Received: 15 July 2024 | Accepted: 18 July 2024

DOI: 10.1111/myc.13777

ORIGINAL ARTICLE

mycoses Diagnosis, Therapy and Prophylaxis of Fungal Diseases WILEY

Next generation mycological diagnosis: Artificial intelligence-based classifier of the presence of *Malassezia* yeasts in tape strip samples

Martin Köberle | Alexander Zink | Tilo Biedermann | Sebastian Sitaru

Department of Dermatology and Allergy, TUM School of Medicine and Health, Technical University of Munich, Munich, Germany

Correspondence
Sebastian Sitaru, Department of Dermatology and Allergy, TUM School of Medicine and Health, Technical University of Munich, Biedersteiner Str. 29, Munich 80802, Germany.
Email: sebastian.sitaru@tum.de

Abstract

Background: *Malassezia* yeasts are almost universally present on human skin worldwide. While they can cause diseases such as pityriasis versicolor, their implication in skin homeostasis and pathophysiology of other dermatoses is still unclear. Their analysis using native microscopy of skin tape strips is operator dependent and requires skill, training and significant amounts of hands-on time.

Objectives and Methods: To standardise and improve the speed and quality of diagnosis of *Malassezia* in skin tape strip samples, we sought to create an artificial intelligence-based algorithm for this image classification task. Three algorithms, each using different internal architectures, were trained and validated on a manually annotated dataset of 1113 images from 22 samples.

Results: The Vision Transformer-based algorithm performed the best with a validation accuracy of 94%, sensitivity of 94.0% and specificity of 93.5%. Visualisations providing insight into the reasoning of the algorithm were presented and discussed.

Conclusion: Our image classifier achieved very good performance in the diagnosis of the presence of *Malassezia* yeasts in tape strip samples of human skin and can therefore improve the speed and quality of, and access to this diagnostic test. By expanding data sources and explainability, the algorithm could also provide teaching points for more novice operators in future.

KEYWORDS
artificial intelligence, classifier, diagnosis, *Malassezia*, microscopy

1 | INTRODUCTION

Almost all humans worldwide are colonised by *Malassezia* spp., a genus of lipid-dependent Basidiomycota yeasts.[1,2] While under some circumstances, they can cause infections in the form of pityriasis (tinea) versicolor (PV) or *Malassezia* folliculitis, for the most part, the role of their presence in the skin microbiome and for other skin conditions such as seborrheic dermatitis remains to be elucidated.[3–6]

These yeasts can be observed as globose to ovoid, unipolar-budding cells and as filaments of cylindrical cells. Brightfield microscopy of native preparations is the preferred diagnostic test in routine dermatological care.[7] It is fast and causes almost no consumable costs. With PV, tape strips are especially suitable, as they recover a relatively large area of stratum corneum skin scales in intact spatial arrangement, on top of which *Malassezia* grow, while potash lye maceration is of limited benefit, since there is no growth inside the

scales (as opposed to dermatophytes). The presence of either no or sporadic single yeast cells, as opposed to clusters of *Malassezia*, typically of both morphologies ('spaghetti and meatballs'), is well suited for the discrimination between no or commensal colonisation, and pathogenic overgrowth.

However, the manual performance of microscopic analyses requires significant hands-on time by an experienced examiner, and interpretation is still operator-dependent, creating an unmet need for automated and objective analysis of these samples. In digital medicine, artificial intelligence (AI)-based algorithms can be trained to classify images with very good accuracy, sometimes even surpassing individual human performance.[8–10] Usually, algorithms pre-trained on huge image datasets are fine-tuned using a specific, annotated (labelled) dataset for the intended purpose.[9,11] Different algorithm classes with different internal architectures exist for image classification. Notable milestones for convolutional neural networks include ResNet, XCeption and EfficientNet,[12–14] while vision transformers (ViT) are a relatively new class of algorithms offering one of the best classification performances while also preserving time and computer resources when training.[15]

Therefore, in this study, we sought to create and validate an AI image classifier for detecting the presence of *Malassezia* yeast in native tape strip samples, comparing different algorithm architectures, to standardise and improve the speed and quality of *Malassezia* diagnostics, as well as to provide teaching input to more novice evaluators.

2 | MATERIALS AND METHODS

For routine diagnostics in suspected PV at our department, tape strips are obtained by repeatedly glueing adhesive tape strips on affected skin and pulling them off. The tape strips are then glued onto standard glass slides for native microscopy without any further preparation (Figure 1). For this study, 1113 field of views of 22 of such tape strip samples on glass slides were considered. The samples were from the torso ($n=18$), face ($n=3$) and upper arm ($n=1$). No synthetic or other image sources were used. Images were captured using a Keyence BZ-X810 microscope (Keyence Corp, Osaka, Japan) in brightfield mode with a 40× NA 0.95 PlanAPO lens to create 1920×1440 pixel 8-bit grayscale LZW compressed TIF images with the BZ-X800 Viewer software. Original TIFs were slightly sharpened (amount: 21, radius: 3, threshold: 6) in ACDsee Ultimate 8 (ACD Systems, Victoria, Canada) and manually labelled by an experienced mycologist to either positive or negative for *Malassezia* yeasts. This annotation is defined as the ground truth. The resulting dataset of 539 negative and 574 positive images was randomly split image-wise to 80% training and 20% validation data.

Three algorithms utilising three different backbones were trained using this data with help of the TensorFlow Keras Framework v2.15.0 (Python v3.10.13) using one nVidia RTX4090 24GB GPU. The hyperparameters for training were the same for all models: batch size 16, 200 epochs, Adam optimizer with default parameters, binary crossentropy as loss function. All backbones were initialised with their respective pre-trained weights. As a classification head, one fully connected layer of 256 neurons with ReLU activation, followed by one fully connected layer of 1 neuron with sigmoid activation, was used. Data augmentation was performed, consisting of a random crop from 500×500 pixels (load size) to the respective input sizes of the networks, random brightness (0.3× to 1.3× the original brightness), random contrast (0.3× to 1.3× the original contrast), random flip (both axes). The software and more details about the algorithm are published at https://github.com/ssitaru/malassezia-classifier.

For Algorithm A, a vision transformer (ViT) backend with 12 layers and patch size 32×32 was used (ViTB/32 implemented by the vit_keras package v0.1.2). The image size was 384×384 pixels. For Algorithm B, an XCeption backend implemented by the Keras application layer was used. The image size was 299×299 pixels. For Algorithm

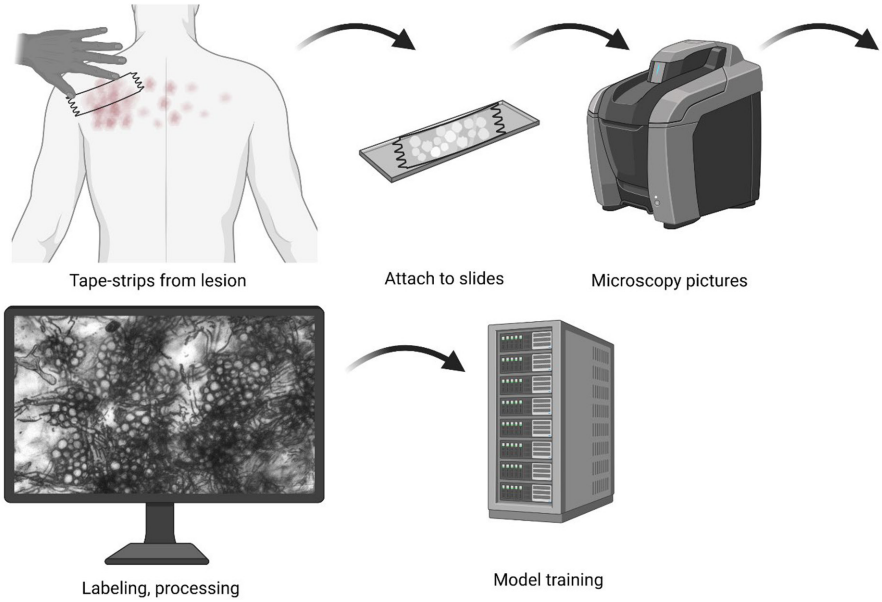

FIGURE 1 Schematic study workflow.

C, an EfficientNetB7 backend, implemented by the Keras application layer, with an image size of 224×224 pixels was used. After each training epoch, validation was performed. Checkpoints were saved, when a new maximum validation accuracy was observed. Additionally, for Algorithm A, the internal attention masks were obtained using the functions in vit_keras and postprocessed by linear colour mapping to the 'jet' colormap using opencv-python v4.10.0.84.

Finally, validation data (predicted labels, scores) were evaluated using R v4.4.0 and mltest v1.0.1 and precrec v0.14.4 libraries.

The workflow scheme was created with BioRender.com.

Since the Bavarian hospital law permits the use of routine patient data for research, informed consent and IRB approval were not necessary (BayKrG article 27 §4). Data were not transmitted to any third-party service or machine.

3 | RESULTS

After training for 200 epochs, the algorithm performance on the validation dataset was evaluated: Algorithm C reached a peak accuracy of 82% with a sensitivity of 86.3% and a specificity of 87.0%, while Algorithm B reached a peak accuracy of 89.2% with a sensitivity of 90.2% and a specificity of 89.8%. Algorithm A reached the best performance with a peak validation accuracy of 94.1%, a sensitivity of 94.0%, specificity of 93.5%, and an area under the receiver operator characteristic curve (AUC-ROC) of 0.99. Of note, all training accuracies were higher than the respective validation accuracies (data not shown).

Example images, their true label (ground truth) and the prediction of Algorithm A are shown in Figure 2. In Panel A, no hyphae or yeast cells can be seen, which has also been correctly identified by the algorithm. In Panel B, a falsely positively classified image, no hyphae or yeast cells can be observed. However, some linear structures, which probably correspond to thickened keratin lamellae are seen in the middle. These 'mosaic hyphae' are also the most common artefact leading to false positive results by human observers. In Panel C, some hyphae can be seen, for example at the bottom (arrow) and very few yeast cells are present in the middle of the image, which was annotated as showing potentially pathogenic accumulation of *Malassezia*. The algorithm did not correctly classify this image. The last image, Panel D, was correctly identified as positive by the algorithm. Hyphae and some single round cells can be identified in the middle of the image (arrow).

To gain insights into the black box of the neural network of Algorithm A, we obtained the attention mask for an example image (Figure 3). There are clear hyphae throughout the image, as well as nests of yeast cells in the middle, top, and left lower corner of the picture. When comparing this to the overlayed attention map, yellow being a higher attention score, it is remarkable that while almost all of the areas with hyphae have high scores, the areas with yeast cells are not fully covered, for example in the lower left corner.

4 | DISCUSSION

In our study, we created and validated three AI classifier algorithms for the diagnosis of the presence of *Malassezia* yeast in human tape strip samples. When comparing their performance, the ViT-backed Algorithm A performed the best with an impressive validation accuracy of 94.1% and a ROC-AUC of 0.99.

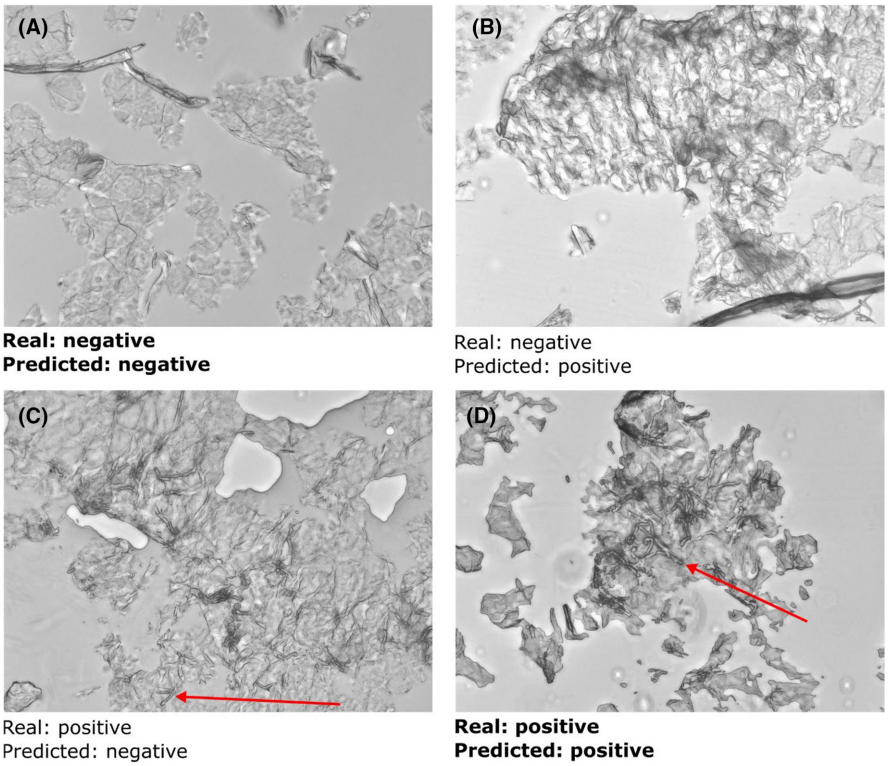

FIGURE 2 Example images from the dataset and their labels. The figure shows four microscopical brightfield images of human skin tape strip specimens for the analysis of *Malassezia* spp. presence. Below each image the real (ground-truth) label, as well as the algorithm prediction is indicated. (A, D) were correctly classified, while (B, C) were not. The red arrows point to hyphae and yeast cells, which confirms the presence of *Malassezia*. The prediction was performed using the Vision Transformer-based Algorithm A.

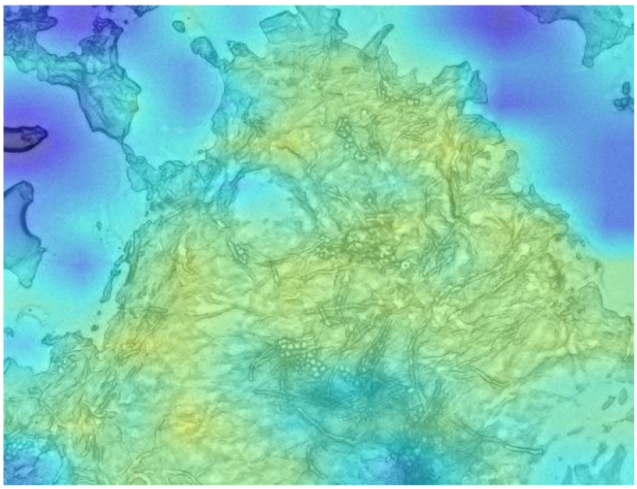

FIGURE 3 Attention mask for an example image. Malassezia-positive original image (ground truth), as used for input to Algorithm A, overlayed with an attention mask, where blue corresponds to lower attention scores (i.e., lower contribution of these pixels to the final decision), and yellow to higher attention scores (i.e., higher contribution of these pixels to the final decision). The correct final decision of the algorithm was positive.

To detect the presence of *Malassezia* spp. in support of the diagnosis of PV and other skin diseases, direct microscopy is the method of choice. This is because other methods are either unreliable, like examination of patients with Wood's light or fungal culture on media supplemented with a long chain fatty acid source, or are not able to distinguish between commensal or pathogenic colonisation (polymerase chain reaction, sequencing). On the other hand, identification on species level, that is rarely possible in native preparations, is also usually not necessary for diagnostic purposes.[16]

Interestingly, but not very surprisingly, each more modern algorithm class performed better than the last (in the order of EfficientNet, XCeption, ViT, which corresponds to the order in which they were published). This suggests that AI algorithm performance in real-world settings closely tracks the more theoretical performance on publicly available datasets, such as ImageNet, which are usually used to train and compare them in a first instance.

When creating and using AI algorithms, explainability (XAI) is still a major challenge and could pose a barrier to widespread adoption, especially in higher-risk areas such as medicine.[17] ViT algorithms use so-called attention mechanisms, special layers within the network, whose contents can relatively easily be visualised on top of the original image.[18] These attention maps are indicative of the weight of individual pixels for the final decision and can therefore help understand how the algorithms arrive at their classification.[18] Our analysis of the attention masks of Algorithm A suggest that the algorithm gives more weight to hyphae ('spaghetti') than cells ('meatballs') for the classification. This is of course in contrast to the usual diagnostic criteria of the presence of *Malassezia* yeasts, which require both hyphae and cells for positivity. However, in PV, the filamentous form (hyphae) is often dominating,[19] while mostly yeast cells are found in pustules of *Malassezia* folliculitis patients.[7] In seborrheic dermatitis conflicting findings have been published.[20,21] Since our positive samples were taken from PV patients, this deviation from textbook knowledge likely reflects the specifics of the particular disease investigated.

The limitations of our study include the limited number of included samples, from which pictures have been captured with one device. Only samples from patients with PV were included. To improve reliability, generalisability and objectiveness, a higher number of samples from multiple data sources and classified by several annotators should be included in future. A higher number of training pictures from different diseases with proven or suspected involvement of *Malassezia* would also enable a more nuanced classification (e.g. elevated number of yeasts, presence of *Malassezia* hyphae, fulminant spaghetti and meatball structures) and enable a disease specific validation. Similarly, images could be annotated for segmentation of pathogenic structures rather than simple classification, which also enables more detailed analysis (e.g. quantification of yeasts and hyphae) and better explainability.

Also, our data only included one preparation method for this diagnostic test. Variations include skin scrapings instead of tape strips, additional keratolysis (e.g. with KOH), or staining of fungal components (e.g. Gram stain, lactophenol cotton blue or even fluorescence staining with Calcofluor white), before microscopy, which were not present in our dataset and therefore, algorithm performance for these data types is unknown.

Still, our AI-based algorithm offers impressive performance for the diagnosis of the presence of *Malassezia* spp. in native human skin tape strip specimens, and therefore could not only improve the speed and quality of the diagnostic process while reducing labour costs, but also provide teaching input to more novice operators. It furthermore highlights the tremendous potential for AI assistance in visual pattern recognition in dermatological laboratory workflow, while simplifying the interpretation of basic diagnostic procedures like tape stripping, which hopefully will allow easier access to this method and consequently accelerated research into these common, but still elusive facultative pathogens and their meaning for common skin conditions as well as skin homeostasis.

AUTHOR CONTRIBUTIONS

Martin Köberle: Conceptualization; methodology; data curation; formal analysis; visualization; writing – original draft; writing – review and editing. **Alexander Zink:** Conceptualization; supervision; resources; writing – review and editing; funding acquisition. **Tilo Biedermann:** Conceptualization; supervision; project administration; resources; writing – review and editing; funding acquisition. **Sebastian Sitaru:** Conceptualization; methodology; data curation; software; visualization; writing – review and editing; writing – original draft; investigation; validation.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Sebastian Sitaru https://orcid.org/0000-0001-7324-9139

REFERENCES

1. Prohic A, Jovovic Sadikovic T, Krupalija-Fazlic M, Kuskunovic-Vlahovljak S. Malassezia species in healthy skin and in dermatological conditions. *Int J Dermatol*. 2016;55(5):494-504. doi:10.1111/ijd.13116
2. Jagielski T, Rup E, Ziółkowska A, Roeske K, Macura AB, Bielecki J. Distribution of Malassezia species on the skin of patients with atopic dermatitis, psoriasis, and healthy volunteers assessed by conventional and molecular identification methods. *BMC Dermatol*. 2014;14:3. doi:10.1186/1471-5945-14-3
3. Chebil W, Rhimi W, Haouas N, et al. Virulence factors of Malassezia strains isolated from pityriasis versicolor patients and healthy individuals. *Med Mycol*. 2022;60(8):myac060. doi:10.1093/mmy/myac060
4. Theelen B, Cafarchia C, Gaitanis G, Bassukas ID, Boekhout T, Dawson TL. Malassezia ecology, pathophysiology, and treatment. *Med Mycol*. 2018;56(suppl_1):S10-S25. doi:10.1093/mmy/myx134
5. Tsai YC, Wang JY, Wu YH, Wang YJ. Clinical differences in pediatric and adult *Malassezia* folliculitis: retrospective analysis of 321 cases over 9 years. *J Am Acad Dermatol*. 2019;81(1):278-280. doi:10.1016/j.jaad.2019.03.014
6. Pedrosa AF, Lisboa C, Gonçalves RA. Malassezia infections: a medical conundrum. *J Am Acad Dermatol*. 2014;71(1):170-176. doi:10.1016/j.jaad.2013.12.022
7. Henning M a S, Hay R, Rodriguez-Cerdeira C, et al. Position statement: recommendations on the diagnosis and treatment of Malassezia folliculitis. *J Eur Acad Dermatol Venereol*. 2023;37(7):1268-1275. doi:10.1111/jdv.18982
8. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*. 2018;29(8):1836-1842. doi:10.1093/annonc/mdy166
9. Sitaru S, Oueslati T, Schielein MC, et al. Automatic body part identification in real-world clinical dermatological images using machine learning. *J Dtsch Dermatol Ges*. 2023;21(8):863-869. doi:10.1111/ddg.15113
10. Kim Y, Bu S, Tao C, Bae KT, HALT Polycystic Kidney Disease Study Group. Deep learning-based automated imaging classification of ADPKD. *Kidney Int Rep*. 2024;9(6):1802-1809. doi:10.1016/j.ekir.2024.04.002
11. Schielein MC, Christl J, Sitaru S, et al. Outlier detection in dermatology: performance of different convolutional neural networks for binary classification of inflammatory skin diseases. *J Eur Acad Dermatol Venereol*. 2023;37(5):1071-1079. doi:10.1111/jdv.18853
12. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2017 Accessed October 31, 2021. IEEE. http://arxiv.org/abs/1610.02357
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015. doi:10.48550/arXiv.1512.03385
14. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. 2020. doi:10.48550/arXiv.1905.11946
15. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is worth 16×16 words: transformers for image recognition at scale. 2021. doi:10.48550/arXiv.2010.11929
16. Saunte DML, Gaitanis G, Hay RJ. Malassezia-associated skin diseases, the use of diagnostics and treatment. *Front Cell Infect Microbiol*. 2020;10:112. doi:10.3389/fcimb.2020.00112
17. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform*. 2021;113:103655. doi:10.1016/j.jbi.2020.103655
18. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. 2023. doi:10.48550/arXiv.1706.03762
19. Łabędź N, Navarrete-Dechent C, Kubisiak-Rzepczyk H, Bowszyc-Dmochowska M, Pogorzelska-Antkowiak A, Pietkiewicz P. Pityriasis Versicolor—a narrative review on the diagnosis and management. *Lifestyles*. 2023;13(10):2097. doi:10.3390/life13102097
20. Li J, Feng Y, Liu C, et al. Presence of Malassezia hyphae is correlated with pathogenesis of seborrheic dermatitis. *Microbiol Spectr*. 2022;10(1):e0116921. doi:10.1128/spectrum.01169-21
21. Wikramanayake TC, Borda LJ, Miteva M, Paus R. Seborrheic dermatitis—looking beyond Malassezia. *Exp Dermatol*. 2019;28(9):991-1001. doi:10.1111/exd.14006

How to cite this article: Köberle M, Zink A, Biedermann T, Sitaru S. Next generation mycological diagnosis: Artificial intelligence-based classifier of the presence of *Malassezia* yeasts in tape strip samples. *Mycoses*. 2024;67:e13777. doi:10.1111/myc.13777