RESEARCH ARTICLE

Deep learning approach to describe and classify fungi microscopic images

Bartosz Zieliński[1,2]*, **Agnieszka Sroka-Oleksiak**[3,4], **Dawid Rymarczyk**[1,2],
Adam Piekarczyk[1], **Monika Brzychczy-Włoch**[4]

1 Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland, **2** Ardigen, Kraków, Poland, **3** Department of Mycology, Chair of Microbiology, Faculty of Medicine, Jagiellonian University Medical College, Kraków, Poland, **4** Department of Molecular Medical Microbiology, Chair of Microbiology, Faculty of Medicine, Jagiellonian University Medical College, Kraków, Poland

* bartosz.zielinski@uj.edu.pl

Abstract

Preliminary diagnosis of fungal infections can rely on microscopic examination. However, in many cases, it does not allow unambiguous identification of the species due to their visual similarity. Therefore, it is usually necessary to use additional biochemical tests. That involves additional costs and extends the identification process up to 10 days. Such a delay in the implementation of targeted therapy may be grave in consequence as the mortality rate for immunosuppressed patients is high. In this paper, we apply a machine learning approach based on deep neural networks and bag-of-words to classify microscopic images of various fungi species. Our approach makes the last stage of biochemical identification redundant, shortening the identification process by 2-3 days, and reducing the cost of the diagnosis.

Introduction

Yeast and yeast-like fungi are a component of natural human microbiota [1]. However, as opportunistic pathogens, they can cause surface and systemic infections [2]. The leading causes of the fungal infections are impaired function of the immune system and imbalanced microbiota composition in the human body. Other factors of fungal infections include steroid treatment, invasive medical procedures, and long-term antibiotic treatment with a broad spectrum of antimicrobial agents [3–5].

The standard procedure in mycological diagnostics begins with collecting various types of test materials like swabs, scraps of skin lesions, urine, blood, or cerebrospinal fluid. Next, the clinical materials (marked as **B** in Fig 1) are directly cultured on special media, while the blood and cerebrospinal fluid samples (marked as **A** in Fig 1) require prior cultivation in automated closed systems for additional 2-3 days. Material incubates under specific temperature conditions (usually for 2-4 days in case of yeast-like fungi). The initial identification of fungi bases on the assessment of the cells' shapes observed under the microscope as well as the growth rate, type, shape, color, and the smell of the colonies. Such analysis allows the assignment to fungi type; however, identification of the species is usually impossible due to the significant
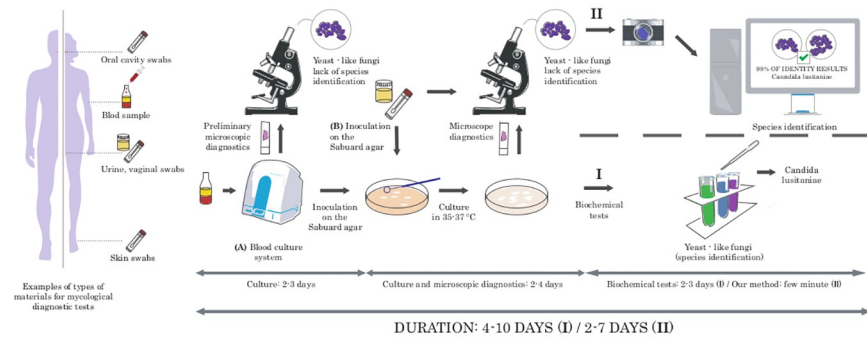
Fig 1. Standard and computer-aided mycological diagnosis. Standard mycological diagnostics (I) require analysis
with biochemical tests. As a result, the entire diagnostic process can last 4-10 days. In our computer-aided approach
(II), biochemical tests are replaced with a machine learning approach that predicts fungi species based only on
microscopic images. It shortens the diagnosis by 2-3 days.

https://doi.org/10.1371/journal.pone.0234806.g001

similarity between them. Because of that, further analysis consisting of biochemical tests, is
necessary. As a result, the entire diagnostic process from the moment of culture to species
identification can last 4-10 days (see Fig 1).

In this paper, we apply a machine learning approach based on deep neural networks and
bag-of-words approaches to classify microscopic images of various fungus species. As a result,
the last stage of biochemical identification is unnecessary, which shortens the identification
process by 2-3 days and reduces the cost of diagnosis. It allows accelerating the decision about
the introduction of an appropriate antifungal drug, which prevents the progression of the dis-
ease and shortens the time of patient recovery.

According to our best knowledge, there are no other methods for classifying fungi species
based only on microscopic images. Existing methods involve techniques such as morphologi-
cal identification of a type of fungi [6], fluorescence in situ hybridization (FISH) [7], biochemi-
cal techniques, molecular approaches, such as PCR [8], and sequencing [9]. However, all of
them are costly. On the other hand, our method bases on basic microbiological staining
(Gram staining) and a simple microscope equipped with a camera, and takes only a few min-
utes, which makes it easily applicable in many laboratories.

The paper is structured as follows. First, we introduce a fungus database and describe a clas-
sification method based on deep neural networks and bag-of-words methods. Then, we pres-
ent experimental setup, results, and conclusion.

Materials and methods

Materials

One of the most common fungal infections is candidiasis [5], mainly caused by *Candida albi-
cans* (50-70% of cases) [10]. Other species responsible for the diseases are *Candida glabrata* [2,
3], *Candida tropicalis* [4], *Candida krusei* [11], and *Candida parapsilosis* [3, 4]. In high-risk
patients, severe infections can also be caused by *Cryptococcus neoformans* [12] and *Saccharo-
myces phylum* [13]. Taking those facts into consideration, we prepared database, which con-
sists of five yeast-like fungal strains: *Candida albicans* ATCC 10231 (CA), *Candida glabrata*
ATCC 15545 (CG), *Candida tropicalis* ATCC 1369 (CT), *Candida parapsilosis* ATCC 34136
(CP), and *Candida lustianiae* ATCC 42720 (CL); two yeast strains: *Saccharomyces cerevisae*
ATCC 4098 (SC) and *Saccharomyces boulardii* ATCC 74012 (SB); and two strains belonging to
the Basidiomycetes: *Maalasezia furfur* ATCC 14521 (MF) and *Cryptococcus neoformans*
ATCC 204092 (CN). All strains are from the American Type Culture Collection. The species
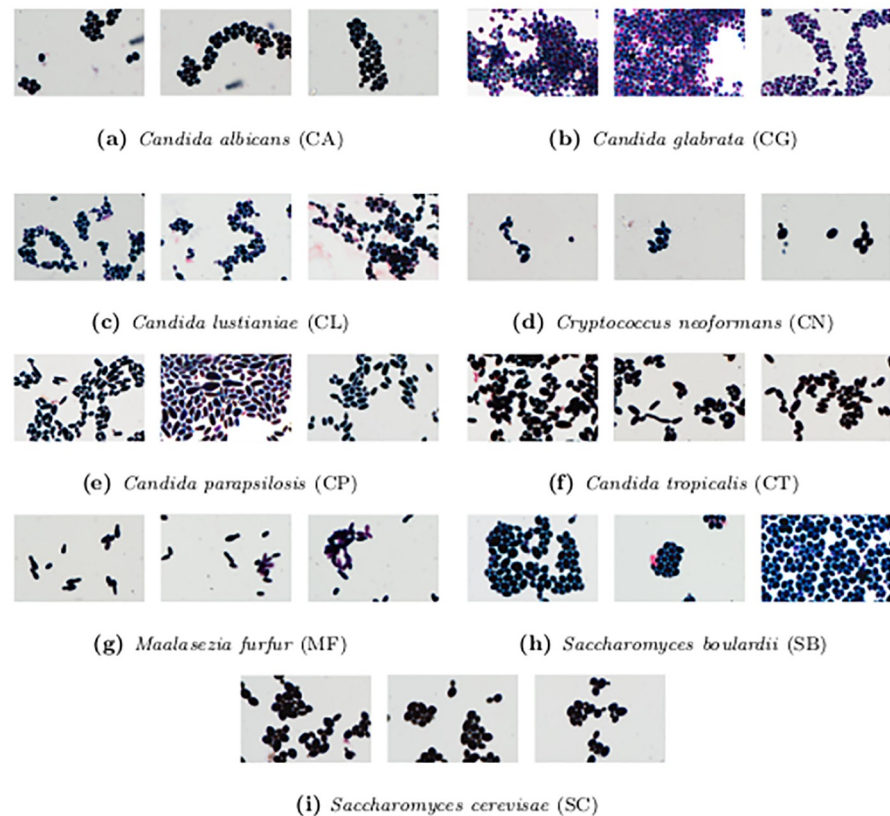
Fig 2. Sample images from DIFaS database. Three random images for each of the strains from DIFaS database.

in our database highly overlap with the most common fungal infections; however, they are not identical due to the limitations of our repository.

The strains were cultured on Sabouraud agar at 37°C for 48h (together with olive oil in the case of Maalaseizia furfur). After this time, microscopic preparations were made (2 preparations for each fungal strain) and stained with Gram method. Images were taken using an Olympus BX43 microscope with 100 times a super-apochromatic objective under oil-immersion. The photographic documentation was then produced with an Olympus BP74 camera and CellSense software (Olympus).

Altogether, our Digital Images of Fungus Species database (DIFaS) contains 180 images (9 strains × 2 preparations × 10 images) of resolution 3600 × 5760 × 3 with 16-bits intensity range in every pixel. In Fig 2, we present three random thumbnails for each of the registered strains.

Method

Deep Neural Networks (DNN) have shown human-level performance in case of large amounts of training data; however, they are limited when it comes to the application on small datasets due to the large numbers of parameters. Therefore, in this work, we consider two types of domain adaptation, both based on DNN features initially pre-trained on a different task (i.e., instance classification [14]). As a baseline method, we fine-tune the classifier's block of the well-known network architectures, i.e., AlexNet [15], DenseNet169 [16], InceptionV3 [17], and ResNet [18] (with frozen features' block). As we present in results, such architectures are
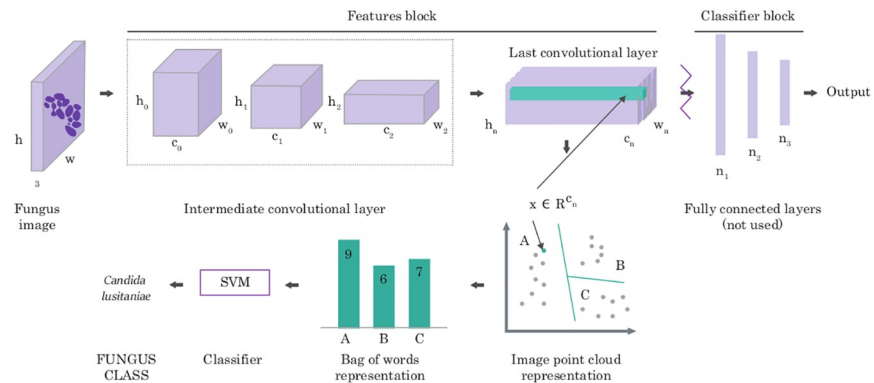
Fig 3. Deep bag-of-words algorithm for mycological diagnosis. The multi-step algorithm produces robust image features using previously trained deep neural network, aggregates them using one of the bag-of-words approaches, and classifies them with Support Vector Machine.

not optimal due to the small training set (see S1 and S3 Tables). Hence, we propose to apply the deep bag-of-words multi-step algorithm shown in Fig 3. In contrast to baseline methods, which utilize "shallow" Neural Network to previously calculated features, our strategies aggregate those features using one of the bag-of-words approaches and then classify them with Support Vector Machine (SVM). Such a policy, previously applied to texture recognition [19] and bacteria colony classification [20], is more accurate than the baseline methods; however, it is not well known. Therefore, to make this paper self-contained, below, we describe its successive steps.

To generate robust **image representation**, AlexNet [15], InceptionV3 [17] or ResNet [18] pre-trained on ImageNet [14] database are used. Another option would be to use conventional handcrafted descriptors (like ORB [21] or DSIFT [22]); however, they are usually outperformed by deep features. Considered network architectures consist of two parts: convolutional layers, which are responsible for extracting image features (so-called features' block), and fully connected layers, which are responsible for the classification (so-called classifier's block). Classifier's block cannot be directly used because it was trained for other types of images; however, features' block encodes more general, reusable information. Therefore, removing the classifier block from the network and preserving convolutional layers allows us to generate robust image features. In the case of AlexNet, we obtain a set of points in 256-dimensional space, whose number depends on the input image's resolution (e.g., in case of resolution 500×500 pixels, 169 points ($13 \cdot 13$) are generated).

Since the classified patches are always of the same size, their features' blocks could be used directly by the classifier. It, however, would lead to vast data dimensionality (i.e., the feature vector of size 43264), which according to our experiments, results in the lack of generalization, primarily due to the relatively small size of the training set (100 images). Therefore, to obtain a more reliable representation of patches, we **pool** the acquired set of points using Bag of Words, BoW [23, 24], or its more expressive modification called Fisher Vector, FV [25]. The idea behind both of them is to aggregate a set of points (representing the patch) with a so-called codebook. The codebook is usually generated from the subset of training data in an unsupervised manner using a clustering algorithm (e.g., k-Means or Expectation Maximization [26]). Given a codebook, the set of 256-dimensional points obtained with AlexNet for a particular image is encoded by assigning points to the nearest codeword. In traditional Bag of Words, this encoding leads to a codeword histogram, i.e., a histogram for which each

codeword contains points closest to this codeword. In the case of the Fisher Vector, the clusters are replaced with a Gaussian Mixture Model (GMM), and the representation encodes the log-likelihood gradients with respect to the parameters of this model. In this paper, we will use notations *deep Bag of Words* and *deep Fisher Vector* to refer to those two types of pooling methods. To make this article self-contained, we recall definitions of BoW and FV in S1 Appendix.

As a result of pooling, one fixed-size vector is obtained for each of the analyzed patches, which can be **classified** with any machine learning methods to distinguish between various fungus species. We decided to use Support Vector Machine and Random Forest classifiers for this step.

Experimental setup and results

For the experiments, we split our DIFaS database (9 strains × 2 preparations × 10 images) into two subsets, so that both of them contain images of all strains, but from different preparation. It is because each preparation has its characteristics, and according to our previous studies [20], using images from the same preparation both in training and test set can result in over-stated accuracy. As an example, let us consider the background-size, which depends on the size of the colony moved by inoculation loop from Sabouraud agar to preparation. Because there are only two preparations for each species in the dataset, the classifier could end up learning clinically irrelevant background-size instead of relevant fungus features. Therefore, images from particular preparation should not be shared between training and test set. We decided to use 2-fold cross-validation (one fold with 90 images from the first preparations and the second fold with 90 images from the second preparations). Moreover, we decided to classify patches instead of the whole image (see Image preprocessing for details) and introduce additional class corresponding to the background (BG) to compensate for the preparation characteristic on the final result. Nevertheless, we report accuracy for both patch-level and scan-level classification (the latter with majority voting).

For each fold, we optimize the following parameters using internal 5-fold cross-validation: number of clusters in BoW $\in$ [5, 10, 20, 50, 100, 200, 500]; number of clusters in FV $\in$ [5, 10, 20, 50]; SVM kernel $\in$ [*linear*, *RBF*]; SVM $C \in$ [1, 10, 100, 1000]; SVM $\gamma \in$ [0.001, 0.0001]. As the evaluation metric for grid search optimization, we use the accuracy classification score. Best results were obtained for FV with 10 clusters and SVM with *RBF* kernel, $C = 1$, and $\gamma = 0.0001$.

We performed all the experiments on a workstation with one 12 GB GPU and 256 GB RAM. On average, feature extraction, pooling, and classification take from 1 to 2 hours when training deep Fisher Vector. Such performance was possible thanks to the adaptation of the VLFeat library [27]. For comparison, the fine-tuning of the well-known architectures takes from 70 to 85 hours (see Table 1). Processing time in case of baseline methods was measured by multiplying the average time of an epoch by the number of epochs till the early stopping (i.e., the increase in validation loss). In the case of deep bag-of-words approaches, processing time was computed as a sum of all three steps of the algorithm (i.e., obtaining image representation, pooling, and classification).

The remaining part of this section is structured as follows. First, we describe image preprocessing, including contrast stretching and background removal. Then, we describe the results obtained for patch-based classification using deep bag-of-words approaches and compare them with the well-known network architectures. To explain the outcomes of deep BoW, we introduce an in-depth explanatory analysis of the obtained codebooks together with the microbiological feedback. We continue this investigation for a deep FV approach. Finally, we present

Table 1. Test accuracy of patch-based classification averaged over two runs (for two subsets described in Experimental setup).

Method	CA	CG	CL	CN	CP	CT	MF	SB	SC	BG	Total	Training time (s)
AlexNet	78.6 ± 4.3	**80.0 ± 1.4**	55.8 ± 1.4	63.4 ± 14.7	75.0 ± 7.9	35.0 ± 6.4	71.6 ± 16.6	67.9 ± 5.0	72.1 ± 2.1	90.0 ± 0.1	71.6 ± 2.4	250600
DenseNet169	67.9 ± 17.9	72.1 ± 0.7	53.6 ± 0.7	56.3 ± 13.3	60.0 ± 12.7	81.4 ± 0.7	68.7 ± 5.6	85.0 ± 5.0	68.8 ± 8.6	81.2 ± 2.0	72.9 ± 0.6	271600
InceptionV3	67.3 ± 1.5	55.0 ± 2.2	59.3 ± 6.4	67.0 ± 18.2	64.3 ± 1.4	81.4 ± 2.9	61.1 ± 11.1	55.0 ± 0.7	89.3 ± 1.3	84.7 ± 3.1	69.9 ± 1.9	309400
ResNet18	91.4 ± 3.6	67.1 ± 5.7	67.1 ± 5.7	61.9 ± 7.5	65.0 ± 0.7	69.8 ± 2.8	54.8 ± 12.8	93.6 ± 3.2	**93.5 ± 1.2**	93.1 ± 0.5	75.9 ± 2.6	263900
ResNet50	86.4 ± 3.5	57.9 ± 2.1	89.3 ± 3.5	61.8 ± 11.5	60.0 ± 7.1	64.5 ± 6.0	66.4 ± 19.3	79.3 ± 0.7	64.3 ± 12.4	90.4 ± 2.7	73.9 ± 2.6	276500
AlexNet BoW RF	87.3 ± 9.7	39.0 ± 0.3	88.3 ± 7.0	64.0 ± 17.0	79.0 ± 11.0	79.3 ± 10.7	55.1 ± 3.3	93.3 ± 2.0	81.3 ± 6.0	92.6 ± 1.0	76.7 ± 1.0	156
InceptionV3 BoW RF	44.3 ± 7.7	51.0 ± 13.7	60.3 ± 0.3	34.2 ± 17.8	28.0 ± 0.7	40.7 ± 11.3	13.3 ± 1.4	39.7 ± 11.7	36.3 ± 8.3	83.5 ± 2.0	44.6 ± 0.3	155
ResNet18 BoW RF	54.3 ± 2.3	39.7 ± 8.3	80.7 ± 12.7	46.2 ± 20.1	72.0 ± 3.3	61.3 ± 11.3	42.1 ± 2.9	58.7 ± 8.7	68.0 ± 4.0	91.6 ± 2.0	62.7 ± 2.4	158
AlexNet BoW SVM	92.3 ± 3.0	44.3 ± 18.3	88.7 ± 9.3	62.2 ± 22.2	**83.0 ± 5.7**	71.0 ± 15.0	76.1 ± 11.8	85.0 ± 4.3	76.3 ± 7.0	89.5 ± 3.6	77.6 ± 1.2	5124
InceptionV3 BoW SVM	44.3 ± 10.3	32.3 ± 5.7	57.7 ± 2.3	33.3 ± 17.6	28.3 ± 1.0	39.7 ± 6.3	13.0 ± 1.7	34.7 ± 8.0	32.0 ± 2.0	78.8 ± 2.1	40.8 ± 1.3	5153
ResNet18 BoW SVM	59.0 ± 3.7	32.3 ± 5.7	84.0 ± 11.3	54.7 ± 22.0	66.0 ± 0.0	62.0 ± 4.0	39.7 ± 1.0	59.3 ± 2.0	73.0 ± 12.3	92.9 ± 1.8	63.4 ± 0.7	5195
AlexNet FV RF	83.3 ± 10.7	54.3 ± 31.7	81.7 ± 0.3	49.3 ± 32.0	78.0 ± 12.0	73.0 ± 15.7	76.5 ± 5.8	89.3 ± 2.0	74.7 ± 6.7	88.0 ± 0.9	75.8 ± 0.4	166
InceptionV3 FV RF	40.7 ± 9.3	53.3 ± 11.3	60.3 ± 5.0	37.3 ± 21.7	27.7 ± 6.3	47.7 ± 13.0	16.5 ± 0.2	35.3 ± 13.3	32.0 ± 10.7	81.4 ± 4.1	44.6 ± 0.7	168
ResNet18 FV RF	63.7 ± 2.3	37.0 ± 5.7	78.7 ± 12.7	52.5 ± 21.1	73.3 ± 4.0	64.3 ± 5.0	61.8 ± 6.8	61.0 ± 9.7	69.3 ± 7.3	92.7 ± 2.0	66.4 ± 2.1	167
AlexNet FV SVM	**93.7 ± 2.3**	53.7 ± 16.3	**90.7 ± 4.0**	59.6 ± 15.2	77.7 ± 14.3	**87.7 ± 9.7**	**82.8 ± 6.9**	**97.3 ± 1.3**	81.3 ± 10.0	91.1 ± 2.5	**82.4 ± 0.2**	1541
InceptionV3 FV SVM	46.0 ± 14.0	45.0 ± 20.3	58.3 ± 3.0	42.2 ± 20.4	24.0 ± 5.3	43.7 ± 3.7	13.0 ± 1.1	26.7 ± 5.0	76.7 ± 3.7	41.3 ± 1.9	41.3 ± 1.9	1511
ResNet18 FV SVM	71.3 ± 11.3	35.3 ± 1.3	59.3 ± 10.3	51.6 ± 31.1	77.0 ± 8.3	76.7 ± 4.7	57.5 ± 0.5	73.3 ± 4.7	77.7 ± 9.7	**94.5 ± 1.3**	71.3 ± 1.5	1535

https://doi.org/10.1371/journal.pone.0234806.t001

results obtained for scan-based classification, computed by aggregating patch-based scores. The code implemented in Python with PyTorch library is available at https://github.com/bziiuj/fungus.

Image preprocessing

DIFaS database contains 180 images of relatively high resolution and intensity range (from 0 to 65535); however, the actual pixel values are usually between 0 and 1000 (see Fig 4a). Therefore, in the first step of preprocessing, we compute the lower and upper-intensity limits (separately for every image) and use them for contrast stretching (see Fig 4b). Moreover, images are scaled to the range [0, 1].

To overcome the issues with preparation characteristic (e.g., background-size), as the second step of preprocessing, we extract and classify only image patches with the reasonable foreground to background proportions (FBP), so patches with a rational number of foreground pixels. To obtain foreground-background segmentation on the pixel level, we apply thresholding (with threshold equal 0.5) to a grayscaled and blurred version of the scanned image. Such a simple segmentation is sufficient and works for all the images from the dataset (see S1 and S2 Figs) because the background is always much brighter than the areas with fungi cells. We tested three possible options of FBP: 1: 2, 1: 1, and 2: 1 (see Fig 4d–4f). Based on empirical
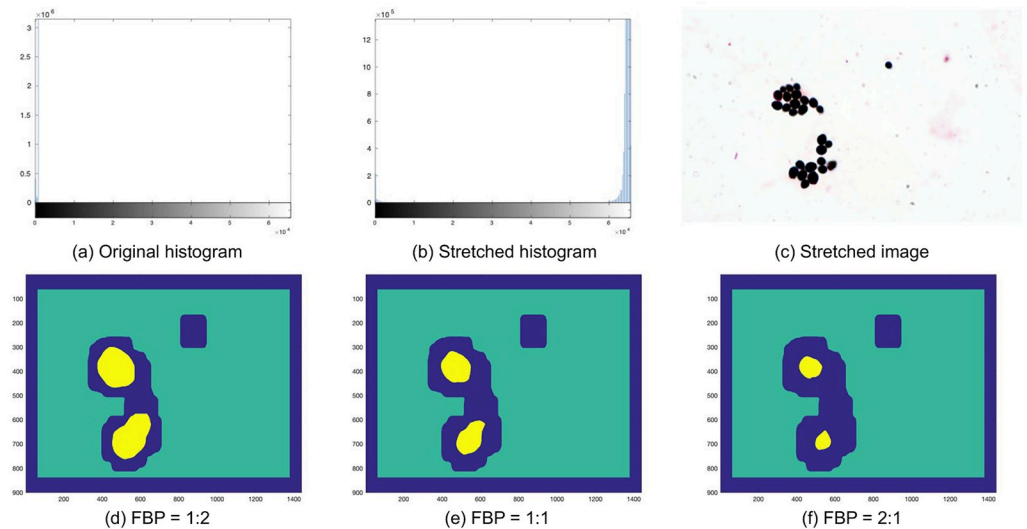
Fig 4. Example image, its histogram, and foreground-background mask. Original (a) and stretched (b) histogram of the 16-bits image; the stretched image itself (c); and its foreground-background masks with the various foreground to background proportions (d-f). Center locations of foreground and background patches are marked as yellow and green, respectively, while blue color corresponds to the area between them (omitted during classification).

https://doi.org/10.1371/journal.pone.0234806.g004

studies, we decided to use FBP equal 2: 1, which gains around 1.5% comparing to the other options. As a result, we obtain rough segmentations with approximated locations of foreground patches (those with FBP greater than 2: 1) and background patches (those with FBP smaller than 1: 100). Additionally, we experimented with two image scales: the original images and images scaled by factor 0.5 (with bicubic interpolation), concluding that the latter gains around 4% comparing to the former.

Patch-based classification

In this experiment, we use baseline models (well-known network architectures) as well as deep Bag of Words and deep Fisher Vector models to classify each patch of the image separately. As baseline models, we fine-tune the classifier's block of the well-known network architectures, such as AlexNet [15], DenseNet169 [16], InceptionV3 [17], and ResNet [18] for 100 epochs (with frozen features' block). Every baseline model was previously pre-trained on the ImageNet database [14]. Before running all the experiments, we experimentally chose the optimal FBP (2: 1), patch size (500 × 500 pixels), and image scale (0.5) using grid search optimization. The number of foreground patches overlapped by less than 50% oscillates between 2000 and 3000, depending on the fold. Moreover, the number of patches significantly varies depending on the strains (see S1 and S3 Tables). Therefore, we apply data augmentation (rotations, mirror reflection, and random noise) for better regularization.

The overall comparison of tested methods is presented in Table 1. One can observe that deep Fisher Vector works better than all the other techniques, including deep Bag of Words. However, its accuracy drops dramatically in the case of *Candida glabrata* (CG) and *Cryptococcus neoformans* (CN). In the case of CN it is most probably caused by a reduced number of samples, while in the case of CG due to its more substantial variance in the arrangement, appearance, and quantity (especially between two preparations, see S1 and S2 Figs). Moreover, CG images are hard to classify due to partial discoloration (pink color instead of purple) and vast overlapping of cells. As a result, CG is often classified as *Candida lustianiae* (CL)
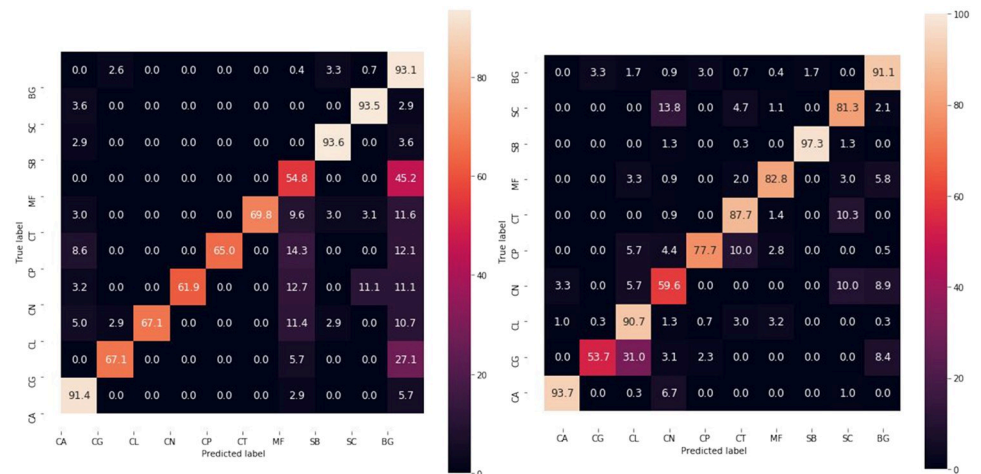
Fig 5. Test confusion matrices. Normalized test confusion matrices for the best baseline (ResNet18) and the best deep bag-of-words approach (deep Fisher Vector).

belonging to the same genus (see confusion matrix in Fig 5b). However, the classification error should decrease if the biological material of microscopic preparation has the smallest possible density with separated cells, as overlapping is the leading cause of blurriness.

To further understand the reason for incorrect classification, we prepare a qualitative confusion matrix for deep Fisher Vector to show examples of correctly and incorrectly classified patches (see Fig 6). We observe a high morphological similarity between misclassified species belonging to genus *Candida*, *Cryptococcus*, and *Saccharomyces*, especially if the preparation
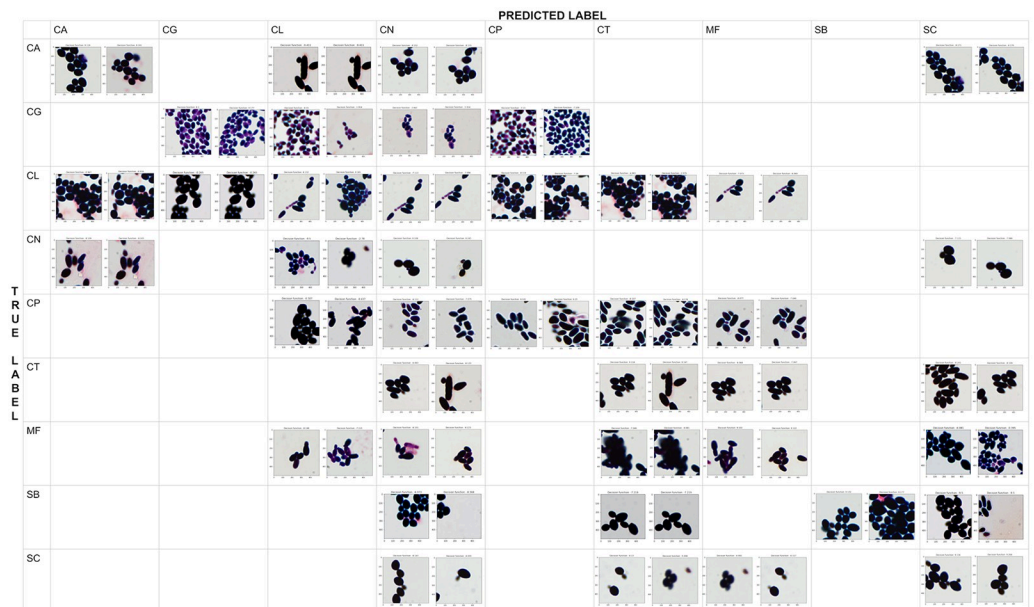

Fig 6. The qualitative confusion matrix for deep Fisher Vector (with AlexNet and SVM). Each cell contains at most two patches of a "true" strain (represented by rows) that were classified as "predicted" strain (given by columns). Notice that the patches highly overlap, and they are often misclassified due to the artifacts in the background (such as purple trail).

with the biological material is discolored. Moreover, one can notice that deep Fisher Vector can return two different results for two highly overlapped patches from the same scan. It is usually caused by the artifacts in the background, such as purple trail in *Candida lustianiae* (CL), predicted as *Cryptococcus neoformans* (CN), or *Maalasezia furfur* (MF) in Fig 6. The other incorrect classifications appear due to the small number of incomplete (fragmented) cells (like *Candida glabrata* (CG) predicted as CN).

Analysis of deep Bag of Words clusters

In this section, we first analyze deep Bag of Words pooling step by visualizing clusters using the patches nearest to their centroids. Then, based on those patches, we introduce a description of the considered species using properties pre-defined by the microbiologists. Finally, we present the mean deep BoW for every species. To make our analysis clearer, in this section, we limit deep BoW to 10 clusters, although its optimal number obtained with grid search optimization is 50. Moreover, the presented properties are introduced only to explain the intrinsic rules of the method. They are not used in the automatic classification, which requires only a scan image as an input.

Ten nearest neighbors of ten deep BoW centroids obtained with k-Means algorithm are presented in Fig 7. One can observe that they share common features and, therefore, can be
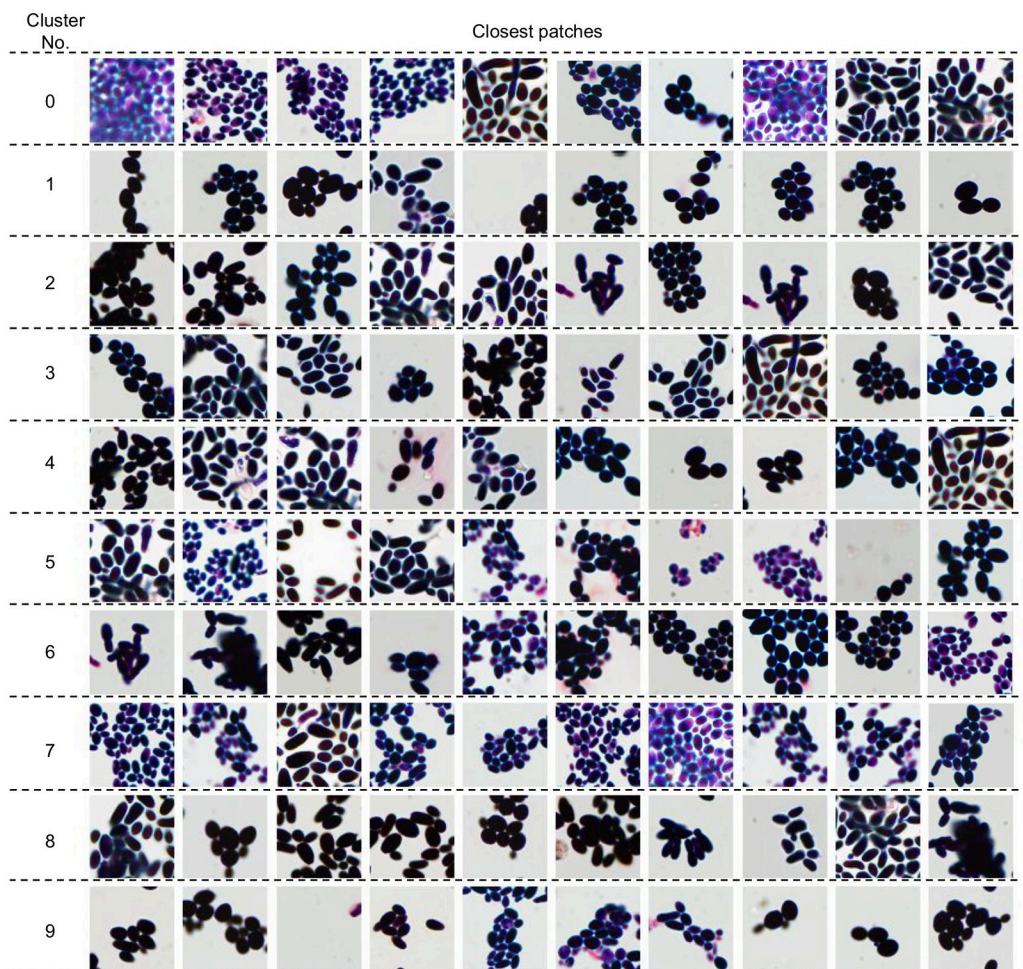

Fig 7. Ten nearest neighbors of deep Bag of Words centroids.

Table 2. A visual description of deep Bag of Words centroids from Fig 7.

Cluster No.	Brightness	Size	Shape	Arrangement	Appearance	color	Quantity
0	bright	small	oval longitudinal	regular	grouped fragmentary	black pink	high
1	dark	medium	oval circular	irregular	grouped	black	low
2	dark	large	longitudinal variform	irregular	grouped fragmentary	black	medium
3	dark	medium	variform oval	irregular	grouped fragmentary	black	medium
4	dark	large	longitudinal	irregular	grouped fragmentary	black blue	medium
5	bright	small	longitudinal oval	irregular	grouped	blue purple	medium
6	dark	medium	longitudinal oval	irregular	grouped fragmentary	black	medium
7	bright	small	longitudinal oval	irregular regular	grouped fragmentary	purple	medium
8	dark	medium	longitudinal oval	irregular	grouped fragmentary	black	high
9	dark	medium	oval	irregular	grouped	black	low

used to determine which visual properties are essential for the classifier. We consider the following properties (see Table 2): brightness (dark or bright), size (small, medium or large), shape (circular, oval, longitudinal or variform), arrangement (regular or irregular), appearance (singular, grouped or fragmentary), color (pink, purple, blue or black), and quantity (low, medium or high). As a result, the standard set of parameters used to describe the species (size, shape, arrangement, and appearance) was significantly extended.

To investigate which visual properties are essential for the classifier, we calculate mean deep Bag of Words representation for every species (see Fig 8) and then examine how the visual information about their main clusters corresponds to the knowledge of a microbiologist. The main conclusions she drew are as follows:

- species of the genus *Candida* mainly belong to cluster 2 with black cells of medium or large size, and oval or longitudinal shape;

- *Maalasezia furfur* has been assigned to clusters 0, 2, 5 and 8, mostly representing the black and longitudinal shape of various size;

- *Saccharomyces boulardii* and *Saccharomyces cerevisiae* are mainly described by clusters 1, 2, 4 and 8, which are characterized by black color, medium or large size and longitudinal shape;

- *Candida tropicalis* and *Saccharomyces cerevisiae* have very similar mean Bag of Words, which confirms high morphological similarity described in [28], i.e., size 3.0-8.0 × 5.0-10 μm, oval shape, elongated, and occurring singly or in small groups).

Analysis of deep Fisher Vector and SVM classifier

In this section, we first analyze the power of deep Fisher Vector representation using the t-SNE algorithm [29] by projecting it on a 2D surface. Then, we analyze classifier certainty based on the scores obtained for various patches.
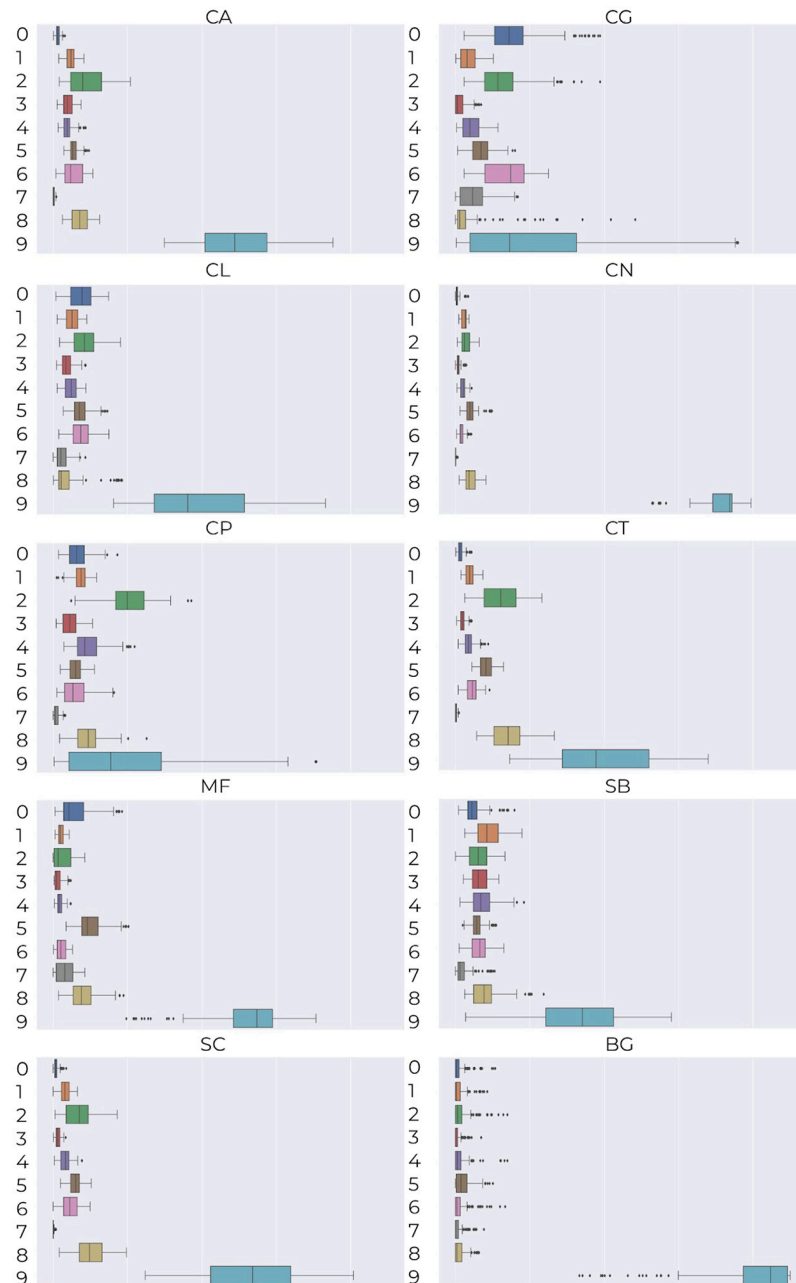
Fig 8. Mean deep Bag of Words for individual species together with the variance.

Projection of high-dimensional deep Fisher Vector to 2D using the t-SNE algorithm is presented in Fig 9. One can observe that classes are generally well separated in the case of AlexNet and ResNet18 architectures. Nevertheless, species of the same genus are not more coherent than the other species, in contrast to what we expected. Moreover, one can observe that InceptionV3 fails domain adaptation in the case of microbiological images, which explains the results in Table 1. Direct explanation of this behavior is hard, due to the neural network's black-box character. However, it is consistent with the results reported in [30–32], which suggest that ResNet representations are more robust than those obtained from InceptionV3.

TSNE visualization of fungus representation from Fisher Vector with representation from different NNs

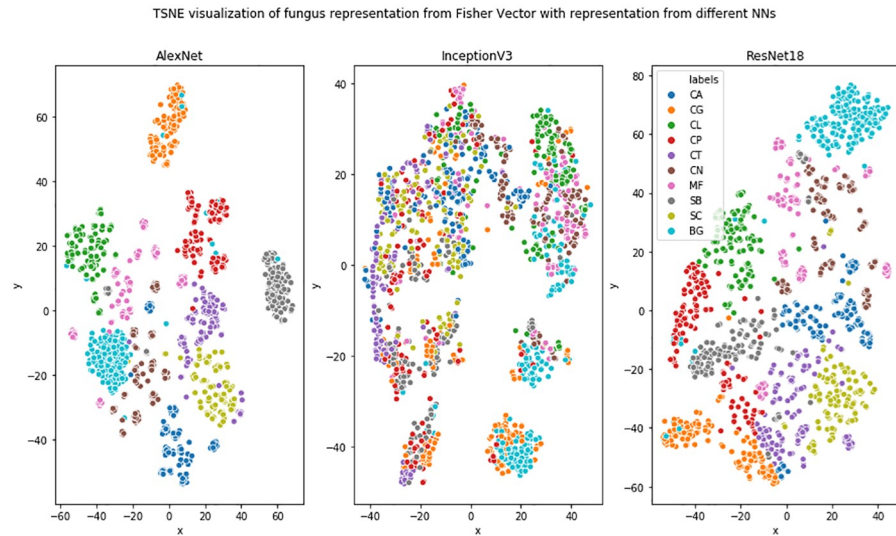**Fig 9. Projection of high-dimensional deep Fisher Vector to 2D using the t-SNE algorithm.**

https://doi.org/10.1371/journal.pone.0234806.g009

The second task in this section was to analyze classifier certainty. For this purpose, we investigate the distance of patches' representations from the classifier hyperplane, which roughly corresponds to how sure the classifier is of its decisions. Most left and right patches in Fig 10 are correctly classified with high probability, while the ones in the middle are ambiguous. The most representative fungal *Malassezia furfur* (MF) cells have oval, longitudinal shape, and often occur in the budding form, in which the daughter cells are as wide as the parent cells. While in the case of *Saccharomyces cerevisae* (SC), fungal cells characterize with round shapes, more significant in relation to *Candida albicans* (CA), which are arranged individually or in small groups.
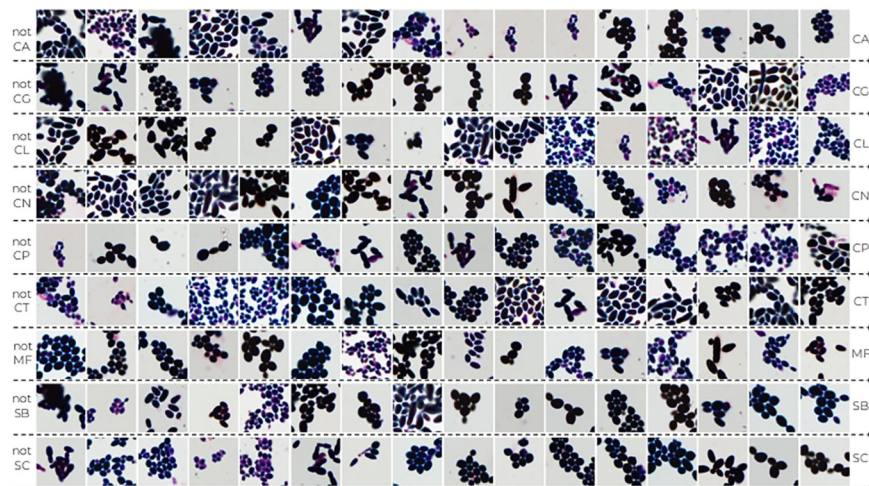

Fig 10. Classifier certainty for deep Fisher Vector. From left to right, one can observe the most negative and most positive patches according to the classifier of the particular species.

https://doi.org/10.1371/journal.pone.0234806.g010

Table 3. Test accuracy of scan-based classification obtained by aggregating patch-based classification and determining the most frequent prediction.

Method	CA	CG	CL	CN	CP	CT	MF	SB	SC	Total
AlexNet	95.5 ± 5.0	**95.0 ± 5.0**	70.0 ± 10.0	57.0 ± 6.9	95.0 ± 5.0	45.0 ± 10.0	70.0 ± 10.0	75.0 ± 5.0	85.0 ± 5.0	77.3 ± 4.2
DenseNet169	80.0 ± 10.0	85.0 ± 5.0	45.0 ± 5.0	57.0 ± 21.2	70.0 ± 10.0	**100.0 ± 0.0**	85.0 ± 15.0	95.0 ± 5.0	75.0 ± 10.0	77.6 ± 6.6
InceptionV3	50.0 ± 10.0	50.0 ± 10.0	80.0 ± 0.0	78.6 ± 14.3	55.0 ± 10.0	60.0 ± 10.0	50.0 ± 10.0	85.0 ± 5.0	85.0 ± 5.0	65.9 ± 4.9
ResNet18	**100.0 ± 0.0**	75.0 ± 5.0	**100.0 ± 0.0**	78.6 ± 6.9	50.0 ± 10.0	70.0 ± 0.0	70.0 ± 10.0	95.0 ± 5.0	80.0 ± 10.0	78.3 ± 5.4
ResNet50	**100.0 ± 0.0**	85.0 ± 15.0	**100.0 ± 0.0**	57.0 ± 6.9	50.0 ± 10.0	45.0 ± 15.0	80.0 ± 10.0	95.0 ± 5.0	75.0 ± 5.0	78.1 ± 8.3
AlexNet BoW RF	**100.0 ± 0.0**	75.0 ± 25.0	**100.0 ± 0.0**	75.0 ± 15.0	**100.0 ± 0.0**	90.0 ± 10.0	90.0 ± 10.0	**100.0 ± 0.0**	**100.0 ± 0.0**	92.2 ± 4.4
InceptionV3 BoW RF	80.0 ± 20.0	75.0 ± 25.0	**100.0 ± 0.0**	45.0 ± 15.0	40.0 ± 0.0	60.0 ± 30.0	0.0 ± 0.0	70.0 ± 10.0	45.0 ± 15.0	57.2 ± 2.8
ResNet18 BoW RF	**100.0 ± 0.0**	70.0 ± 10.0	**100.0 ± 0.0**	45.0 ± 15.0	**100.0 ± 0.0**	**100.0 ± 0.0**	80.0 ± 10.0	90.0 ± 10.0	**100.0 ± 0.0**	87.2 ± 1.7
AlexNet BoW SVM	**100.0 ± 0.0**	65.0 ± 25.0	**100.0 ± 0.0**	70.0 ± 10.0	**100.0 ± 0.0**	**100.0 ± 0.0**	85.0 ± 5.0	**100.0 ± 0.0**	**100.0 ± 0.0**	91.1 ± 2.2
InceptionV3 BoW SVM	70.0 ± 10.0	55.0 ± 15.0	**100.0 ± 0.0**	40.0 ± 20.0	45.0 ± 5.0	55.0 ± 15.0	0.0 ± 0.0	55.0 ± 15.0	40.0 ± 20.0	51.1 ± 2.3
ResNet18 BoW SVM	**100.0 ± 0.0**	60.0 ± 20.0	**100.0 ± 0.0**	65.0 ± 0.0	**100.0 ± 0.0**	90.0 ± 10.0	60.0 ± 0.0	95.0 ± 5.0	**100.0 ± 0.0**	85.6 ± 2.2
AlexNet FV RF	**100.0 ± 0.0**	65.0 ± 35.0	**100.0 ± 0.0**	55.0 ± 5.0	**100.0 ± 0.0**	90.0 ± 10.0	95.0 ± 5.0	**100.0 ± 0.0**	**100.0 ± 0.0**	89.4 ± 2.2
InceptionV3 FV RF	65.0 ± 5.0	**95.0 ± 5.0**	**100.0 ± 0.0**	50.0 ± 10.0	30.0 ± 10.0	75.0 ± 25.0	5.0 ± 5.0	45.0 ± 25.0	45.0 ± 35.0	56.7 ± 3.3
ResNet18 FV RF	95.0 ± 5.0	60.0 ± 0.0	**100.0 ± 0.0**	65.0 ± 5.0	**100.0 ± 0.0**	**100.0 ± 0.0**	95.0 ± 5.0	95.0 ± 5.0	95.0 ± 5.0	89.4 ± 1.7
AlexNet FV SVM	**100.0 ± 0.0**	75.0 ± 25.0	**100.0 ± 0.0**	**75.0 ± 15.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**95.0 ± 5.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**93.9 ± 3.9**
InceptionV3 FV SVM	75.0 ± 25.0	60.0 ± 0.0	**100.0 ± 0.0**	55.0 ± 5.0	45.0 ± 15.0	85.0 ± 5.0	5.0 ± 5.0	25.0 ± 15.0	45.0 ± 25.0	55.0 ± 5.6
ResNet18 FV SVM	**100.0 ± 0.0**	60.0 ± 0.0	**100.0 ± 0.0**	45.0 ± 15.0	95.0 ± 5.0	**100.0 ± 0.0**	**95.0 ± 5.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	88.3 ± 2.7

https://doi.org/10.1371/journal.pone.0234806.t003

Scan-based classification

To analyze classification score for the whole scan (instead of just patches, like in previous sections), we predict classification for all foreground patches of one scan and aggregate them to obtain the most frequently predicted species. As presented in Table 3, deep Fisher Vector performs better than the other methods, also in this case, obtaining 15.6% better accuracy than the best baseline method (ResNet18).

Conclusions and future work

In this paper, we apply deep neural networks and bag-of-words approaches to classify microscopic images of various fungi species. According to our experiments, the combination of features from deep neural networks with Fisher Vector works better than fine-tuning the classifier's block of the well-known network architectures and has the potential to be successfully used by microbiologists in their daily practice.

A large part of this paper is dedicated to the explainability of deep bag-of-words approaches to increase the trust in deep neural networks. For this purpose, we introduce an in-depth visual description of the properties pre-defined by the microbiologists. We hope that it will help to understand similarities and differences between fungi species better.

In our experiment, we assumed that images are obtained from the same laboratory and with the same scanner (details are presented in the Materials). However, in our opinion, this method could be easily extended to more diversified datasets by using additional preprocessing steps, which unify the input data. Due to the lack of data for such experiments, we did not cover this issue in the current article; however, it is planned for future research. Moreover, we would like to extend the DIFaS database so that it contains more preparations for all species, also gathered from other laboratories and scanners. Finally, we plan to prepare scans containing more than one species, as the automatic classification of such images would help to exclude the culture phase from the microbiological pipeline.

Supporting information

S1 Appendix.
(PDF)

S2 Appendix.
(PDF)

S1 Table. The number of foreground patches overlapped by less than 50% for the images from S1 Fig.
(PDF)

S2 Table. The number of background patches overlapped by less than 50% for the images from S1 Fig.
(PDF)

S3 Table. The number of foreground patches overlapped by less than 50% for the images from S2 Fig.
(PDF)

S4 Table. The number of background patches overlapped by less than 50% for the images from S2 Fig.
(PDF)

S1 Fig. Images of the first preparations from the DIFaS database and their foreground-background masks.
(TIF)

S2 Fig. Images of the second preparations from the DIFaS database and their foreground-background masks.
(TIF)

Author Contributions

Conceptualization: Bartosz Zieliński, Agnieszka Sroka-Oleksiak, Monika Brzychczy-Włoch.

Data curation: Agnieszka Sroka-Oleksiak.

Formal analysis: Bartosz Zieliński.

Funding acquisition: Bartosz Zieliński.

Investigation: Bartosz Zieliński, Agnieszka Sroka-Oleksiak, Dawid Rymarczyk.

Methodology: Bartosz Zieliński, Dawid Rymarczyk, Adam Piekarczyk, Monika Brzychczy-Włoch.

Project administration: Bartosz Zieliński.

Resources: Bartosz Zieliński, Dawid Rymarczyk.

Software: Dawid Rymarczyk, Adam Piekarczyk.

Supervision: Bartosz Zieliński, Monika Brzychczy-Włoch.

Validation: Bartosz Zieliński, Agnieszka Sroka-Oleksiak, Dawid Rymarczyk, Adam Piekarczyk.

Visualization: Dawid Rymarczyk, Adam Piekarczyk.

Writing – original draft: Bartosz Zieliński, Agnieszka Sroka-Oleksiak, Dawid Rymarczyk, Adam Piekarczyk, Monika Brzychczy-Włoch.

References

1. Maiken CA. Candida and candidaemia. Susceptibility and epidemiology. Danish medical journal. 2013; 60 11:B4698.

2. Rodrigues C, Silva S, Henriques M. Candida glabrata: a review of its features and resistance. European Journal of Clinical Microbiology & Infectious Diseases. 2014; 33 5:673–688. https://doi.org/10.1007/s10096-013-2009-3

3. Trofa D, Gácser A, Nosanchuk JD. Candida parapsilosis, an emerging fungal pathogen. Clinical Microbiology Reviews. 2008; 21 4:606–625. https://doi.org/10.1128/CMR.00013-08 PMID: 18854483

4. Silva S, Negri M, Henriques M, Oliveira R, Williams DW, Azeredo J. Candida glabrata, Candida parapsilosis and Candida tropicalis: biology, epidemiology, pathogenicity and antifungal resistance. FEMS Microbiology Reviews. 2012; 36(2):288–305. https://doi.org/10.1111/j.1574-6976.2011.00278.x PMID: 21569057

5. Silveira FP, Husain S. Fungal infections in solid organ transplantation. Medical Mycology. 2007; 45 (4):305–320. https://doi.org/10.1080/13693780701200372 PMID: 17510855

6. Papagianni M. Characterization of fungal morphology using digital image analysis techniques. J Microb Biochem Technol. 2014; 6(4):189. https://doi.org/10.4172/1948-5948.1000142

7. Lakner A, Essig A, Frickmann H, Poppert S. Evaluation of fluorescence in situ hybridisation (FISH) for the identification of Candida albicans in comparison with three phenotypic methods. Mycoses. 2012; 55(3):e114–e123. https://doi.org/10.1111/j.1439-0507.2011.02154.x PMID: 22126597

8. Ferrer C, Colom F, Frasés S, Mulet E, Abad JL, Alió JL. Detection and identification of fungal pathogens by PCR and by ITS2 and 5.8 S ribosomal DNA typing in ocular infections. Journal of clinical microbiology. 2001; 39(8):2873–2879. https://doi.org/10.1128/JCM.39.8.2873-2879.2001 PMID: 11474006

9. Raja HA, Miller AN, Pearce CJ, Oberlies NH. Fungal identification using molecular tools: a primer for the natural products research community. Journal of natural products. 2017; 80(3):756–770. https://doi.org/10.1021/acs.jnatprod.6b01085 PMID: 28199101

10. Lam P, Kok S, Lee K, Hung Lam K, Hau D, Wong WY, et al. Sensitization of Candida albicans to terbinafine by berberine and berberrubine. Biomedical Reports. 2016; 4:449–452. https://doi.org/10.3892/br.2016.608 PMID: 27073630

11. Abbas J, P Bodey G, A Hanna H, Mardani M, Girgawy E, Abi-Said D, et al. Candida krusei Fungemia: An Escalating Serious Infection in Immunocompromised Patients. Archives of Internal Medicine. 2000; 160:2659–2664. https://doi.org/10.1001/archinte.160.17.2659 PMID: 10999981

12. Saha DC, Goldman DL, Shao X, Casadevall A, Husain S, Limaye AP, et al. Serologic evidence for reactivation of cryptococcosis in solid-organ transplant recipients. Clinical and Vaccine Immunology. 2007; 14:1550–1554. https://doi.org/10.1128/CVI.00242-07 PMID: 17959819

13. Enache-Angoulvant A, Hennequin C. Invasive Saccharomyces Infection: A Comprehensive Review. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2006; 41:1559–68. https://doi.org/10.1086/497832

14. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee; 2009. p. 248–255.

15. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–1105.

16. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–4708.

17. 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; 2016. p. 2818–2826.

18. 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; 2016. p. 770–778.

19. Cimpoi M, Maji S, Vedaldi A. Deep filter banks for texture recognition and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3828–3836.

20. Zieliński B, Plichta A, Misztal K, Spurek P, Brzychczy-Włoch M, Ochońska D. Deep learning approach to bacterial colony classification. PloS one. 2017; 12(9):e0184554. https://doi.org/10.1371/journal.pone.0184554 PMID: 28910352

21. Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF. In: Computer Vision (ICCV), 2011 IEEE international conference on. IEEE; 2011. p. 2564–2571.

22. Liu C, Yuen J, Torralba A. Sift flow: Dense correspondence across scenes and its applications. IEEE transactions on pattern analysis and machine intelligence. 2011; 33(5):978–994. https://doi.org/10.1109/TPAMI.2010.147 PMID: 20714019

23. McCallum A, Nigam K, et al. A comparison of event models for naive bayes text classification. In: AAAI-98 workshop on learning for text categorization. vol. 752; 1998. p. 41–48.

24. Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos. In: Ninth IEEE International Conference on Computer Vision, 2003. IEEE; 2003. p. 1470–1477.

25. Perronnin F, Dance C. Fisher kernels on visual vocabularies for image categorization. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE; 2007. p. 1–8.

26. Nasrabadi NM. Pattern recognition and machine learning. Journal of electronic imaging. 2007; 16 (4):049901. https://doi.org/10.1117/1.2819119

27. Vedaldi A, Fulkerson B. VLFeat: An Open and Portable Library of Computer Vision Algorithms; 2008. http://www.vlfeat.org/.

28. Krzyściak P, Skóra M, B Macura A. Atlas grzybow chorobotworczych czlowieka. MedPharm; 2010.

29. Maaten Lvd, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008; 9(Nov):2579–2605.

30. Gupta R., Motwani K. Linear Models for Video Memorability Prediction Using Visual and Semantic Features. MediaEval. 2018.

31. Nieto D., Brill A., Kim B., Humensky T.B. Exploring deep learning as an event classification method for the Cherenkov Telescope Array. arXiv preprint arXiv:1709.05889 (2017).

32. Bechmann A. Keeping it real: From faces and features to social values in deep learning algorithms on social media images. Proceedings of the 50th Hawaii international conference on system sciences. 2017.