IET Cyber-Systems and Robotics

WILEY

ZHEJIANG UNIVERSITY PRESS | IET The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

MASA-Net: Multi-Aspect Channel–Spatial Attention Network With Cross-Layer Feature Aggregation for Accurate Fungi Species Identification

Indranil Bera[1] | Rajesh Mukherjee[1] | Bidesh Chakraborty[2]

[1]Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, West Bengal, India | [2]Department of Computer Science and Engineering (AIML), Haldia Institute of Technology, Haldia, West Bengal, India

Correspondence: Bidesh Chakraborty (bidesh.me@hithaldia.ac.in)

Received: 27 April 2025 | **Revised:** 12 July 2025 | **Accepted:** 28 July 2025

Funding: The authors received no specific funding for this work.

Keywords: attention | convolutional neural network (CNN) | cross-layer feature aggregation | fungi Classification | transfer learning

ABSTRACT

Accurate identification of fungal species is essential for effective diagnosis and treatment. Traditional microscopy-based methods are often subjective and time-consuming. Deep learning has emerged as a promising tool in this domain. However, existing deep learning models often struggle to generalise in the presence of class imbalance and subtle morphological differences, which are common in fungal image datasets. This study proposes MASA-Net, a deep learning framework that combines a fine-tuned DenseNet201 backbone with a multi-aspect channel–spatial attention (MASA) module. The attention mechanism refines spatial and channel-wise features by capturing multi-scale spatial patterns and adaptively emphasising informative channels. This enhances the network's ability to focus on diagnostically relevant fungal structures while suppressing irrelevant features. The MASA-Net is evaluated on the DeFungi dataset and demonstrates superior performance in terms of accuracy, precision, recall and *F*1-score. It also outperforms established attention mechanisms such as squeeze-and-excitation networks (SE) and convolutional block attention module (CBAM) under identical conditions. These results highlight MASA-Net's robustness and effectiveness in addressing class imbalance and structural variability, offering a reliable solution for automated fungal species identification.

1 | Introduction

Fungi are essential for ecological balance. They play a vital role in the recycling of nutrients, and contribute to various natural processes [1]. Since their recognition as a separate kingdom in 1969 [2], fungi have been found to exist in a wide range of forms —from unicellular yeasts to complex multicellular organisms. They are also very much beneficial in industries such as drug manufacturing, food fermentation and agriculture [3]. However, their pathogenic potential poses significant threats to human health, causing diseases affecting multiple organs. Severe fungal infections may lead to fatal complications, which highlight the urgent need for effective prevention and diagnostic strategies [3–7]. The COVID-19 pandemic further underscored this need. It has become a growing concern globally, with over 300 million people affected annually, resulting in an estimated 1.5 million deaths [8, 9].

Despite the growing awareness and severity of fungal infections, effective diagnosis remains a challenge. Current methods, such as culturing, are often slow and lack accuracy. These traditional approaches require processing samples, such as blood, skin, or nail scrapings, and culturing them for up to 31 days, followed by morphological analysis [10, 11]. The lengthy turnaround time

and potential for false negatives can significantly delay treatment, leading to negative patient outcomes. Furthermore, these manual processes rely heavily on trained mycologists, which can introduce variability and errors, particularly when handling a high volume of samples.

To address these limitations, computer vision and deep learning techniques offer the potential to revolutionise fungal infection diagnosis. By providing faster and more reliable results, these methods can enable earlier detection and treatment.

Furthermore, the rising prevalence of superficial fungal infections caused by microscopic fungi underscores the urgent need for advanced diagnostic tools [12]. AI-based computer-aided system has shown high potential in faster and accurate identification and diagnosis [13], due to advancement in artificial intelligence in recent decade [14].

In particular, deep learning algorithms have shown promising capacity in automated classification. This has not only improved diagnostic accuracy but also reduced human error and accelerated the analysis of large datasets [5, 15–17]. Specifically, convolutional neural networks (CNNs) and vision transformers (ViTs) have already demonstrated their ability to accurately identify various fungal species, aiding in both medical and agricultural contexts [18–23]. Gaikwad et al. [24] demonstrated that CNN models can achieve high accuracy rates in detecting fungi in plant leaves.

Building on this, ViTs in particular have exhibited advantages over CNNs and other traditional approaches by effectively handling complex datasets, leading to superior performance in fungal species classification [25]. Moreover, deep learning models have demonstrated improved sensitivity and specificity compared to conventional diagnostic methods in the detection of fungal hyphae in microscopic images [26].

However, even with these improvements, there are still gaps that are not yet addressed by current deep learning approaches. Specifically, the complexity of the existing models often necessitates substantial computational resources, and the inherent class imbalance presented in many fungal datasets poses a significant challenge to model generalisation.

In this study, we propose a novel framework that combines the representational power of pre-trained DenseNet201 with a custom attention mechanism to dynamically prioritise discriminative features while suppressing irrelevant background noise. We have evaluated the performance of the proposed architecture in the DeFungi dataset [27]. The results underscore the potential of the proposed framework to advance computer-aided diagnosis (CAD) systems for early mycological assessment.

2 | Related Works

The increasing use of deep learning in identification-based applications, such as fungal detection and classification, signifies a notable trend. Recent advancements in image recognition, particularly through deep learning, have significantly contributed to the development of fine grained visual categorisation (FGVC), which focuses on distinguishing the visually similar categories. In fungal species classification, this task is especially challenging due to the presence of subtle morphological differences in the classes. To address this, researchers have increasingly explored methods that emphasise the identification of discriminative regions within an image. Such efforts have led to the adoption of region focused models, as demonstrated in the works of Zhang et al. [28], Huang et al. [29] and Feng et al. [30], which illustrate the advantages of the localised feature extraction in improving FGVC performance. Moreover, visual attention mechanisms have demonstrated a significant ability to emphasise critical regions in this context. Because of the potential in identifying most discriminative region, visual attention experienced a surge in popularity within recent research studies, as seen in Refs. [31, 32].

CNNs have emerged as a powerful tool in the fine-grained classification by demonstrating excellent performance in tasks such as identification. Their ability to accurately identify subtle differences between species makes them highly effective for fungal image classification. In Ref. [33], Gaikwad et al. demonstrated superior performance of CNN in classifying fungi affecting apple and mango plant leaves using images from an existing plant pathology dataset. Their study highlights the effectiveness of deep learning to facilitate timely intervention and management of fungal diseases in crops. The combined use of CNNs with visual attention mechanisms and multiple neural networks or bilinear models has become a prevalent strategy for enhancing feature extraction [34, 35]. Researchers also adopted transfer learning methodologies by strategically employing pre-trained models in pursuit of further enhancement in accuracy [36, 37].

In parallel, the development of specialised datasets remains an important aspect for advancing research in this field. Picek et al. [25] contributed significantly with the creation of the Danish Fungi 2020 dataset. The author also performed a rigorous comparative analysis of CNNs and ViT and demonstrated superior performance of ViT over the considered CNN frameworks.

Koo et al. [26] developed a deep learning model utilising regional CNNs for detecting fungal hyphae in microscopic images. Their model demonstrated high sensitivity, achieving 95.2% at 100× magnification and 99.6% at 40× magnification, along with high specificity of 100% at 100× magnification and 86.6% at 40× magnification. This significantly improved detection capabilities compared to the traditional methods for diagnosing fungal infections.

Gao et al. [38] proposed a novel approach by integrating microscopy with a ResNet-50 deep learning model. The integration was specifically designed to improve fungal detection in skin and nail samples. The study showed high sensitivity and specificity, with 99.5% and 95.2% sensitivity and 91.4% and 100% specificity for skin and nails, respectively.

Rahman et al. [39] explored the classification of pathogenic fungi with various deep CNN architectures and reported the superiority of DenseNet model in terms of performance. This study demonstrated the ability of CNNs to differentiate between 89 fungal genera in microscopic images.

Alvarez et al. [40] explored the effectiveness of pre-trained models in fungal classification and concluded that the VGG16 model with pre-trained weights showed improved performance.

To address class imbalance, varying image resolution and background noise in fungal image classification with CNN models, Nawarathne et al. [41] focused on data augmentation and pre-processing. Their study assessed 13 pre-trained models at various image resolutions, revealing that the BigTransfer (BiT) model achieved an impressive accuracy of 87.32% when using a combination of original and high-resolution images. This outcome underscores the significance of optimising image processing strategies for improved classification accuracy.

These studies highlight the potential for deep learning in fungal classification. However, further enhancement in accuracy and reliability, especially with complex data, remains necessary. To address these challenges, we propose a novel deep learning framework that leverages the transfer learning and a custom attention mechanism and aims to advance mycological diagnostics.

3 | Methodology

The proposed framework, shown in Figure 1, combines deep feature extraction with an advanced attention mechanism to focus on critical spatial and channel information for improved classification. The model consists of two primary stages: (1) extracting feature with base CNN and (2) refining these features with selective focus on relevant spatial and channel information using a multi-scale channel spatial attention (MASA) module.

3.1 | Base Model and Feature Extraction

To extract robust and hierarchical features, the framework employs a pre-trained DenseNet201 [42] network as a robust feature extractor. This base CNN is fine-tuned to adapt to the specific task, leveraging both the specialised features learnt from task-specific training and the rich patterns captured from the ImageNet dataset. Although ViTs have demonstrated strong performance in fine-grained classification, they typically require large datasets and high computational resources. These requirements limit their suitability for fungal microscopy tasks. In contrast, DenseNet201 provides a good balance between depth, efficiency and generalisation, especially in small and domain-specific datasets. By combining DenseNet201 with the MASA module, we enhance both feature richness and spatial focus. This makes the approach effective for fungal species classification.

Let the input image be represented as $\boldsymbol{I} \in \mathbb{R}^{H \times W \times 3}$, where H and W denote the height and width of the image, and the 3 channels represent the RGB colour space. The backbone of this model is DenseNet201 [42], a CNN that applies multiple layers of convolutional operations to the input image. Each convolutional layer computes features as follows:

$$\boldsymbol{F}^{(l)} = \sigma\left(\sum_{k=1}^{K} \boldsymbol{W}_k^{(l)} * \boldsymbol{F}^{(l-1)} + b_k^{(l)} \right), \tag{1}$$

where $\boldsymbol{F}^{(l-1)}$ and $\boldsymbol{F}^{(l)}$ are the input and output feature maps at the l^{th} layer, $\boldsymbol{W}_k^{(l)}$ represents the weight matrix of the k^{th} convolutional filter at the l^{th} layer, $b_k^{(l)}$ is the bias term, $*$ denotes the convolution operation, and σ is the activation function.

The DenseNet201 backbone processes the input image, and transforms it into a deep feature representation, defined as follows:

$$\boldsymbol{F} = N(I) \in \mathbb{R}^{h \times w \times c}, \tag{2}$$

where $N(I)$ denotes the DenseNet201 feature extractor, and h, w and c are the spatial dimensions and channel depth of the resulting feature map.

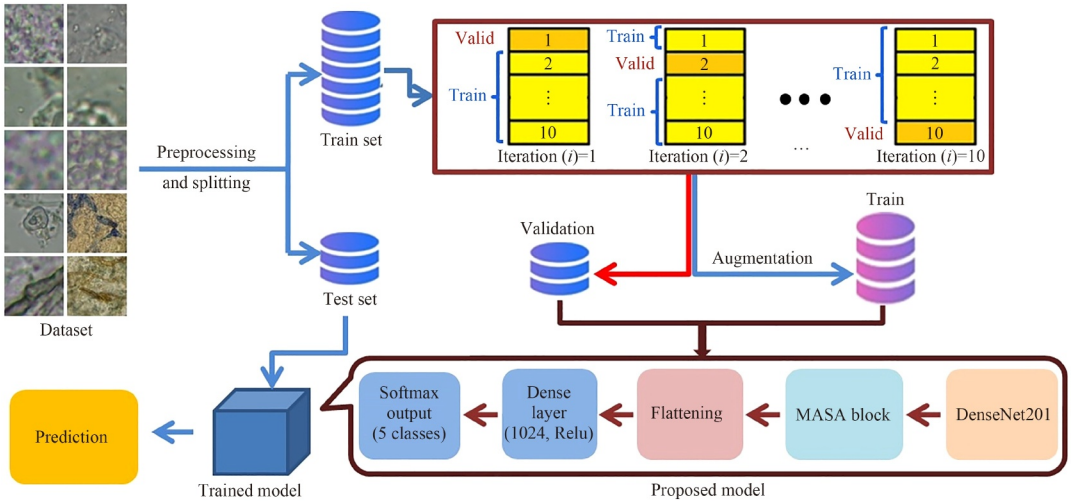

FIGURE 1 | Proposed classification framework. The model employs a pre-trained DenseNet201 backbone integrated with the multi-aspect channel–spatial attention (MASA) block, followed by flattening, dropout, dense layers and a softmax classifier for final predictions.

3.2 | MASA

The base CNN captures rich features, but lacks an explicit mechanism to highlight spatially localised and semantically discriminative regions. To overcome this limitation, we introduce the MASA module. This module aims to improve the network's ability to focus selectively on the most relevant spatial regions and feature channels associated with fungal characteristics. It consists of two attention components designed to address the complex visual characteristics of fungal microscopy images. The spatial attention mechanism selectively emphasises regions that are likely to contain diagnostic morphological patterns, which are often nonuniformly distributed and vary in shape and size across species. To capture these patterns at multiple levels of granularity, the spatial branch employs a multi-scale approach using parallel convolutional filters with varied receptive fields.

In contrast, the channel attention mechanism focuses on enhancing feature discriminability by adaptively reweighting the importance of individual channels. This is particularly relevant in fungal classification, where subtle morphological and textural variations across species are captured in different feature channels of the learnt representation.

Together, these two components work to suppress irrelevant background information while amplifying diagnostically salient features. A structural overview of the MASA module is presented in Figure 2, with implementation details described in the following subsections.

3.2.1 | Spatial Attention Mechanism

The spatial attention mechanism begins by computing both average pooling and max pooling across the spatial dimensions of the input feature map, $\boldsymbol{F} \in \mathbb{R}^{h \times w \times c}$, to capture distinct spatial patterns. These pooled outputs are concatenated to form a combined spatial representation, denoted as $\boldsymbol{f}_c$. To capture multi-scale spatial features, the concatenated representation is passed through three convolutional layers with kernel sizes of 1×1, 3×3 and 5×5. Each convolutional layer generates a spatial attention map at its respective scale, highlighting significant regions in the input feature map.

The resulting multi-scale attention maps are concatenated, and passed through a 1×1 convolution to generate a unified spatial attention map, $\boldsymbol{A}_s$. This spatial attention map is applied to the input feature map via element-wise multiplication $(\cdot)$, resulting in the spatially refined feature map, $\boldsymbol{F}_{\text{spatial}}$, as defined below:

$$\boldsymbol{f}_a = P_{\text{avg}}(\boldsymbol{F}), \quad \boldsymbol{f}_m = P_{\max}(\boldsymbol{F}), \tag{3}$$

$$\boldsymbol{f}_c = [\boldsymbol{f}_a \oplus \boldsymbol{f}_m], \tag{4}$$

$$\boldsymbol{A}_s = \sigma(\text{Conv}_{\text{multi}}(\boldsymbol{f}_c)), \tag{5}$$

$$\boldsymbol{F}_{\text{spatial}} = \boldsymbol{F} \cdot \boldsymbol{A}_s, \tag{6}$$

where P_{avg} and $P_{\max}$ denotes the average and max pooling operation, respectively; $\boldsymbol{f}_a$ and $\boldsymbol{f}_m$ is the average- and max-pooled feature map, respectively; $\oplus$ represents concatenation, and $\text{Conv}_{\text{multi}}$ refers to the multi-scale convolutional operation.

3.2.2 | Channel Attention Mechanism

The channel attention mechanism focuses on refining channel-wise features within $\boldsymbol{F}_{\text{spatial}}$. First, global average pooling (GAP) and global max pooling (GMP) are applied to extract global contextual information. These pooled outputs are passed through two shared fully connected (dense) layers: a reduction layer to compress the feature representation and an expansion layer to restore the original dimensionality. The outputs from GAP and GMP pathways are concatenated ($\boldsymbol{D}_c$), passed through a sigmoid activation function and reshaped to form the channel attention map, $\boldsymbol{A}_c$, which is then applied to $\boldsymbol{F}_{\text{spatial}}$ via element-wise multiplication:

$$\boldsymbol{f}_{\text{GAP}} = P_{\text{GAP}}(\boldsymbol{F}_{\text{spatial}}), \quad \boldsymbol{f}_{\text{GMP}} = P_{\text{GMP}}(\boldsymbol{F}_{\text{spatial}}), \tag{7}$$

$$\boldsymbol{D}_c = \text{Dense}_2(\text{Dense}_1(\boldsymbol{f}_{\text{GAP}})) + \text{Dense}_2(\text{Dense}_1(\boldsymbol{f}_{\text{GMP}})), \tag{8}$$

$$\boldsymbol{A}_c = \sigma(\boldsymbol{D}_c), \tag{9}$$

$$\boldsymbol{F}_c = \boldsymbol{F}_{\text{spatial}} \cdot \boldsymbol{A}_c. \tag{10}$$

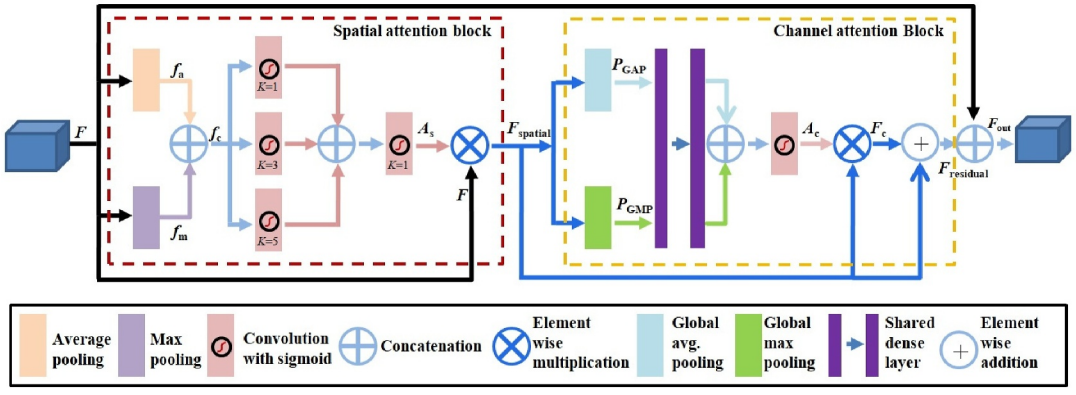

FIGURE 2 | Workflow of the proposed multi-aspect channel–spatial attention (MASA) block, illustrating the sequential spatial and channel attention mechanisms designed to enhance feature representations.

Here, P_{GAP} and P_{GMP} denotes the global average and max pooling operation, respectively, while $Dense_1$ and $Dense_2$ represents the reduction and expansion fully connected layer, respectively. The resulting channel attention map is denoted by $\boldsymbol{F}_c$.

To retain critical spatial information and enhance the learnt feature representation, a residual combination strategy is adopted. Specifically, the $\boldsymbol{F}_c$ is combined with the spatially attended feature map $\boldsymbol{F}_{spatial}$ via element-wise addition, resulting in a residual-enhanced representation $\boldsymbol{F}_{residual}$. Subsequently, a global skip connection is introduced by adding the original input $\boldsymbol{F}$ to $\boldsymbol{F}_{residual}$, producing the final output feature map $\boldsymbol{F}_{out}$.

$$\boldsymbol{F}_{residual} = \boldsymbol{F}_{spatial} + \boldsymbol{F}_c, \tag{11}$$

$$\boldsymbol{F}_{out} = \boldsymbol{F}_{residual} + \boldsymbol{F}, \tag{12}$$

Finally, the refined feature map $\boldsymbol{F}_{out}$ through a dropout layer and dense classification layers, outputting class probabilities as follows:

$$\boldsymbol{y} = \text{softmax}(\boldsymbol{F}_{out}) \tag{13}$$

where $\boldsymbol{y}$ represents the final classification probabilities, emphasising the model's focus on critical spatial and channel regions while maintaining input fidelity through skip connections.

In contrast to existing mechanisms such as squeeze-and-excitation (SE) [43] and convolutional block attention module (CBAM) [44], the proposed MASA module introduces two key enhancements. First, although SE and CBAM employ single-scale or fixed-kernel spatial operations, MASA adopts a multi-scale structure using parallel convolutional layers with varying receptive fields. This enables the extraction of structural features across multiple spatial resolutions. Such multi-scale representation is particularly important for capturing the diverse morphological patterns observed in fungal species. Second, unlike CBAM, which processes spatial and channel modulation sequentially, MASA combines their outputs through a residual framework. This integration preserves the original input context while facilitating more effective feature recalibration. Together, these architectural choices improve the model's capacity to emphasise diagnostically informative regions, particularly under challenging conditions such as class imbalance and morphological variation.

4 | Experiments and Results

4.1 | Dataset Description

To evaluate the performance of the proposed model, this study used the DeFungi dataset [27]. This publicly available dataset comprises microscopic fungal images suitable for the analysis of superficial fungal infections resulting from dermatophyte fungi, yeasts and moulds. The images were manually categorised by a subject matter expert into five classes: tortuous septate hyaline hyphae (TSH), beaded arthroconidial septate hyaline hyphae (BASH), groups or mosaics of arthroconidia (GMA), septate hyaline hyphae with chlamydioconidia (SHC) and broad brown hyphae (BBH). As detailed by Sopo et al. [27], the original images underwent a process of automated cropping and patching to create a final collection of 9114 image patches. To ensure consistent input dimensions for our model, all images were resized to 224 × 224 pixels. Samples of the five fungal types from the DeFungi dataset are displayed in Figure 3, and the distribution of samples for each class is illustrated in Figure 4.

4.2 | Experimental Setup

The experimental evaluation was conducted on Kaggle's cloud-based computing platform, utilising a Tesla P100-PCIE-16 GB GPU to accelerate training. The computational environment included an Intel Xeon CPU (2.00 GHz, 4 cores, 31 GB RAM) running on a 64-bit architecture. The model was implemented in Python 3.10.14 using TensorFlow as a deep learning framework to support training and optimisation.

The training configuration, including key hyperparameters such as learning rate, batch size, number of epochs, dropout rate and optimiser type, is detailed in Table 1. These values were empirically determined through preliminary experiments to ensure robust convergence and optimal model performance.

4.3 | Data Partitioning

To comprehensively assess the performance of the model, two different data-splitting strategies were adopted. First, an 80–20 train-test partition was used. The dataset was randomly divided into 80% train set and 20% test set. This partitioning scheme facilitated direct performance evaluation on an independent test set, ensuring a reliable benchmark for model evaluation. Second, a 90–10 partition with 10-fold cross-validation was applied. The data set was initially divided into 90% training data (8201 samples) and 10% test data (913 samples). To further evaluate the robustness and generalisation of the model, the training set was subjected to 10-fold cross-validation. In each iteration, one fold was designated as the validation set, whereas the remaining nine folds were utilised for training. Following model training, the final evaluation was performed on the independent test set. The mean performance across all folds was reported to ensure statistical reliability in performance estimation. The study employed two evaluation methodologies to ensure a rigorous and unbiased assessment of the proposed model. This further enabled meaningful comparisons with the existing approaches.

4.4 | Implementation Specification

The proposed framework employed a pre-trained DenseNet201 model as the base CNN, fine-tuned for the specific classification task. Layers in block 11, 12, 13 and 14 of DenseNet201 were set as trainable, whereas the remaining layers were frozen to retain features previously learnt from the ImageNet dataset. The input colour images were normalised to the range of [0, 1] by rescaling

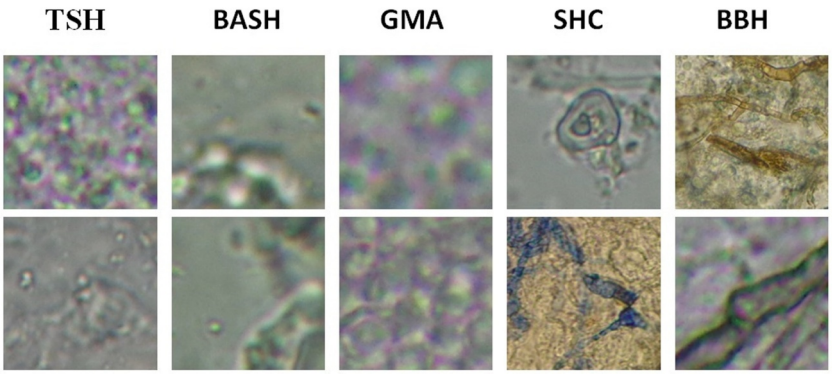

FIGURE 3 | Sample images from the DeFungi dataset, demonstrating the five classes used for classifying superficial fungal infections: TSH, BASH, GMA, SHC and BBH. BASH, beaded arthroconidial septate hyaline hyphae; BBH, broad brown hyphae; GMA, groups or mosaics of arthroconidia; SHC, septate hyaline hyphae with chlamydioconidia and TSH, tortuous septate hyaline hyphae.

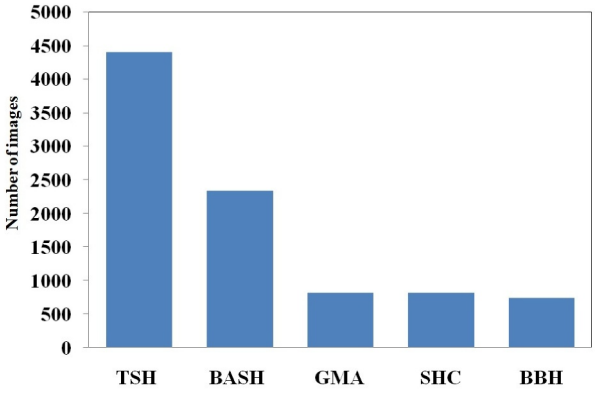

FIGURE 4 | Class-wise sample distribution in the DeFungi dataset. BASH, beaded arthroconidial septate hyaline hyphae; BBH, broad brown hyphae; GMA, groups or mosaics of arthroconidia; SHC, septate hyaline hyphae with chlamydioconidia and TSH, tortuous septate hyaline hyphae.

TABLE 1 | Training configuration and hyperparameter settings for the proposed MASA-Net model.

Hyperparameter	Configuration
Dropout probability	0.5
Learning rate	0.0001
Batch size	16
No. of epoch	150
Optimiser	Adam

the pixel values. Standard augmentation techniques are applied, including random zoom (± 20%), random rotation (± 30°), horizontal and vertical flips, shear transformations (± 20%) and random width and height shifts (± 20%). Each image was then cropped to a fixed size of 224 × 224.

4.5 | Performance Evaluation Metrics

To comprehensively evaluate model performance, both overall and class-specific metrics were employed. Overall performance was quantified using macro-averaged accuracy, precision, recall and *F*1-score, providing a balanced assessment across all classes. Class-specific metrics were utilised to gain insights into the model's performance on individual classes.

Overall accuracy (A): The overall accuracy, reflecting the proportion of correctly classified samples, was determined using the following formula:

$$A = \frac{\sum_{i=1}^{N} Y_{\mathrm{TP}_i}}{\sum_{i=1}^{N} \left(Y_{\mathrm{TP}_i} + Y_{\mathrm{FP}_i} + Y_{\mathrm{FN}_i}\right)}, \tag{14}$$

where Y_{TP_i}, Y_{TN_i}, Y_{FP_i} and Y_{FN_i} represent the number of true positive, true negative, false positive and false negative predictions for class i, respectively. N denotes the total number of classes.

Class-specific metrics: For each class i, the following metrics were calculated:

$$P_i = \frac{Y_{\mathrm{TP}_i}}{Y_{\mathrm{TP}_i} + Y_{\mathrm{FP}_i}}, \tag{15}$$

$$R_i = \frac{Y_{\mathrm{TP}_i}}{Y_{\mathrm{TP}_i} + Y_{\mathrm{FN}_i}}, \tag{16}$$

$$F_i = 2 \times \frac{P_i \times R_i}{P_i + R_i}, \tag{17}$$

$$S_i = \frac{Y_{\mathrm{TN}_i}}{Y_{\mathrm{TN}_i} + Y_{\mathrm{FP}_i}}, \tag{18}$$

where P_i, R_i, F_i and S_i represent the precision, recall, *F*1-score and specificity for class i, respectively.

Macro-averaging (M_{macro}): Macro-averaged metrics were computed as the arithmetic mean of the corresponding class-specific metrics, enabling a balanced overall assessment:

$$M_{\text{macro}} = \frac{1}{N}\sum_{i=1}^{N} M_i, \tag{19}$$

where M_i is the corresponding class-specific metric.

4.6 | Result Analysis

The efficiency of the proposed framework was evaluated through a comprehensive assessment of its performance using two rigorous methods: a random 80–20 train-test split and 10-fold cross-validation.

4.6.1 | Performance Comparison on Random Split

As shown in Table 2, MASA-Net emerged as the top performer in most metrics, achieving the highest overall accuracy of 93.85%, average precision of 96.02% and average *F*1-score of 94.56%. However, in terms of average recall, MASA-Net achieved 93.26%, which is slightly lower than that of MobileNetV3Small [47] (93.80%) but marginally higher than MeFunX [45] (93.14%). It is also worth noting that, unlike ViT-based hybrid models, which struggled with different metrics, MASA-Net consistently delivered overall robust performance.

In addition to this, Table 3 demonstrates superior class-specific performance of the proposed MASA-Net across most fungi types. For TSH, it achieved a precision of 92.32%, recall of 96.94% and *F*1-score of 94.57%. Similarly, for SHC and BBH, MASA-Net delivered exceptional results with *F*1-scores of 97.85% and 96.53%, respectively, showcasing its robustness across different fungi types. The confusion matrix illustrated in Figure 5 provides additional insights into the classification performance of the proposed architecture.

Overall, these findings highlight the superiority of MASA-Net in achieving high accuracy and balanced performance across various metrics and fungi types, setting a new benchmark for classification tasks in this domain.

4.6.2 | Performance on 10-Fold Cross Validation

The robustness and generalisability of the proposed MASA-Net were further evaluated through 10-fold cross-validation. The training and validation curves (Figure 6) demonstrate that the model quickly achieved high training accuracy and maintained strong validation accuracy throughout the epochs. The validation performance after approximately 30 epochs indicates that the model generalised well to unseen data, with only minor fluctuations.

The results of this evaluation are presented in Tables 4–6. As shown in Table 5, MASA-Net achieved a mean accuracy of 92.00% with a low standard deviation of 0.0096 across the 10 folds, indicating consistent performance and stability. The mean precision and *F*1-score were 94.71% and 93.31%, respectively, further demonstrating the reliability of the proposed framework. Table 5 compares the mean 10-fold cross-validation accuracy of MASA-Net and the existing models, including ResNet50, Vgg16 and InceptionV3 based on performance reported in ref. [27]. MASA-Net significantly outperformed these models, with an average accuracy of 92.00% compared to 83.04%, 85.04%, and 82.80% for ResNet50, Vgg16 and InceptionV3, respectively. Furthermore, MASA-Net exhibited a lower standard deviation compared to other models, underscoring its superior stability and reliability. Figure 7 showcases the consistently high performance of 10-fold cross validation, with specificity showing the least variation.

Therefore, these results solidify the robustness and generalisation of MASA-Net, demonstrating its consistent performance across diverse data subsets and validating its potential for reliable deployment in real-world applications.

TABLE 2 | Comparison of deep learning models performance on DeFungi dataset, demonstrating the superior performance of proposed MASA-Net.

Model	Overall accuracy	Avg. precision	Avg. recall	Avg. *F*1-score
VGG16	83.37	88.34	80.61	83.38
InceptionV3	72.37	74.04	68.61	70.56
ResNet	87.12	87.20	88.07	87.39
EfficientNet	89.09	91.43	88.54	89.86
MERNet [45]	90.35	91.01	88.13	87.44
MeFunX [45]	92.49	94.15	93.14	93.62
ViT+VGG16 [46]	85.43	89.35	82.10	85.80
ViT+InceptionV3 [46]	86.53	88.85	85.27	87.20
ViT+ResNet50 [46]	90.13	91.22	89.71	90.91
MobileNetV3Small [47]	92.89	95.40	**93.80**	94.40
MASA-Net	**93.85**	**96.02**	93.26	**94.56**

Note: The best results are in bold.
Abbreviations: Avg., the average; MASA, multi-aspect channel–spatial attention.

TABLE 3 | Category wise comparison of deep learning models performance on DeFungi dataset.

Model	Fungi type	Precision (%)	Recall (%)	*F*1-score (%)	Accuracy (%)
VGG16	TSH	80.93	96.37	87.98	83.77
	BASH	81.67	62.96	71.10	
	GMA	94.34	60.98	74.07	
	SHC	90.30	90.85	90.58	
	BBH	94.44	91.89	93.15	
InceptionV3	TSH	74.98	86.04	80.13	72.37
	BASH	58.48	50.96	54.46	
	GMA	69.81	45.12	54.81	
	SHC	85.82	73.78	79.34	
	BBH	81.13	87.16	84.04	
ResNet	TSH	88.62	91.03	89.81	87.12
	BASH	85.30	75.80	80.27	
	GMA	71.86	87.20	78.79	
	SHC	96.69	89.02	92.70	
	BBH	93.51	97.30	95.36	
EfficientNet	TSH	88.88	93.42	91.09	89.09
	BASH	83.78	80.73	82.22	
	GMA	93.57	79.88	86.18	
	SHC	94.97	92.07	93.50	
	BBH	95.97	96.62	96.30	
MERNet [45]	TSH	93.14	93.57	92.87	90.35
	BASH	82.75	89.76	84.38	
	GMA	92.90	79.12	83.24	
	SHC	94.30	85.63	86.59	
	BBH	91.95	92.57	90.13	
MeFunX [45]	TSH	**93.60**	92.96	93.28	92.49
	BASH	86.24	**89.94**	88.05	
	GMA	94.84	**89.63**	92.16	
	SHC	98.10	94.51	96.27	
	BBH	97.99	**98.65**	**98.32**	
MASA-Net	TSH	92.32	**96.94**	**94.57**	**93.85**
	BASH	**92.38**	88.22	**90.25**	
	GMA	**98.65**	89.02	**93.59**	
	SHC	**98.15**	**97.55**	**97.85**	
	BBH	**98.58**	94.56	96.53	

Note: The best results are in bold.
Abbreviations: BASH, beaded arthroconidial septate hyaline hyphae; BBH, broad brown hyphae; GMA, groups or mosaics of arthroconidia; SHC, septate hyaline hyphae with chlamydioconidia; TSH, tortuous septate hyaline hyphae.

4.6.3 | Ablation Study

To evaluate the impact of the proposed MASA block on the performance of our fine-tuned DenseNet201 model, we conducted an ablation study by comparing the model's performance with and without the MASA block. The results are summarised in Table 7. The results indicate that incorporating the MASA block significantly enhances the model's performance across several metrics. Specifically, we observed an increase in overall accuracy from 92.54% to 94.13% when the MASA block was included, demonstrating its effectiveness in improving classification outcomes. Furthermore, the average precision improved from 94.80% to 96.17%, suggesting that the MASA block helps better distinguish between classes, leading to fewer false positives. The average sensitivity also showed a positive change, increasing from 93.09% to 93.94%, indicating that more true positive cases were correctly identified with the inclusion of the proposed attention module. In addition, the average specificity

improved slightly from 97. 71% to 98. 00%, reflecting a reduction in false negatives, further confirming its beneficial impact on the reliability of the model. The average *F*1 score also increased from 93.83% to 95.00%, clearly highlighting an overall improvement in the balance between the precision and recall.

In addition to the baseline comparison, we extended the ablation study by comparing it against two widely used attention mechanisms: SE and CBAM, using the same DenseNet201 backbone and identical training conditions. The performance metrics, summarised in Table 8, demonstrate that MASA consistently outperforms both SE and CBAM across all evaluation criteria, including overall accuracy, precision, sensitivity, specificity and *F*1-score.

Specifically, MASA achieved the highest overall accuracy of 94.13%, surpassing SE (93.77%) and CBAM (93.03%). Notably, MASA also produced the highest average *F*1-score (95.00%), indicating a more balanced performance in terms of precision and recall. These results affirm the advantage of MASA's multi-scale spatial encoding and combined channel recalibration over conventional single-path attention strategies.

A class-wise comparison is presented in Table 9, which further highlights MASA's strength in handling inter-class variability. MASA achieved superior or comparable classification performance across most fungal species, particularly for the TSH and SHC classes, where it outperformed all other variants. These improvements confirm the proposed module's capability to capture subtle discriminative patterns and reduce misclassifications across diverse fungal morphologies.

Overall, these findings underscore the importance of the MASA block in enhancing model performance, and provide insights into its role in improving classification accuracy and reliability across various metrics.

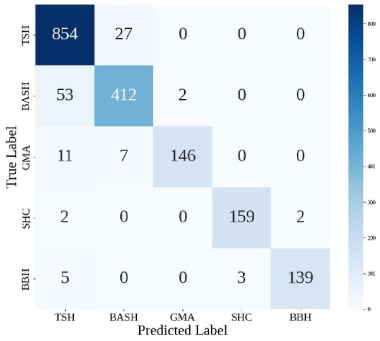

FIGURE 5 | Confusion matrix of MASA-Net evaluated on the DeFungi dataset using the 80–20 train–test partition. BASH, beaded arthroconidial septate hyaline hyphae; BBH, broad brown hyphae; GMA, groups or mosaics of arthroconidia; SHC, septate hyaline hyphae with chlamydioconidia and TSH, tortuous septate hyaline hyphae.

4.6.4 | Feature Map Visualisation

The *t*-distributed stochastic neighbour embedding (*t*-SNE) [48] visualizations, generated using the fold-9 test set of the DeFungi dataset, demonstrate the efficacy of the proposed MASA-Net as illustrated in Figure 8. Figure 8a serves as a baseline, visualising the feature embeddings extracted from a pre-trained DenseNet201 model without fine-tuning. As expected, the feature clusters appear widely dispersed with significant inter-class overlap, indicating that the pre-trained model does not effectively align with the target dataset, resulting in suboptimal feature separability. Figure 8b,c illustrate the impact of different stages of MASA-Net on feature learning for visualisation. Figure 8b presents the feature distribution after fine-tuning DenseNet201, revealing more distinct clusters compared to Figure 8a. This suggests that fine-tuning enables the model to better capture dataset-specific patterns. However, some degree of class overlap remains, indicating the need for further refinement in feature discrimination. Figure 8c shows the representations of the characteristics obtained after the MASA block. This final visualisation demonstrates well-separated clusters with minimal overlap, highlighting the attention module's effectiveness in refining feature learning. The improved

TABLE 4 | Overall performance summary of proposed MASA-Net based on 10-fold cross-validation.

Metric	Mean	Standard deviation (sample)
Accuracy	92.00	0.0096
Precision	94.71	0.0077
Recall	92.13	0.0073
Specificity	97.31	0.0029
*F*1-score	93.31	0.0069

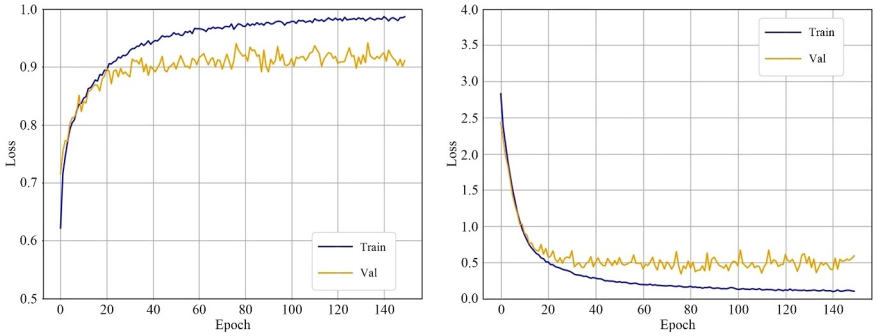

FIGURE 6 | Training and validation accuracy and loss curves for MASA-Net across epochs for Fold-9 (90–10 split) of the DeFungi dataset.

TABLE 5 | Performance comparison of 10-fold cross-validation results for MASA-Net against other models on the DeFungi dataset.

Model	Accuracy (mean)	Standard deviation (%, population)
ResNet50 [27]	83.04	1.97
VGG16 [27]	85.04	1.86
InceptionV3 [27]	82.80	3.51
MASA-Net	**92.00**	**0.91**

Note: The best results are in bold.

TABLE 6 | Details performance of proposed MASA in each fold of the 10-fold cross validation.

Fold	Overall accuracy	Avg. precision	Avg. sensitivity	Avg. specificity	Avg. *F*1-score
Fold-1	92.70	94.96	92.28	97.46	93.47
Fold-2	92.09	94.46	91.74	97.32	92.96
Fold-3	91.36	93.78	92.11	97.26	92.88
Fold-4	91.47	95.33	91.34	97.08	93.16
Fold-5	91.23	94.28	92.10	97.04	93.11
Fold-6	91.83	94.78	91.97	97.34	93.31
Fold-7	91.33	94.66	91.78	97.02	93.09
Fold-8	91.08	93.50	91.49	97.10	92.40
Fold-9	94.13	96.17	93.94	98.00	95.00
Fold-10	92.80	95.16	92.54	97.47	93.69

Abbreviation: Avg., average.

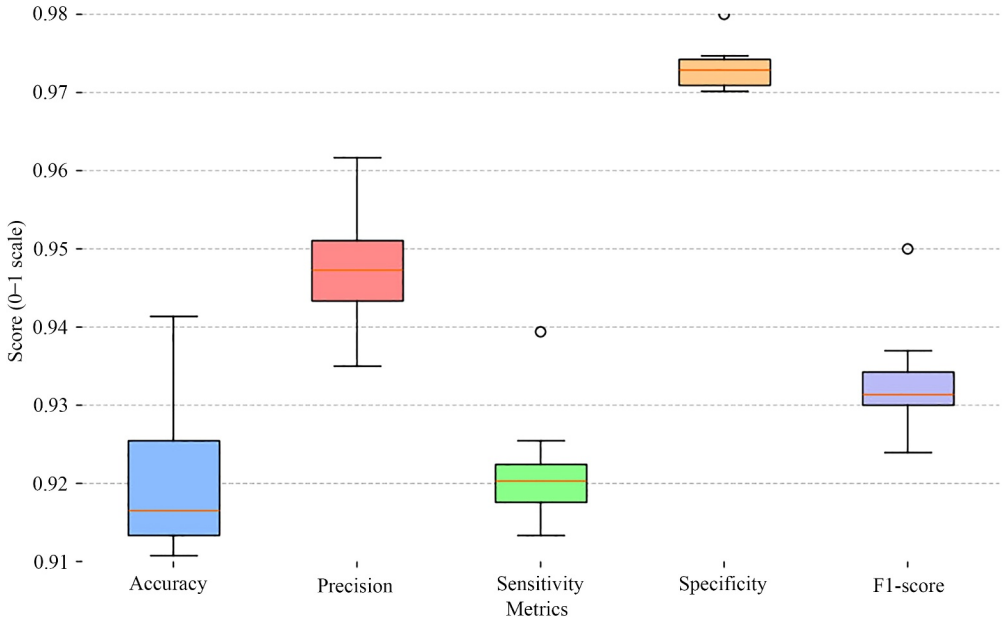

FIGURE 7 | Box plot illustrating the distribution of the proposed MASA-Net's performance across 10-fold cross-validation on the DeFungi dataset.

TABLE 7 | Overall performance of fine-tuned DenseNet201 with and without MASA Block on DeFungi dataset.

Metric	Finetuned DenseNet201 without MASA block	Finetuned DenseNet201 with MASA block
Overall accuracy	92.54	94.13
Average precision	94.80	96.17
Average sensitivity	93.09	93.94
Average specificity	97.71	98.00
Average *F*1-score	93.83	95.00

TABLE 8 | Performance comparison of MASA block with existing attention mechanisms (CBAM and SE) on the DeFungi dataset using the same DenseNet201 backbone.

Attention module	Overall accuracy (%)	Avg. precision (%)	Avg. sensitivity (%)	Avg. specificity (%)	Avg. *F*1-score (%)
SE block	93.77	95.59	93.39	97.94	94.44
CBAM	93.03	92.73	93.06	97.84	92.85
Proposed MASA	**94.13**	**96.17**	**93.94**	**98.00**	**95.00**

Note: The best results are in bold.
Abbreviation: Avg., average.

TABLE 9 | Class-specific performance comparison of different attention mechanisms integrated with DenseNet201 on DeFungi dataset.

Class	Baseline (no attention)	Baseline with SE	Baseline with CBAM	Baseline with MASA
TSH	91.92	96.46	95.96	**96.97**
BASH	**92.86**	88.57	86.19	88.10
GMA	86.30	89.04	**90.41**	**90.41**
SHC	95.89	95.89	**97.26**	**97.26**
BBH	**98.48**	96.97	95.45	96.97

Note: The best results are in bold.
Abbreviations: BASH, beaded arthroconidial septate hyaline hyphae; BBH, broad brown hyphae; CBAM, convolutional block attention module; GMA, groups or mosaics of arthroconidia; MASA, multi-scale channel spatial attention; SE, squeeze-and-excitation; SHC, septate hyaline hyphae with chlamydioconidia; TSH, tortuous septate hyaline hyphae.

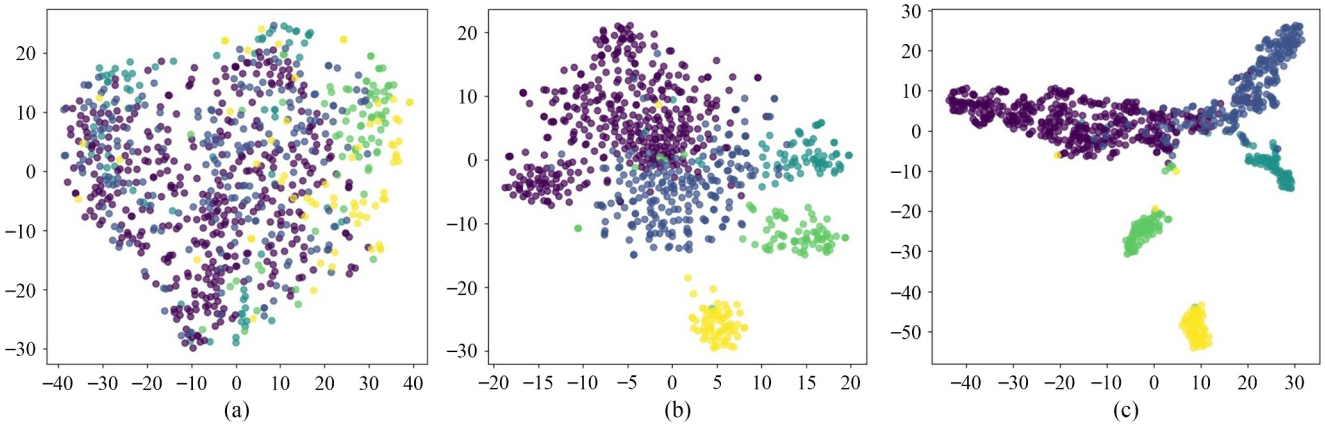

FIGURE 8 | *t*-SNE plots for visualisations of feature maps on the DeFungi test dataset: (a) DenseNet201 without fine-tuning. (b) Features extracted from the fine-tuned DenseNet201 base model. (c) Features obtained from the complete proposed model, incorporating the fine-tuned DenseNet201 with a custom attention block.

class separability suggests that MASA enhances spatial and channel-wise feature interactions, leading to a more discriminative and robust feature representation.

Overall, the comparison among the Figure 8a–c demonstrates a progressive enhancement of feature representations, with MASA-Net achieving superior class separability through fine-tuning and attention-based refinement.

5 | Discussion

The experimental results strongly indicate that MASA-Net significantly improves the classification performance compared to that of the traditional CNN architectures and hybrid transformer-based models. The integration of the proposed attention mechanism effectively improves feature representation, leading to the superior precision, recall and *F*1-scores across all fungi classes. A comparison with the existing models highlights that MASA-Net outperforms the state-of-the-art architectures, including MobileNetV3Small and MeFunX, particularly in challenging fungi categories such as BASH and SHC. The high recall rates suggest that MASA-Net is effective in correctly identifying fungal species without excessive false negatives, ensuring a high level of reliability for critical applications. The 10-fold cross-validation results confirm the stability of MASA-Net as evidenced by the minimal variance across different folds. This suggests that the model maintains its performance across different data partitions, reducing the risk of overfitting. The high specificity values indicate that MASA-Net minimises false positives, ensuring reliable classification. Therefore, MASA-Net demonstrates superior classification accuracy and generalisability.

Furthermore, the ablation study demonstrates that incorporating specific components such as the MASA block can lead to significant improvements in model performance, validating its importance in our classification framework. Moreover, the extended ablation study, which compared MASA with other widely used attention mechanisms such as SE and CBAM, further confirms its effectiveness. MASA consistently achieved superior performance across all evaluation metrics, including in challenging fungal classes. These findings underscore the effectiveness of the proposed sequential weighted attention mechanism in improving feature extraction and classification performance, making MASA-Net a strong candidate for practical applications in fungal species identification.

6 | Conclusion

This study addressed the critical need for improved diagnostic tools in mycology, particularly in the face of increasing fungal infections and the limitations of traditional diagnostic methods. The core innovation of proposed architecture lies in its integration of a pre-trained DenseNet201 backbone with a custom-designed attention mechanism. This attention mechanism dynamically prioritises discriminative fungal features, effectively suppressing irrelevant background noise and enabling the model to focus on crucial morphological details.

The experimental results strongly demonstrate that our framework significantly improves classification performance compared to that of the traditional CNN architectures and hybrid transformer-based models. Furthermore, the high recall rates strengthen the effectiveness in correctly identifying fungal species without excessive false negatives. This ensures a high level of reliability for critical applications. Furthermore, the 10-fold cross-validation results indicates a consistent performance across data subsets and demonstrate the stability of the proposed framework with minimal variance across folds. Therefore, the findings highlight the framework's improved feature extraction and classification performance, making the proposed architecture a viable solution.

Acknowledgements

The authors would like to express their sincere gratitude to Haldia Institute of Technology for providing the necessary support and infrastructure to carry out this research. This study is a collaborative effort between the Department of Computer Science and Engineering (CSE) and the Department of Computer Science and Engineering with specialisation in Artificial Intelligence and Machine Learning (CSE [AIML]). The authors deeply appreciate the encouragement and the conducive research environment fostered by the institution.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The DeFungi dataset utilised in this investigation is publicly available and can be accessed from the following sources: (1) https://www.kaggle.com/datasets/camilovu/lemm-pre-processed, (2) https://www.kaggle.com/datasets/camilovu/lemm-raw.

References

1. J. W. Deacon, *Fungal Biology* (John Wiley & Sons, 2005), https://doi.org/10.1002/9781118685068.

2. R. H. Whittaker, "New Concepts of Kingdoms of Organisms: Evolutionary Relations Are Better Represented by New Classifications Than by the Traditional Two Kingdoms," *Science* 163, no. 3863 (1969): 150–160, https://doi.org/10.1126/science.163.3863.150.

3. C. L. Duddington, *Micro-Organisms as Allies. The Industrial Use of Fungi and Bacteria*. 1st U.S. edition (Macmillan, 1961).

4. T. M. Butt, C. Jackson, and N. Magan, *Fungi as Biocontrol Agents: Progress, Problems and Potential* (CABI Publishing, 2001), https://doi.org/10.1079/9780851993560.0000.

5. P. Badiee and Z. Hashemizadeh, "Opportunistic Invasive Fungal Infections: Diagnosis & Clinical Management," *Indian Journal of Medical Research* 139, no. 2 (2014): 195–204, https://doi.org/10.4103/0971-5916.127585.

6. D. Huang, D. He, L. Gong, et al., "A Prediction Model for Hospital Mortality in Patients With Severe Community-Acquired Pneumonia and Chronic Obstructive Pulmonary Disease," *Respiratory Research* 23, no. 1 (2022): 250, https://doi.org/10.1186/s12931-022-02181-9.

7. T. Lahmer, P. M. Peçanha-Pietrobom, R. M. Schmid, and A. L. Colombo, "Invasive Fungal Infections in Acute and Chronic Liver Impairment: A Systematic Review," *Mycoses* 65, no. 2 (2022): 140–151, https://doi.org/10.1111/myc.13403.

8. M. Nucci and K. A. Marr, "Emerging Fungal Diseases," *Clinical Infectious Diseases* 41, no. 4 (2005): 521–526, https://doi.org/10.1086/432060.

9. M. L. Rodrigues and J. D. Nosanchuk, "Fungal Diseases as Neglected Pathogens: A Wake-Up Call to Public Health Officials," In *Advances in Clinical Immunology, Medical Microbiology, COVID-19, and Big Data* (Jenny Stanford Publishing, 2021), 399–411, https://doi.org/10.1201/9781003180432.

10. P. P. Bosshard, "Incubation of Fungal Cultures: How Long Is Long Enough?," *Mycoses* 54, no. 5 (2011): e539–e545, https://doi.org/10.1111/j.1439-0507.2010.01977.x.

11. M. Pihet, N. Clément, C. Kauffmann-Lacroix, et al., "Diagnosis of Dermatophytosis: An Evaluation of Direct Examination Using Mycet-Color® and MycetFluo®," *Diagnostic Microbiology and Infectious Disease* 83, no. 2 (2015): 170–174, https://doi.org/10.1016/j.diagmicrobio.2015.06.022.

12. B. Havlickova, V. A. Czaika, and M. Friedrich, "Epidemiological Trends in Skin Mycoses Worldwide," supplement, *Mycoses* 51, no. s4 (2008): 2–15, https://doi.org/10.1111/j.1439-0507.2008.01606.x.

13. C. Shen, W. Li, Q. Xu, et al., "Interactive Medical Image Segmentation With Self-Adaptive Confidence Calibration," *Frontiers of Information Technology & Electronic Engineering* 24, no. 9 (2023): 1332–1348, https://doi.org/10.1631/FITEE.2200299.

14. J. Xu, Y. Luo, J. Wang, et al., "Artificial Intelligence-Aided Rapid and Accurate Identification of Clinical Fungal Infections by Single-Cell Raman Spectroscopy," *Frontiers in Microbiology* 14 (2023): 1125676, https://doi.org/10.3389/fmicb.2023.1125676.

15. K. Kristensen, L. M. Ward, M. L. Mogensen, and S. L. Cichosz, "Using Image Processing and Automated Classification Models to Classify Microscopic Gram Stain Images," *Computer Methods and Programs in Biomedicine Update* 3 (2023): 100091, https://doi.org/10.1016/j.cmpbup.2022.100091.

16. Y. Zhang, H. Jiang, T. Ye, and M. Juhas, "Deep Learning for Imaging and Detection of Microorganisms," *Trends in Microbiology* 29, no. 7 (2021): 569–572, https://doi.org/10.1016/j.tim.2021.01.006.

17. M. E. Mital, R. R. Tobias, H. Villaruel, et al., "Transfer Learning Approach for the Classification of Conidial Fungi (Genus Aspergillus) Thru Pre-Trained Deep Learning Models," in *2020 IEEE REGION 10 CONFERENCE (TENCON)* (IEEE, 2020), 1069–1074, https://doi.org/10.1109/TENCON50793.2020.9293803.

18. H. Mohammad-Rahimi, R. Rokhshad, S. Bencharit, J. Krois, and F. Schwendicke, "Deep Learning: A Primer for Dentists and Dental Researchers," *Journal of Dentistry* 130 (2023): 104430, https://doi.org/10.1016/j.jdent.2023.104430.

19. O. Duzyel, M. S. Catal, C. E. Kayan, A. Sevinc, and A. Gumus, "Adaptive Resizer-Based Transfer Learning Framework for the Diagnosis of Breast Cancer Using Histopathology Images," *Signal, Image and Video Processing* 17, no. 8 (2023): 4561–4570, https://doi.org/10.1007/s11760-023-02692-y.

20. B. Chakraborty, R. Mukherjee, and S. Das, "Gemstone Classification Using Deep Convolutional Neural Network," *Journal of The Institution of Engineers (India): Series B* 105, no. 4 (2024): 773–785, https://doi.org/10.1007/s40031-024-01003-4.

21. A. Ahmad, D. Saraswat, and A. El Gamal, "A Survey on Using Deep Learning Techniques for Plant Disease Diagnosis and Recommendations for Development of Appropriate Tools," *Smart Agricultural Technology* 3 (2023): 100083, https://doi.org/10.1016/j.atech.2022.100083.

22. R. Raj, M. Sarkar, R. Mukherjee, and B. Chakraborty, "Yogasana Classification Using Deep Neural Network: A Unique Approach," *CLEI Electronic Journal* 28, no. 1 (2025): 8–1, https://doi.org/10.19153/cleiej.28.1.8.

23. V. M. Tiryaki, "Mass Segmentation and Classification From Film Mammograms Using Cascaded Deep Transfer Learning," *Biomedical Signal Processing and Control* 84 (2023): 104819, https://doi.org/10.1016/j.bspc.2023.104819.

24. S. S. Gaikwad, S. Rumma, and M. Hangarge, "Fungi Classification Using Convolution Neural Network," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, no. 10 (2021): 4563–4569, https://doi.org/10.17762/turcomat.v12i10.5200.

25. L. Picek, M. Šulc, J. Matas, et al., "Danish Fungi 2020-Not Just Another Image Recognition Dataset," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2022), 1525–1535, https://doi.org/10.1109/WACV51458.2022.00334.

26. T. Koo, M. H. Kim, and M. S. Jue, "Automated Detection of Superficial Fungal Infections From Microscopic Images Through a Regional Convolutional Neural Network," *PLoS One* 16, no. 8 (2021): e0256290, https://doi.org/10.1371/journal.pone.0256290.

27. C. J. P. Sopo, F. Hajati, and S. Gheisari, "DeFungi: Direct Mycological Examination of Microscopic Fungi Images," *arXive preprint* (2021), arXiv:2109.07322, https://doi.org/10.48550/arXiv.2109.07322.

28. N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-Based R-CNNs for Fine-Grained Category Detection," in *European Conference on Computer Vision* (Springer, 2014), 834–849, https://doi.org/10.1007/978-3-319-10590-1_54.

29. S. Huang, Z. Xu, D. Tao, and Y. Zhang, "Part-Stacked CNN for Fine-Grained Visual Categorization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016), 1173–1182, https://doi.org/10.1109/CVPR33180.2016.

30. Z. Feng, K. Fu, and Q. Zhao, "Learning to Focus and Discriminate for Fine-Grained Classification," in *2019 IEEE International Conference on Image Processing (ICIP)* (IEEE, 2019), 415–419, https://doi.org/10.1109/ICIP.2019.8803005.

31. T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The Application of Two-Level Attention Models in Deep Convolutional Neural Network for Fine-Grained Image Classification," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), 842–850, https://doi.org/10.1109/CVPR.2015.7298685.

32. X. Liu, T. Xia, J. Wang, Y. Yang, F. Zhou, and Y. Lin, "Fully Convolutional Attention Networks for Fine-Grained Recognition," *arXive preprint* (2016), arXiv:1603.06765, https://doi.org/10.48550/arXiv.1603.06765.

33. S. S. Gaikwad, S. Bhalerao, S. S. Rumma, and M. Hangarge, "Classification of Microscopic Images of Fungi Using Deep Learning Models," supplement, *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology* 13, no. S2 (2021): 105–110, https://doi.org/10.18090/samriddhi.v13spli02.1.

34. T. Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN Models for Fine-Grained Visual Recognition," in *2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2015), 1449–1457, https://doi.org/10.1109/ICCV.2015.170.

35. L. Zhang, S. Huang, W. Liu, and D. Tao, "Learning a Mixture of Granularity-Specific Experts for Fine-Grained Categorization," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, 2019), 8331–8340, https://doi.org/10.1109/ICCV.2019.00842.

36. Y. Cui, Y. Song, C. Sun, A. Howard, and S. Belongie, "Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), 4109–4118, https://doi.org/10.1109/CVPR.2018.00432.

37. J. Ngiam, D. Peng, V. Vasudevan, S. Kornblith, Q. V. Le, and R. Pang, "Domain Adaptive Transfer Learning With Specialist Models," *arXive preprint* (2018), arXiv:1811.07056, https://doi.org/10.48550/arXiv.1811.07056.

38. W. Gao, M. Li, R. Wu, et al., "The Design and Application of an Automated Microscope Developed Based on Deep Learning for Fungal Detection in Dermatology," *Mycoses* 64, no. 3 (2021): 245–251, https://doi.org/10.1111/myc.13209.

39. M. A. Rahman, M. Clinch, J. Reynolds, et al., "Classification of Fungal Genera From Microscopic Images Using Artificial Intelligence," *Journal of Pathology Informatics* 14 (2023): 100314, https://doi.org/10.1016/j.jpi.2023.100314.

40. M. A. V. Álvarez, L. Sopó, C. J. P. Sopo, F. Hajati, and S. Gheisari, "P456 Defungi: Direct Mycological Examination of Microscopic Fungi Images," supplement, *Medical Mycology* 60, no. S1 (2022): myac072P456, https://doi.org/10.1093/mmy/myac072.P456.

41. U. Nawarathne and H. Kumari, "Classification of Fungi Images Using Different Convolutional Neural Networks," in *2023 8th International Conference on Information Technology Research (ICITR)* (IEEE, 2023), 1–6, https://doi.org/10.1109/ICITR61062.2023.10382784.

42. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2017), 4700–4708, https://doi.org/10.1109/CVPR.2017.243.

43. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), 7132–7141, https://doi.org/10.1109/CVPR.2018.00745.

44. S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," In *Proceedings of the European Conference on Computer Vision (ECCV)* (Springer, 2018), 3–19, https://doi.org/10.1007/978-3-030-01234-2_1.

45. S. Rawat, B. Bisht, V. Bisht, N. Rawat, and A. Rawat, "MeFunX: A Novel Meta-Learning-Based Deep Learning Architecture to Detect Fungal Infection Directly From Microscopic Images," *Franklin Open* 6 (2024): 100069, https://doi.org/10.1016/j.fraope.2023.100069.

46. S. I. Ahmed and A. O. Haque, "Microscopic Fungi Classification Using Vision Transformer Guided by Transfer Learning Approach," in

2023 26th International Conference on Computer and Information Technology (ICCIT) (IEEE, 2023), 1–6, https://doi.org/10.1109/ICCIT60459.2023.10441604.

47. J. Bhimavarapu, A. Chinta, S. V. Movva, and J. P. Jampani, "Fungi Classification: Enhancing Diagnosis Using Deep Learning," in *2024 2nd World Conference on Communication & Computing (WCONF)* (IEEE, 2024), 1–6, https://doi.org/10.1109/WCONF61366.2024.10692154.

48. L. Van der Maaten and G. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research* 9, no. 11 (2008), https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.