

A Deep Learning Approach for Histopathological Diagnosis of Onychomycosis: Not Inferior to Analogue Diagnosis by Histopathologists

Florence DECROOS¹, Sebastian SPRINGENBERG², Tobias LANG², Marc PÄPPER², Antonia ZAPF³, Dieter METZE⁴, Volker STEINKRAUS¹ and Almut BÖER-AUER^{1,4}

¹Dermatologikum Hamburg, ²Mindpeak GmbH, ³Institut für Medizinische Biometrie und Epidemiologie, Hamburg, and ⁴Department of Dermatology, Münster University, Münster, Germany

Onychomycosis is common. Diagnosis can be confirmed by various methods; a commonly used method is the histological examination of nail clippings. A deep learning system was developed and its diagnostic accuracy compared with that of human experts. A dataset with annotations for fungal elements was used to train an artificial intelligence (AI) model. In a second dataset (n=199) the diagnostic accuracy of the AI was compared with that of dermatopathologists. The results show a non-inferiority of the deep learning system to that of analogue diagnosis (non-inferiority margin 5%) with respect to specificity and the area under the receiver operating characteristic curve (AUC). The AI achieved an AUC of 0.981. One limitation of this system is the need for a large number of training images. The AI had difficulty recognizing spores and confused serum or aggregated bacteria with fungal elements. Use of this deep learning system in dermatopathology routine might help to diagnose onychomycosis more efficiently.

Key words: artificial intelligence; deep learning; onychomycosis; dermatopathology.

Accepted Aug 17, 2021; Epub ahead of print Aug 18, 2021

Acta Derm Venereol 2021; 101: adv00532.

Corr: Florence Decroos, Dermatologikum Hamburg, Stephansplatz 5, DE-20354 Hamburg, Germany. E-mail: f.decroos@dermatologikum.de

Onychomycosis is a fungal infection of the nails, which occurs in approximately 10% of the general population, with half of affected people being over 70 years of age. Patients present with discoloured nails, thickened nail plate, and onycholysis (1). The infectious agents differ depending on geographical region; in North America and Europe dermatophytes are the causative pathogens in 65% of cases, followed by yeasts in 21%, and non-dermatophyte moulds in 13% (2). The predominant dermatophyte causing onychomycosis is *Trichophyton rubrum* (3).

Diagnosis and appropriate treatment of onychomycosis is important, as it may prevent progression to tinea pedis and help to avoid severe complications, such as erysipelas (4). Not uncommonly it is clinically relatively difficult to differentiate onychomycosis from other nail diseases, such as nail psoriasis or nail involvement in lichen planus. However, the treatment of onychomycosis is different

SIGNIFICANCE

Onychomycosis is a common nail infection. One diagnostic method is the histopathological examination of nail clippings, which is labour intensive. Use of artificial intelligence is emerging in medicine, but it is not yet used for the histological diagnosis of onychomycosis. A deep learning system was developed for diagnosis of onychomycosis using scanned sections of nail clippings. In 199 cases the diagnostic accuracy of the artificial intelligence was compared with that of dermatopathologists. The system can be used to assist dermatopathologists and can reduce the workload in everyday routine. Similar systems may also be developed to detect fungal organisms in skin biopsies for the diagnosis of tinea.

from that of inflammatory diseases of the nail apparatus. If the clinical presentation is extensive or refractory to topical therapies, systemic treatments, such as terbinafine and itraconazole, are indicated, but possible side-effects and interactions with other medications need to be considered and the correct diagnosis has to be proven (5).

There are several methods used to diagnose mycoses: potassium hydroxide (KOH) preparation or fluorescent-assisted direct microscopy of nail scrapings, culture, histological examination of formalin-fixed, paraffin-embedded nail clippings stained with periodic acid–Schiff reaction (PAS stain) represent conventional, widely used methods, while molecular diagnostics using PCR techniques and mass spectrometry are more recent approaches, which are not yet widely available (6). Compared with a conventional diagnostic standard method (KOH), the sensitivity of the molecular diagnostic method PCR-based sequencing, was high, at 97% (7). A combination of culture and histopathological examination with PAS stain reaches 94% sensitivity. However, routine histopathological examination with PAS stain was 85% sensitive vs culture on Sabouraud agar with chloramphenicol and cycloheximide (Mycosel agar) with 32% sensitivity. Histological examination of nail clippings still appears to be a very sensitive method for diagnosing onychomycosis, which is easily performed, atraumatic, and the results are available relatively quickly (8).

In recent years, artificial intelligence (AI) methods for histopathological image analysis have become increa-

singly popular (9). These methods are typically based on deep learning (9–12), a subfield of machine learning. Deep learning systems most often use so-called convolutional neural networks (CNNs), which comprise several layers of artificial neurones or nodes (13) connected to each other via weights. While traditional computer vision techniques involve designing hand-crafted analysis rules, and often fail to generalize to the variances present in medical images, machine learning approaches learn diagnostically relevant features by looking at the statistics present in training images. In the last decade, they have been shown to perform with high accuracy in a wide variety of tasks and domains in healthcare, including clinical imaging, electronic health records, genomics and health applications for mobile devices (14–17).

The first application of deep learning to histopathology was the detection of mitoses in haematoxylin and eosin (HE)-stained sections (18). Many other applications have followed in general histopathology, from detecting lymph node metastases in breast cancer (19) over gastric carcinoma (20) to the detection of nuclei in tissue (21). In dermatopathology, there have been few approaches so far; e.g. for the histopathological diagnosis of melanoma and basal cell carcinoma (22, 23).

With regard to the clinical diagnosis of onychomycosis and AI, a group in Seoul studied the accuracy of binary classification of clinical photographs of nails as showing or not showing onychomycosis (24). They used almost 50,000 clinical images of previously diagnosed cases to train a deep learning system. In a study of clinical diagnosis of onychomycosis, their AI system outperformed dermatologists. A different later study also focused on clinical photographs of nails (25). However, in addition to clinical assessment, a second diagnostic procedure is needed for the diagnostic process to unequivocally prove the infection in the tissue. Accurate identification of the micro-organism, e.g. by microscopy of histopathological PAS-stained sections of nail clippings, is required before the prescription of systemic treatment which may have severe side-effects.

Given the increasing importance of use of AI in dermatology and pathology, and the high frequency of onychomycosis in dermatological practice, together with an increasing workload of pathologists worldwide, this study aimed to develop AI for recognition of fungal organisms in scanned whole-slide images of PAS-stained sections of nail clippings. To assess the performance of the AI it was compared with the diagnoses made by dermatopathologists working on the same PAS-stained sections using a conventional microscope.

MATERIALS AND METHODS

Samples

This monocentric study, conducted in the histopathological department of the Dermatologikum Hamburg, used 2 different datasets.

The first dataset comprised 528 cases and was used for training the AI. The second dataset comprised 199 cases and was used to study the diagnostic accuracy of the AI and compare it with the performance of 4 dermatopathologists with different levels of experience. The samples were collected retrospectively and chosen randomly from the years 2018 to 2020. Inclusion criteria were a clinical suspicion of onychomycosis and histopathological analysis of a nail clipping as diagnostic method. Exclusion criteria were composed of air under the cover slip, a displaced cover slip, and no visible nail fragment on the PAS-stained slide. The study material ($n=199$) came from 104 male and 95 female patients and the mean age was 54 (range 18–89) years. Specimens came from toenails in 169 cases (right foot $n=70$, left foot $n=70$, both feet collective sample $n=29$) and from fingernails in 12 cases (right hand $n=6$, left hand $n=5$, both hands collective sample $n=1$). No site was given in 18 cases.

All patients had given informed consent to the procedure of nail clipping and diagnosis by microscopy at the time of clinical presentation. This study was approved by the local ethics committee (Ärztchamber Hamburg, Hamburg, Germany) (processing number: WF-178/20).

The nail samples were fixed in formalin and processed in paraffin. Sections of 4- μm thickness were stained with Artisan Link Pro Special Staining System, an *in vitro* diagnostic device for automated special stains on formalin-fixed, paraffin-embedded tissue sections. (serial number ALP812080; Agilent Technologies Deutschland GmbH, Waldbronn, Germany). The Artisan Periodic Acid Schiff Stain Kit (number AR165; Firma Agilent Technologies; Deutschland GmbH, Waldbronn, Germany) was used according to the manufacturer's instructions.

Scanning

Samples were scanned with a resolution of 0.25 $\mu\text{m}/\text{pixel}$, resulting in digital whole-slide images (WSI) of the tissue (approximate size 100,000 \times 100,000 pixels). The scanners 3DHitech Panoramic SCAN II and 3DHitech Panoramic P1000 (Sysmex, Hamburg, Germany) were used with the following scanning profile: Scanner Panoramic SCAN II, 3D Histech, Objective type 20 \times , Output resolution 37 \times , Focus, distance 15FOVs, Source Brightfield, Specimen threshold Auto // 220, Multilayer Mode Extended focus, Levels 7, Step Size 5 (0.2 μm), Image quality Good, Compression JPEG, Bit, depth 8 bit and stitching ON. Scans were fully automatic, with subsequent manual checks that all tissue had been covered. For 8 object plates with little material, Roche Ventana DP 200 (Roche Diagnostics, Washington, USA) was used, because it had a higher performance in focusing on small areas of tissue. Sections were anonymized for the scanning process.

Deep learning model

A CNN model was designed for image analysis of histopathological nail tissue. A CNN consists of a series of layers of computational units that calculate a cascade of image features of increasing complexity. The most abstract features are used to compute the final analysis. The parameters (weights) of the computational units are optimized during AI training, based on example images.

The CNN architecture is similar to VGG-13 (26), but introduces dilation to the convolution operations. Dilation allows the kernel size to be increased without increasing the amount of parameters needed. This enables the incorporation of more fine-grained information from larger context fields at the individual convolution stages without increasing the model complexity. The CNN used in the current study consists of 13 convolution layers with kernels of size 4 and a dilation factor increasing with the network depth.

The CNN takes a "patch", a small image of size 256 \times 256 pixels, as input, processes it and provides a probability estimate between

0% and 100% about the patch containing onychomycosis. The magnification used for patches is $0.25 \mu\text{m}/\text{pixel}$. To make predictions for a full whole-slide image (with a size of approximately $100,000 \times 100,000$ pixels), the WSI is split into patches (**Fig. 1**). The CNN makes predictions for all patches. These predictions are aggregated to obtain a final decision for the WSI. Aggregation is computed by taking the 10 patches with the highest onychomycosis probabilities into account, calculating their average probability and subsequently applying a threshold in order to obtain a binary decision for the WSI. The optimal threshold is determined on the validation data of the training dataset: it is set to the value that achieves the optimal split between positive and negative validation cases, based on their aggregated probabilities.

Data and annotations

A dataset of 528 WSI was used to train the CNN. Samples consisted of 286 positive (with fungus) and 242 negative (without fungus) cases. Positiveness was determined by an experienced dermatopathologist. The positive cases include 11 cases that were initially marked as negative, but had to be corrected to be positive after the AI found small hyphae structures that were later confirmed by the human experts (there were no cases that were initially marked as positive and had to be corrected to negative.)

For training, negative patches were taken at random from the negative WSI. Positive WSI contain patches both with and without hyphae. To determine positive patches from the positive WSI, hyphae structures were annotated on positive samples by a dermatology resident, using a few dots per hypha. Annotations were not exhaustive, but only a small number of hyphae (1–30) were marked on each positive case. Annotations were made using the software Automated Slide Analysis Platform (ASAP) 1.9 (Computational Pathology Group, Radboud University Medical Center, Netherlands, <https://computationalpathologygroup.github.io/ASAP/>). In a later additional, so-called active learning, step, the CNN identified patches in the positive samples that did not yet have annotations, and about which it was particularly uncertain

whether onychomycosis was present. These patches were labeled by the human expert, using an unpublished proprietary patch classification software tool, and then added to the training dataset. All annotations were reviewed and confirmed by an experienced dermatopathologist, resulting in annotations being confirmed independently by 2 human experts.

Deep learning training procedure

For the patch-based approach, training WSI were partitioned into patches by overlaying a grid and extracting patches with a resolution of 256×256 pixels (**Fig. 2**). As it is difficult and labour intensive to annotate positive patches in large numbers, a self-supervised learning approach was adopted. In a pre-learning step, this enables the learning of a good pre-initialization of the CNN parameters, consisting of general image features without the need for any annotations (27, 28). In a subsequent supervised learning step, the CNN is fine-tuned based on the annotated patches. For training, the data were split into a development (80%) and a validation (20%) set with slide holdout. The validation set was used to optimize the learning procedure. The Adam optimizer, with a learning rate of 0.0001 and rotation and colour image augmentations, was used. Background areas without tissue in training images were removed using standard computer vision filtering (Otsu mask).

Study evaluation

For evaluation of the AI performance, 199 new samples were collected from Dermatologikum Hamburg, Hamburg, Germany. This dataset does not include any samples that were used for AI training. The ground truth (the gold standard) as to whether samples were positive or negative was determined by 2 dermatopathologists, who reviewed the PAS-stained sections independently and agreed on all diagnoses, dividing them into negative and positive classes. For 49 uncertain negative cases, PCR examination for dermatophyte-specific DNA on the same paraffin block was used to confirm negativity. Cases were anonymized with numbers

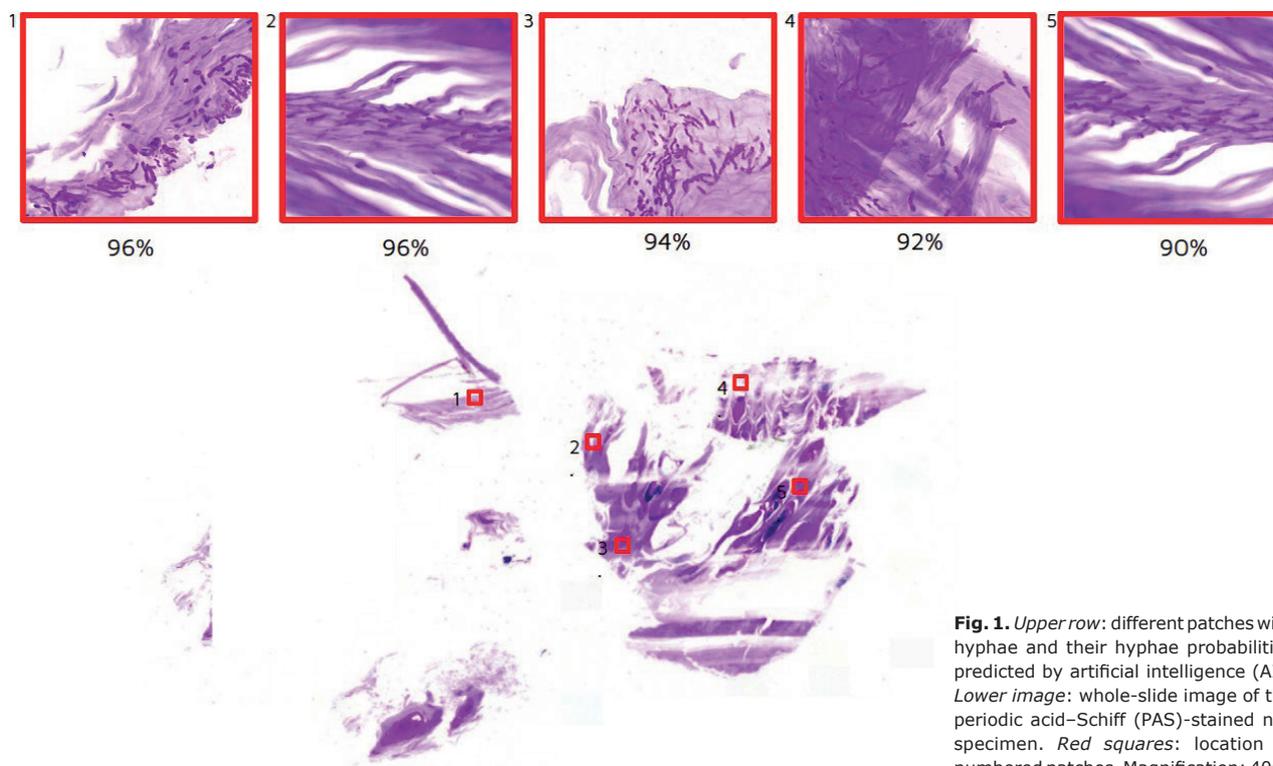


Fig. 1. Upper row: different patches with hyphae and their hyphae probabilities predicted by artificial intelligence (AI). Lower image: whole-slide image of the periodic acid-Schiff (PAS)-stained nail specimen. Red squares: location of numbered patches. Magnification: 400x

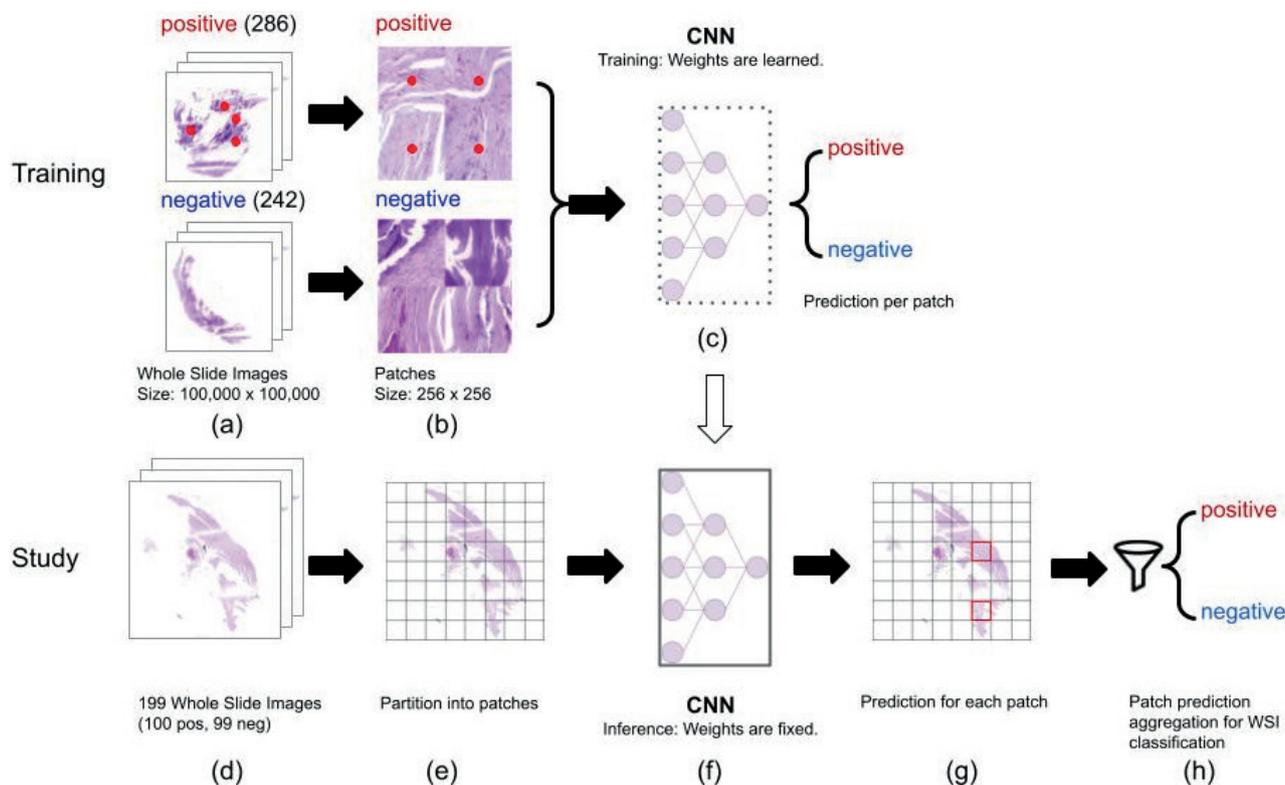


Fig. 2. Overview over artificial intelligence (AI) training and study. (a) A total of 528 periodic acid–Schiff-stained whole-slide images (WSI) were used for AI training. Markers for hyphae are shown as *red dots* (only for visualization of the annotations; red dots are not part of the actual training images). (b) Positive and negative patches were extracted for training from the WSI. (c) A convolutional neural network (CNN) was trained that predicts for a patch whether onychomycosis is present. (d) A total of 199 new WSI were used to study the performance of the AI in comparison with human experts. (e) To make an AI prediction for a WSI, the WSI was partitioned into patches. (f) The CNN is now fixed. (g) It predicts a probability for each patch independently. (h) All patch predictions were aggregated into a single probability estimate for the WSI.

chosen randomly by a computer. The WSI were analysed by the AI. For comparison, the original tissue slides were analysed with a conventional microscope by 4 different dermatopathologists of different levels of experience. Two dermatopathologists are full-time dermatopathologists with several years of experience, and the other 2 recently completed their specialist training in dermatopathology. None of the dermatopathologists had participated in the annotation process; and none of them was involved in the determination of the ground truth.

The aim of this confirmatory diagnostic accuracy study was to determine whether diagnosis of onychomycosis by software using AI is not inferior to an analogous diagnosis by dermatopathologists.

Statistical analysis

For a given case, the AI estimates a probability for the presence of onychomycosis. Using a cut-off value of 91% (previously established in the training dataset), these results are dichotomized into positive ($>91\%$) or negative ($\leq 91\%$). The primary hypothesis is that the diagnostic accuracy of AI, as measured by the area under the curve (AUC), sensitivity and specificity, is non-inferior to the diagnostic accuracy of human diagnosticians. The non-inferiority limit was set *a priori* at 5% absolute. Because sensitivity and specificity are co-primary endpoints, the type I error does not need to be corrected for multiplicity here. However, since the AUC is considered as an additional primary endpoint, the type I error was corrected according to the Bonferroni method. This results in a 2-sided type I error of 2.5% for all 3 hypotheses and corresponding 2-sided 97.5% confidence intervals for primary analyses.

First, for descriptive analysis the receiver operating characteristics curve (ROC curve) for the continuous AI probability was calculated. Furthermore, AUCs are given with associated 2-sided 95% logit confidence intervals. After dichotomizing the AI probability (with 91% as cut-off; see above), sensitivity and specificity were calculated for the different diagnoses, again with the corresponding 2-sided 95% logit confidence intervals.

For the primary analysis, a non-parametric factorial model with the 2 fixed factors method (AI vs analogous) and dermatopathologist (nested under method=analogous) was used (29). From this model, the mean sensitivity, specificity, and AUC of the 4 dermatopathologists were calculated. For dichotomous evaluation the AUC is equal to the arithmetic mean of sensitivity and specificity as an aggregated measure. Then, the differences in the sensitivities, specificities, and AUC (mean analogous minus AI) and the corresponding 2-sided 97.5% Wald confidence intervals were calculated. If the upper limit of the respective confidence interval was below the non-inferiority margin of 5% absolute, the corresponding non-inferiority null hypothesis could be rejected. In secondary analyses, the predictive values with corresponding 2-sided 95% logit confidence intervals were calculated. For the analysis, SPSS 25 (Statistical Package for Social Science, IBM (Armonk, New York, USA)) and the statistical software R (30) including the R package “diagnostic 0.4.2” (31) were used.

RESULTS

Of the total 199 samples, onychomycosis was present in 101 and not present in 98; a prevalence of 51%. There

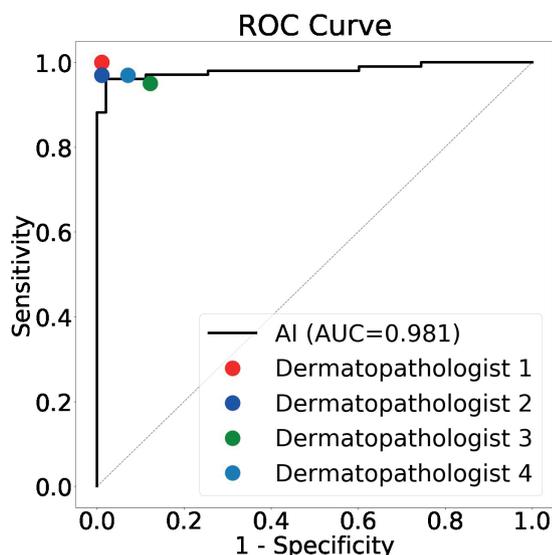


Fig. 3. The receiver operating characteristic (ROC) curve shows that the onychomycosis probabilities predicted by artificial intelligence (AI) allow it to sort cases on an accuracy level comparable to that of human experts.

were no missing values. **Fig. 3** shows the ROC curve of the AI probability and the pairs of sensitivity and specificity of the dermatopathologists. **Table I** shows the individual AUCs, sensitivities and specificities with corresponding 2-sided 95% logit confidence intervals (CI lower, CI upper). This shows that there are 2 dermatopathologists (1 and 2) who achieve a higher AUC and specificity, while 2 dermatopathologists (3 and 4) achieve a lower AUC and specificity than the AI, while all dermatopathologists achieved a higher sensitivity than the AI.

The results of the primary analyses are shown in **Table II**. The AI achieved an AUC operator curve of 98.1% (CI 96.1–99.8%) based on AI probabilities. After dichotomizing the AI, the mean analogous sensitivity of the dermatopathologists was 3.2% higher than the sensitivity of the AI (CI –2.3%; 8.8%), while the mean analogous specificity was 3.3% lower than the specificity of the AI (CI –7.8%; 1.2%). The AUC of the dichotomized AI and analogue AUC was similar (difference –0.1%, CI –3.6%; 3.5%). This means that non-inferiority of the deep learning system to the analogue diagnosis by histopathologists was shown with respect to AUC and specificity, but not with respect to sensitivity. **Fig. 4** shows patches that were misclassified by the AI.

Table II. Differences in sensitivities and specificities with corresponding 2-sided 97.5% Wald confidence intervals (low, up)

	AI with dichotomized probabilities	Analogous (mean)	Difference (An. – AI)	Diff. low	Diff. up
Sensitivity	94.1%	97.3%	3.2%	–2.3%	8.8%
Specificity	98.%	94.6%	–0.033	–7.8%	1.2%
AUC	0.9601	0.9596	–0.001	–0.036	0.035

AI: artificial intelligence; AUC: area under the curve.

In **Table III**, the results of the secondary analyses regarding the positive predictive value (PPV) and the negative predictive value (NPV) are provided. Interrater agreements of dermatopathologists among each other and with respect to the AI model are shown in **Table IV**.

DISCUSSION

Onychomycosis is a common nail infection. Nail clippings for microscopic diagnosis make up a considerable number of specimens in dermatopathology laboratories. Searching hyphae in PAS stained sections is a tedious task if the number of micro-organisms in the specimen is low. Moreover, as standard of care, multiple PAS-stained sections often have to be examined. This takes time, which a dermatopathologist could better use for more difficult diagnoses, e.g. in melanocytic lesions or inflammatory skin diseases, where weighting of criteria and integration of clinical context requires human expertise. With the increasing importance of AI in medicine and a high workload in pathology laboratories, AI-assisted systems for diagnosis are an interesting approach. Detecting onychomycosis represents a difficult task from an image recognition perspective. Target structures can occur only very sparsely in a whole-slide image, meaning that they can easily be missed. Furthermore, artefacts and bacteria particles can resemble hyphae and thus introduce false-positive cases that need to be identified correctly.

The current study developed an AI model, based on a CNN, for recognition of fungal organisms in scanned whole-slide images of PAS-stained sections of nail clippings, using training data compiled by 2 human experts. Seeking to develop a robust AI model that generalizes well across different whole-slide images and does not overfit on target structure characteristics, self-supervised learning was first performed to gain a good initialization of the network weights of the CNN. The resulting

Table I. Areas under the curve (AUCs), sensitivities and specificities with corresponding 2-sided 95% logit confidence intervals (low, up)

	AUC	AUC low	AUC up	Sens., %	Sens. low, %	Sens. up, %	Spec., %	Spec. low, %	Spec. up, %
Dermatopathologist 1	0.995	0.965	0.999	100	100	100	99	93.11	99.9
Dermatopathologist 2	0.980	0.948	0.993	97	91.1	99	99	93.1	99.9
Dermatopathologist 3	0.914	0.866	0.946	95	88.6	97.9	87.8	79.6	92.9
Dermatopathologist 4	0.949	0.909	0.973	97	91.1	99	92.9	85.7	96.6
AI with probabilities	0.981	0.961	0.998	–	–	–	–	–	–
AI with dichotomized probabilities	0.960	0.922	0.980	94.1	87.4	97.3	98	92.2	99.5

AI: artificial intelligence; Sens: sensitivity; Spec: specificity. Non bold is the corresponding 2-sided 95% logit confidence intervals.

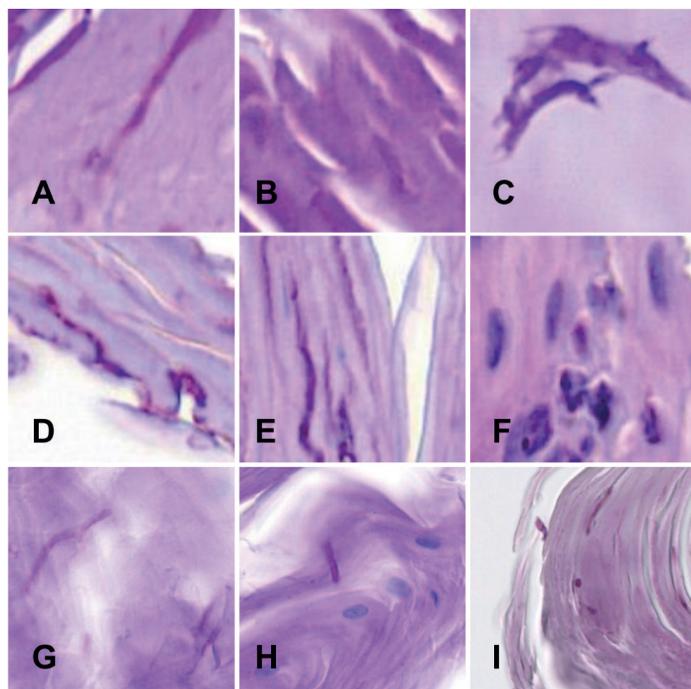


Fig. 4. Illustration of patches that were misclassified by artificial intelligence: (A–F) false-positives; (G–I) false-negatives. (A and B) Serum. (C) Overlay of nail fragments. (D and E) Bacteria in linear arrangement. (F) Remnants of neutrophils in zones of parakeratosis. (G and H) Hyphae were few and the area was not completely focused in the scan. (I) Hyphae were mainly crosscut. Magnification 400x.

network was then fine-tuned with supervised learning, based on the annotated training data.

The performance of the AI was evaluated on a second test set, consisting of entirely new cases. To set the “ground truth”, this test set was reviewed by 2 experienced dermatopathologists, and a large proportion of negative cases were also confirmed with PCR testing. The diagnoses made by the AI were compared with the diagnoses made by 4 other dermatopathologists of different levels of experience examining the same PAS-stained sections using a conventional microscope. A slight discordance was found between the dermatopathologists, which corresponded to their different levels of experience. Two dermatopathologists with several years of experience performed slightly better than the AI, while the AI performance was slightly superior to the other 2

Table III. Predictive values with corresponding 2-sided 95% logit confidence intervals

	PPV, %	PPV low, %	PPV up, %	NPV, %	NPV low, %	NPV up, %
Dermatopathologist 1	99.0	93.5	99.9	100.0	100.0	100.0
Dermatopathologist 2	99.0	93.4	99.9	97.0	91.6	99.0
Dermatopathologist 3	88.9	83.8	92.5	94.5	89.6	97.2
Dermatopathologist 4	93.3	88.1	96.4	96.8	91.4	98.9
AI with dichotomized probabilities	97.9	92.8	99.4	94.1	88.5	97.1

PPV: positive predictive value; NPV: negative predictive value. Non-bold is the corresponding 2-sided 95% logit confidence intervals.

dermatopathologists who had recently completed dermatopathology training. The results show that the AI was non-inferior to the dermatopathologists with regards to specificity and accuracy (AUC). However, non-inferiority regarding sensitivity could not be proven. We conclude that this AI system is equal to human experts.

High specificity of the AI is important, because, as mentioned previously, a diagnosis of onychomycosis may require systemic treatment with terbinafine or itraconazole (5), which can have severe side-effects. Onychomycosis should always be proven by demonstration of the micro-organism by a confirmatory method prior to systemic treatment, and this confirmation can be established by our AI. The AI is forced to present patches containing hyphae with the highest probability, even if present in a very small fragment. In a practical setting, the AI developed in this study could be used as a screening tool, presenting patches in the slide with areas suspicious for hyphae, which could then quickly be confirmed by a dermatopathologist. Such a system would also be of help to residents in training or less experienced dermatopathologists, who naturally have higher rates of error.

Only one study has previously addressed AI-assisted histopathological diagnosis of onychomycosis (32). A main difference from the current study was that neural network scanning was used to provide assistance for the pathologists, while, in our system, the AI learned to diagnose cases autonomously. Moreover, the technical background has evolved significantly since then, enabling higher quality of scanning and the use of deep learning. A recent study showed that a deep learning algorithm can surpass human knowledge in diagnosing melanoma, and that one of the reasons may be that AI can identify image features that are disregarded or not discernible by the human expert (33). This hypothesis may also be valid for the current study, because, in some cases, the algorithm outperformed the humans. This was already seen during the training phase, in which the AI identified suspicious areas in supposedly negative slides, which in 2 supplementary control steps were confirmed by an experienced dermatopathologist as positive.

A further advantage of the use of AI is the time saving, since traditional microscopic examination is time-consuming. An example is a deep learning approach for diagnosis of malaria (34). A similar binary approach

Table IV. Inter-rater agreements of dermatopathologists among each other and with respect to the artificial intelligence (AI) model

Dermatopathologist	2	3	4	AI
1	98.0%	90.5%	94.0%	95.0%
2	-	91.5%	93.0%	94.0%
3	-	-	90.5%	88.4%
4	-	-	-	91.0%

Bold highlights the AI in contrast to the pathologists.

was used to determine benign or malignant epithelial neoplasia, where accuracy and workflow efficiencies could be improved by computer-aided diagnostics (35).

Limitations

One of the limitations of using AI for analysing medical images is the need for a large number of training images (24). The current study training set comprised 528 whole-slide images; however, it was only feasible to annotate a small fraction of hyphae in each positive case. Self-supervised learning was used to counterbalance this. It was noticed that the AI had no difficulty correctly identify the hyphae, which are basically elongated pink structures. However, the AI could not easily recognize the spores, which are found more commonly in infections with *Candida albicans*. Another difficulty was that the AI confounded serum particles or aggregation of bacteria with fungal elements; however, beginner histopathologists encounter the same difficulties. When hyphae are cut transversely, they are seen as tiny round structures, which can be simulated by serum or bacterial aggregations. In this situation, a conventional microscope offers the advantage of focusing through the 4- μ m cut, which helps to identify hyphae correctly. Sometimes overlays and dirt or overstaining posed difficulties for the AI. Since all sections came from the same laboratory, stainings were relatively even. When applied to sections stained in different laboratories, the AI could initially encounter difficulties. Finally, the cost of the AI equipment may be a limitation; the scanner and software cannot be afforded by every laboratory.

Conclusion

Previous studies of AI in dermatopathology have shown that CNNs can be applied in various domains, such as routine diagnostics, education and research. However, more prospective studies are required to confirm the previous findings (26).

By using a common disease, AI was successfully trained in solving the binary problem of detecting onychomycosis in nail clippings. This study showed that the use of AI on whole-slide images is statistically non-inferior regarding AUC and specificity compared with dermatopathologists using a conventional analogous setting with a microscope. A potential application of the current AI model could be as support software that screens whole-slide images and highlights suspicious areas to the pathologist. These results may also lead to other possible uses of AI in dermatopathology. PAS stains used routinely in every biopsy of an inflammatory dermatosis could be screened for fungal organisms using a similar AI system. Important advantages are the possibilities of time saving and the chance to reduce incorrect diagnosis due to time pressure, high workload, or lack of experience in dermatopathological routine. The current

study demonstrates that AI systems provide promising opportunities as assisting tools in dermatopathology. A follow-up study is planned, to evaluate the current AI model on PAS-stained slides from other laboratories and for use of both nail clippings and skin.

The authors have no conflicts of interest to declare.

REFERENCES

1. Westerberg DP, Voyack MJ. Onychomycosis: current trends in diagnosis and treatment. *Am Fam Physician GP* 2013; 88: 762–770.
2. Reinel D. Non-dermatophyte fungi in onychomycosis-epidemiology and consequences for clinical practice. *Mycoses* 2021; 64: 694–700.
3. Lipner SR, Scher RK. Onychomycosis: clinical overview and diagnosis. *J Am Acad Dermatol* 2019; 80: 835–851.
4. Vanhootheghem O, Szepietuk G, Paurobally D, Heures F. Chronic interdigital dermatophytic infection: a common lesion associated with potentially severe consequences. *Diabetes Res Clin Pract* 2011; 91: 23–25.
5. Lipner SR, Scher RK. Onychomycosis: treatment and prevention of recurrence. *J Am Acad Dermatol* 2019; 80: 853–867.
6. Ghannoum M, Mukherjee P, Isham N, Markinson B, Rosso JD, Leal L. Examining the importance of laboratory and diagnostic testing when treating and diagnosing onychomycosis. *Int J Dermatol Venereol* 2018; 57: 131–138.
7. Bao F, Fan Y, Sun L, Yu Y, Wang Z, Pan Q, et al. Comparison of fungal fluorescent staining and ITS rDNA PCR-based sequencing with conventional methods for the diagnosis of onychomycosis. *J Eur Acad Dermatol Venereol* 2018; 32: 1017–1021.
8. Lawry MA, Haneke E, Strobeck K, Martin S, Zimmer B, Romano PS. Methods for diagnosing onychomycosis: a comparative study and review of the literature. *JAMA* 2000; 283: 1112–1116.
9. Chang HY, Jung CK, Woo JI, Lee S, Cho J, Kim SW, et al. Artificial intelligence in pathology. *J Pathol Transl Med* 2019; 53: 1–12.
10. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115–118.
11. Sargül M, Ozyildirim BM, Avci M. Differential convolutional neural network. *Neural Netw* 2019; 116: 279–287.
12. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med* 2019; 25: 24–29.
13. Du-Harpur X, Watt FM, Luscombe NM, Lynch MD. What is AI? Applications of artificial intelligence to dermatology. *Br J Dermatol* 2020; 183: 423–430.
14. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2018; 19: 1236–1246.
15. Chin YPH, Hou ZY, Lee MY, Chu HM, Wang HH, Lin YT, et al. A patient-oriented, general-practitioner-level, deep-learning-based cutaneous pigmented lesion risk classifier on a smartphone. *Br J Dermatol*, 2020; 182: 1498–1500.
16. Tschandl P, Argenziano G, Razzmaria M, Jap J. Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features. *Br J Dermatol* 2019; 181: 155–165.
17. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol* 2019; 180: 373–381.
18. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. *Med Image Comput Assist Interv* 2013; 16: 411–418.
19. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al diagnostic as-

- assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; 318: 2199–2210.
20. Sharma H, Zerbe N, Klempert I, Hellwich O, Hufnagl P. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Comput Med Imaging Graph* 2017; 61: 2–13.
 21. Höfener H, Homeyer A, Weiss N, Molin J, Lundström C. F, Hahn HK. Deep learning nuclei detection: A simple approach can deliver state-of-the-art results. *Comput Med Imaging Graph* 2018; 70: 43–52.
 22. Hogarty DT, Su JC, Phan K, Attia M, Hossny M, Nahavandi S, et al. Artificial intelligence in dermatology – where we are and the way to the future: a review. *Am J Clin Dermatol* 2020; 21: 41–47.
 23. Jiang YQ, Xiong JH, Li HY, Yang XH, Yu WT, Gao M, et al. Recognizing basal cell carcinoma on smartphone-captured digital histopathology images with a deep neural network. *Br J Dermatol* 2020; 182: 754–762.
 24. Han SS, Park GH, Lim W, Kim MS, Na JI, Park I, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS One* 2018; 13: e0191493.
 25. Aishwarya, Goel A, Nijhawan R. A deep learning approach for classification of onychomycosis nail disease. *Proc ICETIT* 2019; 605: 1112–1118.
 26. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations, ICLR, 2015*; arXiv:1409.1556v4 [cs.CV].
 27. Wells A, Patel S, Lee JB, Motaparathi K. Artificial intelligence in dermatopathology: Diagnosis, education, and research. *J Cutan Pathol* 2021; 48: 1061–1068.
 28. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. *International conference on machine learning. PMLR* 2020; 119: 1597–1607.
 29. Brunner EZA. *Nonparametric ROC analyses for diagnostic trials*. Newark, NJ: John Wiley & Sons; 2014: p. 471–483.
 30. Team RDC. *A language and environment for statistical computing*. R Foundation for statistical computing 2008. [Accessed March 30, 2021] Available from: <http://www.R-project.org>.
 31. Siemens W. *Covariate adjusted nonparametric estimation of sensitivity and specificity in clustered data*. Master thesis Heidelberg University, Institut für Medizinische Biometrie; 2017.
 32. Boon ME, Berger TH, Middag-Broekman AJ, Kok LP. Optimized quality of histologic images allows the use of neural network scanning in diagnosis of fungal infection of abnormal nails. *Anal Quant Cytol Histol* 2006; 28: 78–86.
 33. Hekler A, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur J Cancer* 2019; 118: 91–96.
 34. Yang F, Poostchi M, Yu H, Zhou Z, Silamut K, Yu J, et al. Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE J Biomed Health Inform* 2020; 24: 1427–1438.
 35. Olsen TG, Jackson BH, Feeser TA, Kent M. N, Moad JC, Krishnamurthy S, et al. diagnostic performance of deep learning algorithms applied to three common diagnoses in dermatopathology. *J Pathol Inform* 2018; 9: 32.