Volume 13, Issue 1, Page 152-160, 2024 https://doi.org/10.46810/tdfd.1442556 Research Article

BİNGÖL UNIVERSITY

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science

www.dergipark.gov.tr/tdfd

Classification of Microscopic Fungi Images Using Vision Transformers for Enhanced Detection of Fungal Infections

Abdurrahman GUMUS[1*]

[1] Izmir Institute of Technology, Faculty of Engineering, Electrical-Electronics Engineering, Izmir, Türkiye
Abdurrahman GUMUS ORCID No: 0000-0003-2993-5769

**Corresponding author: abdurrahmangumus@iyte.edu.tr*

(Received: 24.02.2024, Accepted: 19.03.2024, Online Publication: 26.03.2024)

Keywords
Vision transformer, Swin transformer, Image classification, Fungal infection, Microscopic images

Abstract: Fungi play a pivotal role in our ecosystem and human health, serving as both essential contributors to environmental sustainability and significant agents of disease. The importance of precise fungi detection cannot be overstated, as it underpins effective disease management, agricultural productivity, and the safeguarding of global food security. This research explores the efficacy of vision transformer-based architectures for the classification of microscopic fungi images of various fungal types to enhance the detection of fungal infections. The study compared the pre-trained base Vision Transformer (ViT) and Swin Transformer models, evaluating their capability in feature extraction and fine-tuning. The incorporation of transfer learning and fine-tuning strategies, particularly with data augmentation, significantly enhances model performance. Utilizing a comprehensive dataset with and without data augmentation, the study reveals that the Swin Transformer, particularly when fine-tuned, exhibits superior accuracy (98.36%) over the ViT model (96.55%). These findings highlight the potential of vision transformer-based models in automating and refining the diagnosis of fungal infections, promising significant advancements in medical imaging analysis.

Mantar Enfeksiyonlarının Gelişmiş Tespiti İçin Görüntü Dönüştürücüleri Kullanılarak Mikroskobik Mantar Görüntülerinin Sınıflandırılması

Anahtar Kelimeler
Görüntü dönüştürücüler, Swin dönüştürücü, Görüntü sınıflandırma, Mantar enfeksiyonu, Mikroskobik görseller

Öz: Mantarlar, hem çevresel sürdürülebilirliğe temel katkıda bulunarak, hem de önemli hastalık etmenleri olarak hizmet ederek, ekosistemimizde ve insan sağlığında kritik bir rol oynamaktadırlar. Mantarların hassas olarak tespiti, etkili hastalık yönetimi, tarımsal verimlilik ve küresel gıda güvenliğinin korunması açısından önemlidir. Bu araştırma, çeşitli mantar türlerinin mikroskobik görüntülerinin sınıflandırılmasında görüntü dönüştürücü tabanlı mimarilerin etkinliğini keşfetmekte ve mantar enfeksiyonlarının tespitini geliştirmeyi amaçlamaktadır. Çalışma, önceden eğitilmiş temel görüntü dönüştürücü (ViT) ve Swin dönüştürücü modellerini karşılaştırmış, özellik çıkarma ve ince ayarlanma yeteneklerini değerlendirmiştir. Nakil öğrenme ve ince ayar stratejilerinin, özellikle veri artırımı ile birlikte, model performansını önemli ölçüde artırdığı belirlenmiştir. Veri artırımı yapılmış ve yapılmamış kapsamlı bir veri setini kullanarak yapılan çalışma, ince ayar yapıldığında Swin dönüştürücünün (%98,36), ViT modeline kıyasla (%96,55) üstün doğruluk sergilediğini ortaya koymuştur. Bu bulgular, mantar enfeksiyonlarının tanısını otomatikleştirmede ve iyileştirmede vizyon dönüştürücü tabanlı modellerin potansiyelini vurgulamakta, tıbbi görüntüleme analizinde önemli ilerlemeler vaat etmektedir.

1. INTRODUCTION

Fungi are critical to addressing global challenges, significantly impacting both ecosystems and human health. They exist in diverse forms, from complex multicellular organisms to single-celled entities, thriving mostly on land in soil or plant matter. Fungi's dual nature is evident in their contributions to medicine, agriculture, and ecological processes through antibiotic production, food fermentation, and nutrient cycling, contrasted with their potential to cause diseases in humans, animals, and plants, leading to significant health and economic concerns [1–3].

Fungal infections pose a substantial challenge to global health, impacting millions of people each year across various demographics and geographic locations. Fungal diseases, ranging from minor infections like athlete's foot to severe systemic conditions, pose threats to humans, animals, and crops, impacting economic stability and food security [2,4–6]. Accurate fungi identification is crucial for enhancing human well-being, promoting effective disease treatment, optimizing agricultural practices, and preventing disease outbreaks. Early and precise diagnosis is key to effective treatment, reducing the risk of severe outcomes and ensuring appropriate care. Particularly, microscopic fungi are major culprits behind superficial infections, making precise identification crucial for successful treatment [7–9].

Traditional diagnostic methods, reliant on microscopy and culture techniques, face challenges in speed and specificity, leading to a demand for more advanced solutions. Automated fungi classification, leveraging deep learning algorithms, offers significant advantages in processing efficiency, diagnostic accuracy, and cost reduction. It enables rapid analysis of large image datasets, facilitating timely and informed decisions in healthcare and agriculture. Automated classification mitigates the risk of human error and supports research by providing consistent, reproducible results, which is critical for understanding disease patterns and contributing to public health efforts [7,8,10,11].

Recent advances in deep learning have shown promise in different fields [12–14], especially in medical image analysis [15–18]. The integration of deep learning techniques for fungal detection, especially through microscopic imaging, has demonstrated the ability to accurately identify various fungal species, highlighting the technology's potential in both medical and agricultural fields [8,9,19]. A multitude of studies have applied convolutional neural networks (CNNs) and, more recently, vision transformers, demonstrating their efficacy in distinguishing fungal species with high accuracy. These advances offer profound implications for understanding the morphological diversity of fungi and their impact on human health and food security [20–22].

For instance, the work of S. S. Gaikwad et al. highlights the application of CNN models to categorize fungi affecting apple plant leaves, achieving an impressive 88.9% accuracy using images from an accessible plant pathology dataset. This study illustrates the potential of deep learning in agricultural disease management by facilitating early detection and treatment of fungal infections in crops [20]. Similarly, L. Picek et al. introduced the Danish Fungi 2020 dataset, a comprehensive collection aiding in the fine-grained classification of fungal species. Their research underscores the challenges posed by highly unbalanced class distributions and complex class hierarchies in fungal identification. By comparing CNN models and vision transformers, they demonstrated the superior performance of vision transformers, with an accuracy of 80.45% and a notable reduction in classification error, showcasing the potential of these models in handling complex, fine-grained classification tasks [21]. Koo et al. developed a deep learning model with a regional convolutional neural network to detect fungal hyphae in microscopic images, achieving high sensitivity (95.2% for 100× and 99% for 40× magnification models) and specificity (100% for 100× and 86.6% for 40× magnification models), suggesting significant improvements over conventional fungal infection diagnostics [23]. Gao et al. (2021) developed an automated microscope coupled with a deep learning model, primarily ResNet-50, to enhance fungal detection in dermatological samples. The system demonstrated high sensitivity (99.5% for skin, 95.2% for nails) and specificity (91.4% for skin, 100% for nails), showcasing the potential to significantly improve efficiency and accuracy in fungal diagnostics in dermatology [24]. In another vein, M. A. Rahman et al. explored the classification of pathogenic fungi using deep CNN models across several well-known architectures. Their study, which achieved top accuracy of 65.35% with the DenseNet model, underscores the diverse capabilities of CNNs in processing microscopic images to distinguish among 89 different fungal genera, thereby contributing to faster and more accurate diagnostic processes [22].

C. J. P. Sopo, F. Hajati, and S. Gheisari's work further exemplifies the role of deep learning in fungi classification, experimenting with different CNN models and training approaches. Their findings, particularly the high performance of the VGG16 model under transfer learning, highlight the effectiveness of leveraging pre-trained models to enhance classification accuracy in specialized domains such as mycology [25]. Nawarathne et al. explored the classification of fungi images using various CNNs, addressing challenges like class imbalance through data augmentation and employing multiple preprocessing techniques. By testing thirteen pre-trained CNN models across different image resolutions, the research finds the BigTransfer (BiT) model, particularly with a mix of original and high-resolution images, to outperform others with an accuracy of 87.32%, with optimal precision, recall, and F1-score [26]. Cinar et al. investigated the application of deep learning techniques for detecting fungal infections from microscopic images, employing CNNs and transfer learning for accurate classification of various fungal species. Through data augmentation and fine-tuning, the study achieves significant improvements in accuracy as 97.19%, showcasing the potential of deep learning in enhancing diagnostic processes for fungal infections [27]. These studies exemplify the continuous evolution of methodologies in the classification and identification of fungal infections, promising to improve diagnostic processes through increased efficiency and reduced costs, with profound implications for global health, agriculture, and ecological preservation.

Our study delves into the application of deep learning techniques, specifically leveraging vision transformers within a framework that utilizes the DeFungi dataset, transfer learning, and data augmentation to enhance the accuracy of fungal detection from microscopic images.

The aim is to improve the identification and classification of fungal infections, thereby supporting clinical decision-making with rapid, accurate, and scalable diagnostic tools.

The remaining of the paper is organized as follows. In Section 2, the details of the methodology are given, such as dataset, data augmentation, vision transformer architectures, transfer learning, and experimental details. In Section 3, the results of the proposed methods and corresponding discussions are presented. The conclusion of the study is reported in the last section.

2. MATERIAL AND METHOD

In the Methodology section, an approach employing vision transformers for classifying fungi images into five distinct categories is detailed. This section describes the process of dataset selection and preparation, the application of data augmentation techniques to enhance model performance, the experimental setup, and the architecture of the vision transformers utilized. Additionally, the application of transfer learning and fine-tuning processes aimed at enhancing the accuracy of the model is described.

2.1. Dataset

The DeFungi dataset utilized for this study was acquired from the University of California Irvine (UCI) Machine Learning Repository [25]. The images in the dataset were provided by a mycological laboratory in Colombia. The dataset contains images depicting superficial fungal infections attributed to yeasts, molds, and dermatophyte fungi. These images underwent a detailed classification process, being manually sorted into five distinct categories with the assistance of domain experts to ensure the categorization's accuracy and relevance. Subsequently, automated coding procedures were employed to crop and patch these images. The finalized dataset includes a total of 9,114 images, distributed across five categories. The fungi image samples from the dataset are shown in Figure 1.

2.2. Data Augmentation

Data augmentation was implemented to enhance the diversity of the training dataset, especially to increase the representation of underrepresented classes. The augmentation aimed to balance the number of images across all categories, as depicted in Table 1, which illustrates the image counts in each class before and after augmentation. The augmentation procedure involved four specific transformations applied to the images in the dataset such as vertical flip, horizontal flip, and 45-degree rotations.

For each class (H1, H2, H3, H5, and H6), the number of images was augmented to match a target value. The target for classes H2, H3, H5, and H6 was set to approximately match the initial count of H1, which had the most images at 4404.

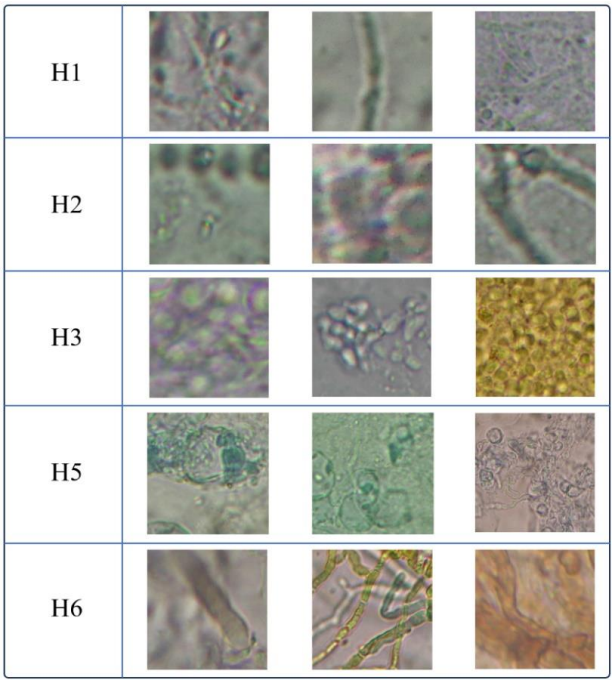

Figure 1. Representative images illustrating five distinct classes within the fungi dataset. H1) Tortuous septate hyaline hyphae (TSH), H2) Beaded arthroconidial septate hyaline hyphae (BASH), H3) Groups or mosaics of arthroconidia (GMA), H5) Septate hyaline hyphae with chlamydioconidia (SHC), and H6) Broad brown hyphae (BBH).

This target was selected to ensure that each class had a similar number of images, thus preventing class imbalance that could bias the model training. The augmentation script dynamically generated the required number of images by applying the aforementioned transformations consecutively to existing images until the target count was reached. The generated images were saved with a unique naming convention indicating the class, the type of transformation, and a sequence number to ensure uniqueness. This data augmentation process allowed the creation of additional training data, which was essential for training robust models capable of generalizing well across different presentations of fungal classes. It increased the total number of images in the dataset from 9,114 to 22,004, significantly enriching the dataset and providing more variability for model training.

Table 1. Number of images for different classes before and after data augmentation.

Classes	Original Data	With Data Augmentation
H1	4404	4404
H2	2334	4400
H3	819	4400
H5	818	4400
H6	739	4400
Total	**9114**	**22004**

2.3. Vision Transformer Architectures

In this study, we explore the application of vision transformer models for image classification tasks. Vision transformers represent an adaptation of the transformer architecture, originally designed for natural language processing [28], to the domain of computer vision.

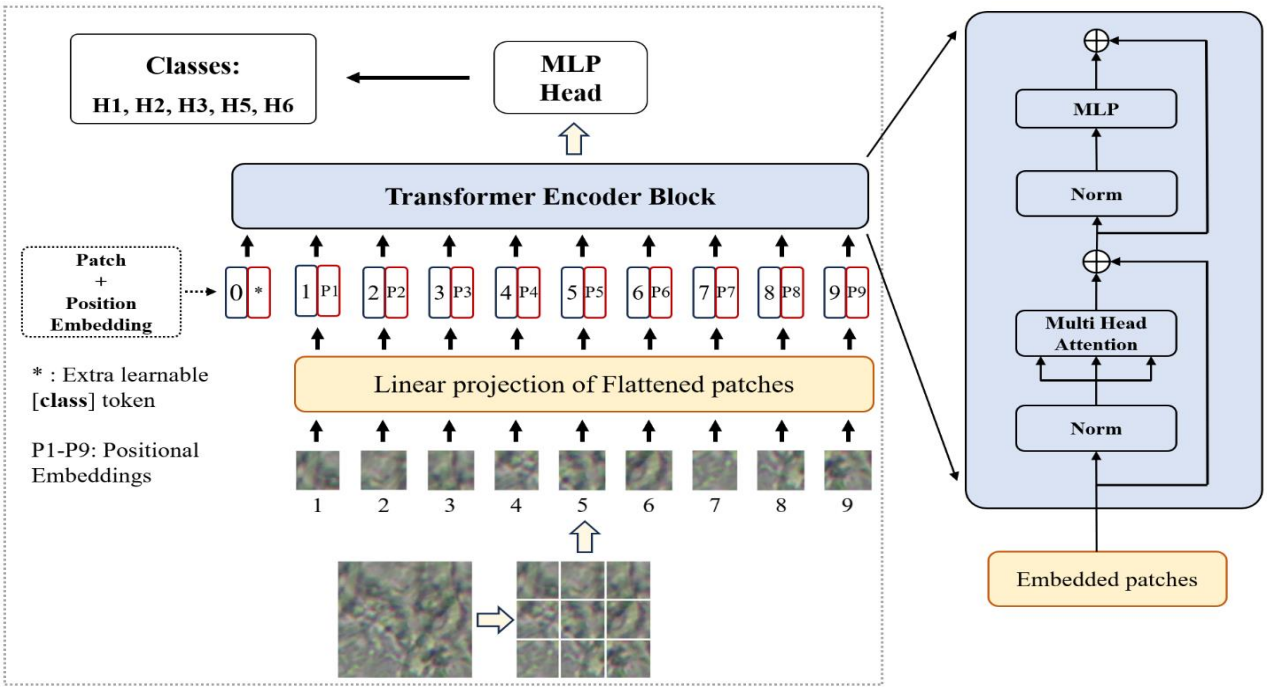

Figure 2. Schematic representation of the Vision Transformer (ViT) model utilized in the study, illustrating the transformer architecture adapted for the task of classifying fungi images into five categories, highlighting the sequence of image patches, positional embeddings, multi-head attention, and the fully connected layers.

Unlike traditional convolutional neural networks (CNNs), vision transformers divide an image into a series of patches and process these patches as a sequence, employing self-attention mechanisms to capture global dependencies within the image [29,30].

In the first model, the base Vision Transformer (ViT) model is utilized [29]. This approach segments an image into fixed-size patches, linearizes these patches, and then processes them through a series of transformer blocks that include multi-headed self-attention and position-wise fully connected layers. The base ViT model treats the image as a sequence of patches to learn representations, relying on the self-attention mechanism to weigh the importance of each patch relative to others in the image. The architecture of the ViT model is shown in Figure 2.

Additionally, the Swin Transformer is employed, which is a variation of the standard ViT that introduces a hierarchical structure through the use of shifted windows [31]. This design allows the Swin Transformer to efficiently manage computational resources by focusing on smaller sections of the image at lower levels of the hierarchy before progressively merging these sections at higher levels. The Swin Transformer's unique approach to partitioning the image and processing it in stages enables it to adaptively focus on different scales of image features, potentially offering advantages in capturing both local and global image contexts. Figure 3 shows the Swin Transformer architecture.

2.4. Transfer Learning and Fine-Tuning

Transfer learning is a machine learning methodology where a model trained on one task is repurposed for a second related task [32]. This approach is often utilized in image classification, with models pre-trained on large and diverse datasets, such as ImageNet, to leverage learned features and patterns applicable across various visual domains [33]. By starting with models that have already learned a broad representation of images, the training process can be accelerated, higher performance achieved with less labeled data, and model generalization improved.

Fine-tuning represents a specific application of transfer learning, where the pre-trained model is further adjusted to suit the target task more closely. This process involves unfreezing some or all of the layers of the pre-trained

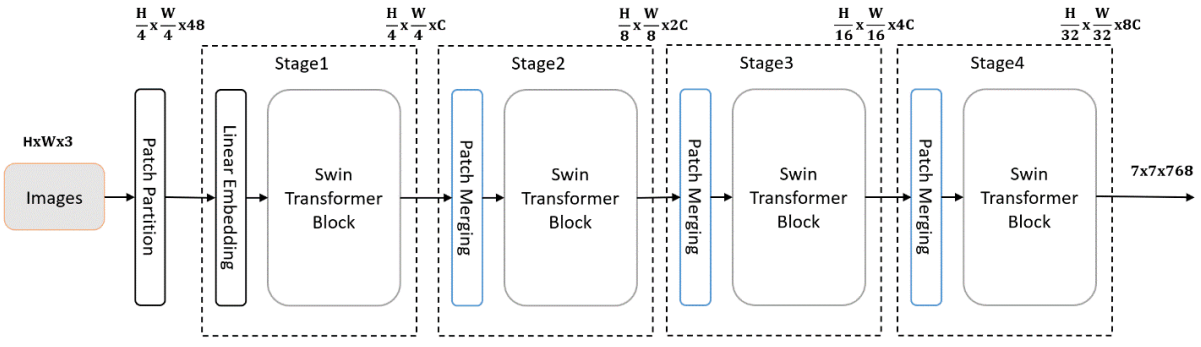

Figure 3. Illustration of the Swin Transformer utilized for image classification, detailing the mechanism of shifted windows and embedded patch processing for feature extraction, as applied to the categorization of complex image data.

model and continuing the training process on the new dataset. The extent of the fine-tuning can vary, from adjusting just the final layers to retraining the entire network. This allows the model to adapt the generic features learned from the initial dataset to the nuances of the new task, enhancing performance on the target domain [32,34].

In this study, both the base ViT and Swin Transformer models were employed, leveraging their pre-training on the ImageNet dataset [35] to harness a rich feature set relevant to the image classification tasks at hand. Two principal approaches were taken: feature extraction and fine-tuning. In feature extraction, the pre-trained models were used as feature extractors. Images were passed through the models to obtain high-level feature representations from the layers preceding the output layer. These features were then used as input to fully connected layers designed for the classification tasks in this study. This method benefited from the deep and complex representations learned by the models without necessitating extensive retraining. The second approach involved fine-tuning the pre-trained models on the specific fungi dataset of this study. This involved making minor adjustments to the model parameters to better align with the specific features and distributions of the task. After fine-tuning, features extracted from these adjusted models were fed into fully connected layers for classification.

2.5. Experimental Details

The Python programming language was used for the preprocessing and training operations. The vision transformer models are implemented in the PyTorch framework. Computational experiments were performed on a workstation with the following properties: Intel i5-12600k CPU, 64 GB RAM, and NVIDIA RTX 3090 Ti 24 GB GPU.

Two distinct approaches for image classification were implemented utilizing the base ViT and Swin Transformer models. The first approach involves fine-tuning, where the pretrained vision transformers are adapted by replacing their classification head with a new fully connected layer tailored to our specific number of output classes. This method allows the entire network to learn and adjust to our dataset during training. The second approach is feature extraction, where vision transformer models are utilized to generate feature representations of the images, effectively freezing the pretrained layers and only training a series of new fully connected layers for classification. This custom network consists of dense layers with ReLU activations, batch normalization, and dropout for regularization, concluding with a softmax layer for output. The training involved optimizing these layers with an Adam optimizer and categorical cross-entropy loss, using the features extracted from vision transformers as input. Hold-out validation is utilized to assess the performance of image classification models. Data is allocated as 70% for training, 15% for validation, and the remaining 15% for testing.

To assess the performance of our image classification models, four metrics are employed: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly identified images across the dataset, providing an overall effectiveness of the model. Precision quantifies the accuracy of positive predictions, indicating the model's ability to minimize false positives. Recall assesses the model's capability to correctly identify all relevant instances, reflecting its sensitivity. F1-score combines precision and recall into a single metric, offering a balanced view of the model's performance by accounting for both the precision's and recall's contributions.

3. RESULTS AND DISCUSSION

The detection of fungi plays a pivotal role in addressing key global challenges, impacting both human health and ecosystems. Fungi, which range from simple single-celled organisms to complex multicellular forms, are significant for their contributions to medicine, agriculture, and ecological balance, but also pose serious health and economic risks through disease [1–3]. Accurate identification of fungi is essential for effective disease management, enhancing agricultural efficiency, and safeguarding food security. Traditional diagnostic methods often fall short in speed and accuracy, underscoring the need for advanced, automated classification techniques. Leveraging deep learning-based algorithms for fungi detection promises greater diagnostic precision, efficiency, and cost-effectiveness, crucial for early and accurate diagnosis, ultimately supporting healthcare, agriculture, and ongoing research efforts [7–9].

In this study, the efficacy of vision transformer models for classifying medical images, specifically for fungi diagnosis, was assessed. The findings suggest that vision transformer models are potentially more adept at capturing detailed and nuanced features from complex medical images than conventional approaches, indicating their capacity to improve diagnostic accuracy. This paper outlines the model's accuracy rates, feature extraction capabilities, and the outcomes of fine-tuning, alongside discussing the progress in fungi image classification facilitated by the use of vision transformer models.

The comparison between Convolutional Neural Networks (CNNs) and vision transformer models reveals unique strengths and challenges for each technology. CNNs excel in identifying spatial features in images due to their strong image analysis capabilities and are more computationally efficient, particularly with smaller datasets or images. However, they face challenges in modeling long-range dependencies, though there are methods to address this issue. Thanks to their localized processing and shared weight architecture, CNNs typically offer better generalization on smaller datasets. Conversely, vision transformers are capable of identifying complex global patterns by analyzing images as a whole, which requires more computational resources, especially for processing larger images.

Table 2. Model performance metrics for feature extraction and fine-tuning. All values are percentages.

Data Augmentation	Method	Model	Accuracy	Precision	Recall	F1-Score
No	Feature Extraction	Swin Transformer	82.32	82.43	82.32	81.40
No	Feature Extraction	ViT	86.70	86.56	86.70	86.48
No	Fine Tuning	Swin Transformer	90.87	90.84	90.87	90.80
No	Fine Tuning	ViT	88.39	88.44	88.39	88.35
Yes	Feature Extraction	Swin Transformer	92.61	92.71	92.61	92.57
Yes	Feature Extraction	ViT	94.34	94.35	94.34	94.33
Yes	Fine Tuning	Swin Transformer	98.36	98.38	98.36	98.36
Yes	Fine Tuning	ViT	96.55	96.55	96.55	96.54

Their use of self-attention mechanisms allows them to effectively handle long-range dependencies. Additionally, vision transformers tend to scale better with larger datasets compared to CNNs [30,36–38].

The architectural distinctions between the models play a pivotal role in their ability to manage the complexities inherent in various classification challenges. The base ViT model, with its straightforward approach to treating images as sequences of patches, demonstrates remarkable efficiency in capturing global image features. However, its performance indicates potential limitations in scenarios requiring a nuanced understanding of local features due to its uniform treatment of image patches. In contrast, the Swin Transformer's hierarchical, shifted window approach introduces a level of adaptability and efficiency not present in the base ViT model. By processing images in stages and allowing for variable-sized representation, the Swin Transformer exhibits a more nuanced capability to balance between local and global feature recognition. This structural difference notably enhances the model's performance on complex image classification tasks, where the interplay of local and global image features is critical. A key advantage of the Swin Transformer is its linear computational complexity, achieved by applying self-attention in a localized manner. On the other hand, the ViT model faces quadratic computational complexity due to its global application of self-attention across all patches.

The comprehensive results presented in Table 2, in conjunction with the visual examples from Figure 1, shed light on the ability of ViT and Swin Transformer models to handle the complexity and variability inherent in microscopic fungi images. The distinct textural and morphological features across the five classes, as depicted in Figure 1, underscore the challenges faced in accurate fungi classification.

When the models were not fine-tuned nor exposed to data augmentation, the ViT model displayed a commendable aptitude for feature extraction with an accuracy of 86.70%. This suggests an innate capability of ViT model to discern proper features even in unaltered datasets. The Swin Transformer lagged slightly behind with an 82.32% accuracy rate. However, once fine-tuned, the Swin Transformer's accuracy improved remarkably to 90.87%, outperforming the fine-tuned ViT model's accuracy of 88.39%.

This improvement hints at the Swin Transformer's ability to adapt its architecture more effectively upon learning from the dataset's specific characteristics.

The introduction of data augmentation substantially increased the number of images across all categories, with the total image count rising from 9,114 to 22,004. This enhancement in dataset volume, particularly for underrepresented classes such as H3, H5, and H6, has contributed to the notable performance gains observed in the models. The feature extraction capacity of the Swin Transformer enhanced to 92.61% accuracy, while the ViT model reached 94.34%, indicating a significant positive impact of data augmentation in preparing the models to recognize and generalize from the diversified visual data. Notably, the fine-tuning of both models with data augmentation yielded the most impressive outcomes, with the Swin Transformer reaching an accuracy of 98.36% and the ViT model an accuracy of 96.55%. These results reflect the models' enhanced ability to classify complex patterns observed in the varied images, achieving high precision, recall, and F1-scores uniformly across classes.

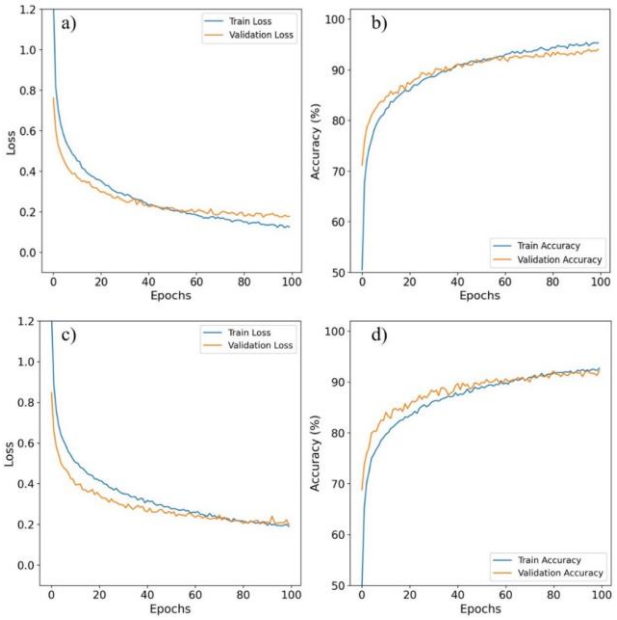

Figure 4. Training performance comparison for the ViT and Swin Transformer models during the feature-extraction approach with augmented data. Subfigures (a) and (b) illustrate the accuracy and loss plots for the ViT model, while subfigures (c) and (d) show the same for the Swin Transformer model.

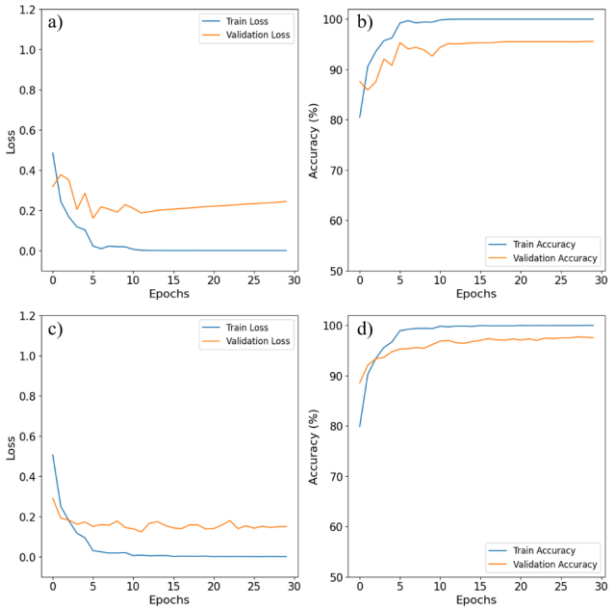

Figure 5. Training performance comparison for the ViT and Swin Transformer models during the fine-tuning approach with augmented data. Subfigures (a) and (b) depict the accuracy and loss plots for the ViT model, whereas subfigures (c) and (d) present these plots for the Swin Transformer model.

The use of transfer learning and fine-tuning in this study is justified by their demonstrated ability to enhance model performance, especially in situations with limited labeled data. By beginning with models that have learned comprehensive visual representations, the need for extensive computational resources and time required to train complex models from scratch is reduced. The nuanced textural differences and class-specific features highlighted in Figure 1 may explain the models' improved performance when fine-tuned with data augmentation, as the process likely aids the models in learning to differentiate subtle variations and complex patterns within and across the classes.

Figures 4 and 5 display the training performance of the ViT and Swin Transformer models with augmented data. In feature extraction (Figure 4), both models exhibit high validation accuracy, with ViT reaching convergence swiftly, indicative of efficient feature transfer. For fine-tuning (Figure 5), the Swin Transformer slightly outperforms ViT in validation accuracy, suggesting a better adaptation to the fungi classification task. Figure 6 depicts confusion matrices for ViT and Swin Transformer models with augmented data. Both models exhibit high classification accuracy across all classes, with improved precision upon fine-tuning, as shown by denser diagonals in Figures 6c and 6d. Fine-tuning particularly enhances the Swin Transformer's ability to distinguish between the more confusable classes.

In conclusion, the exemplary performance of the Swin Transformer and ViT models, especially when fine-tuned with augmented data, demonstrates their robustness and adaptability in classifying high-variability microscopic images. Such findings are promising for the field of automated medical diagnosis, suggesting that with sufficient training and data enhancement, these models can potentially serve as

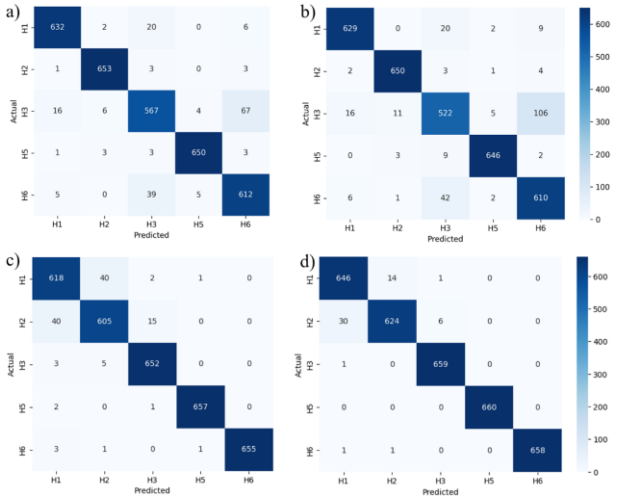

Figure 6. Confusion matrices demonstrating classification accuracy with augmented data for (a, c) the ViT model and (b, d) the Swin Transformer model, with (a, b) illustrating results from feature extraction approach and (c, d) fine-tuning approach, respectively.

reliable tools for accurate fungi classification, ultimately aiding in the prompt and precise diagnosis of fungal diseases.

4. CONCLUSION

Understanding the intricate role of fungi within our ecosystems and their impact on human health emphasizes the need for nuanced approaches to their detection and management. The development of advanced diagnostic tools stands as a critical step towards mitigating the risks fungi pose, while also leveraging their beneficial properties for environmental and medical applications. This research underscores the transformative potential of employing vision transformer-based models in the classification of microscopic fungi images. The Swin Transformer model, especially with fine-tuning and data augmentation, emerges as particularly effective, demonstrating a notable accuracy improvement (98.36%). These results not only advocate for the integration of vision transformer technologies in medical diagnostics but also mark a promising avenue for future research. Further exploration into optimizing these models for broader diagnostic applications and investigating their performance in diverse medical imaging contexts could lead to significant enhancements in automated disease detection, offering a new horizon for precision medicine and healthcare.

REFERENCES

[1] Lange L. The importance of fungi and mycology for addressing major global challenges. IMA Fungus 2014;5:463–71. https://doi.org/10.5598/imafungus.2014.05.02.10.

[2] Almeida F, Rodrigues ML, Coelho C. The still underestimated problem of fungal diseases worldwide. Front Microbiol 2019;10:1–5. https://doi.org/10.3389/fmicb.2019.00214.

[3] Ravikant KT, Gupte S, Kaur M. A Review on Emerging Fungal Infections and Their Significance.

J Bacteriol Mycol Open Access 2015;1:39–41. https://doi.org/10.15406/jbmoa.2015.01.00009.

[4] Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: Human fungal infections. Sci Transl Med 2012;4:1–9. https://doi.org/10.1126/scitranslmed.3004404.

[5] Grosjean P, Weber R. Fungus balls of the paranasal sinuses: A review. Eur Arch Oto-Rhino-Laryngology 2007;264:461–70. https://doi.org/10.1007/s00405-007-0281-5.

[6] Hernandez H, Martinez LR. Relationship of environmental disturbances and the infectious potential of fungi. Microbiol (United Kingdom) 2018;164:233–41. https://doi.org/10.1099/mic.0.000620.

[7] Kristensen K, Ward LM, Mogensen ML, Cichosz SL. Using image processing and automated classification models to classify microscopic gram stain images. Comput Methods Programs Biomed Updat 2023;3:100091. https://doi.org/10.1016/j.cmpbup.2022.100091.

[8] Zhang Y, Jiang H, Ye T, Juhas M. Deep Learning for Imaging and Detection of Microorganisms. Trends Microbiol 2021;29:569–72. https://doi.org/10.1016/j.tim.2021.01.006.

[9] Kumar S, Arif T, Alotaibi AS, Malik MB, Manhas J. Advances Towards Automatic Detection and Classification of Parasites Microscopic Images Using Deep Convolutional Neural Network: Methods, Models and Research Directions. Arch Comput Methods Eng 2023;30:2013–39. https://doi.org/10.1007/s11831-022-09858-w.

[10] Tahir MW, Zaidi NA, Rao AA, Blank R, Vellekoop MJ, Lang W. A fungus spores dataset and a convolutional neural network based approach for fungus detection. IEEE Trans Nanobioscience 2018;17:281–90. https://doi.org/10.1109/TNB.2018.2839585.

[11] Mital ME, Ruzcko Tobias R, Villaruel H, Maningo JM, Kerwin Billones R, Vicerra RR, et al. Transfer learning approach for the classification of conidial fungi (genus aspergillus) thru pre-trained deep learning models. IEEE Reg 10 Annu Int Conf Proceedings/TENCON 2020;2020-Novem:1069–74. https://doi.org/10.1109/TENCON50793.2020.9293803.

[12] Mohammad-Rahimi H, Rokhshad R, Bencharit S, Krois J, Schwendicke F. Deep learning: A primer for dentists and dental researchers. J Dent 2023;130:104430. https://doi.org/10.1016/j.jdent.2023.104430.

[13] Demir HO, Parlat SZ, Gumus A. Ethereum Blockchain Smart Contract Vulnerability Detection Using Deep Learning. ISAS 2023 - 7th Int Symp Innov Approaches Smart Technol Proc 2023:1–5. https://doi.org/10.1109/ISAS60782.2023.1039179.

[14] Kayan CE, Yuksel Aldogan K, Gumus A. Intensity and phase stacked analysis of a Φ-OTDR system using deep transfer learning and recurrent neural networks. Appl Opt 2023;62:1753. https://doi.org/10.1364/ao.481757.

[15] Ahmad A, Saraswat D, El Gamal A. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. Smart Agric Technol 2023;3:100083. https://doi.org/10.1016/j.atech.2022.100083.

[16] Aslani S, Jacob J. Utilisation of deep learning for COVID-19 diagnosis. Clin Radiol 2023;78:150–7. https://doi.org/10.1016/j.crad.2022.11.006.

[17] Duzyel O, Catal MS, Kayan CE, Sevinc A, Gumus A. Adaptive resizer-based transfer learning framework for the diagnosis of breast cancer using histopathology images. Signal, Image Video Process 2023;17:4561–70. https://doi.org/10.1007/s11760-023-02692-y.

[18] Müjdat Tiryaki V. Mass segmentation and classification from film mammograms using cascaded deep transfer learning. Biomed Signal Process Control 2023;84:104819. https://doi.org/10.1016/j.bspc.2023.104819.

[19] Zieliski B, Sroka-Oleksiak A, Rymarczyk D, Piekarczyk A, Brzychczy-Woch M. Deep learning approach to describe and classify fungi microscopic images. PLoS One 2020;15:1–16. https://doi.org/10.1371/journal.pone.0234806.

[20] Gaikwad SS, Rumma SS, Hangarge M. Fungi Classification using Convolution Neural Network. Turkish J Comput Math Educ 2021;12:4563–9.

[21] Picek L, Sulc M, Matas J, Jeppesen TS, Heilmann-Clausen J, Lassoe T, et al. Danish Fungi 2020 - Not Just Another Image Recognition Dataset. Proc - 2022 IEEE/CVF Winter Conf Appl Comput Vision, WACV 2022 2022:3281–91. https://doi.org/10.1109/WACV51458.2022.00334.

[22] Rahman MA, Clinch M, Reynolds J, Dangott B, Meza Villegas DM, Nassar A, et al. Classification of fungal genera from microscopic images using artificial intelligence. J Pathol Inform 2023;14. https://doi.org/10.1016/j.jpi.2023.100314.

[23] Koo T, Kim MH, Jue MS. Automated detection of superficial fungal infections from microscopic images through a regional convolutional neural network. PLoS One 2021;16:1–11. https://doi.org/10.1371/journal.pone.0256290.

[24] Gao W, Li M, Wu R, Du W, Zhang S, Yin S, et al. The design and application of an automated microscope developed based on deep learning for fungal detection in dermatology. Mycoses 2021;64:245–51. https://doi.org/10.1111/myc.13209.

[25] Sopo C, Hajati F, Gheisari S. Defungi: Direct mycological examination of microscopic fungi images. Arxiv 2021. https://doi.org/10.48550/arXiv.2109.07322.

[26] Nawarathne UMMPK, Kumari HMNS. Classification of Fungi Images Using Different Convolutional Neural Networks. 2023 8th Int Conf Inf Technol Res 2023.

[27] Cinar I, Taspinar YS. Detection of Fungal Infections from Microscopic Fungal Images Using Deep Learning Techniques. Proc Int Conf Adv Technol 2023. https://doi.org/10.58190/icat.2023.12.

[28] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. Adv Neural Inf Process Syst 2017;30.
[29] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. ICLR 2021 - 9th Int. Conf. Learn. Represent., 2021.
[30] Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in Vision: A Survey. ACM Comput Surv 2022;54. https://doi.org/10.1145/3505244.
[31] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. 2021 IEEE/CVF Int Conf Comput Vis 2021.
[32] Iman M, Arabnia HR, Rasheed K. A Review of Deep Transfer Learning and Recent Advancements. Technologies 2023;11:1–14. https://doi.org/10.3390/technologies11020040.
[33] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. Proc IEEE 2021;109:43–76. https://doi.org/10.1109/JPROC.2020.3004555.
[34] Öztürk C, Taşyürek M, Türkdamar MU. Transfer learning and fine-tuned transfer learning methods' effectiveness analyse in the CNN-based deep learning models. Concurr Comput Pract Exp 2023;35:1–26. https://doi.org/10.1002/cpe.7542.
[35] Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conf Comput Vis Pattern Recognit 2010:248–55. https://doi.org/10.1109/cvpr.2009.5206848.
[36] Maurício J, Domingues I, Bernardino J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. Appl Sci 2023;13. https://doi.org/10.3390/app13095521.
[37] Liu Y, Zhang Y, Wang Y, Hou F, Yuan J, Tian J, et al. A Survey of Visual Transformers. IEEE Trans Neural Networks Learn Syst 2023;PP:1–21. https://doi.org/10.1109/TNNLS.2022.3227717.
[38] Jamil S, Jalil Piran M, Kwon OJ. A Comprehensive Survey of Transformers for Computer Vision. Drones 2023;7:1–27. https://doi.org/10.3390/drones7050287.