReLU activation (negative slope 0.2). The channel depth is progressively reduced through the sequence:  $512 \rightarrow 256 \rightarrow 128 \rightarrow 64 \rightarrow 32$ . The final output layer applies a convolution to reduce the channels to 1 and employs a Tanh activation function. The model is optimized using the Adam optimizer with a learning rate of  $1 \times 10^{-5}$  and a batch size of 32. Training is conducted for 50 epochs per fold in a 5-fold cross-validation setup. In our implementation, the negative expected log-likelihood term from the negative ELBO is approximated using the L1 reconstruction error between the input image and its reconstruction. The practical training objective therefore combines the L1 reconstruction error and the KL divergence, as shown in equation (3).

$$\mathcal{L}_{VAE} = \sum |x - \hat{x}| + \beta \cdot D_{KL}(q(z|x) \| p(z)), \quad (3)$$

where  $x$  represents the original input data,  $\hat{x}$  is the reconstructed output,  $\beta$  is the regularization coefficient (set to  $\beta = 1$ ),  $D_{KL}$  denotes the Kullback-Leibler divergence, and  $z$  is the latent representation.

The second variant is the  $\beta$ -VAE, which utilizes the identical architectural backbone, optimizer, and training schedule as the standard VAE described above, with the distinction that  $\beta = 3$  in the objective function.

### 3.3.2. Convolutional Autoencoder

The third model is a deterministic Convolutional Autoencoder (CAE). To ensure a fair comparison, the CAE shares the identical architectural backbone as the VAE and  $\beta$ -VAE variants. This includes the use of the ResNet-18 encoder (modified for grayscale input) and the five-stage upsampling decoder utilizing nearest-neighbour interpolation to prevent checkerboard artifacts. However, the CAE removes the probabilistic sampling mechanism. The flattened output of the encoder (dimensions  $512 \times 8 \times 8$ ) is mapped directly to the bottleneck vector  $z$  of size  $d = 32$  via a fully connected layer, which is then projected back to the feature dimensions to feed the decoder. Since there is no probabilistic latent distribution, the KL divergence term is removed. The model is trained solely on the reconstruction objective ( $\mathcal{L}_{CAE} = \frac{1}{N} \sum |x - \hat{x}|$ ), using the Adam optimizer. Consistent with the other models, the final output layer employs a Tanh activation function. The training duration (50 epochs, with added early stopping to avoid overfitting) remains consistent with the variational models; however, the learning rate is set to  $1 \times 10^{-4}$  (increased from the VAE's  $1 \times 10^{-5}$  to facilitate faster convergence in the absence of the KL regularization term).

### 3.3.3. Performance Evaluation

The performance of the proposed encoder-decoder architectures will be further analysed to determine the model that best succeeds in accurately reconstructing images belonging to the reference class  $c_{ref}$  while, at the same time, failing to reconstruct images of the  $c_{other}$ .

The best-performing encoder-decoder model will be further integrated in the OD-Abs one-class anomaly detector.

For comparing two images (in our case, an input image and the output image reconstructed by an encoder-decoder model  $M$ ), several metrics from the image processing literature are used:

- *mean squared error* (MSE) is the mean squared difference between the pixel in one image and its corresponding pixel in another image;
- the *structural similarity index* (SSIM) (Wang et al., 2004) is a perceptual metric that evaluates the visual similarity between two images based on luminance, contrast, and structure;
- the *signal-to-noise ratio* (SNR) is the ratio of the power of the meaningful signal to the power of the corrupting background noise.

For highly similar images, low values for MSE and high values for SSIM and SNR are expected.

Denoting by  $M$  an encoder-decoder model and by  $\mathcal{D}$  a dataset of images of abscesses (fungal or bacterial), aggregated values are computed by averaging the metric between an image and its reconstruction obtained by  $M$ , i.e.

$$\begin{aligned} \text{MSE}_{\mathcal{D}} &= \frac{\sum_{I \in \mathcal{D}} \text{MSE}(I, \tilde{I})}{|\mathcal{D}|}, & \text{SSIM}_{\mathcal{D}} &= \frac{\sum_{I \in \mathcal{D}} \text{SSIM}(I, \tilde{I})}{|\mathcal{D}|}, \\ \text{SNR}_{\mathcal{D}} &= \frac{\sum_{I \in \mathcal{D}} \text{SNR}(I, \tilde{I})}{|\mathcal{D}|}. \end{aligned}$$

### 3.4. Building the OD-Abs Classification Model

As illustrated in Fig. 1, our proposed OD-Abs approach uses an encoder-decoder model  $M$  that is trained on a dataset of images belonging to the reference class  $C_{ref}$  to encode information about the structure of the trained images. For training the model, 60% of the images will be used, 20% for the validation of the model, and the remaining 20% will be used for testing.

The model  $M$  is trained through self-supervision to encode the structure of the majority class images, using an *error* function that penalizes the model; this function is expressed as the difference between the provided input image and the output image reconstructed through the decoder component.

Using the encoder-decoder model  $M$  selected following the evaluation introduced in Section 3.3 and trained as previously detailed, the OD-Abs classifier will produce an approximation  $p_{ref}^M$  of the probability function  $p_{ref}: \text{Img} \rightarrow [0, 1]$  expressing the probability that an abscess image  $I \in \text{Img}$  belongs to the reference class ( $c_{ref}$ ).

The learned function  $p_{ref}^M$  will be used by the OD-Abs classifier to predict if an abscess image belongs to the fungal or bacterial class; the probability that an image  $I$  belongs to the bacterial class is computed as  $p_{other}^M(I) = 1 - p_{ref}^M(I)$ .

Let us denote by  $\tilde{I}_M$  the output of the trained model  $M$  for an input image  $I$  (i.e. the reconstruction of  $I$ ) and by  $RE(I, \tilde{I}_M)$  the reconstruction error obtained for  $I$  ( $L_1$  dis-

tance between the input and the reconstruction). The probabilities  $p_{ref}^M(I)$  and  $p_{other}^M(I)$  estimated by the OD-Abs classifier using the encoder-decoder model  $M$  are computed as shown in equation (4).

$$p_{ref}^M(I) = \frac{\tau_M}{\tau_M + RE(I, \tilde{I}_M)}, \quad p_{other}^M(I) = \frac{RE(I, \tilde{I}_M)}{\tau_M + RE(I, \tilde{I}_M)}. \quad (4)$$

The idea behind computing the equation (4) is that the encoder-decoder model  $M$  can accurately reconstruct the images sampled from the same distribution as the training images (i.e. images of the reference class). Moreover, the model  $M$  should not be able to accurately reconstruct through its learned latent representation an input image that is structurally dissimilar to the training images, e.g. the images that belong to the minority class.

In equation (4),  $\tau_M$  is a hyperparameter of the OD-Abs classifier and represents the *threshold* applied on the reconstruction error of an input image  $I$  to decide if it belongs to the class  $c_{ref}$  (i.e. if it is similar enough to the fungal class represented by the encoder-decoder model  $M$ ). One observes from equation (4) that if  $RE(I, \tilde{I}_M) < \tau_M$ , then  $p_{ref}^M(I) > 0.5$ ; thus,  $I$  will be classified as belonging to the fungal class  $c_{ref}$ . Otherwise, if  $RE(I, \tilde{I}_M) \geq \tau_M$ , then  $p_{ref}^M(I) \leq 0.5$ , then  $I$  will be considered dissimilar to the class represented by the model  $M$  and classified as belonging to the bacterial class  $c_{other}$ . It should also be noted that if  $RE(I, \tilde{I}_M) = 0$ , then  $p_{ref}^M(I) = 1$ , while if  $RE(I, \tilde{I}_M) = \tau_M$ , then  $p_{ref}^M(I) = 0.5$ . Moreover, the value of  $p_{ref}^M(I)$  decreases as  $RE(I, \tilde{I}_M)$  increases.

Two variants are considered for computing the hyperparameter  $\tau$  after completing the training of OD-Abs.

**V1** For a specific encoder-decoder model  $M$  we are computing the *mean* ( $\mu(M)$ ) and the *standard deviation* ( $\sigma(M)$ ) of the reconstruction errors  $RE(I, \tilde{I}_M)$  computed for all the images from the training dataset. If  $\mathcal{D}$  is the training dataset and  $n$  represents the number of images from the training set ( $n = |\mathcal{D}|$ ), then

$$\mu(M) = \frac{\sum_{I \in \mathcal{D}} RE(I, \tilde{I}_M)}{n} \quad \text{and} \quad \sigma(M) = \sqrt{\frac{\sum_{I \in \mathcal{D}} (RE(I, \tilde{I}_M) - \mu(M))^2}{n}}.$$

The first method for estimating the threshold  $\tau$  is to compute it as the mean of the reconstruction errors of the training images plus twice the standard deviation of the errors, i.e.  $\tau_M = \mu(M) + 2\sigma(M)$ . The intuition behind this computation is that the rule “mean plus twice the standard deviation of the values from a population” is important in statistical distributions, particularly in the normal distribution. Even though the maximum of a population is not necessarily located at the  $\mu \pm 2\sigma$ , this formula gives an estimate of a good enough “upper bound” of the population values, as in many normal distributions, most values lie within  $\mu \pm 2\sigma$ .

**V2** The second variant optimizes the decision threshold  $\tau$  on a validation set after training the OD-Abs classifier. To this end, we construct a tuning set comprising 20% of the total normal samples (the validation split of the current fold) and 50% of the available anomaly samples, with the remaining anomalies reserved for testing. The optimal threshold is selected via a grid search over candidate thresholds derived from the

Precision–Recall (PR) curve computed on the tuning set. Specifically, reconstruction error scores are used to generate PR-curve thresholds, for which the F1-score is evaluated, and the threshold maximizing the F1-score is chosen.

### 3.5. Testing the OD-Abs Classifier

After training, the OD-Abs classifier is tested on both images of fungi and bacteria from abscess to assess its predictive performance.

As previously described in Section 3.1, based on the number of abscess images available for the “fungal” and “bacterial” classes, the majority class is considered the *reference* class and used for training the OCC model. Besides the dataset of abscess images of the reference class used for training, validation, and testing the OD-Abs model, a dataset of abscess images of the minority class will be used for testing.

As shown in Section 3.4, the training dataset is randomly split into training-validation-testing. Due to the randomness of the splitting process, a *cross-validation* testing methodology is used. More specifically, the data set of images of the majority class is randomly split into training-validation-testing. For a specific split, the OD-Abs classifier will be tested on a dataset obtained by combining 20% of the training images that were not used for training and validation, and the dataset of images of the minority class.

For measuring the performance of OD-Abs on a testing set, the following evaluation measures are computed (Gu *et al.*, 2009), considering the “fungal” class as the negative one and the “bacterial” class as the positive one: *precision* ( $\text{Prec} = \frac{TP}{TP+FP}$ ), *sensitivity* (*recall*, or *true positive rate*,  $\text{Sens} = \frac{TP}{TP+FN}$ ), *specificity* (*true negative rate*,  $\text{Spec} = \frac{TN}{TN+FP}$ ), *F1 score* ( $\text{F1} = \frac{2 \cdot \text{Prec} \cdot \text{Sens}}{\text{Prec} + \text{Sens}}$ ) and *Area Under the ROC Curve* ( $\text{AUC} = \frac{\text{Spec} + \text{Sens}}{2}$ ).

We note that for computing the AUC measure, we consider the case of classifiers that provide the output class directly (Fawcett, 2006). In this case, the ROC (Receiver Operating Characteristic) curve is a quadrilateral obtained by linking the points (0, 0) and (1, 1) to the (1 – Spec, Sens) point. According to the literature, AUC is one of the most effective evaluation metrics for assessing the performance of imbalanced classifiers (Fawcett, 2006). For each of the previously described performance metrics, the performance of OD-Abs averaged over the five runs is reported together with the 95% confidence interval (CI) (Brown *et al.*, 2001) of the mean.

For classifying an input abscess image as having a fungal or bacterial origin, the trained OD-Abs model performs the steps described in Section 3.4 and summarized in Algorithm 1.

## 4. Experimental Evaluation

To complete the answer to RQ1, an experimental evaluation of the OD-Abs classifier is further conducted. The predictive performance of the classifier is assessed following the testing methodology introduced in Section 3.5. The code is made publicly available at (Toader, 2025).

**Algorithm 1** The classification stage of OD-Abs

---

**function** CLASSIFICATION( $M, \tau_M, I$ )
**Require:**

- $M$  – the trained encoder-decoder model
- $\tau_M$  – the decision threshold (Section 3.4)
- $I$  – the query image that has to be classified

**Ensure:**

returns  $c_{ref}$  if the predicted class for  $I$  is the reference (majority) class and  $c_{other}$  (the minority) otherwise

---

```

// compute the reconstruction error obtained by the model  $M$  for the image  $I$ 
 $RE(I, \tilde{I}_M) \leftarrow$  the reconstruction error computed for  $I$  using the model  $M$ 
// compute the probability that  $I$  belongs to the class  $c_{ref}$ 
 $p_{ref}^M(I) \leftarrow \frac{\tau_M}{\tau_M + RE(I, \tilde{I}_M)}$ 
if  $p_{ref}^M(I) \leq 0.5$  then
    /*  $I$  is predicted as belonging to the minority class (an anomaly of the majority
       class) */
    return  $c_{other}$ 
else
    /*  $I$  is predicted as belonging to the reference class */
    return  $c_{ref}$ 
end if
end function

```

---

## 4.1. Datasets

This section describes the datasets used in the experiments: the labelled dataset  $Df$  containing images of fungi from abscesses and the labelled dataset  $Db$  containing images of bacteria from abscesses.

For  $Df$  we used the DeFungi dataset (Sopo *et al.*, 2021), a publicly available repository of 9114 manually labelled microscopic images. These images were obtained from direct mycological examinations of fungal infections and processed into  $500 \times 500$  pixel patches. The dataset is organized into five classes corresponding to morphological patterns observed under microscopy: Tortuous septate hyaline hyphae (TSH), Beaded arthroconidial septate hyaline hyphae (BASH), Groups or mosaics of arthroconidia (GMA), Septate hyaline hyphae with chlamyidioconidia (SHC), and Broad brown hyphae (BBH).

For  $Db$  we employed both the DIBaS (Digital Images of Bacterial Species) (Zieliński *et al.*, 2017) dataset and a subset of the DeepBacs (Spahn *et al.*, 2022) image collection, more specifically the raw images from the mixed segmentation dataset. The DIBaS dataset contains approximately 640 bacteria Gram-stained microscopic images covering 32 bacterial species. The DeepBacs dataset that we used contains 190 brightfield and fluorescence images covering three species: *S. aureus*, *E. coli* and *B. subtilis*.

Table 1  
Datasets used in this study and their characteristics.

| Dataset   | Source           | Domain   | Classes | Images | Resolution                                                                                             | Microscopy                             |
|-----------|------------------|----------|---------|--------|--------------------------------------------------------------------------------------------------------|----------------------------------------|
| <i>Df</i> | DeFungi          | Fungi    | 5       | 9114   | $500 \times 500$                                                                                       | Direct light microscopy                |
| <i>Db</i> | DIBaS + DeepBacs | Bacteria | 33      | 833    | $2048 \times 1532$ (643)<br>$1024 \times 1024$ (124)<br>$256 \times 256$ (56)<br>$512 \times 512$ (10) | Gram-stained Bright-field/Fluorescence |

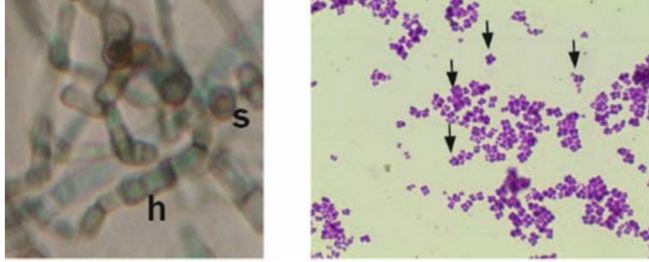


Fig. 2. Example of a fungi image from DeFungi Dataset (left) and bacteria image from DeepBacs dataset (right).

Figure 2 comparatively illustrates a fungi image (the left side of the image) and a bacteria image (the right side of the image). One observes that fungi appear as filamentous hyphae (marked with **h** in the left-side image), comprised of long, tube-like strands with rounded segments identified as possible spores or conidia (marked with **s** in left-side image), in contrast to the bacterial cells that look like clustered spherical cells typical of Gram-positive cocci (marked with arrows in the right-side image).

We note that in our experiment the bacterial class is underrepresented, and thus the reference class is the majority “fungal” class ( $c_{ref} = Df$ ), while the minority class is the “bacterial” class ( $c_{other} = Db$ ).

According to the methodology introduced in Section 3, the OD-Abs classifier is trained on *Df* and learns a decision boundary by representing the characteristics of the fungal class. At the query time, the images that resemble the learned representation are predicted as belonging to the “fungal” class; the images that fall outside the decision boundary (i.e. deviate significantly from the learned representation) are considered outliers and predicted as “bacterial”.

#### 4.2. Preliminary Experiment

The performance of the three encoder-decoder models introduced in Section 3.3 is comparatively analysed to decide the best-performing encoder-decoder model to be integrated in the OD-Abs one-class classifier.

The dataset containing images of fungi from abscesses (*Df*) and the dataset of images of bacteria from abscesses (*Db*) will be used for evaluating the performance of the three

Table 2  
Performance metrics including sensitivity index.

| Model        | Sensitivity index $\uparrow$ | Fungal data ( $D_f$ ) |                 |                | Bacterial data ( $D_b$ ) |                   |                  |
|--------------|------------------------------|-----------------------|-----------------|----------------|--------------------------|-------------------|------------------|
|              |                              | MSE $\downarrow$      | SSIM $\uparrow$ | SNR $\uparrow$ | MSE $\uparrow$           | SSIM $\downarrow$ | SNR $\downarrow$ |
| VAE          | 2.875                        | 0.009                 | 0.787           | 22.769         | 0.077                    | <b>0.405</b>      | 12.765           |
| $\beta$ -VAE | <b>3.205</b>                 | 0.009                 | <b>0.788</b>    | 22.730         | <b>0.078</b>             | 0.408             | <b>12.685</b>    |
| CAE          | 2.156                        | <b>0.002</b>          | 0.784           | <b>22.999</b>  | 0.018                    | 0.424             | 13.263           |

encoder-decoder models proposed. From a one-class outlier detection perspective, the best model is considered the one that provides an accurate reconstruction of images of fungi but fails to reconstruct images of bacteria.

Table 2 presents the values of the metrics computed on  $D_f$  and  $D_b$ . The best values obtained for the metrics are highlighted. We note that we search for higher similarities between the fungal images and their reconstruction (lower  $MSE_{D_f}$ , higher  $SSIM_{D_f}$  and  $SNR_{D_f}$ ), while lower similarities between the bacterial images and their reconstruction (higher  $MSE_{D_b}$ , lower  $SSIM_{D_b}$  and  $SNR_{D_b}$ ). In order to evaluate the model’s ability to distinguish between the two classes, we utilize the Sensitivity Index (SI), defined as  $d' = \frac{|\mu_{D_f} - \mu_{D_b}|}{\sqrt{\frac{1}{2}(\sigma_{D_f}^2 + \sigma_{D_b}^2)}}$ . A higher SI value indicates superior class separability. The best values for the evaluation metrics are highlighted.

Table 2 reveals a trade-off between reconstruction fidelity and separation capability across the three models. While CAE achieves the best reconstruction metrics on the target fungal data (lowest  $MSE_{D_f}$  of 0.002 and highest  $SNR_{D_f}$ ), it fails to sufficiently differentiate the bacterial samples, shown by the low  $MSE_{D_b}$  and the lowest Sensitivity Index ( $d' = 2.156$ ). This suggests that while CAE is a good compressor, it lacks the necessary regularizing constraints to distinguish between the two biological classes effectively.

In contrast, the variational approaches provide better separability. Although the standard VAE offers a balanced performance, the  $\beta$ -VAE model proves to be the superior version. It yields the highest error and lowest SNR when reconstructing bacterial images ( $MSE_{D_b} = 0.078$ ) while maintaining comparable performance on the fungal class. Most importantly,  $\beta$ -VAE achieves the highest Sensitivity Index ( $d' = 3.205$ ), indicating the largest statistical distance between the normal and anomalous distributions. Therefore, based on the maximization of class separability,  $\beta$ -VAE is selected as the most suitable model for the proposed anomaly detection framework (see Fig. 3).

#### 4.3. Results

As shown in Section 3.3, the  $\beta$ -VAE model was empirically found to be the best encoder-decoder model in preserving the structural characteristics of images of fungal abscesses, while failing to reconstruct images of bacterial abscesses. Thus, we hypothesize that the OD-Abs classifier with the  $\beta$ -VAE model integrated can accurately encode the fungal class and uncover hidden patterns and structural information useful to effectively determine whether images of abscesses have a fungal or bacterial origin.

To complete the answer to RQ1, the performance of the OD-Abs classifier incorporating the  $\beta$ -VAE encoder-decoder is evaluated on the datasets of images of fungi and

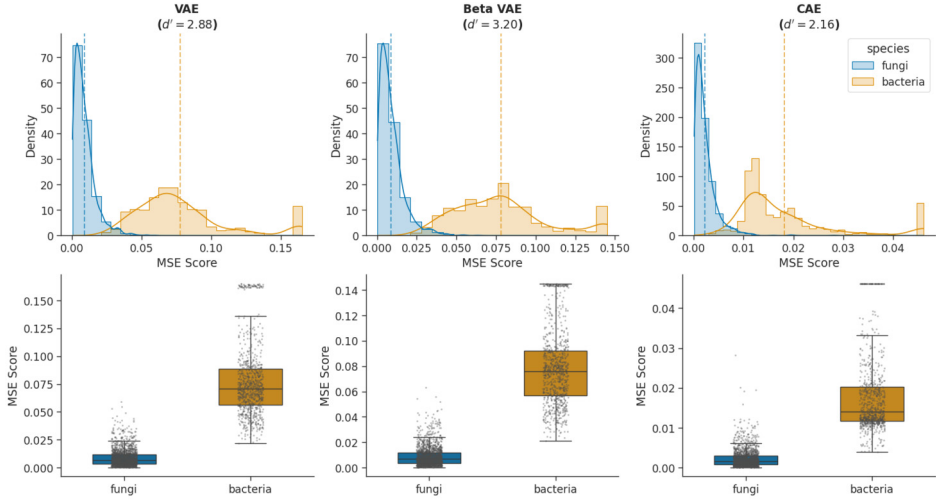


Fig. 3. Distribution of Mean Squared Error (MSE) scores for VAE,  $\beta$ -VAE, and CAE models evaluated on  $D_f$  (blue) and  $D_b$  (orange) datasets. The top row displays density histograms with vertical dashed lines indicating the mean MSE for each class. The bottom row presents corresponding box plots with overlaid data points. The Sensitivity Index ( $d'$ ) is reported for each model title. The plots visually confirm that  $\beta$ -VAE achieves the highest class separability ( $d' = 3.205$ ), showing the distinct gap between the reconstruction errors of the two species.

Table 3  
Experimental results. 95% CIs are used for the mean values of the performance metrics.

| Threshold $\tau_M$ | Prec $\uparrow$                     | Sens $\uparrow$                     | Spec $\uparrow$                     | F1 $\uparrow$                       | AUC $\uparrow$                      |
|--------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| <b>V1</b>          | $0.764 \pm 0.014$                   | <b><math>0.994 \pm 0.002</math></b> | $0.929 \pm 0.006$                   | $0.864 \pm 0.009$                   | <b><math>0.962 \pm 0.002</math></b> |
| <b>V2</b>          | <b><math>0.967 \pm 0.012</math></b> | $0.968 \pm 0.013$                   | <b><math>0.992 \pm 0.003</math></b> | <b><math>0.967 \pm 0.008</math></b> | <b><math>0.980 \pm 0.006</math></b> |

bacteria from abscesses described in Section 4.1. For the hyperparameter  $\tau_M$  (the decision threshold) both variants (V1, V2) described in Section 3.4 are used. Table 3 presents the experimental results obtained following the testing methodology introduced in Section 3.5. The values for the performance metrics are averaged over five runs, and the 95% CI of the mean values is reported. The best values for the performance metrics are highlighted.

Analysing the performance metric values from Table 3, one observes that, in terms of AUC, both variants for selecting the threshold  $\tau_M$  yield high performances with a clear advantage for V2. We consider that variant V2 provides a better balance between the *true positive rate* (TPRate, Sens) and *true negative rate* (TNRate, Spec). We note that while V1 provides a near-perfect TPRate, it exhibits a lower capability to recognize fungal images from abscesses (lower Spec) compared to V2, meaning more fungal images are incorrectly flagged as bacterial/anomalous. Moreover, compared to V1, V2 provides a much lower *false alarm ratio* (FAR, the proportion of images incorrectly predicted as “bacterial” from the total number of predicted bacterial images), indicated by its superior Prec value (0.967 vs 0.764). A low FAR is desirable, as it indicates that fewer fungal images

Table 4  
Comparison to baseline OCC methods. 95% CIs are reported for the results.

| Model            | Prec $\uparrow$          | Sens $\uparrow$          | Spec $\uparrow$          | F1 $\uparrow$            | AUC $\uparrow$           |
|------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <b>OD-Abs</b>    | <b>0.967</b> $\pm$ 0.012 | 0.968 $\pm$ 0.013        | <b>0.992</b> $\pm$ 0.003 | <b>0.967</b> $\pm$ 0.008 | <b>0.980</b> $\pm$ 0.006 |
| <b>OC-SVM</b>    | 0.786 $\pm$ 0.013        | 0.959 $\pm$ 0.008        | 0.940 $\pm$ 0.005        | 0.864 $\pm$ 0.009        | 0.950 $\pm$ 0.005        |
| <b>Deep-SVDD</b> | 0.749 $\pm$ 0.018        | <b>0.986</b> $\pm$ 0.009 | 0.924 $\pm$ 0.007        | 0.851 $\pm$ 0.014        | 0.955 $\pm$ 0.007        |

are incorrectly flagged as bacterial. This is relevant in practice, since fungal infections are generally less frequent than bacterial infections and should therefore not be overlooked.

#### 4.3.1. Comparison to Related Work

Since we did not identify any studies in the literature specifically addressing binary classification between fungal and bacterial microscopic images, we benchmark our solution against alternative one-class classification (OCC) methods that follow the same training paradigm (namely learning exclusively from fungal images while treating bacterial samples as anomalies), and the same testing methodology as for our OD-Abs classifier (Section 3.5). In particular, we evaluate a one-class support vector machine (OC-SVM) (Schölkopf *et al.*, 1999), using the `sklearn` library implementation. The images are pre-processed and flattened into vector representations before learning the decision boundary. In addition, we consider a deep one-class classification method based on Deep SVDD (Ruff *et al.*, 2018). To ensure a fair comparison, this model utilizes a pre-trained ResNet18 as the feature extractor, followed by a projection network. The model is trained to minimize the distance of fungal samples to the hypersphere centre in the latent space, with bacterial images expected to yield larger distances.

The results are presented in Table 4. We mention that the performance of our OD-Abs model is reported when using variant V2 for selecting the threshold  $\tau_M$  (the best variant according to the analysis from Section 4.3). To ensure an accurate comparison, both OC-SVM and Deep-VDD were tested using the testing methodology employed for our OD-Abs model (see Section 3.5). The results in the table represent average values for the performance metrics obtained during cross-validation and are presented together with their 95% CIs. The best results for the performance metrics are highlighted in the table.

From Table 4 one observes that the precision (Prec) of our OD-Abs is significantly higher (with more than 18%) than the precision of the baseline methods. Higher precision values for OD-Abs highlight a false alarm ratio lower than for the OCC classifiers from the literature, more specifically, a lower risk for misclassifying images of fungi from abscesses as being bacteria. Such low false alarm ratios are desirable, mainly for OCC classifiers. In addition, OD-Abs outperforms the baseline OCC models with more than 10% in terms of F1 and more than 2.5% in terms of AUC. However, the probability of detecting the “bacterial” class is slightly lower for our model, being surpassed by at most 1.8%. For testing the statistical significance of the improvement achieved by OD-Abs over the OC-SVM and Deep-SVDD approaches, the sample of performances provided by OD-Abs was tested against the sample of the baseline OCC models’ performance values using a one-tailed paired Wilcoxon signed-rank test (Statistics, 2026; Rosner *et al.*, 2006). A *p*-value

less than 0.01 was obtained, highlighting that OD-Abs achieves a statistically significant improvement over the OCC baseline models from Table 4, at a significance level 0.01.

Addressing RQ2, the comparative results shown in Table 4 confirm that the proposed OD-Abs framework brings a performance improvement over both the OC-SVM and Deep-SVDD baselines. While the existing methods exhibit high sensitivity, they also show significantly lower precision and specificity. In contrast, OD-Abs achieves the highest F1-score of 0.967 and an AUC of 0.980. Therefore, we can conclude that the proposed approach is the best solution in terms of precision-recall trade-off, resulting in a more accurate classification of bacteria and fungi.

#### 4.4. Discussion

To better highlight the good performance of our proposed OD-Abs one-class classification model, we performed an additional comparison to binary classification models. However, this comparison requires important contextualization regarding the learning paradigms employed. Obviously, it is expected that the binary classifiers achieve better predictive performance than the OCC ones, since the former are trained on images from both classes to learn the decision boundary between them, while the latter receive samples from only one class and learn a boundary surrounding the images from the reference class. However, the aim is to see how close or far the performance of OD-Abs is from the baseline binary classifiers.

For comparison, we selected the architectures most commonly applied to the related tasks discussed in Section 2.2, such as species-level classification, macroscopic binary classification, and multiclass microorganism classification. Consequently, we compare our approach with baseline CNN and Transformer-based architectures, specifically ResNet50 (He *et al.*, 2016), VGG16 (Simonyan and Zisserman, 2015), and a base Vision Transformer (ViT) (Dosovitskiy *et al.*, 2021). These models were initialized with pre-trained weights and fine-tuned for five epochs on our fungal and bacterial datasets.

The supervised architectures achieved near-perfect separation, with ResNet50 reaching an AUC of 1.0000, followed closely by ViT (0.9993) and VGG16 (0.9976), surpassing the OD-Abs model. Our OD-Abs one-class classifier is outperformed by the baseline binary classifiers with at most 3.3% in terms of *Prec* and *F1*, 3.2% in terms of *Sens*, and 2% in terms of *AUC*. We have to remark that in terms of *Spec*, OD-Abs outperformed VGG16 by 0.4%, and ViT by 3.8%, having nearly the same performance as ResNet50 (with only 0.8% lower). Thus, our model provides lower misclassification rates for the images of fungi from abscesses (i.e. false positive rates) than VGG16 and ViT, being able to better distinguish the “fungal” class from the “bacterial” class. In this context, we consider the performance achieved by our OD-Abs model notable.

It has to be noted that the baseline supervised classification models benefit from a significant advantage: supervised training, allowing them to explicitly learn the decision boundary between fungal and bacterial features. In contrast, OD-Abs operates in a one-class setting, training exclusively on fungal images without exposure to bacterial samples during training. The fact that OD-Abs achieves a high AUC and F1 scores and a performance very close to the performance of binary classifiers, even outperforming them in

terms of false positive rate, demonstrates the robustness of the one-class hypothesis. It suggests that the structural pattern of fungi is distinct enough that a model can learn to reject bacterial anomalies effectively without ever seeing them.

To conclude, the proposed OD-Abs model succeeds in ensuring a very good balance between preserving the structure of the images from the reference “fungal” class and uncovering the underlying structural relationships within them, and successfully recognizing the “bacterial” class as well. However, the gap in AUC between our custom model and the pre-trained supervised baselines implies that our current architecture may miss fine-grained texture representations.

## 5. Explainability

For answering RQ3, this section explores the explainability part of OD-Abs and analyses the interpretations provided by the reconstruction error map from a biological perspective.

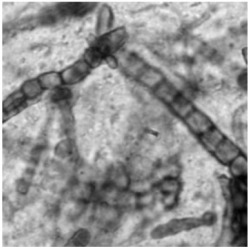
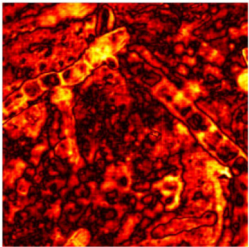
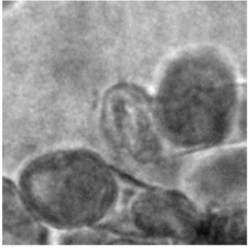
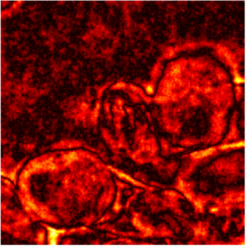
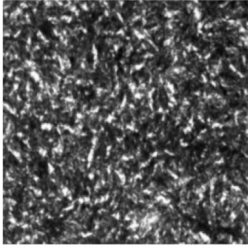
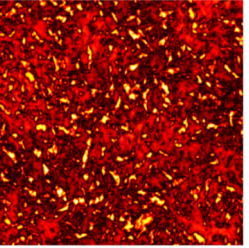
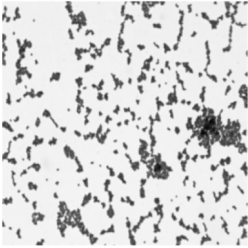
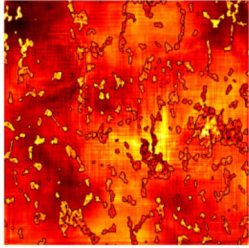
For a more detailed analysis of the OD-Abs classifier and its potential limitations, we selected four cases (samples): **S1** – an image of fungi from an abscess that was misclassified by OD-Abs; **S2** – an image of fungi from an abscess correctly classified by OD-Abs; **S3** – an image of bacteria from an abscess misclassified by OD-Abs; and **S4** – an image of bacteria from an abscess correctly classified by OD-Abs. For each of the samples, the goal is to understand the way the model inferred the output (predicted class) and to correlate these interpretations with the biological interpretation. Table 5 presents a description of the four samples: input image, correct class, description of the sample, class predicted by OD-Abs, and the reconstruction error map for the image.

The input images depicted in Table 5 represent micrographs of fungal and bacterial cells. The **S1** sample consists of fungal hyphae with branching filaments and septa, the **S2** sample consists of oval cells most often seen in yeast, the **S3** sample consists of a low magnification micrograph with bacterial cells hard to distinguish from the grids typically used for electron microscopy investigation, and the **S4** sample consists of irregular shapes of individual bacterial cells, which appear to be cocci or possibly small bacilli, clustered together.

The reconstruction error map represents the pixel-wise difference between the input image and its reconstruction and highlights regions where the model fails to accurately reproduce image structures. In this map, well-reconstructed areas appear in darker, cooler colours (e.g. black or blue), whereas regions that the model finds difficult to reconstruct appear in brighter, warmer colors. Therefore, in our case, anomalous structures (i.e. bacteria) are expected to manifest as yellow or white regions corresponding to high reconstruction error.

The reconstruction error maps shown in Table 5 provide insights into the challenges encountered by the model. The misclassified samples **S1** and **S3** exhibit more diffuse and widespread reconstruction errors, suggesting that the model struggled to capture meaningful structural patterns in these noisier images compared to the correctly classified cases. For **S2**, we can also notice that the high intensity spots seem to correspond with darker

Table 5  
Description of the four samples analysed together with the reconstruction error maps explanations.

| Sample | Input image                                                                         | Correct class | Predicted class | Reconstruction error map                                                             |
|--------|-------------------------------------------------------------------------------------|---------------|-----------------|--------------------------------------------------------------------------------------|
| S1     |    | FUNGAL        | BACTERIAL       |    |
| S2     |    | FUNGAL        | FUNGAL          |    |
| S3     |   | BACTERIAL     | FUNGAL          |   |
| S4     |  | BACTERIAL     | BACTERIAL       |  |

spots in the original image. In sample **S2**, although the image was correctly classified as fungal, several localized bright regions can be observed in the error map, which seem to correspond with regions of high-intensity (white) areas in the original image. For **S4**, we can observe many bright yellow spots corresponding to the other class (in our case, bacteria) and some larger, lighter-coloured areas in the background as well. Overall, these observations suggest that the model has greater difficulty with noisier images and with cases where microorganism boundaries are less sharp and not clearly delineated from the background.

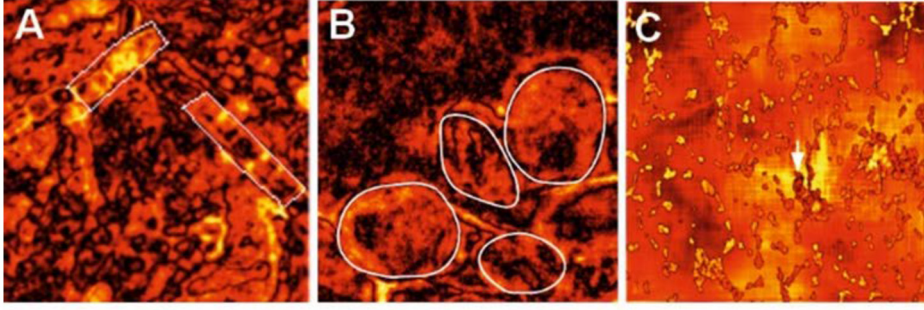


Fig. 4. Biological interpretation for the classification of samples **S1**, **S2**, and **S4** from Table 5. **A**: Incorrect bacterial identification of fungal hyphae in **S1** – marked in white. **B**: Correct identification of yeast cells, as representatives of fungi in **S2** – marked in white. **C**: Correct identification of bacterial cells, except for a small area with a misinterpretation in **S4** – marked with a white arrow.

From a biological perspective, sample **S1** is misclassified as bacterial cells, as the OD-Abs model seems to identify hyphal zones as representing bacterial cells. Figure 4A illustrates the regions (marked in white) that were misclassified as bacterial cells. The second misclassification of OD-Abs is sample **S3**. Sample **S3** appears to be an electron micrograph with cells difficult to distinguish from the uniform background provided by the carbon disk used for loading the samples into the microscope, due to low magnification. Samples **S2** and **S4** are correctly classified as belonging to the “fungal” and “bacterial” classes, respectively. In **S2**, the OD-Abs model correctly identifies oval cells characteristic of yeast (marked in white in Fig. 4B), as well as in **S4**, where it determines correctly irregular individual bacterial cells (except the small region marked with a white arrow in Fig. 4C where the identification is incorrect due to the presence of darker shade pixels).

## 6. Conclusions and Future Work

In conclusion, we presented OD-Abs, a one-class classifier that distinguishes between microscopic images of bacteria or fungi. The model, based on a  $\beta$ -VAE architecture, is trained on solely fungi images and is evaluated on both fungi and bacteria. We also discuss two methods of setting the threshold for the classification, as well as detail our choice of AE architecture.

The research questions can now be answered. In response to RQ1, our results demonstrate that one-class outlier detection can successfully uncover structural distinctions between images of fungi and bacteria from abscesses to a significant extent, as evidenced by an F1-score of 0.967 and an AUC of 0.980. These metrics indicate that the model effectively learned the hidden structural patterns unique to fungal morphology. It is also notable that the proposed OD-Abs (V2) achieves a specificity score of 0.992 and sensitivity of 0.968, therefore showing a lack of bias towards the reference class. This suggests that the structural separation is robust and that the model successfully differentiates fungi from bacteria. Regarding RQ2, our comparative analysis confirms that the proposed OD-Abs

framework yields a statistically significant improvement over OC-SVM and Deep-SVDD baselines. While existing methods demonstrate high sensitivity but lower precision and specificity, OD-Abs achieves a superior F1-score and AUC. To answer RQ3, the analysis of reconstruction residuals confirms that for correctly classified samples, the model's decisions are primarily driven by biologically meaningful features, such as oval cells and individual bacteria. However, the maps also reveal the model's limitations in handling images where cells are more difficult to distinguish from the background due to data collection, as well as its tendency to confuse certain structures, such as the misidentification of hyphal zones as bacteria. These observations demonstrate a strong alignment between the technical and biological perspectives, as the high-error regions in the residual maps correspond, for the most part, to recognizable morphological features and background noise.

To bridge the performance gap with fully supervised baselines, future iterations could expand the framework into a bidirectional system by training a parallel AE with bacteria as the reference class that would allow for the classification of fungi, bacteria, and an "other" class, moving toward a more universal tool for digital pathology. We also plan on utilizing the AE's latent features in a multitask learning setting, combining unsupervised representations with supervised classification heads to boost discriminative power. The experimental results also suggest further exploration of alternative OCC models instead of AE, such as ViT, together with a binary classifier, focusing on images where their predictions differ, to take advantage of the strengths of both models.

## Acknowledgements

This research was supported by the project "Romanian Hub for Artificial Intelligence – HRIA", Smart Growth, Digitization and Financial Instruments Program, 2021–2027, MySMIS no. 351416. The authors thank anonymous reviewers for their useful suggestions and comments that helped to improve the article and the presentation.

## References

- Ahmed, S.I., Haque, A.O. (2023). Microscopic fungi classification using vision transformer guided by transfer learning approach. In: *2023 26th International Conference on Computer and Information Technology (ICCIT)*. IEEE, pp. 2275–2280.
- Alam, S., Sonbhadra, S.K., Agarwal, S., Nagabhushan, P. (2020). One-class support vector classifiers: a survey. *Knowledge-Based Systems*, 196, 105754.
- Bhimavarapu, J., Chinta, A., Movva, S.V., Jampani, J.P. (2024). Fungi classification: enhancing diagnosis using deep learning. In: *2024 2nd World Conference on Communication & Computing (WCONF)*, IEEE, pp. 1–6.
- Brown, L.D., Cat, T.T., DasGupta, A. (2001). Interval estimation for a proportion. *Statistical Science*, 16, 101–133.
- Dey, S., Smarandache, F., Debbarma, R., Majumder, P. (2022). A hybrid IF-FUCOM-GRA approach and its application to determine optimal bacterial concentrations on mortar at optimal curing day. *Informatica*, 34(2), 223–248.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N. (2021). An image is Worth 16 x 16 words: transformers

- for image recognition at scale. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3–7, 2021*, pp. 611–631.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Filatovas, E., Stripinis, L., Orts, F., Paulavičius, R. (2024). Advancing research reproducibility in machine learning through blockchain technology. *Informatica*, 35(2), 227–253.
- Gu, Q., Zhu, L., Cai, Z. (2009). Evaluation measures of the classification performance of imbalanced data sets. In: *International Symposium on Intelligence Computation and Applications (ISICA)*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 461–471.
- Gümiş, A. (2024). Classification of microscopic fungi images using vision transformers for enhanced detection of fungal infections. *Türk Doğa ve Fen Dergisi*, 13(1), 152–160.
- He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A. (2017). beta-VAE: Learning basic visual concepts with a constrained variational framework. In: *International Conference on Learning Representations (Poster)*, pp. 60–81.
- Kingma, D.P., Welling, M. (2014). Auto-Encoding Variational Bayes. In: Bengio, Y., LeCun, Y. (Eds.), *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14–16, 2014, Conference Track Proceedings*, pp. 1–14.
- Mehta, N., Marietta, M., Copelin II, E. (2025). *Intraabominal Abscesses*. StatPearls.
- Mohamed, H., Nagib, A.E., Hany, M. (2024). Comparative study of microscopic fungii classification using transfer learning models. In: *2024 Intelligent Methods, Systems, and Applications (IMSA)*, IEEE, pp. 87–92.
- Mueller-Mang, C., Castillo, M., Mang, T., Cartes-Zumelzu, F., Weber, M., Thurnher, M.M. (2007). Fungal versus bacterial brain abscesses: is diffusion-weighted MR imaging a useful tool in the differential diagnosis? *Neuroradiology*, 49, 651–657.
- Nawarathne, U., Kumari, H. (2023). Classification of fungi images using different convolutional neural networks. In: *2023 8th International Conference on Information Technology Research (ICITR)*, IEEE, pp. 1–6.
- Poomrittigul, S., Chomkwah, W., Tanpatanan, T., Sakorntanant, S., Treebupachatsakul, T. (2022). A comparison of deep learning CNN architecture models for classifying bacteria. In: *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, IEEE, pp. 290–293.
- Prajna, N., Assaf, J., Acharya, N.R., Rose-Nussbaumer, J., Lietman, T.M., Campbell, J.P., Keenan, J.D., Song, X., Redd, T.K. (2025). Multimodal deep learning for differentiating bacterial and fungal keratitis using prospective representative data. *Ophthalmology Science*, 5(2), 100665.
- Prasad, K., Mukhopadhyay, C., Banerjee, B., et al. (2025). Vision transformer based bacteria classification model for Gram-stained direct smear images. *Multimedia Tools and Applications*, 84, 20289–20309.
- Prommakhot, A., Srinonchat, J. (2024). Combining convolutional neural networks for fungi classification. *IEEE Access*, 12, 58021–58030.
- Rahman, A., Gain, M., Rahman, M.S., Saha, S.K., Adhikary, A., Bairagi, A.K., Raha, A.D. (2025). Enhancing microscopic fungi classification through advanced pretraining in deep learning models. In: *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, pp. 1–6.
- Rahman, M.A., Clinch, M., Reynolds, J., Dangott, B., Villegas, D.M.M., Nassar, A., Hata, D.J., Akkus, Z. (2023). Classification of fungal genera from microscopic images using artificial intelligence. *Journal of Pathology Informatics*, 14, 100314.
- Rawson, T.M., Antcliffe, D.B., Wilson, R.C., Abdolrasouli, A., Moore, L.S.P. (2024). Management of bacterial and fungal infections in the ICU: diagnosis, treatment, and prevention recommendations. *Infection and Drug Resistance*, 4(16), 2709–2726.
- Redd, T.K., Prajna, N.V., Srinivasan, M., Lalitha, P., Krishnan, T., Rajaraman, R., Venugopal, A., Acharya, N., Seitzman, G.D., Lietman, T.M., Keenan, J.D., Campbell, J.P., Song, X. (2022). Image-based differentiation of bacterial and fungal keratitis using deep convolutional neural networks. *Ophthalmology Science*, 2(2), 100119.
- Rosner, B., Glynn, R.J., Lee, M.-L.T. (2006). The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics*, 62(1), 185–192.
- Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M. (2018). Deep one-class classification. In: *Proceedings of the 35th International Conference on Machine Learning*, pp. 4393–4402.
- Satvati, M.A., Lakestani, M., Khamnei, H.J., Allahviranloo, T. (2024). Deblurring medical images using a new Grünwald-Letnikov fractional mask. *Informatica*, 35(4), 817–836.

- Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J. (1999). Support vector method for novelty detection. In: *Proceedings of the 13th International Conference on Neural Information Processing Systems, NIPS'99*. MIT Press, Cambridge, MA, USA, pp. 582–588.
- Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In: Bengio, Y., LeCun, Y. (Eds.), *3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings*, pp. 1–14.
- Sopo, C.J.P., Hajati, F., Gheisari, S. (2021). DeFungi: Direct Mycological Examination of Microscopic Fungi Images. [arXiv:2109.07322](https://arxiv.org/abs/2109.07322).
- Spahn, C., Gómez-de-Mariscal, E., Laine, R.F., Pereira, P.M., von Chamier, L., Conduit, M., Pinho, M.G., Jacquemet, G., Holden, S., Heilemann, M., Henriques, R. (2022). DeepBacs for multi-task bacterial image analysis using open-source deep learning approaches. *Communications Biology*, 5(1), 688.
- Sroka-Oleksiak, A., Pardyl, A., Rymarczyk, D., Olechowska-Jarząb, A., Biegun-Drożdż, K., Ochońska, D., Wronka, M., Borowa, A., Gosiewski, T., Adamczyk, M., Telega, H., Zieliński, B., Brzychczy-Włoch, M. (2025). AI-driven rapid identification of bacterial and fungal pathogens in blood smears of septic patients. *Computers in Biology and Medicine*, 199, 111328.
- Statistics (2026). Wilcoxon Signed-Rank Test. <http://www.socscistatistics.com/tests/>. Last accessed 15 January 2026.
- Toader, T.-A. (2025). OD-Abs outlier detector. GitHub. Last accessed 20 December 2025. <https://github.com/teo-toader/OD-Abs>.
- Treebupachatsakul, T., Poomrittigul, S. (2020). Microorganism image recognition based on deep learning application. In: *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, IEEE, pp. 1–5.
- Visitsattapongse, S., Bunkum, M., Pintavirooj, C., Paing, M.P. (2024). A deep learning model for bacterial classification using big transfer (BiT). *IEEE Access*, 12, 15609–15621.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
- Won, Y.K., Lee, H., Kim, Y., Han, G., Chung, T.-Y., Ro, Y.M., Lim, D.H. (2023). Deep learning-based classification system of bacterial keratitis and fungal keratitis using anterior segment images. *Frontiers in Medicine*, 10, 1162124.
- Zawadzki, P. (2020). Deep learning approach to the classification of selected fungi and bacteria. In: *2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE)*, IEEE, pp. 1–4.
- Zieliński, B., Plichta, A., Misztal, K., Spurek, P., Brzychczy-Włoch, M., Ochońska, D. (2017). Deep learning approach to bacterial colony classification. *PloS One*, 12(9), 0184554.

**G. Czibula** is a professor at the Computer Science Department, Faculty of Mathematics and Computer Science of the Babeş-Bolyai University, Romania. She published more than 220 papers in prestigious journals and conference proceedings. Her research interests include machine learning, distributed artificial intelligence and multiagent systems, and bioinformatics.

**T.-A. Toader** is currently a third-year PhD student at the Doctoral School in Mathematics and Computer Science of the Babeş-Bolyai University, in Cluj-Napoca, Romania. Her research interests include computer vision and deep learning.

**C. Mircea** is an assistant professor at the Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeş-Bolyai University, Romania. She has expertise in isolation, identification, and characterization of bacteria and fungi from a wide range of environments.