The Visual Computer (2026) 42:274
https://doi.org/10.1007/s00371-026-04479-x

RESEARCH

Optimization of deep learning models and applications of explainable artificial intelligence classification in some Tricholoma species

Fatih Ekinci[1] · Mehmet Serdar Güzel[2] · Koray Acici[1,3] · Abdullah Aydoğan[2] · Şehmus Altaş[2] · Ömer Burak Altındal[2] · Eda Kumru[4] · Tunc Asuroglu[5,6] · Ilgaz Akata[7]

Received: 19 July 2025 / Accepted: 30 March 2026
© The Author(s) 2026

Abstract

Macrofungi are vital to ecosystem stability and biodiversity conservation, yet traditional species identification methods often require expert input and are prone to error. This study proposes a deep learning-based classification framework for six macrofungi species using advanced convolutional neural network (CNN) architectures. Among the tested models, DenseNet121 achieved a 90% F1-score, while EfficientNetV2-M reached 92% accuracy and 92% precision. AUC scores of 90% or higher were observed for DenseNet, EfficientNet-B4, EfficientNetV2-M, and ShuffleNet, demonstrating robust classification capabilities. Importantly, this study pioneers the integration of explainable artificial intelligence (XAI) in macrofungi classification by employing Grad-CAM. These visualizations revealed that top-performing models focused on biologically relevant features such as the cap, stem, and spore surfaces, whereas less accurate models fixated on background areas, leading to misclassification. The results highlight the importance of dataset size and variability in model performance and underscore the broader applicability of deep learning models beyond macrofungi, including spore-level analysis, plant disease diagnostics, and agricultural classification tasks. By enhancing interpretability and accuracy, this research provides a novel and effective approach for automated fungal identification, contributing significantly to biodiversity monitoring and ecological research.

Keywords Deep learning · Macrofungi classification · Grad-CAM visualization · Convolutional neural networks (CNNs) · Ecological monitoring

1 Introduction

Fungi, one of the most diverse and ecologically significant kingdoms, play a crucial role in maintaining the balance of ecosystems. Macrofungi, in particular, contribute to nutrient

✉ Tunc Asuroglu
tunc.asuroglu@tuni.fi

Fatih Ekinci
fatihekinci@ankara.edu.tr

Mehmet Serdar Güzel
mguzel@ankara.edu.tr

Koray Acici
kacici@ankara.edu.tr

Abdullah Aydoğan
21290200@ogrenci.ankara.edu.tr

Şehmus Altaş
20291316@ogrenci.ankara.edu.tr

Ömer Burak Altındal
21290398@ogrenci.ankara.edu.tr

Eda Kumru
ekumru@ankara.edu.tr

Ilgaz Akata
akata@science.ankara.edu.tr

[1] Institute of Artificial Intelligence, Ankara University, 06100 Ankara, Turkey

[2] Department of Computer Engineering, Faculty of Engineering, Ankara University, 06830 Ankara, Turkey

[3] Department of Artificial Intelligence and Data Engineering, Ankara University, 06100 Ankara, Turkey

[4] Graduate School of Natural and Applied Sciences, Ankara University, 06830 Ankara, Turkey

[5] Faculty of Medicine and Health Technology, Tampere University, 33720 Tampere, Finland

[6] VTT Technical Research Centre of Finland, 33101 Tampere, Finland

[7] Department of Biology, Faculty of Science, Ankara University, 06100 Ankara, Turkey

Published online: 29 April 2026

cycling, organic matter decomposition, and the formation of symbiotic relationships with plants, which are essential for soil fertility and forest ecosystems [1]. These organisms also possess medicinal, agricultural, and industrial applications, further enhancing their importance in biodiversity conservation and scientific research [2].

Tricholoma is a well-established genus within the order Agaricales, encompassing over 1200 species documented globally [3]. Tricholoma species are widespread, particularly in temperate and subtropical regions, where they commonly form ectomycorrhizal associations with coniferous and broadleaf trees [4]. These mushrooms are typically large and fleshy, often attracting foragers, despite many being inedible or toxic. Identification primarily relies on macroscopic features such as cap color, texture, and gill characteristics, with odor and taste providing additional clues. Gills are usually white to off-white but may stain reddish-brown with age or damage. Microscopic variation is minimal, with the genus often recognized by its emarginate or sinuate gill attachment [5]. However, accurate identification remains challenging due to significant morphological variability and the presence of visually similar species. Traditional methods, which depend on expert analysis and microscopy, are time-consuming, labor-intensive, and prone to human error.

This study focuses on the classification of six macrofungi species: Tricholoma atrosquamosum, Tricholoma portentosum, Tricholoma saponaceum, Tricholoma sejunctum, Tricholoma sulphureum, and Tricholoma vaccinum. These species, belonging to the Tricholoma genus, are noted for their complex taxonomy and diverse morphological characteristics. T. atrosquamosum is recognized by its dark-scaled cap, while T. portentosum features a grayish, slimy cap with a mild taste [6]. T. saponaceum is distinguished by its characteristic soapy odor, and T. sejunctum exhibits greenish-yellow hues [7]. T. sulphureum is remarkable for its bright yellow coloring and strong sulfuric scent, whereas T. vaccinum is identified by its reddish-brown cap and fibrous texture [8]. The morphological similarities among these species complicate accurate identification through traditional methods, underscoring the need for automated and efficient classification techniques.

Machine learning models have been used in fungal classification but often underperform due to limited feature extraction and reliance on handcrafted features. Traditional models like support vector machines (SVM), decision trees, and k-nearest neighbors (KNN) struggle with complex, high-dimensional image data [9]. In contrast, convolutional neural networks (CNNs) excel by automatically learning hierarchical features, capturing intricate patterns, and extracting spatial and texture-related information through multiple layers [10]. This enables CNN-based models to outperform traditional approaches by improving accuracy and generalizability [11]. The integration of AI and deep learning has further revolutionized fungal classification by reducing the need for expert input, minimizing errors, and shortening processing times [10]. Additionally, transfer learning and large-scale datasets enhance deep learning models' adaptability and performance. By employing state-of-the-art CNN architectures, this study aims to provide a robust, automated, and scalable solution for macrofungi classification, contributing to biodiversity research and ecological monitoring [11].

This study aims to optimize deep learning models for macrofungi classification, surpassing prior approaches by exclusively utilizing state-of-the-art CNN architectures. As a contribution to the literature, a high-accuracy automated system has been developed for distinguishing visually similar species. Models like DenseNet121, EfficientNetV2-M, and ShuffleNet have demonstrated superior feature extraction and low error rates. Grad-CAM visualizations have clarified the models' decision processes, enhancing reliability and generalizability. This research serves as a key reference for biodiversity and ecological monitoring studies.

Table 1 Names of datasets obtained from open sources, their respective sources, the proportion each dataset contributes to the overall dataset, and the technical details of these sources

Mushroom species name	Photographic contribution by Source	Type of photograph	Resolution (width and height)
Tricholoma atrosquamosum	< 90%	JPG	300
Tricholoma portentosum	< 93%	JPG	300
Tricholoma saponaceum	< 91%	JPG	300
Tricholoma sejunctum	< 94%	JPG	300
Tricholoma sulphureum	< 95%	JPG	300
Tricholoma vaccinum	< 90%	JPG	300

2 Materıals and methods

The dataset used in this study consists of 1,390 images representing six different macrofungi species: *Tricholoma atrosquamosum, Tricholoma portentosum, Tricholoma saponaceum, Tricholoma sejunctum, Tricholoma sulphureum,* and *Tricholoma vaccinum*. These images were collected from Global Core Biodata Resource (www.gbif.org) and contain with a 300 dpi resolutions and JPEG format (Table 1), [12]. The dataset comprises images captured under different lighting conditions, angles, and backgrounds

to ensure robustness and generalizability in model training. This diversity is critical for overcoming challenges related to intra-class variation and enhancing the model's ability to generalize across different environmental conditions. To construct a balanced and effective training process, the dataset was split into three subsets: 60% for training, 20% for validation, and 20% for testing.

This division ensures that the models learn efficiently from a large portion of the data while maintaining a fair evaluation process using unseen images. The validation set plays a crucial role in tuning hyperparameters, preventing overfitting, and ensuring the model generalizes well. The test set serves as an independent evaluation benchmark to measure the final performance of the trained models (Fig. 1).

Since deep learning models perform better with large-scale datasets, we incorporated transfer learning by utilizing pre-trained weights from ImageNet, one of the most extensive image classification datasets available [13]. ImageNet features millions of labeled images across various categories, allowing pre-trained convolutional neural networks (CNNs) to extract rich, hierarchical feature representations. The advantage of using ImageNet pre-trained models is that they already contain learned edge detection, texture patterns, and complex shapes, which can be fine-tuned for macrofungi classification [14]. By leveraging these pre-trained weights, we reduced training time while enhancing accuracy, particularly in feature extraction from mushroom structures [15]. Additionally, augmentation techniques were applied to further increase dataset diversity and improve the model's ability to recognize macrofungi in varying conditions. These augmentations included random rotation, horizontal and vertical flipping, contrast adjustments, and slight cropping, which simulate real-world variations and enhance model robustness [16].

Previous research has emphasized the importance of dataset quality in macrofungi classification. The dataset construction in this study builds upon previous methodologies by ensuring high-resolution, diverse, and well-labeled images. Unlike conventional machine learning approaches that rely on manually engineered features, this study adopts a CNN-based automated feature extraction pipeline, minimizing human bias in feature selection [17]. The dataset's structure aligns with deep learning best practices, ensuring high generalizability and effective model training. By curating a well-balanced dataset with appropriate augmentations and leveraging transfer learning from ImageNet, this study provides a strong foundation for macrofungi classification using deep learning. The next sections will elaborate on the deep learning models and evaluation metrics employed in this research [16, 17].

Deep learning (DL) has revolutionized image classification, providing significant advantages over traditional machine learning techniques. In this study, we utilize 16 state-of-the-art convolutional neural network (CNN) models to classify six different macrofungi species. These models (Table 2) were selected based on their architectural diversity, accuracy, and efficiency in image recognition tasks. Each model has unique characteristics, strengths, and weaknesses, which impact its classification performance. Below is an in-depth overview of the models used in this study [18].

MobileNetV3-L is a lightweight and efficient CNN designed for mobile and embedded devices. It utilizes depthwise separable convolutions, reducing the number of parameters and computations required. This makes it highly suitable for real-time applications while maintaining good accuracy. However, its performance may be slightly lower than heavier architectures when dealing with highly complex datasets like macrofungi classification [19].

Densely connected convolutional network (DenseNet121) introduces dense connections, where each layer is connected to every other layer in a feed-forward manner. This structure promotes feature reuse, reduces the number of parameters, and alleviates the vanishing gradient problem. It performs exceptionally well in image classification but can be computationally expensive due to its complex connectivity [20].

The EfficientNet family applies a compound scaling technique, optimizing depth, width, and resolution simultaneously. EfficientNet-B0 is the smallest and fastest model, while EfficientNet-B4 and EfficientNet-B7 progressively increase in size and accuracy. These models provide a great balance between efficiency and performance, though larger versions like EfficientNet-B7 require significantly more computational resources [21].

Residual networks (ResNet) introduces residual learning through skip connections, allowing very deep networks to be trained efficiently. ResNet50 and ResNet101 offer strong generalization and are widely used in medical imaging and object recognition. However, they tend to be computationally expensive and may not always outperform newer architectures like EfficientNet in specific tasks [22].

Extreme inception (Xception) builds upon inception modules by using depthwise separable convolutions, enhancing computational efficiency. This architecture provides high accuracy with relatively fewer parameters compared to traditional CNNs. However, it can still be resource-intensive for certain applications [23].

VGG-16 and VGG-19 are classical CNN architectures with a deep, simple structure. They use 3×3 convolutional layers stacked in sequence, making them easy to understand and implement. While these models (Table 2) provide good feature extraction, they have a high number of parameters and are computationally demanding, often requiring more memory compared to newer architectures [24].

MaxVit-S is a hybrid architecture combining CNN and transformer-based self-attention mechanisms. This model improves feature extraction by leveraging both spatial and



Fig. 1 This figure displays six macrofungi species from the dataset, illustrating their diversity in texture, shape, and coloration. The species are as follows: **(a)** *Tricholoma atrosquamosum*, **(b)** *Tricholoma portentosum*, **(c)** *Tricholoma saponaceum*, **(d)** *Tricholoma sejunctum*, **(e)** *Tricholoma sulphureum*, and **(f)** *Tricholoma vaccinum*. Each species exhibits distinct morphological traits, contributing to the dataset's variability

Table 2 Performance comparison of deep learning models based on accuracy, precision, recall, F1-score, and specificity

Models	Accuracy	Precision	Recall	F1-Score	Specificity	AUC
DarkNet-53	0.88	0.88	0.88	0.88	0.98	0.88
DenseNet121	**0.90**	**0.90**	**0.90**	**0.90**	**0.98**	**0.90**
EfficientNet-B0	**0.89**	**0.89**	**0.89**	**0.89**	**0.98**	**0.89**
EfficientNet-B4	**0.90**	**0.90**	**0.90**	**0.90**	**0.98**	**0.90**
EfficientNet-B7	0.87	0.87	0.87	0.87	0.97	0.87
EfficientNetV2-M	**0.92**	**0.92**	**0.92**	**0.92**	**0.98**	**0.92**
GoogleNet (InceptionV3)	0.84	0.84	0.84	0.83	0.97	0.84
Maxvit-S	0.85	0.85	0.85	0.85	0.97	0.85
MobileNetV3-L	0.88	0.88	0.88	0.88	0.98	0.88
ResNet101	0.45	0.41	0.45	0.38	0.89	0.43
ResNetV2-50	0.84	0.84	0.84	0.84	0.97	0.84
ShuffleNetV2	**0.90**	**0.90**	**0.90**	**0.90**	**0.98**	**0.90**
SqueezeNet	0.82	0.84	0.82	0.82	0.96	0.83
VGG-16	0.86	0.86	0.86	0.86	0.97	0.86
VGG-19	0.85	0.84	0.84	0.84	0.97	0.84
Xception	0.87	0.87	0.87	0.87	0.97	0.87

Bold color indicates models that have the best performances for each evaluation criteria

channel-wise attention mechanisms, making it well suited for detailed texture recognition, such as macrofungi classification. However, due to its hybrid nature, it requires more computational power than traditional CNNs [25].

DarkNet-53 is the backbone of the You Only Look Once (YOLO) object detection framework. It consists of a series of convolutional layers optimized for real-time image analysis. While it is extremely effective in object detection, it may not always provide the highest accuracy for fine-grained classification tasks like fungal species identification [26].

InceptionV3 utilizes parallel convolutional layers with different kernel sizes, allowing the model to learn features at multiple scales simultaneously. It is efficient in terms of computation and memory usage. However, its complex architecture requires careful tuning and can be harder to optimize compared to more recent models like EfficientNet [27].

ShuffleNet is an extremely lightweight model designed for low-power devices. It uses pointwise group convolutions and channel shuffling to reduce computational costs. While it is highly efficient, it may not perform as well as larger CNNs on complex datasets [28].

SqueezeNet is another lightweight architecture that achieves comparable accuracy to AlexNet while using significantly fewer parameters. It is optimized for speed and memory efficiency but may struggle with highly detailed classification tasks due to its compact nature [29].

EfficientNetV2-M is an improved version of the original EfficientNet series, offering faster training and better accuracy. It refines the compound scaling approach, making it more efficient for large datasets. While highly effective, it requires more memory and computational power than its predecessor, EfficientNet-B7 [30].

Each of these models has its strengths and weaknesses, affecting their suitability for different classification tasks:

Lightweight and fast models (MobileNetV3-L, ShuffleNet, SqueezeNet) offer efficiency but may struggle with highly detailed fungal classification. High-accuracy models (DenseNet121, EfficientNet-B4, EfficientNetV2-M, Xception) excel in capturing intricate features but require more computational resources. Hybrid models (MaxVit-S, GoogleNet) leverage both CNN and transformer-based techniques for enhanced feature extraction but are computationally intensive. Older architectures (VGG-16, VGG-19) are simple and effective but require large amounts of memory. Object detection-based models (DarkNet-53) are optimized for real-time applications rather than fine-grained species classification [29–31].

In this study, all models were optimized under the same conditions, regardless of whether they were lightweight, complex, fast, or computationally intensive. Each architecture was trained using identical preprocessing techniques, hyperparameter tuning strategies, and evaluation metrics to ensure a fair and unbiased comparison. This standardized approach allows for an objective assessment of the models' strengths and limitations, highlighting their true performance in macrofungi classification.

The evaluation of deep learning models plays a crucial role in determining their effectiveness in macrofungi classification. In this study, we employed various performance metrics, optimization techniques, and explainable AI (XAI)

methods to ensure accurate and interpretable results. We also utilized Python, TensorFlow (Keras), and PyTorch as the primary frameworks for training and testing the models. This section provides a comprehensive overview of the evaluation strategies used in our research [31].

To train our deep learning models efficiently, we used the adaptive moment estimation (Adam) optimizer, which combines the advantages of momentum-based and adaptive learning rate optimization techniques [32]. Adam is widely used in deep learning due to its ability to adjust learning rates dynamically for each parameter, leading to faster convergence and better generalization. The learning rate was set to 0.001, ensuring stable training across different CNN architectures [30–32].

Additionally, ImageDataGenerator was employed for real-time image augmentation and batch-wise training, which helped improve the model's ability to generalize across diverse macrofungi species [33]. Data augmentation techniques included random rotation, flipping, cropping, and brightness adjustments, allowing models to learn from a more diverse set of training samples.

The training process was conducted using Python with the TensorFlow (Keras) and PyTorch frameworks, leveraging their highly optimized deep learning libraries. Keras was chosen for its ease of implementation and scalability, while PyTorch was used for its flexibility in dynamic computation graphs and debugging capabilities [31–33].

To measure the performance of each CNN model, we used standard classification metrics, including accuracy, precision, recall, F1-score and specificity. These metrics (1–5) provide insights into how well the models distinguish between different macrofungi species [29–33]. The formulas for each metric, as referenced from the Sensors article, are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TN + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = \frac{2x(PrecisionxRecall)}{Precision + Recall} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

True positives (TP) refer to cases where the model accurately predicts a positive instance, correctly identifying it as belonging to the target class (Eqs. 1–3). True negatives (TN) occur when the model correctly recognizes a negative instance, confirming that it does not belong to the target class (Eqs. 1,2,3, and 5). False positives (FP) arise when the model erroneously classifies a negative instance as positive, resulting in a false alarm (Eqs. 1,2, and 5). Conversely, false negatives (FN) represent instances where the model fails to detect a positive case, incorrectly classifying it as negative (Eqs. 1–3). These classifications play a critical role in evaluating model performance, as they directly impact key metrics such as accuracy, precision, recall, F1-score, and specificity.

Accuracy measures the proportion of correct predictions relative to the total number of predictions, offering a general assessment of the model's classification performance (1). Precision focuses on the reliability of positive predictions by determining how many of the predicted positive instances are actually correct (2). Recall, on the other hand, evaluates the model's ability to correctly identify actual positive cases, reflecting its sensitivity (3). The F1-score serves as a balanced metric that combines precision and recall, making it particularly useful for evaluating models on imbalanced datasets (4). Specificity measures how well the model correctly identifies negative instances, helping to reduce false positives and offering a complementary perspective to recall (5). The receiver operating characteristic (ROC) curve illustrates the trade-off between sensitivity and specificity at various classification thresholds, providing insight into the model's discrimination ability. The area under this curve (AUC-ROC) quantifies the overall performance by measuring the likelihood that the model ranks a randomly chosen positive instance higher than a randomly chosen negative instance, making it a valuable indicator of classification effectiveness, especially in imbalanced datasets [32–34].

These metrics (1–5) were calculated separately for each model to compare their classification performance on macrofungi species. While deep learning models achieve remarkable accuracy, they often function as "black boxes," making it difficult to understand their decision-making processes. To address this issue, we employed explainable AI (XAI) techniques, specifically Gradient-weighted Class Activation Mapping (Grad-CAM), to visualize and interpret the model's predictions [34]. Grad-CAM generates heatmaps that highlight the regions of an input image that contributed most to the model's classification decision [35]. These heatmaps allow us to verify whether the models are focusing on the relevant structural features of the macrofungi, such as texture, cap shape, and color variations.

By integrating XAI methods, this study enhances transparency in deep learning-based macrofungi classification, ensuring that predictions are made based on meaningful visual cues rather than spurious correlations [36]. This approach provides greater confidence in model reliability, particularly for biodiversity monitoring and ecological research applications. Adam optimizer was used for efficient model training with a learning rate of 0.001. Python,

TensorFlow (Keras), and PyTorch were the primary frameworks for implementation. ImageDataGenerator facilitated real-time augmentation and batch-wise training, improving model generalization. Model performance was evaluated using accuracy, precision, recall, F1-score, and specificity with formulas adopted from the Sensors article. Explainable AI (XAI) techniques, particularly Grad-CAM, were used to visualize decision-making, ensuring model interpretability and reliability [33–36].

The chi-square statistical significance test was applied to assess whether the performance of the models was statistically significant. The chi-square (χ^2) statistical significance test is commonly used to evaluate whether the observed differences between categorical variables are statistically significant. For a multi-class classification problem, the formula 6 for the chi-square test statistic is (6):

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{6}$$

where r is the number of rows, categories of the actual labels, c is the number of columns, categories of the predicted labels, O_{ij} is the observed frequency, and E_{ij} is the expected frequency, calculated as (7):

$$E_{ij} = \frac{\sum_{k=1}^{c} O_{ik} \cdot \sum_{k=1}^{r} O_{kj}}{N} \tag{7}$$

where N is the total number of observations.

The test helps determine whether there is a significant association between the actual and predicted classes. If the p value from the chi-square test is below a predefined significance level (e.g., $\alpha = 0.05$), we reject the null hypothesis, indicating that there is a significant difference between the predicted and actual distributions. To obtain the p value for the chi-square test, after computing the chi-square test statistic, the degrees of freedom (df) are determined as follows (8):

$$df = (r-1) * (c-1) \tag{8}$$

Finally, the p value (p) is obtained by finding the probability of observing a χ^2 statistic at least as extreme as the one calculated, under the chi-aquare distribution with the computed degrees of freedom (9, 10):

$$p = P(\chi^2 > \chi^2_{calculated} | df) \tag{9}$$

$$p = 1 - F_{\chi^2}(\chi^2 > \chi^2_{calculated}, df) \tag{10}$$

where F_{χ^2} is the chi-square cumulative distribution function (CDF). The CDF for the chi-square distribution is not typically calculated manually but instead obtained from statistical tables.

3 Results

This section presents the findings from experiments conducted on macrofungi classification utilizing advanced deep learning models. The primary objective of the study was to assess the effectiveness of various convolutional neural network (CNN) architectures in accurately classifying six distinct macrofungi species. A comprehensive evaluation was performed by analyzing the performance of 16 CNN models based on key metrics, including accuracy, precision, recall, F1-score, specificity, and AUC score. Grad-CAM visualizations were employed to elucidate the decision-making processes of the models, providing deeper insights into their feature extraction capabilities. Additionally, real-time image augmentation was implemented using the ImageDataGenerator framework to enhance model generalization across diverse macrofungi samples. Unlike previous studies that integrated both machine learning and deep learning techniques, this research exclusively leverages state-of-the-art CNN models to develop a robust and scalable approach to macrofungi classification. The subsequent sections provide a detailed comparative analysis of the models, highlighting critical observations and their implications for biodiversity research.

According to the results in Table 2, DenseNet121, EfficientNet-B0, EfficientNet-B4, EfficientNetV2-M, and ShuffleNetV2 have demonstrated the highest performance, achieving the best accuracy and F1-score values. EfficientNetV2-M achieved the highest accuracy at 92%, while DenseNet121 stood out with 90% accuracy, 90% precision, and 98% specificity. EfficientNet-B0, EfficientNet-B4, and ShuffleNetV2 also performed well with high accuracy and balanced metric scores. Additionally, EfficientNetV2-M, along with DenseNet121, EfficientNet-B0, EfficientNet-B4, and ShuffleNet, achieved the highest AUC scores compared to other models, demonstrating their superior ability to distinguish macrofungi species effectively. The strong performance of these models can be attributed to their deep-layered architectures and optimized learning processes (Table 2). Based on the results presented in Table 2, the performance of all 16 models will be analyzed individually to provide a detailed evaluation of their effectiveness in macrofungi classification.

DenseNet121 achieved 90% accuracy, an F1-score of 0.90, and a specificity of 0.98, making it one of the best-performing models in this study. Unlike traditional CNNs, DenseNet121 employs densely connected layers, which

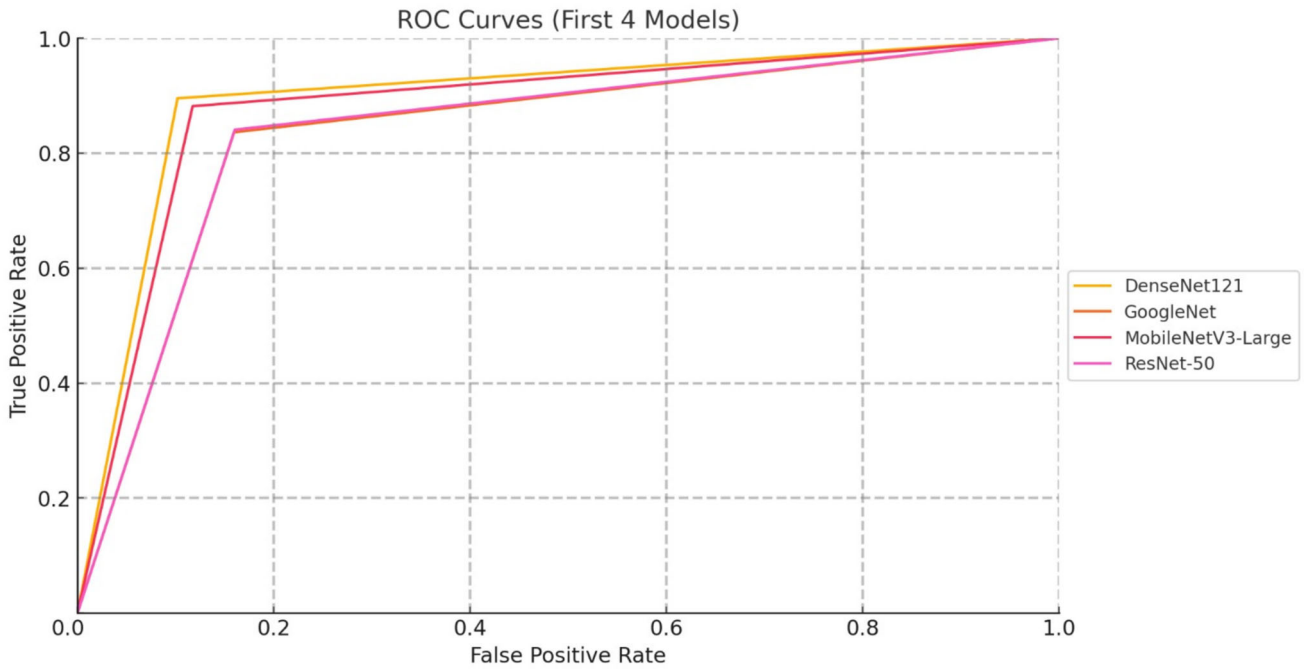


Fig. 2 ROC Curves of DenseNet121, GoogleNet, MobileNetV3-Large, and ResNet50

improve gradient flow, reduce redundancy, and enhance feature reuse, leading to better learning and information retention. This architectural advantage allows DenseNet121 to capture intricate fungal patterns more effectively, particularly in cases where species share similar visual characteristics. Compared to models like DarkNet-53 and EfficientNet-B4, DenseNet121 demonstrates superior classification accuracy and generalization ability, while also maintaining computational efficiency. However, despite its excellent performance, its complex connectivity results in higher memory consumption, making training slower compared to lightweight models such as MobileNetV3-L and ShuffleNetV2. With an AUC of 0.90, it exhibits strong class separability, reinforcing its high reliability in distinguishing macrofungi species (Table 2). ROC curve analysis confirms that DenseNet121 follows an ideal classification curve, showing minimal misclassification risk. Overall, DenseNet121 stands out as one of the most effective models for macrofungi classification, striking an excellent balance between accuracy, efficiency, and robustness (Fig. 2). GoogleNet (InceptionV3) achieved 84% accuracy, an F1-score of 0.83, and a specificity of 0.97, positioning itself among the higher-performing models in macrofungi classification. Its inception modules, which allow multi-scale feature extraction, enable the model to analyze features at different levels of abstraction, improving its ability to distinguish between visually similar fungal species. However, despite its efficiency in extracting fine-grained features, its performance is slightly lower than DenseNet121 and EfficientNet models, which benefit from more advanced feature reuse and residual connections. With an AUC of 0.84, GoogleNet maintains strong class separability but does not reach the exceptional classification ability of EfficientNetV2-M or DenseNet121 (Table 2). ROC curve analysis indicates that while GoogleNet is a reliable model, it does not achieve the same level of precision as the top-tier architectures. Despite this, its lightweight structure results in shorter training times and lower memory consumption, making it an appealing choice for applications where computational efficiency is a priority. Overall, GoogleNet (InceptionV3) is a strong and efficient model for macrofungi classification, but it is slightly outperformed by newer architectures that incorporate more advanced optimization techniques (Fig. 2). MobileNetV3-L achieved 88% accuracy, an F1-score of 0.88, and a specificity of 0.98, demonstrating impressive performance for a lightweight, mobile-optimized model. Designed for efficiency, it offers a strong balance between classification accuracy and computational cost, making it a suitable choice for real-time applications and low-power environments. Despite its compact architecture, MobileNetV3-L effectively captures key fungal features, performing comparably to deeper models like DarkNet-53 while maintaining faster inference times. However, compared to EfficientNetV2-M and DenseNet121, it lacks the advanced feature reuse and optimization techniques that contribute to their superior classification accuracy. With an AUC of 0.88, it demonstrates strong class separability, although it does not reach the top-tier performance seen in EfficientNet models. (Table 2). ROC curve analysis suggests that while

MobileNetV3-L performs well, it is outperformed by models with deeper feature extraction capabilities, particularly in distinguishing closely related fungal species. Nonetheless, its high efficiency and relatively strong accuracy make it an excellent choice for macrofungi classification in applications where computational constraints are a concern (Fig. 2). ResNetV2-50 achieved 84% accuracy, an F1-score of 0.84, and a specificity of 0.97, positioning itself as a significantly improved version of ResNet101 for macrofungi classification. By incorporating batch normalization before activation and enhanced residual connections, ResNetV2-50 effectively mitigates the training inefficiencies seen in its predecessor, resulting in better feature extraction and generalization. Compared to ResNet101, which performed poorly with only 45% accuracy, this optimized version demonstrates that architectural refinements are more impactful than simply increasing model depth. With an AUC of 0.84, ResNetV2-50 exhibits strong separability between fungal classes, though it does not reach the classification efficiency of top-performing models like EfficientNetV2-M and DenseNet121 (Table 2). ROC curve analysis confirms that while ResNetV2-50 is a solid choice, it lags behind models with more advanced optimization techniques, particularly those with densely connected layers or depthwise separable convolutions. Despite this, its balance between depth and efficiency makes it a competitive model, outperforming many traditional architectures while remaining more computationally feasible than deeper networks. Overall, ResNetV2-50 delivers reliable classification performance, proving that well-optimized architectures can outperform deeper but less refined models in macrofungi classification (Fig. 2).

ResNet101 achieved 45% accuracy, an F1-score of 0.38, and a specificity of 0.89, making it one of the poorest-performing models in macrofungi classification. Despite its deep residual architecture, designed to mitigate vanishing gradient issues, the model failed to generalize well, resulting in a high rate of misclassifications. Compared to ResNetV2-50, which is a more optimized version, ResNet101 performed significantly worse, highlighting that simply increasing depth does not guarantee better classification performance. One of the main reasons for its poor results is its reliance on large-scale datasets for optimal feature learning, whereas the macrofungi dataset used in this study may not have been sufficient to fully utilize its hierarchical feature extraction capabilities. Additionally, with an AUC of 0.43, ResNet101 struggles to differentiate between fungal classes effectively, reinforcing its weak performance in this classification task (Table 2). ROC curve analysis confirms that ResNet101's classification ability is significantly lower than that of EfficientNet models, DenseNet121, and even lightweight architectures like MobileNetV3-L. The high parameter count of the model also results in longer training times and increased computational costs, making it inefficient for this particular dataset. Overall, ResNet101 fails to deliver meaningful classification performance in macrofungi identification, struggling with both generalization and computational efficiency, making it one of the least effective models in this study (Fig. 3). EfficientNet-B0 achieved 89% accuracy, an F1-score of 0.89, and a specificity of 0.98, demonstrating strong classification performance while maintaining computational efficiency. As the smallest variant in the EfficientNet family, it balances accuracy and resource consumption effectively. Its lightweight architecture makes it a suitable option for scenarios where computational power is limited, while still achieving results comparable to deeper models. However, despite its high accuracy, it slightly underperforms compared to EfficientNet-B4, EfficientNetV2-M, and DenseNet121, suggesting that additional depth can enhance generalization in macrofungi classification. The AUC value of 0.89 indicates strong discriminative ability, yet deeper models within the EfficientNet family show marginal improvements (Table 2). ROC curve analysis reveals that EfficientNet-B0 significantly outperforms weaker models like ResNet101, but its classification ability is slightly inferior to larger EfficientNet variants such as B4 and B7. While it efficiently extracts fungal features, its relatively shallow depth may limit its ability to capture more complex distinctions among visually similar species. Overall, EfficientNet-B0 offers a strong balance between performance and efficiency, making it a viable choice for macrofungi classification, particularly in resource-constrained environments (Fig. 3). EfficientNet-B4 achieved 90% accuracy, an F1-score of 0.90, and a specificity of 0.98, making it one of the top-performing models in macrofungi classification. As a deeper variant of the EfficientNet family, it builds upon the efficiency of EfficientNet-B0 while offering superior feature extraction and generalization. Its increased depth allows for better representation of complex fungal structures, leading to more precise classification, particularly for species with subtle visual differences. Compared to its smaller counterpart, EfficientNet-B0, it demonstrates a noticeable improvement in accuracy and F1-score, reinforcing the benefit of additional layers in fine-grained classification tasks. However, despite its high accuracy, it performs almost identically to EfficientNet-B7, suggesting that further increasing depth beyond this point does not necessarily yield substantial gains. With an AUC of 0.90, EfficientNet-B4 maintains strong separability between fungal classes, though it is slightly outperformed by EfficientNetV2-M, the best model in this study (Table 2). ROC curve analysis confirms that EfficientNet-B4 excels in classification while maintaining computational efficiency, placing it ahead of many alternative architectures. Overall, EfficientNet-B4 is a powerful model that balances performance and efficiency, making it an excellent choice for macrofungi classification without the excessive computational cost of deeper networks. (Fig. 3) EfficientNet-B7 achieved 87% accuracy, an F1-score

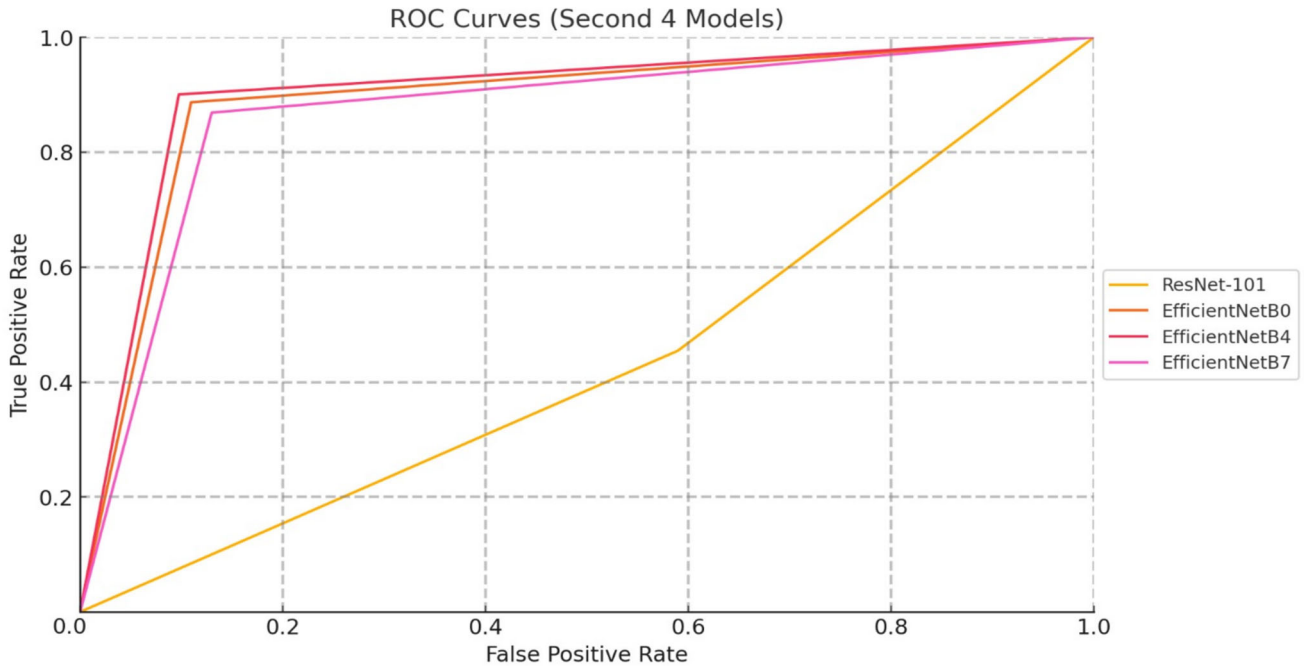


Fig. 3 ROC Curves of ResNet101, EfficientNet-B0, EfficientNet-B4, and EfficientNet-B7

of 0.87, and a specificity of 0.97, demonstrating strong classification performance but slightly underperforming compared to EfficientNet-B4 and EfficientNetV2-M. As the deepest variant in the EfficientNet-B series, it was expected to provide superior feature extraction and generalization, yet its performance remains nearly identical to EfficientNet-B4, indicating diminishing returns with increasing depth. While its deep architecture allows for enhanced feature learning, it also results in higher computational costs and longer inference times, making it less efficient than models that achieve similar results with fewer parameters. With an AUC of 0.87, EfficientNet-B7 exhibits strong class separation capabilities but fails to outperform the best models in this study (Table 2). ROC curve analysis shows that its classification ability is comparable to EfficientNet-B4, reinforcing the idea that deeper networks do not always guarantee significant improvements in fine-grained classification tasks. While EfficientNet-B7 remains a highly capable model for macrofungi classification, its increased computational burden compared to EfficientNet-B4 and its inability to surpass EfficientNetV2-M make it a less optimal choice in resource-limited environments (Fig. 3).

EfficientNetV2-M achieved 92% accuracy, an F1-score of 0.92, and a specificity of 0.98, making it the best-performing model in this study for macrofungi classification. As an advanced version of the original EfficientNet architecture, it incorporates optimized feature extraction, improved training efficiency, and enhanced depth scaling, allowing it to generalize better across visually similar fungal species. Its superior classification ability is reflected in its AUC of 0.92, the highest among all models, indicating exceptional separability between fungal classes. Compared to EfficientNet-B7, which has a deeper structure but slightly lower performance, EfficientNetV2-M demonstrates that architectural refinements are more impactful than simply increasing depth. Additionally, it surpasses DenseNet121 and EfficientNet-B4, both of which also achieved high classification performance but lacked the same level of efficiency and adaptability (Table 2). ROC curve analysis confirms that EfficientNetV2-M follows the ideal classification curve, positioning it as the most effective model in terms of both accuracy and reliability. While it does require more computational power than lightweight models like MobileNetV3-L, its ability to capture intricate fungal patterns with high precision and recall makes it the optimal choice for macrofungi classification in applications where high accuracy is prioritized (Fig. 4). DarkNet-53 achieved 88% accuracy, an F1-score of 0.88, and a specificity of 0.98, demonstrating strong performance in macrofungi classification. Originally designed for object detection within the YOLO framework, it was not specifically optimized for fine-grained classification tasks, yet it performed better than expected in distinguishing fungal species. Its high accuracy and F1-score indicate

that it correctly classified most species, while its specificity of 0.98 confirms a low false positive rate, making it a reliable classifier. However, despite these strengths, DarkNet-53 slightly lags behind top-tier models like EfficientNetV2-M and DenseNet121, as seen in its AUC value of 0.88, which suggests slightly weaker class separability. Additionally, its deep architecture results in higher

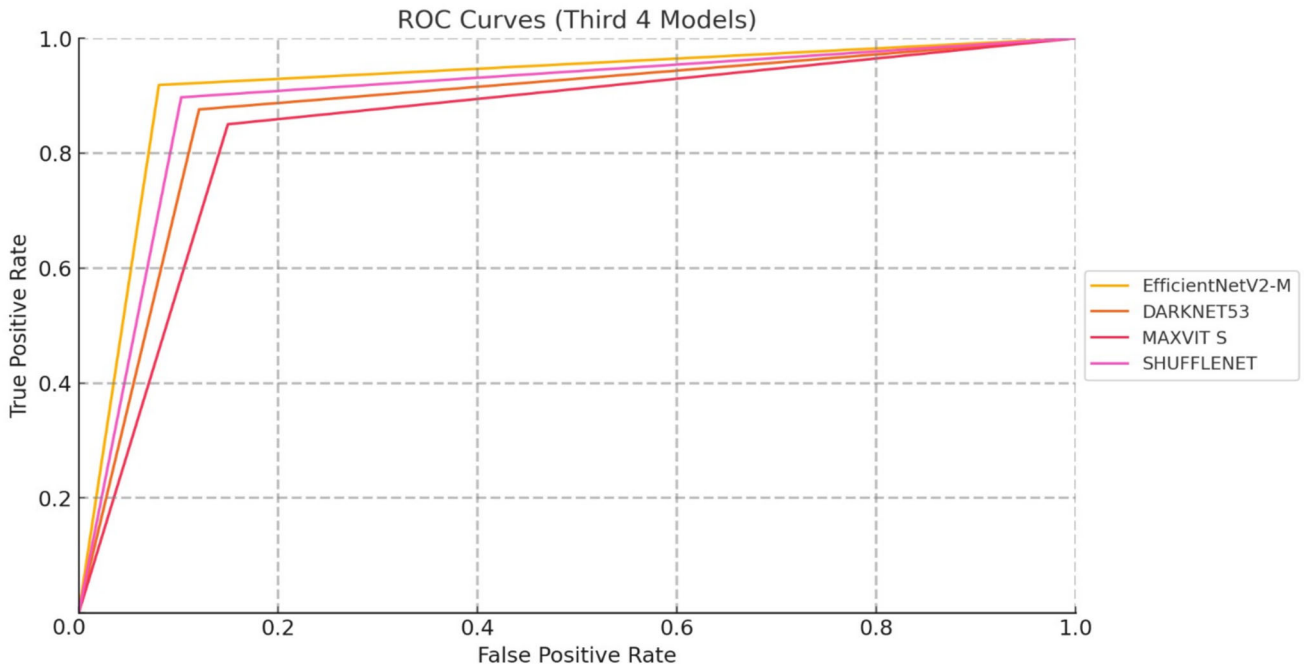


Fig. 4 ROC Curves of EfficientNetV2-M, DarkNet-53, MaxVit-S, and ShuffleNet

computational costs and slower inference times, making it less efficient than lightweight alternatives like MobileNetV3-L. The model's depth may also lead to overfitting, especially in smaller datasets, potentially reducing its generalization capability (Table 2). ROC curve analysis confirms that DarkNet-53 remains competitive with models like MaxVit-S and ShuffleNetV2, despite being outperformed by the best models in this study. Overall, DarkNet-53 is a strong and reliable model for macrofungi classification, but its computational expense and inference speed must be considered when selecting an optimal model for specific applications (Fig. 4). MaxVit-S achieved 85% accuracy, an F1-score of 0.85, and a specificity of 0.97, making it a competitive model for macrofungi classification. Its hybrid architecture, which combines CNNs with Transformer-based self-attention mechanisms, allows it to effectively recognize fungal features across multiple scales. This multi-scale feature extraction capability enhances its ability to distinguish visually similar species, giving it an edge over traditional CNN models in certain scenarios. However, despite its strong classification performance, MaxVit-S underperforms compared to EfficientNetV2-M and DenseNet121, both of which utilize more optimized feature extraction techniques while maintaining computational efficiency. With an AUC of 0.85, MaxVit-S demonstrates good separability between classes, but its classification ability does not reach the near-perfect separation observed in EfficientNetV2-M (Table 2). ROC curve analysis confirms that MaxVit-S remains a strong contender, though it falls slightly behind models with more refined feature reuse mechanisms. Additionally, its higher computational requirements compared to standard CNNs make it less suitable for low-resource environments, limiting its applicability where inference speed is a concern. Overall, MaxVit-S delivers strong classification accuracy and benefits from Transformer-based feature learning, but its computational cost and slight underperformance relative to the best models must be considered in real-world applications (Fig. 4). ShuffleNetV2 achieved 90% accuracy, an F1-score of 0.90, and a specificity of 0.98, making it one of the top-performing lightweight models in macrofungi classification. Designed for computational efficiency, ShuffleNetV2 utilizes group convolutions and channel shuffling techniques, which significantly reduce computational costs while maintaining high classification accuracy. Compared to heavier models like EfficientNet-B7 and ResNet101, it offers a strong trade-off between performance and efficiency, making it an ideal choice for real-time and mobile applications. With an AUC of 0.90, ShuffleNetV2 demonstrates strong class separability, outperforming traditional architectures such as VGG-16 and ResNetV2-50 (Table 2). ROC curve analysis confirms that ShuffleNetV2's classification ability is comparable to EfficientNet-B4, further highlighting its ability to retain high accuracy while using fewer parameters. Despite its lightweight architecture, it performs better than deeper but less optimized models, reinforcing the importance of efficient feature extraction over mere network depth. Overall, ShuffleNetV2 is a highly efficient and well-optimized model for macrofungi classification, proving that even lightweight

architectures can achieve high performance when designed effectively (Fig. 4).

VGG-16 achieved 86% accuracy, an F1-score of 0.86, and a specificity of 0.97, demonstrating that older deep learning architectures can still perform competitively in macrofungi classification. Built on a simple yet deep sequential architecture with 3×3 convolutional layers, VGG-16 effectively extracts hierarchical features, making it capable of capturing fine-grained fungal patterns. However, despite its high accuracy, its lack of architectural optimizations such as residual connections, batch normalization, or depthwise separable convolutions makes it less efficient than modern architectures like EfficientNetV2-M and DenseNet121, both of which achieve higher classification performance while using fewer parameters. Compared to VGG-19, its slightly shallower design results in faster inference times without a significant drop in accuracy, suggesting that increasing depth beyond this point does not necessarily enhance classification capability. With an AUC of 0.86, VGG-16 demonstrates strong class separability, though it falls behind newer architectures that integrate advanced feature reuse mechanisms (Table 2). ROC curve analysis confirms that while VGG-16 remains a strong performer, it does not match the efficiency and generalization capabilities of state-of-the-art models. Additionally, its high parameter count increases memory consumption, making it less suitable for resource-limited environments. Overall, VGG-16 is a reliable and well-established model for macrofungi classification, but its lack of modern architectural enhancements prevents it from reaching the top-tier performance of more optimized networks (Fig. 5). Xception achieved 87% accuracy, an F1-score of 0.87, and a specificity of 0.97, making it one of the top-performing models in macrofungi classification. Built upon depthwise separable convolutions, Xception significantly reduces computational complexity while maintaining strong feature extraction capabilities, allowing it to generalize well across visually similar fungal species. Compared to traditional CNN architectures like VGG-19 and ResNet101, Xception achieves higher classification accuracy with fewer parameters, making it a more efficient model for deep learning tasks. Despite its strong performance, it slightly underperforms compared to EfficientNetV2-M and DenseNet121, both of which incorporate more advanced architectural optimizations that improve feature reuse and gradient flow. With an AUC of 0.87, Xception demonstrates high class separability, though it does not surpass the highest-performing models in this study (Table 2). ROC curve analysis confirms that while Xception is highly effective, it does not reach the level of models with optimized training techniques, such as EfficientNetV2-M, which benefits from enhanced scaling and adaptive feature extraction. Additionally, while Xception is more efficient than traditional CNNs, it still requires significant computational resources compared to lightweight models like MobileNetV3-L and ShuffleNetV2. Overall, Xception is a well-balanced and powerful model for macrofungi classification, offering strong accuracy and efficiency, but it is slightly outperformed by the most optimized architectures in this study (Fig. 5). VGG-19 achieved 85% accuracy, an F1-score of 0.84, and a specificity of 0.97, demonstrating strong classification performance despite being an older deep learning architecture. As a deeper variant of VGG-16, it maintains the same sequential 3×3 convolutional structure, but with additional layers intended to improve feature extraction. However, the increase in depth does not yield a significant performance boost, as its accuracy is only 1% higher than VGG-16, while its computational cost and memory requirements are considerably higher. Compared to modern architectures such as EfficientNetV2-M and DenseNet121, which utilize more advanced feature reuse mechanisms and optimization techniques, VGG-19 falls behind in both efficiency and generalization ability. With an AUC of 0.84, it exhibits good class separability but lags behind top-tier models, which achieve higher accuracy with more refined architectures (Table 2). ROC curve analysis indicates that while VGG-19 remains a capable model, it is outperformed by networks that integrate batch normalization, residual connections, or depthwise separable convolutions. Additionally, its high parameter count and lack of computational optimizations make it less suitable for large-scale applications or environments where inference speed is a priority. Overall, VGG-19 is a solid model for macrofungi classification, but its increased depth does not provide enough performance gains to justify its higher computational demands compared to VGG-16 or more modern architectures (Fig. 5). SqueezeNet achieved 82% accuracy, an F1-score of 0.82, and a specificity of 0.96, making it a moderately strong lightweight model for macrofungi classification. Designed to maintain AlexNet-level accuracy with significantly fewer parameters, SqueezeNet utilizes "fire modules" to reduce the total number of parameters while preserving classification capability. Despite its compact architecture, its performance remains competitive with deeper models, outperforming ResNet101 and VGG-19, which have significantly more parameters but fail to generalize as effectively. However, compared to ShuffleNetV2 and EfficientNet variants, SqueezeNet underperforms slightly in both accuracy and separability, as indicated by its AUC of 0.83 (Table 2). ROC curve analysis confirms that while SqueezeNet maintains reasonable classification capability, it does not match the efficiency of newer lightweight models like ShuffleNetV2, which offers higher accuracy with comparable computational cost. Additionally, while its small parameter count makes it highly efficient, its feature extraction ability is not as advanced as depthwise separable convolutions or dense connectivity-based architectures, which limits its ability to distinguish between visually similar macrofungi species. Overall, SqueezeNet is a solid

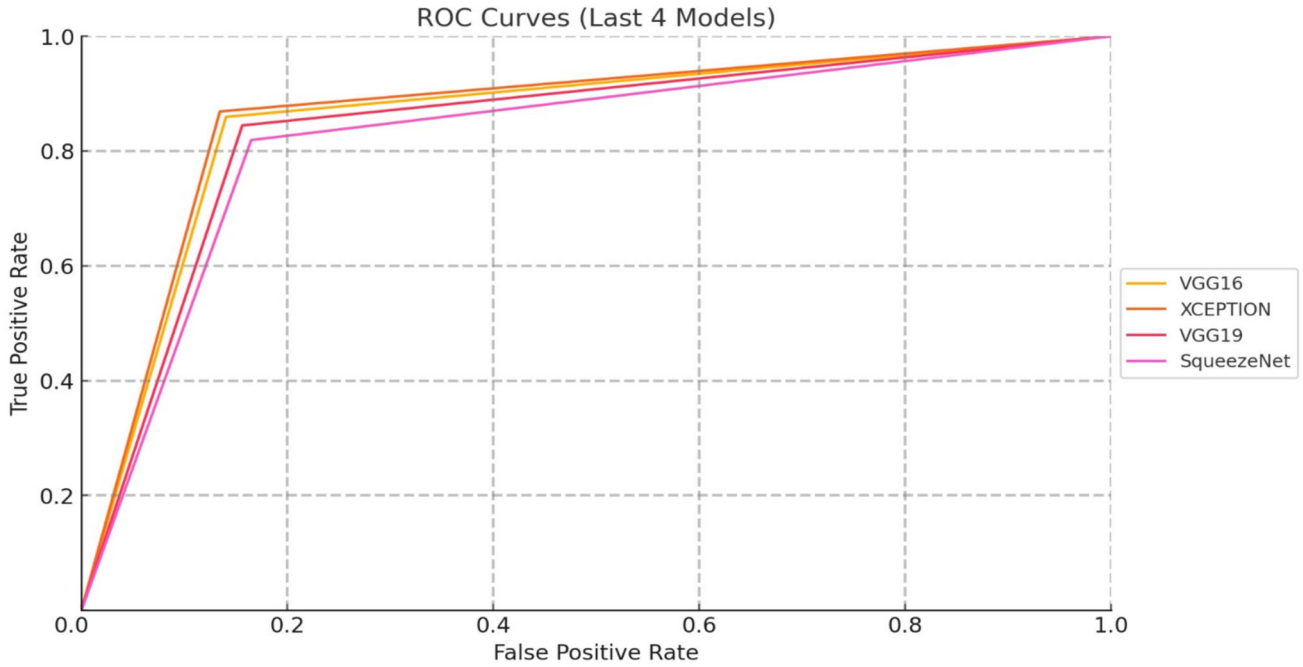


Fig. 5 ROC Curves of VGG-16, VGG-19, Xception, and SqueezeNet

lightweight model for macrofungi classification, offering efficiency and reasonable accuracy, but it is outperformed by more modern and optimized architectures in this study (Fig. 5).

The results of this study highlight a clear distinction between high-performing and low-performing models in macrofungi classification. DenseNet121, EfficientNet-B0, EfficientNet-B4, EfficientNetV2-M, and ShuffleNetV2 emerged as the most successful models, achieving the highest accuracy and F1-scores (Table 2). These models effectively leveraged deep hierarchical feature extraction, residual connections, or attention-based mechanisms, allowing them to generalize better despite the limited dataset size.

On the other hand, ResNet101 exhibited the weakest performance among all models, achieving an accuracy of only 45% and an F1-score of 0.38. This significantly lower performance can be attributed to multiple factors. First, as a deep architecture, ResNet101 typically requires a large and diverse dataset to fully leverage its hierarchical feature extraction capabilities. Given the relatively limited dataset used in this study, it likely suffered from overfitting, struggling to generalize well to new samples. Additionally, its high parameter count may have contributed to inefficient learning on this dataset, leading to suboptimal feature representation and classification. Unlike its shallower counterpart, ResNetV2-50, which demonstrated better efficiency and generalization, ResNet101 failed to balance depth with effective feature learning, ultimately resulting in poor classification performance (Table 2).

For the remaining models, while their performance was slightly lower than that of the top-performing architectures, the differences were not drastically large. Models such as VGG-16, VGG-19, SqueezeNet, and MaxVit-S achieved competitive results, demonstrating accuracy levels between 82 and 86% with balanced precision and recall values. Their performance suggests that although they did not reach the highest classification accuracy, they still managed to effectively differentiate fungal species with reasonable reliability. This indicates that model optimization and architectural design play a crucial role, sometimes even outweighing the sheer depth of a network. The relatively close performance gap between most models highlights that a variety of deep learning approaches can be viable for macrofungi classification, depending on the computational constraints and accuracy requirements (Table 2).

Ultimately, this study underscores the importance of choosing the right CNN architecture for macrofungi classification, with models that balance depth, efficiency, and feature extraction proving to be the most effective. Further improvements could be made by expanding the dataset size, fine-tuning hyperparameters, and leveraging additional augmentation techniques to enhance model generalization.

The graphs (Figs. 6 and 7) illustrate the training and validation accuracy of all models over 20 epochs, providing insights into their learning behavior and generalization capabilities. Figure 6 shows that DenseNet121, EfficientNet-B0, EfficientNet-B4, EfficientNetV2-M, and ShuffleNetV2 achieved the highest training accuracy, nearing 100% within

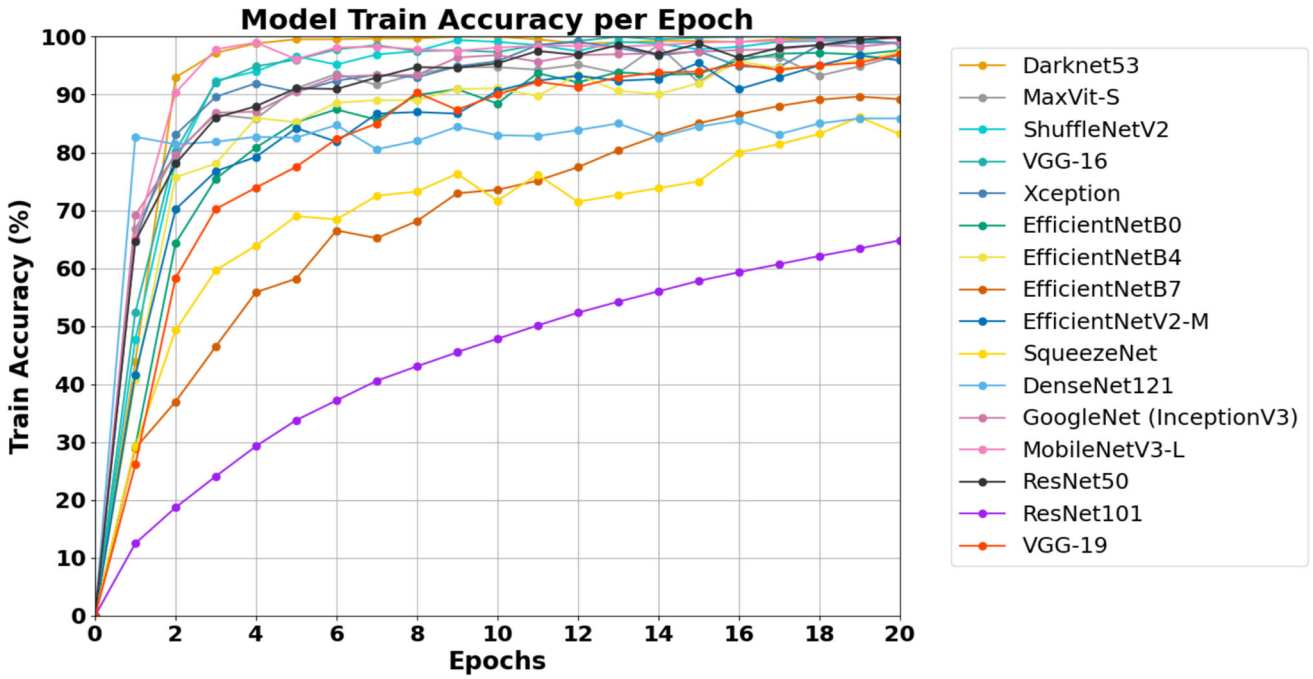


Fig. 6 Training accuracy of deep learning models over 20 epochs

the initial epochs, indicating effective learning and robust feature extraction. In contrast, models like the ResNet101 and SqueezeNet exhibited lower training accuracy, likely due to their lightweight architectures limiting their ability to capture complex features. Figure 7 highlights validation accuracy, revealing that although the top-performing models maintained strong performance, they did not achieve the near-perfect accuracy seen in training, suggesting a degree of overfitting and reduced generalization to unseen data.

On the other hand, models such as EfficientNet-B0, EfficientNet-B4, EfficientNetV2-M, ShuffleNetV2, and SqueezeNet showed poor validation accuracy, suggesting that they not only struggled to learn the training data but also failed to generalize effectively. Interestingly, VGG-16 and VGG-19 showed a moderate level of improvement over time but still lagged behind the highest-performing models. One key observation is that some deep models, such as ResNet101, did not perform as well as their optimized counterparts (e.g., ResNetV2-50). This further reinforces the importance of model efficiency in achieving strong classification performance with a limited dataset. The high training accuracy but relatively lower validation accuracy for certain models, such as DenseNet121 and MaxVit-S, suggests that a larger dataset or additional regularization techniques could help mitigate overfitting and improve generalization performance.

Overall, these graphs highlight the trade-offs between model complexity and generalization ability, demonstrating that while deeper architectures tend to learn faster, they require careful tuning to avoid overfitting. The results suggest that efficient architectures with strong feature extraction capabilities are the best candidates for macrofungi classification.

The provided graphs (Figs. 8, 9, 10, and 11) illustrate the training loss and validation loss of all models across 10 epochs, offering insights into their optimization behavior and generalization ability. Figure 4 shows that DenseNet121, MaxVit-S, Xception, ResNetV2-50, and GoogleNet (InceptionV3) achieved the steepest decline in training loss, indicating that these models effectively minimized the error during training. Their loss values dropped significantly within the first few epochs, suggesting that they learned the patterns in the training dataset efficiently. On the other hand, models such as EfficientNet-B0, EfficientNet-B4, EfficientNetV2-M, ShuffleNetV2, and SqueezeNet exhibited minimal loss reduction, implying that they struggled to learn the training data effectively. Their loss curves remained relatively high, indicating that they did not extract meaningful features from the dataset as efficiently as the top-performing models. VGG-16 and VGG-19, while showing some improvement, did not reach the same level of optimization as DenseNet121 or MaxVit-S.

Figure 9 provides deeper insights into model generalization. While DenseNet121, MaxVit-S, ResNetV2-50, and GoogleNet (InceptionV3) achieved relatively low validation loss, the gap between their training and validation loss suggests some degree of overfitting. These models successfully minimized loss on the training set but could not maintain

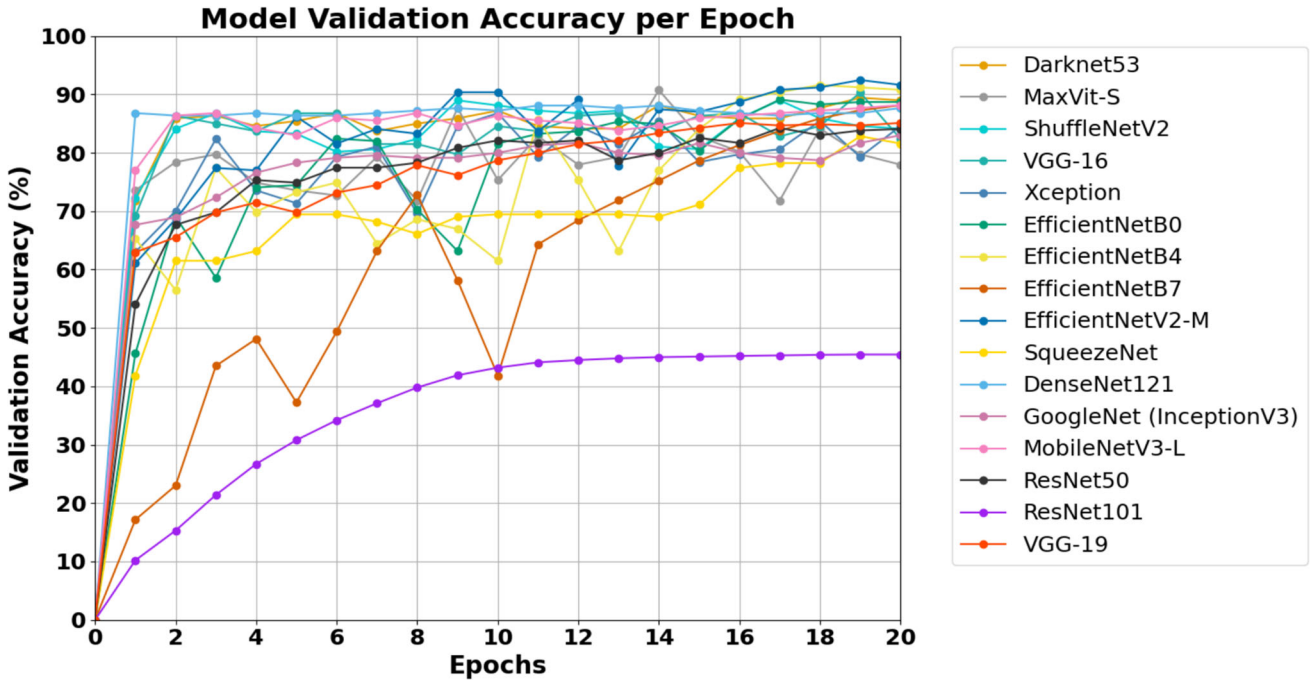


Fig. 7 Validation accuracy of deep learning models over 20 epochs

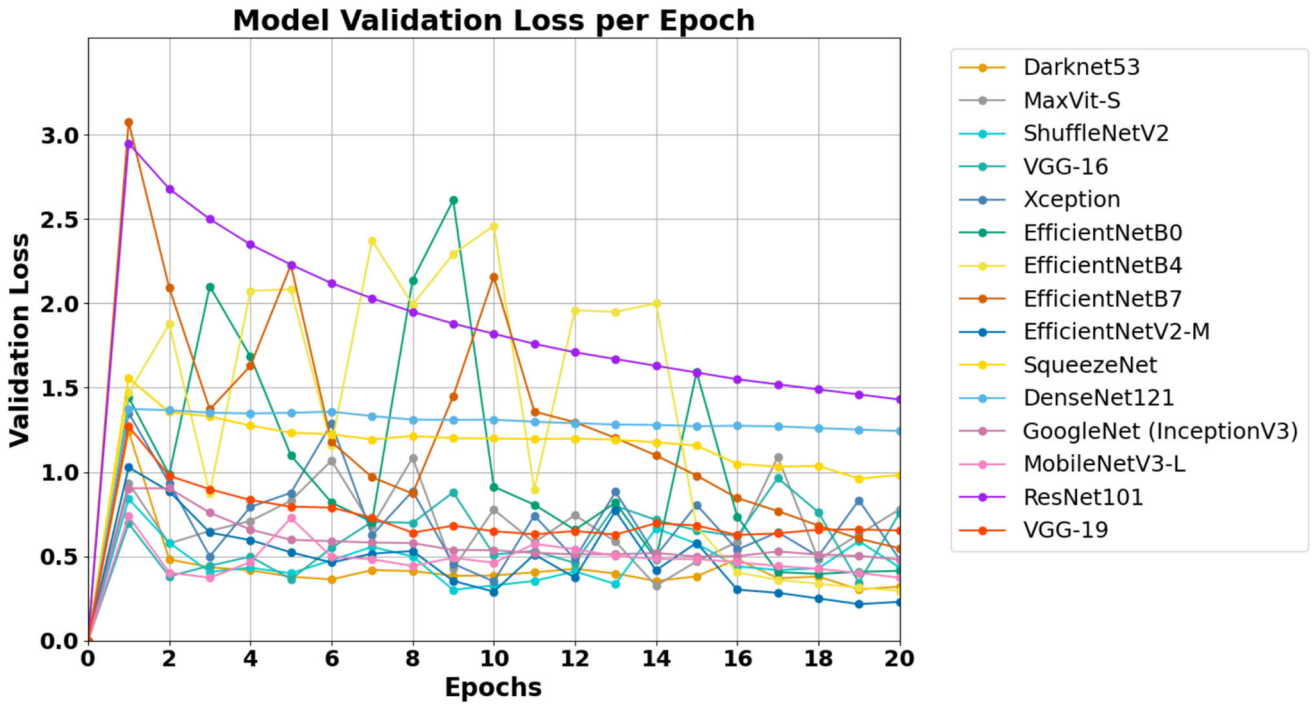


Fig. 8 Validation loss of deep learning models over 20 epochs

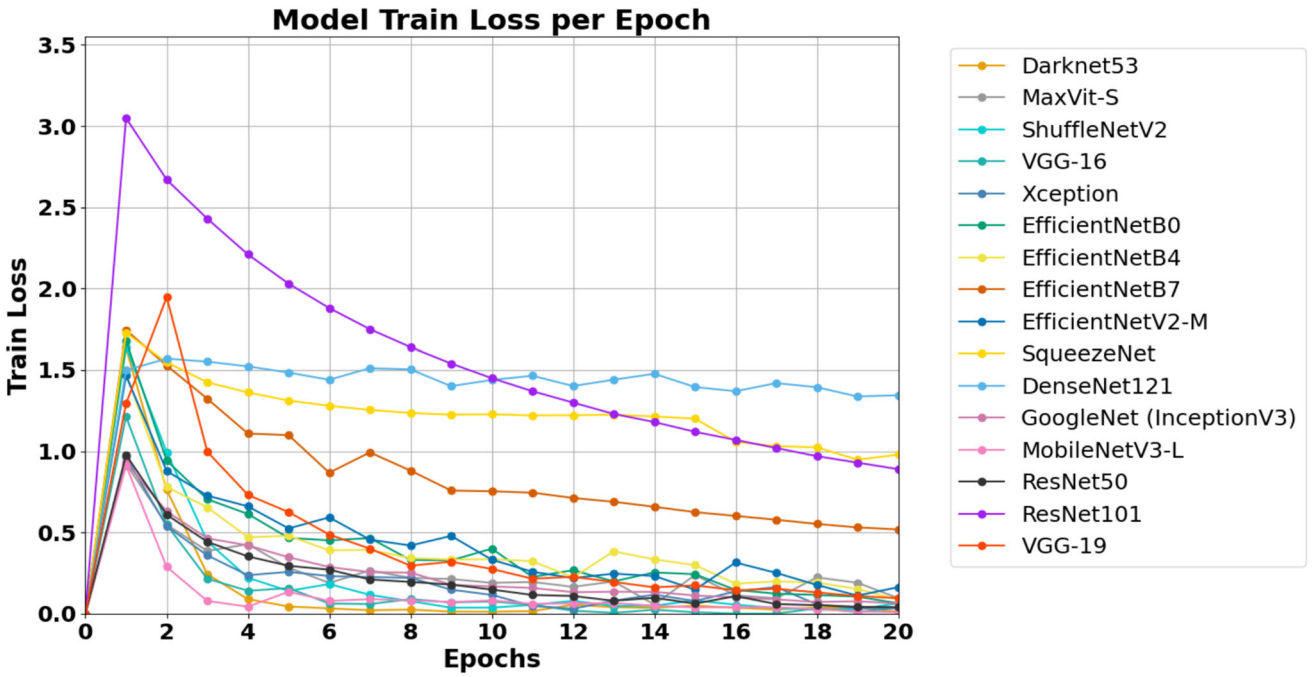


Fig. 9 Training loss of deep learning models over 20 epochs

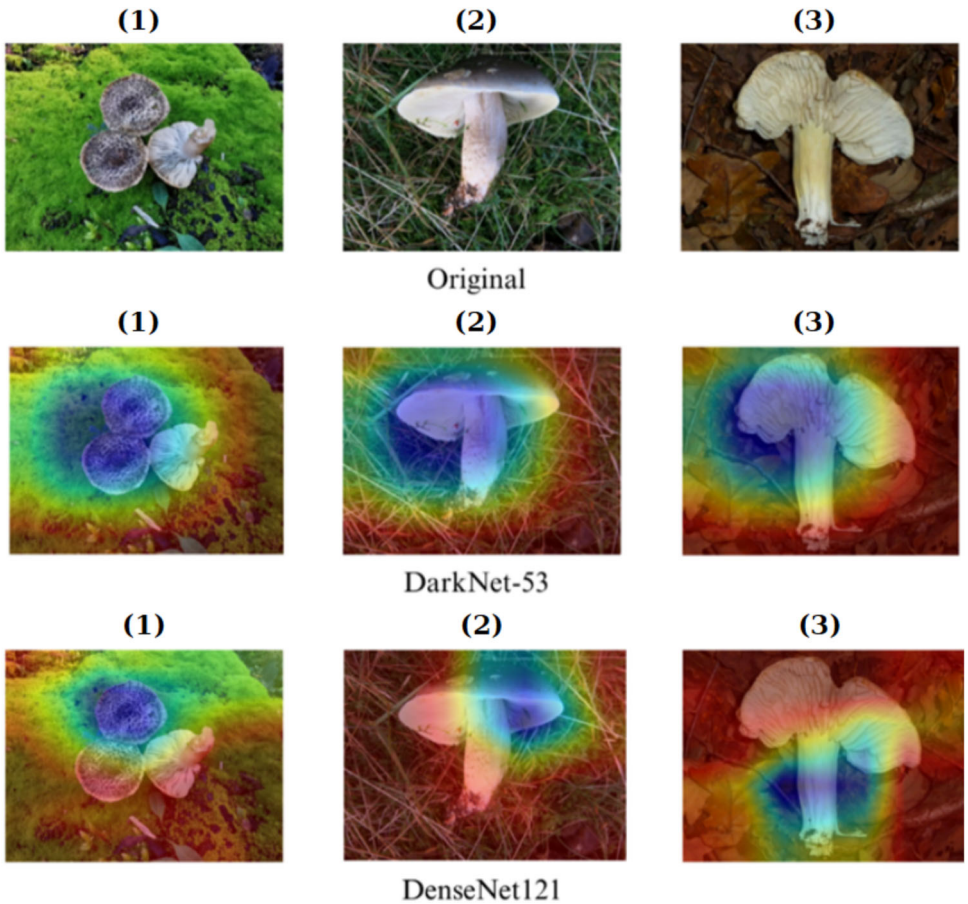


Fig. 10 Grad-CAM visualizations for macrofungi classification using DarkNet-53 and DenseNet121 models. The top row presents the original images of ***(1)****Tricholoma atrosquamosum*, ***(2)****Tricholoma saponaceum*, and ***(3)****Tricholoma sejunctum* from left to right. The second and third rows display the corresponding activation maps for DarkNet-53 and DenseNet121, respectively. These visualizations highlight the most influential regions used by the models for classification

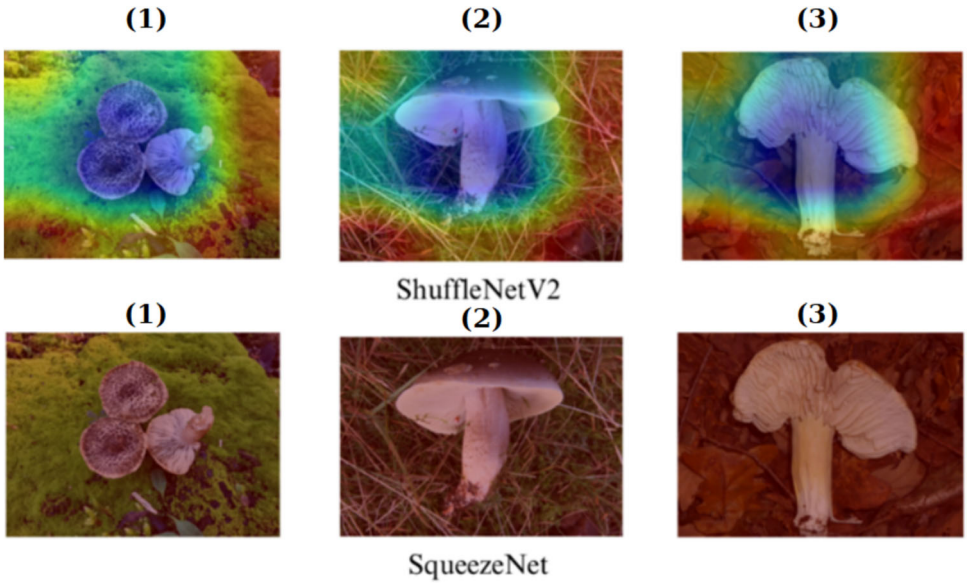


Fig. 11 Grad-CAM visualizations for macrofungi classification using ShuffleNetV2 and SqueezeNet models. The images correspond to three different macrofungi species: ***(1)****Tricholoma atrosquamosum*, ***(2)****Tricholoma saponaceum*, and ***(3)****Tricholoma sejunctum* from left to right. The top row presents the activation maps generated by ShuffleNetV2, while the bottom row displays the corresponding activation maps for SqueezeNet. These visualizations highlight the most influential regions considered by each model during classification

the same level of performance on the validation set, indicating that their learned features were more specialized to the training data rather than generalizing well to unseen data.

Conversely, EfficientNet series, ShuffleNetV2, and SqueezeNet exhibited consistently high validation loss, suggesting that they not only struggled to learn the training data but also failed to generalize well to the validation set. This is evident from the relatively flat loss curves, indicating poor learning capability. ResNet101 and VGG-16 showed moderate loss reduction, but their relatively higher validation loss compared to their optimized versions suggests that they required more data to fully utilize their deep architectures.

One critical observation is that some deep models, such as ResNet101 and VGG-16, exhibited higher validation loss compared to their optimized counterparts (e.g., ResNetV2-50 and DenseNet121). This reinforces the importance of model efficiency in achieving strong classification performance, particularly with a limited dataset. The high training loss reduction but slower validation loss decrease for models like DenseNet121 and MaxVit-S suggests that additional regularization techniques, such as dropout or data augmentation, could be beneficial in improving generalization.

Overall, these graphs (Figs. 6, 7, 8, and 9) highlight the trade-offs between model complexity and generalization, showing that while deeper architectures tend to minimize training loss effectively, their ability to generalize depends heavily on dataset size and proper optimization strategies. The results suggest that efficient architectures with strong feature extraction capabilities are best suited for macrofungi classification, as they achieve lower loss values while maintaining a reasonable balance between training and validation performance. To enhance model interpretability, we utilized explainable AI (XAI) techniques, specifically Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM helps visualize the most important image regions influencing model decisions by generating heatmaps, making it a valuable tool for understanding deep learning predictions.

This method is crucial for identifying biases, assessing model reliability, and improving classification accuracy. By applying Grad-CAM, we observed that higher-performing models focused on biologically relevant fungal structures, while weaker models often highlighted irrelevant areas, leading to misclassifications. In the following section, we present three original test images, followed by Grad-CAM visualizations for each of the 16 models (total of 48 images). These results will provide deeper insights into how different architectures process macrofungi images.

The Grad-CAM visualizations (Figs. 10, 11, 12, 13, 14, and 15) provide insights into the regions where each model focused while making classification decisions. High-performing models such as DenseNet121, EfficientNet-B0, EfficientNet-B4, EfficientNetV2-M, and ShuffleNetV2 effectively highlighted the key fungal structures, including the cap, stem, and gill patterns. These models exhibited strong spatial awareness, ensuring that their predictions were influenced by biologically relevant features rather than background noise. Conversely, weaker models like SqueezeNet and ResNet101 displayed inconsistent or diffused attention maps. Many of these models focused on irrelevant areas, such as the background or nondistinctive parts of the mushrooms, leading to potential misclassifications. Overall, these visualizations (Figs. 10, 11, 12, 13, 14, and 15) demonstrate the importance of explainable AI (XAI) in understanding model behavior. Grad-CAM helps evaluate whether models

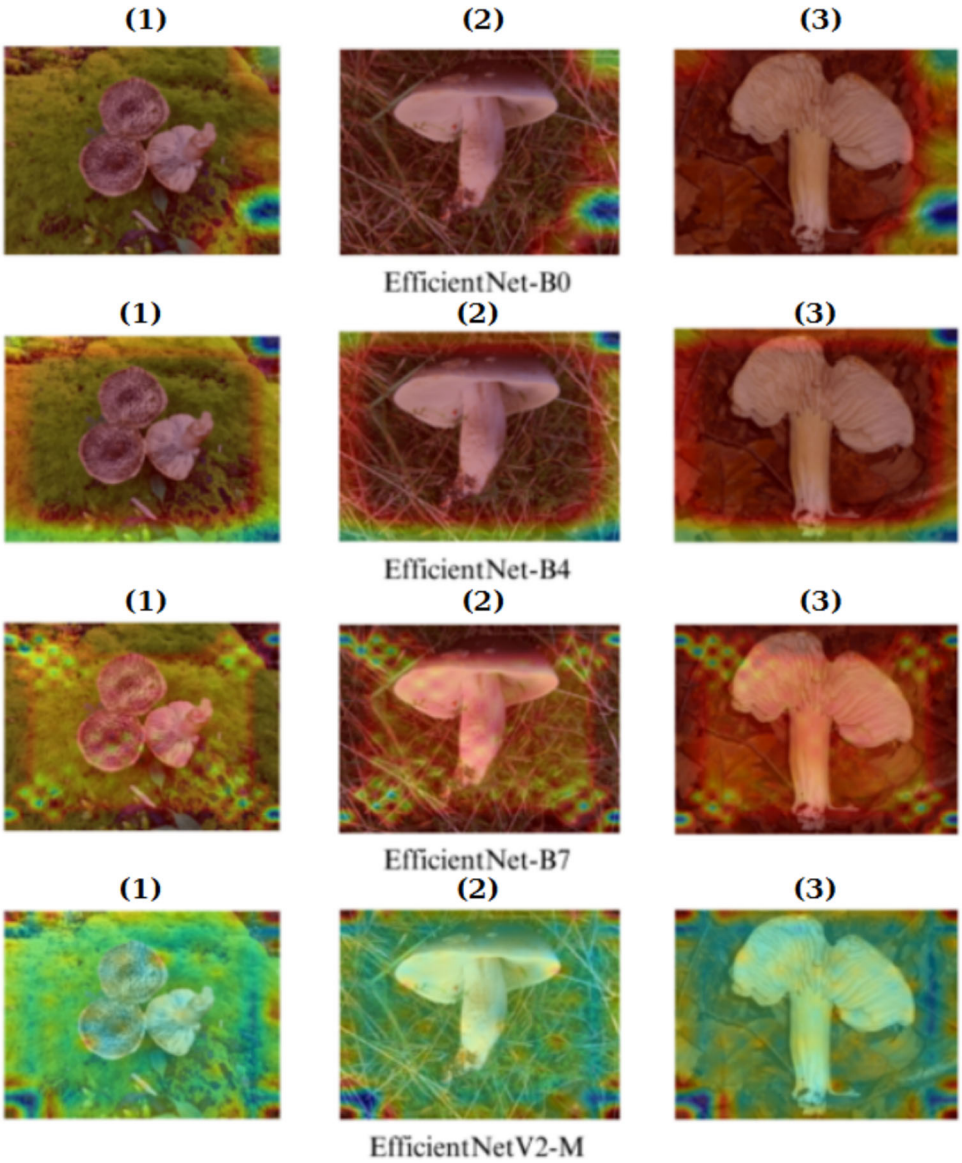


Fig. 12 Grad-CAM visualizations for macrofungi classification using EfficientNet models. The images correspond to three different macrofungi species: *(1) Tricholoma atrosquamosum*, *(2) Tricholoma saponaceum*, and *(3) Tricholoma sejunctum* from left to right. Each row represents a different EfficientNet variant, namely EfficientNet-B0, EfficientNet-B4, EfficientNet-B7, and EfficientNetV2-M, respectively. These visualizations highlight the key regions identified by each model during classification

rely on meaningful features or suffer from attention misalignment, guiding future improvements in model architecture and dataset refinement.

The p values obtained from the confusion matrices for each model using the chi-square significance test are presented in Appendix A, Table 3. In Table 3, the arrows indicate the model with the higher accuracy metric. The absence of an arrow in certain table cells signifies that the accuracy metrics of the two compared models are equal. Bold p values indicate a statistically significant difference between the confusion matrices of the two models, whereas non-bold p values suggest that there is no statistically significant difference, meaning that the performance difference between the two models is likely due to chance. As seen in Table 3, the performance differences between the models are not statistically significant. However, when all models are compared to ResNet101, the performance differences observed against ResNet101 are statistically significant. In other words, the superiority of the models over ResNet101 is not due to chance.

4 Discussion

This study systematically evaluates the effectiveness of various deep learning architectures for macrofungi classification, revealing significant performance disparities among the models [37]. The findings indicate that DenseNet121[38], EfficientNet-B0[39], EfficientNet-B4, EfficientNetV2-M[40], and ShuffleNetV2[41] were the most effective models, achieving high classification accuracy, precision, recall, F1-score, specificity, and AUC score

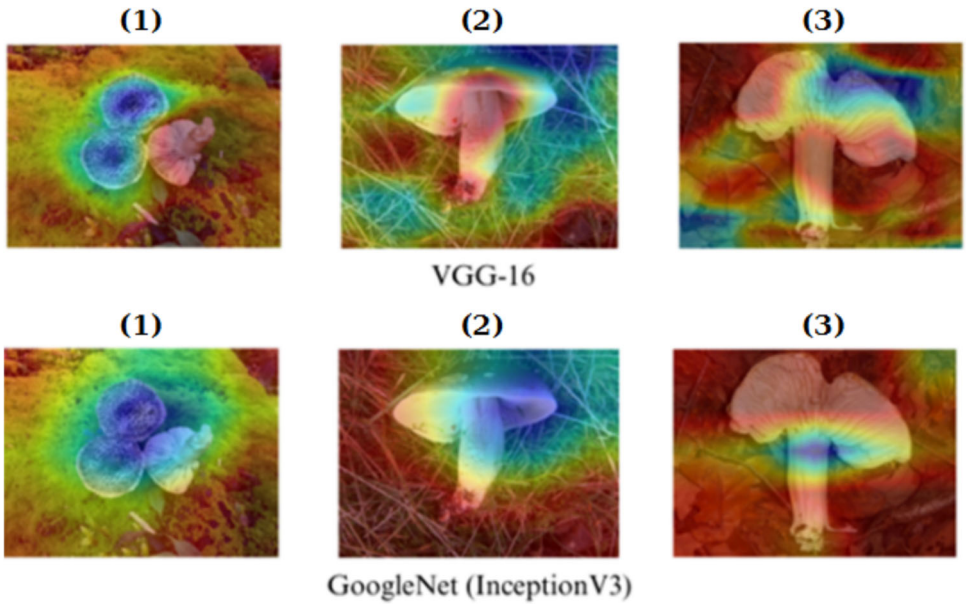


Fig. 13 Grad-CAM visualizations for macrofungi classification using VGG-16 and GoogleNet (InceptionV3) models. The images correspond to three different macrofungi species: ***(1)****Tricholoma atrosquamosum*, ***(2)****Tricholoma saponaceum*, and ***(3)****Tricholoma sejunctum* from left to right. The first row presents activation maps generated by VGG-16, while the second row displays those produced by GoogleNet (InceptionV3)

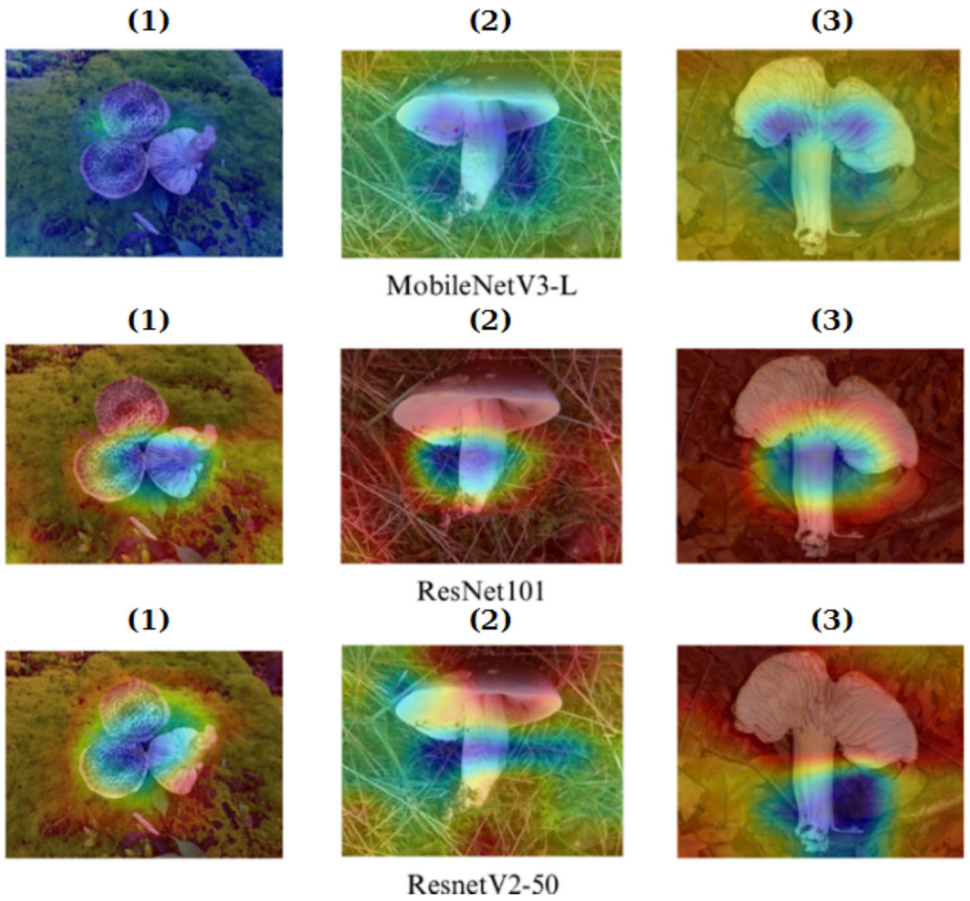


Fig. 14 Grad-CAM visualizations for macrofungi classification using MobileNetV3-L, ResNet101, and ResNetV2-50 models. The images correspond to three different macrofungi species: ***(1)****Tricholoma atrosquamosum*, ***(2)****Tricholoma saponaceum*, and ***(3)****Tricholoma sejunctum* from left to right. The first row presents activation maps generated by MobileNetV3-L, the second row displays those produced by ResNet101, and the third row corresponds to ResNetV2-50

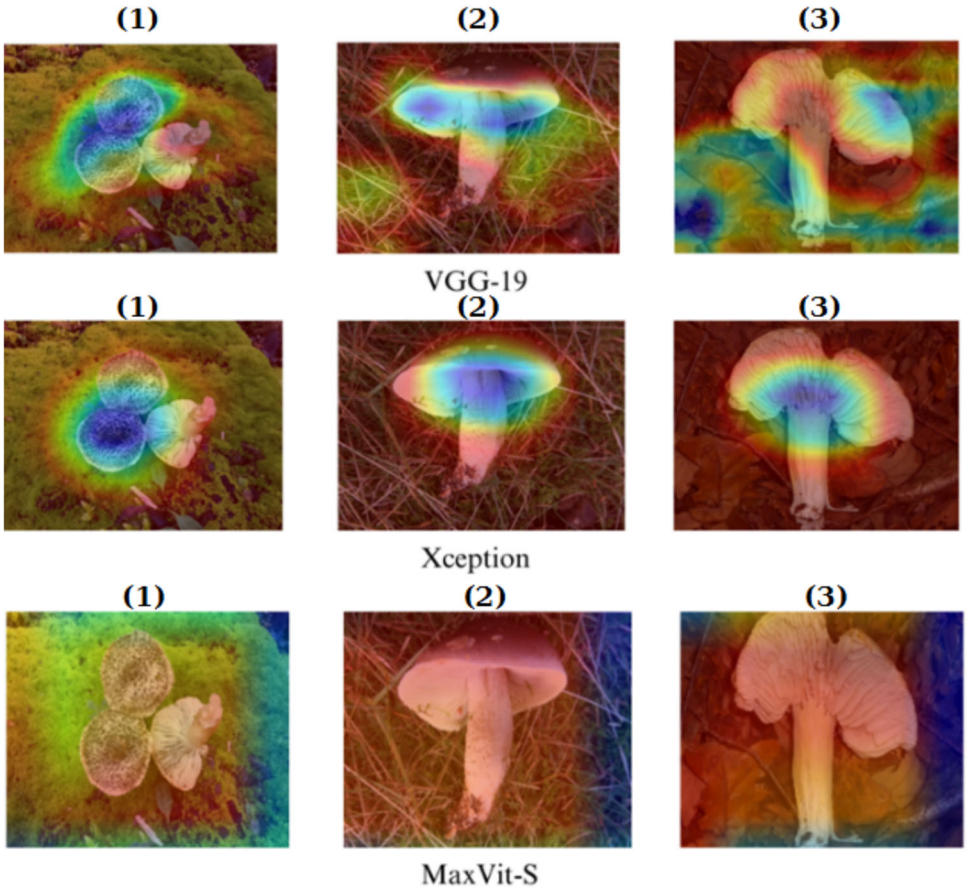


Fig. 15 Grad-CAM visualizations for macrofungi classification using VGG-19, Xception, and MaxVit-S models. The images correspond to three different macrofungi species: ***(1)****Tricholoma atrosquamosum*, ***(2)****Tricholoma saponaceum*, and ***(3)****Tricholoma sejunctum* from left to right. The first row presents activation maps generated by VGG-19, the second row displays those produced by Xception, and the third row corresponds to MaxVit-S. These visualizations highlight the most influential regions considered by each model during classification

well-localized feature extraction in Grad-CAM visualizations. These models outperformed their counterparts due to their ability to capture fine-grained fungal structures, generalize well across different species, and effectively utilize their deep architectural advantages.

The success of the top-performing models in this study can be attributed to their optimized feature extraction mechanisms, efficient connectivity, and advanced architectural designs. EfficientNetV2-M emerged as the most accurate model, achieving an accuracy of 92%, along with 0.92 F1-score and 0.98 specificity. Its superior performance can be attributed to its compound scaling approach, which effectively balances depth, width, and resolution for enhanced feature extraction. Similarly, DenseNet121, with 90% accuracy, excels in maintaining efficient gradient flow through its densely connected layers, facilitating enhanced learning while minimizing redundant computations [42]. EfficientNet-B4, also achieving 90% accuracy, benefits from optimized depth and scaling strategies, ensuring robust generalization while maintaining computational efficiency [43]. ResNetV2-50, with 84% accuracy, leverages residual connections to prevent vanishing gradient issues, making it a stable deep learning model for macrofungi classification. Meanwhile, GoogleNet (InceptionV3), which reached 84% accuracy, benefits from its multi-scale feature extraction via inception modules, allowing it to capture intricate macrofungi patterns at different resolutions.

On the other hand, models such as MaxVit-S, MobileNetV3-L, and SqueezeNet exhibited moderate performance, with accuracy values ranging between 82 and 88%. While these models are designed for efficiency and lightweight applications, their feature extraction capabilities are somewhat limited compared to the top-performing architectures. MaxVit-S, for instance, integrates self-attention mechanisms, improving its focus on essential fungal structures, yet its 85% accuracy suggests that further optimization might be necessary for superior classification [44]. SqueezeNet, which is an ultra-lightweight model optimized for fast inference, struggled with distinguishing between complex fungal features, achieving 82% accuracy, the lowest among the moderate-performing models. These results indicate that while computationally efficient models offer speed advantages, they may require additional fine-tuning or data augmentation to match the performance of deeper architectures.

A particularly significant observation in this study is the weak performance of ResNet101, which achieved only 45% accuracy—substantially lower than other models. Despite

its deep architecture, the model suffered from inefficient feature extraction, likely due to its high parameter count and reliance on large-scale datasets for effective training [45]. The limited dataset size in this study may have led to overfitting, preventing ResNet101 from generalizing effectively to new macrofungi images. Furthermore, its low F1-score of 0.38 and specificity of 0.89 suggest that the model frequently misclassified fungal species, reinforcing the notion that deep networks do not always guarantee superior performance. In contrast, more optimized architectures, such as EfficientNetV2-M, DenseNet121, and ResNetV2-50, outperformed ResNet101 despite having fewer parameters, demonstrating the importance of model efficiency over raw depth.

In addition to ResNet101, VGG-16 and VGG-19, two widely used deep architectures, exhibited relatively lower accuracy compared to other top-performing models [46]. VGG-16 achieved 86% accuracy, while VGG-19 reached 85% accuracy, highlighting their capability to classify macrofungi but also suggesting that their conventional feature extraction pipelines may not be as efficient as modern architectures like EfficientNet or DenseNet. Their large memory requirements and fully connected layers make them computationally demanding, yet their performance in this study was surpassed by more optimized models [47]. Similarly, ShuffleNetV2, which achieved 90% accuracy, performed well compared to other lightweight models but still fell slightly behind the leading models in overall classification effectiveness.

The findings from this study emphasize the crucial role of model selection and architectural optimization in macrofungi classification. The results clearly indicate that certain deep learning models, such as EfficientNetV2-M, DenseNet121, and EfficientNet-B4, are more suited for handling the complexities of macrofungi images due to their robust feature extraction mechanisms and efficiency-driven designs. However, the study also underscores the importance of dataset size, as seen in the underperformance of deeper architectures like ResNet101, which require large-scale datasets to function effectively [48]. These findings suggest that future research should focus on optimizing models not just in terms of depth but also efficiency, ensuring that they are well suited for the available dataset scale.

Beyond the technical evaluation of deep learning architectures, recent studies across diverse AI domains provide complementary perspectives for extending model optimization. For instance, traditional machine learning frameworks used in fake review detection emphasize the critical role of feature selection and data preprocessing, aligning with the importance of dataset curation in visual classification [49]. Likewise, advances in generative AI for trustworthy e-commerce demonstrate that optimized deep learning pipelines can successfully transition to real-world applications, underscoring their scalability beyond image-based tasks [50]. Integrating optimized AI systems with blockchain infrastructures further points to the potential for secure and decentralized deployment of intelligent models, enhancing traceability and trust within digital ecosystems [51]. Similar optimization philosophies are evident in NLP-based trust systems, which suggest that convergence between vision and language modalities could enhance interpretability and consistency in multimodal learning [52]. Moreover, emerging discussions on ethical and explainable AI highlight the necessity of embedding fairness, transparency, and accountability principles within optimization workflows [53]. Finally, the use of large language models for fraud analytics signals a broader shift toward generative and multimodal paradigms, suggesting that hybrid vision–language architectures may represent the next step in advancing adaptability and reasoning capabilities of optimized deep learning models [54].

The dataset used in this study comprises 1,390 images of six Tricholoma species, which may be relatively limited for training deep convolutional neural networks and could contribute to the observed overfitting in ResNet101 (45% accuracy) as well as the mild overfitting detected in top-performing models. Previous fungal classification studies by our group have shown that larger and more diverse datasets significantly enhance model generalization and reduce overfitting. For example, datasets including 1,000 Phallaceae [55], 1582 Mycena and Marasmius [56], 2,800 Discomycetes [57], 1600 Puffball [58], and 1,585 Earthstar [59] images provided more stable and reliable classification outcomes. Therefore, future work will focus on expanding the Tricholoma dataset following similar curation and diversity standards to further strengthen model robustness and ecological inference accuracy.

Beyond classification accuracy, the implications of this study extend to various fields, including biodiversity monitoring, ecological research, and agricultural applications. The ability to accurately classify fungal species using AI-driven models holds significant potential for environmental conservation, early disease detection in plants, and mycological studies. The results of this study contribute valuable insights into the strengths and limitations of different deep learning architectures, offering a solid foundation for future advancements in AI-powered fungal classification. Moving forward, increasing dataset diversity, refining model architectures, and exploring hybrid approaches integrating CNN and transformer-based models will be key strategies to further improve classification performance.

5 Conclusıon

This study conducted a comprehensive evaluation of various deep learning models for macrofungi classification, identifying the most effective architectures. DenseNet121 (F1-score of 90%), EfficientNetV2-M (92% accuracy, 92% precision, and 98% specificity), EfficientNet-B0, EfficientNet-B4 (F1-score of 90%), and ShuffleNet demonstrated superior performance in extracting complex features and capturing subtle structural differences. Their success is attributed to their deep, multilayered architectures and optimized learning strategies. ResNet101 struggled with overfitting and poor generalization due to its depth and the limited dataset size, preventing effective feature extraction. SqueezeNet, as an ultra-lightweight model, lacked sufficient depth for capturing intricate fungal structures. These results highlight the importance of both model efficiency and dataset diversity in achieving optimal classification performance.

This research is notable for being the first to extensively apply explainable artificial intelligence (XAI) techniques in macrofungi classification. Grad-CAM visualizations revealed that high-performing models focused on biologically significant features such as the cap, stem, and spore structures. In contrast, lower-performing models often misclassified samples by focusing on irrelevant regions. These visualizations enhanced the transparency and interpretability of the models' decision-making processes, increasing confidence in their reliability.

It is recommended that the deep learning algorithms used in this study be applied not only to macrofungi but also to the microscopic and spore-level images of other fungal species. Furthermore, these models have potential applications in other domains, such as diagnosing plant diseases, classifying agricultural plant species, and biodiversity monitoring. Future research should focus on expanding dataset diversity, exploring hybrid architectures (e.g., CNN-Transformer models), optimizing hyperparameters, and developing advanced data augmentation methods to enhance model performance and generalization capacity. Additionally, improving model interpretability through explainable AI techniques will be crucial for handling more complex biological datasets.

Appendix A

Table 3 p values for models

Models	DarkNet-53	DenseNet121	EffNet-B0	EffNet-B4	EffNet-B7	EffNetV2	GoogleNet	MaxVit-S	MobileNetV3-L	ResNet50	ResNet101	ShuffleNet	SqueezeNet	VGG-16	VGG-19	Xception
DarkNet-53		9.980e-01 ↑	9.957e-01 ↑	9.994e-01 ↑	9.999e-01 ←	1.000e + 00 ↑	7.105e-01 ←	9.858e-01 ←	9.994e-01	9.621e-01 ←	**0.000e + 00 ←**	9.935e-01 ↑	7.665e-01 ←	9.996e-01 ←	9.830e-01 ←	9.999e-01 ←
DenseNet121			9.996e-01 ←	1.000e + 00	9.901e-01 ←	9.993e-01 ↑	8.973e-01 ←	9.809e-01 ←	9.997e-01 ←	9.836e-01 ←	**0.000e + 00 ←**	9.996e-01	6.343e-01 ←	9.506e-01 ←	9.996e-01 ←	9.997e-01 ←
EffNet-B0				9.989e-01 ↑	9.721e-01 ←	9.996e-01 ↑	4.534e-01 ←	9.974e-01 ←	9.790e-01 ←	8.088e-01 ←	**0.000e + 00 ←**	9.971e-01 ↑	4.743e-01 ←	9.583e-01 ←	9.883e-01 ←	9.993e-01 ←
EffNet-B4					9.971e-01 ←	1.000e + 00 ↑	9.355e-01 ←	9.909e-01 ←	9.971e-01 ←	9.791e-01 ←	**0.000e + 00 ←**	9.832e-01	5.619e-01 ←	9.493e-01 ←	9.987e-01 ←	9.984e-01 ←
EffNet-B7						9.991e-01 ↑	7.303e-01 ←	9.953e-01 ←	9.992e-01 ↑	9.759e-01 ←	**0.000e + 00 ←**	9.819e-01 ↑	4.183e-01 ←	1.000e + 00 ←	9.934e-01 ←	9.996e-01
EffNetV2							5.537e-01 ←	9.889e-01 ←	9.975e-01 ←	9.580e-01 ←	**0.000e + 00 ←**	9.988e-01 ←	4.251e-01 ←	9.971e-01 ←	9.440e-01 ←	9.992e-01 ←
GoogleNet								9.416e-01 ↑	9.925e-01 ↑	9.998e-01	**0.000e + 00 ←**	6.723e-01 ↑	9.112e-01 ←	9.412e-01 ↑	9.496e-01 ↑	6.752e-01 ↑
MaxViT-S									9.724e-01 ↑	9.975e-01 ←	**0.000e + 00 ←**	9.675e-01 ↑	8.248e-01 ←	1.000e + 00 ↑	9.864e-01	9.964e-01 ↑
MobileNetV3-L										9.999e-01 ←	**0.000e + 00 ←**	9.989e-01 ↑	9.745e-01 ←	9.943e-01 ←	9.998e-01 ←	9.990e-01 ←
ResNet50											**0.000e + 00 ←**	8.902e-01 ↑	8.517e-01 ←	9.985e-01 ↑	9.786e-01 ↑	9.451e-01 ↑
ResNet101												**0.000e + 00 ↑**	**1.998e-15 ↑**	**0.000e + 00 ↑**	**0.000e + 00 ↑**	**0.000e + 00 ↑**
ShuffleNet													6.239e-01 ←	9.533e-01 ←	8.721e-01 ←	9.932e-01 ←
SquezeeNet														6.101e-01 ↑	7.927e-01 ↑	4.700e-01 ↑
VGG-16															8.437e-01 ←	9.990e-01 ↑
VGG-19																9.858e-01 ↑
Xception																

Acknowledgements We would like to thank Teus Teknoloji A.Ş. for providing the technical infrastructure that enabled us to obtain data for the artificial intelligence algorithms used in this study.

Authors' contributions Conceptualization was performed by A.A., Ş.A., Ö.B.A., F.E., and E.K.; methodology was performed by M.S.G., F.E., I.A., T.A., and K.A.; software was provided by A.A., Ş.A., Ö.B.A., T.A., and K.A.; validation was carried out by M.S.G., I.A., and F.E.; formal analysis was carried out by A.A., Ş.A., Ö.B.A., T.A., and K.A.; investigation was performed by M.S.G., F.E., E.K., T.A., and K.A.; resources were acquired by A.A., Ş.A., Ö.B.A., M.S.G., F.E., T.A., K.A., I.A., and E.K.; writing—original draft preparation was carried out by A.A., Ş.A., Ö.B.A., F.E., K.A., T.A., and I.A.; writing—review and editing was carried out by M.S.G., I.A., and F.E.; visualization was performed by M.S.G., I.A., K.A., and F.E.; supervision was carried ou by M.S.G, I.A., K.A., T.A., and F.E.

Funding Open access funding provided by Tampere University (including Tampere University Hospital).

Availability of data and materials The dataset analyzed during the current study are available in the GBIF Backbone Taxonomy repository, https://doi.org/10.15468/39omei].

Declarations

Conflict of interest The authors declare no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Niego, A.G.T., Rapior, S., Thongklang, N., Raspé, O., Hyde, K.D., Mortimer, P.: Reviewing the contributions of macrofungi to forest ecosystem processes and services. Fungal Biol. Rev. (2023). https://doi.org/10.1016/j.fbr.2022.11.002
2. Oloyede, F.M., Owolabı, A.A., Akınloye, A.O., Odetunde, E.F., Alade, T.I.: Agricultural biodiversity and human health nexus: A review. Agric. Food Nat. Resour. J. (2024). https://doi.org/10.5281/zenodo.14413345
3. Şen, İ, Alli, H.: Tricholoma (Fr.) Staude in the Aegean region of Turkey. Turk. J. Bot. (2019). https://doi.org/10.3906/bot-1812-52
4. Heilmann-Clausen, J., Christensen, M., Frøslev, T.G., Kjøller, R.: Taxonomy of *Tricholoma* in northern Europe based on ITS sequence data and morphological characters. Persoonia-Molec. Phylog. Evolut. Fungi (2017). https://doi.org/10.3767/003158517X693174
5. Bessette, A.E., Bessette, A.R., Roody, W.C., Trudell, S.A.: Tricholomas of North America: A Mushroom Field Guide, New York. University of Texas Press, USA (2013)
6. Alem, D.: Fungal communities associated with forests in the afromontane region of ethiopia. https://doi.org/10.35376/10324/51840
7. Corbu, V.M., Gheorghe-Barbu, I., Dumbravă, A.Ș, Vrâncianu, C.O., Șesan, T.E.: Current insights in fungal importance a comprehensive review. Microorganisms (2023). https://doi.org/10.3390/microorganisms11061384
8. Pinar, O., Couto, S.R.: Biologically active secondary metabolites from white-rot fungi. Front. Chem. (2024). https://doi.org/10.3389/fchem.2024.1363354
9. Mukhamediev, R.I., Symagulov, A., Kuchin, Y., Zaitseva, E., Bekbotayeva, A., Yakunin, K., Assanov, I., Levashenko, V., Popova, Y., Akzhalova, A., Bastaubayeva, S., Tabynbaeva, L.: Review of some applications of unmanned aerial vehicles technology in the resource-rich country. Appl. Sci. (2021). https://doi.org/10.3390/app112110171
10. Khatri, S., Tripathy, B.K., Kumar, K.V.S.S., Kumar, V., Bharti, S., Gupta, A., Khandelwal, Y.: A deep learning approach for detection of disease in plant leaves. Int. Conf. Comput. Intell. Pattern Recogn. (2022). https://doi.org/10.1007/978-981-99-3734-9_15
11. Mateo, K.C.H., Navarro, L.K.B., Manlises, C.O.: Identification of macro-nutrient deficiency in onion leaves (Allium cepa L.) using convolutional neural network (CNN). In 2022 5th International Seminar on Research of Information Technology and Intelligent Systems *(ISRITI)* (2022). https://doi.org/10.1109/ISRITI56927.2022.10052884
12. In GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2024–12–06.
13. Chakraborty, S., Uzkent, B., Ayush, K., Tanmay, K., Sheehan, E., Ermon, S.: Efficient conditional pre-training for transfer learning In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022). https://doi.org/10.48550/arXiv.2011.10231
14. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2021). https://doi.org/10.48550/arXiv.2012.00364
15. Özbay, E., Özbay, F.A., Gharehchopogh, F.S.: Visualization and classification of mushroom species with multi-feature fusion of metaheuristics-based convolutional neural network model. Appl. Soft Comput. (2024). https://doi.org/10.1016/j.asoc.2024.111936
16. Gong, Y., Hou, Y., Zhang, C., Jiang, M.: Beyond augmentation: Empowering model robustness under extreme capture environments. International Joint Conference on Neural Networks (IJCNN) (2024) . https://doi.org/10.48550/arXiv.2407.13640
17. Parashar, A., Parashar, A., Ding, W., Shabaz, M., Rida, I.: Data preprocessing and feature selection techniques in gait recognition: a comparative study of machine learning and deep learning approaches. Pattern Recognit. Lett. (2023). https://doi.org/10.1016/j.patrec.2023.05.021
18. Archana, R., Jeevaraj, P.E.: Deep learning models for digital image processing: a review. Artif. Intell. Rev. (2024). https://doi.org/10.1007/s10462-023-10631-z
19. Muñoz, B., Remeseiro, B., Aguilar, E.: Towards a lightweight CNN for semantic food segmentation. Iberoamer. Congr. Pattern Recogn. (2024). https://doi.org/10.1007/978-3-031-76607-7_1
20. Yogabalajee, V., Sundaram, K., Kanagaraj, K.: soybean leaf disease classification using enhanced densenet121. In 2024 10th International Conference on Advanced Computing and Communication Systems ICACCS (2024). https://doi.org/10.1109/ICACCS60874.2024.10717087

21. Padhi, J., Korada, L., Dash, A., Sethy, P.K., Behera, S.K., Nanthaamornphong, A.: Paddy leaf disease classification using EfficientNet B4 with compound scaling and Swish activation: A deep learning approach. IEEE Access (2024). https://doi.org/10.1109/ACCESS.2024.3451557
22. Shafiq, M., Gu, Z.: Deep residual learning for image recognition: A survey. Appl. Sci. (Basel) (2022). https://doi.org/10.3390/app12188972
23. Asif, S.: Enhancing pulmonary abnormality detection with an optimized CNN architecture incorporating depth-wise separable convolution and Inception modüle. Evolv. Syst. (2024). https://doi.org/10.1007/s12530-023-09565-2
24. Patil, N.C., Patil, N.J.: A hybrid VGG 19 and capsule network based deep learning model for lung cancer diagnosis using CT scan images. Indian J. Sci. Technol. (2024). https://doi.org/10.17485/IJST/v17i35.1234
25. SriwastawaArul Jothi, A.J.A.: Vision transformer and its variants for image classification in digital breast cancer histopathology: A comparative study. Multim. Tools Appl. (2024). https://doi.org/10.1007/s11042-023-16954-x
26. Nazir, A, Wani, M.A.: You only look once-object detection models: a review In 2023 10th International Conference on Computing for Sustainable Global Development (2023). https://doi.org/10.1109/I2CT51068.2021.9417895
27. Khalil, I., Mehmood, A., Kim, H., Kim, J.: OCTNet: A modified multi-scale attention feature fusion network with InceptionV3 for retinal OCT image classification. Mathematics (2024). https://doi.org/10.3390/math12193003
28. Tesfai, H., Saleh, H., Al-Qutayri, M., Mohammad, M.B., Tekeste, T., Khandoker, A., Mohammad, B.: Lightweight Shufflenet based CNN for arrhythmia classification. IEEE Access (2022). https://doi.org/10.1109/ACCESS.2022.3215665
29. Briouya, H., Briouya, A., Choukri, A.: Surveying Lightweight Neural Network Architectures for Enhanced Mobile Performance. In International Conference on Smart Applications and Data Analysis (2024). https://doi.org/10.1007/978-3-031-77043-2_15
30. Abd El-Aziz, A.A., Mahmood, M.A., Abd El-Ghany, S.: A robust EfficientNetV2-S classifier for predicting Acute Lymphoblastic Leukemia based on cross validation. Symmetry (2024). https://doi.org/10.3390/sym17010024
31. Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., Ranjan, R.: Explainable AI (XAI): Core ideas techniques and solutions. ACM Comput. Surv. (2023). https://doi.org/10.1145/3561048
32. Le-Duc, T., Nguyen-Xuan, H., Lee, J.: Sequential motion optimization with short-term adaptive moment estimation for deep learning problems. Eng. Appl. Artif. Intell. (2024). https://doi.org/10.1016/j.engappai.2023.107593
33. Singh, N., Rajput, G.K., Mittal, P., Madan, P.: Analytical comparison of deep learning algorithms for fracture detection. In 11th International Conference on System Modeling & Advancement in Research Trends (SMART) IEEE (2022). https://doi.org/10.1109/SMART55829.2022.10047112
34. Quach, L.D., Nguyen, K.Q., Nguyen, A.Q., Thai-Nghe, N., Nguyen, T.G.: Explainable deep learning models with gradient-weighted class activation mapping for smart agriculture. IEEE Access (2023). https://doi.org/10.1109/ACCESS.2023.3296792
35. Kumar, S., Abdelhamid, A.A., Tarek, Z.: Visualizing the unseen: exploring grad-cam for interpreting convolutional image classifiers. J. Artif. Intell. Metaheur. (2023). https://doi.org/10.54216/JAIM.040104
36. Bommer, P.L., Kretschmer, M., Hedström, A., Bareeva, D., Höhne, M.M.C.: Finding the right XAI method—a guide for the evaluation and ranking of explainable AI methods in climate science. Artif. Intell. Earth Syst. (2024). https://doi.org/10.48550/arXiv.2303.00652
37. Ozsari, S., Kumru, E., Ekinci, F., Akata, I., Guzel, M.S., Acici, K., Ozcan, E., Asuroglu, T.: Deep learning-based classification of macrofungi: Comparative analysis of advanced models for accurate fungi identification. Sensors (2024). https://doi.org/10.3390/s24227189
38. Chhabra, M., Kumar, R.: A smart healthcare system based on classifier DenseNet 121 model to detect multiple diseases. In Mobile Radio Communications and 5G Networks: Proceedings of Second MRCN (2022). https://doi.org/10.1007/978-981-16-7018-3_23
39. Anand, V., Gupta, S., Koundal, D., Nayak, S. R., Shafi, J., Bhoi, A.K.: Segmentation and classification of skin cancer using K-means clustering and EfficientNetB0 model. In Advances in Communication, Devices and Networking: Proceedings of ICCDN (2022). https://doi.org/10.1007/978-981-19-2004-2_42
40. Devi, R.S., Kumar, V.R., Sivakumar, P.: EfficientNetV2 model for plant disease classification and pest recognition. Comput. Syst. Sci. Eng. (2023). https://doi.org/10.32604/csse.2023.032231
41. Ji, W., Pan, Y., Xu, B., Wang, J.: A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. Agriculture (2022). https://doi.org/10.3390/agriculture12060856
42. Kausalya, K., Chandra, B., Kumar, S.B., Riyaz, A.M.: Enhancing chronic kidney disease diagnosis through Densenet-121 approach. In 2024 4th International Conference on Pervasive Computing and Social Networking (2024). https://doi.org/10.1109/ICPCSN62568.2024.00017
43. Talukder, M.A., Layek, M.A., Kazi, M., Uddin, M.A., Aryal, S.: Empowering covid-19 detection: optimizing performance through fine-tuned efficientnet deep learning architecture. Comput. Biol. Med. (2024). https://doi.org/10.1016/j.compbiomed.2023.107789
44. Pacal, I.: Improved Vision Transformer with Lion Optimizer for Lung Diseases Detection. International Journal of Engineering Research and Development (2024). https://doi.org/10.29137/umagd.1469472
45. Elharrouss, O., Akbari, Y., Almadeed, N., Al-Maadeed, S.: Backbones-review: Feature extractor networks for deep learning and deep reinforcement learning approaches in computer vision. Computer Sci. Rev. (2024). https://doi.org/10.1016/j.cosrev.2024.100645
46. Ali, S., Agrawal, J.: Automated segmentation of brain tumour images using deep learning-based model VGG19 and ResNet 101. Multim. Tools Appl. (2024). https://doi.org/10.1007/s11042-023-16828-2
47. Menghani, G.: Efficient deep learning: A survey on making deep learning models smaller, faster, and better. ACM Comput. Surv. (2023). https://doi.org/10.1145/3578938
48. Asif, S., Wenhui, Y., ur-Rehman, S., ul-ain, Q., Amjad, K., Yueyang, Y., S., J., Awais, M.: Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision. Arch. Comput. Methods Eng. (2024). https://doi.org/10.1007/s11831-024-10148-w
49. Chanda, P., & Ghosh, S. (2024, January). Optimizing workforce efficiency using an artificial ıntelligence approach: a next-gen HR management system. In 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS) (pp. 1416–1421). IEEE. https://doi.org/10.1109/ICETSIS61505.2024.10459590
50. Chanda, P., Singh, P., Chhibber, P., & Ghosh, S. (2023, November). PBFT Consensus Mechanism for Human Resource Offboarding Procedures Based on Blockchain Technology. In International Conference on Trends in Computational and Cognitive Engineering (pp. 411–429). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-1923-5_32
51. Ghosh, S., & Singh, A. (2021). Image classification using deep neural networks: emotion detection using facial ımages. In Machine

Learning and Data Analytics for Predicting, Managing, and Monitoring Disease (pp. 75–85). IGI Global. https://doi.org/10.4018/978-1-7998-7188-0.ch006
52. Ghosh, S., Arya, S., & Singh, A. (2021). Plant disease detection using machine learning approaches: A survey. In Machine learning and data analytics for predicting, managing, and monitoring disease (pp. 122–130). IGI Global. https://doi.org/10.4018/978-1-7998-7188-0.ch009
53. Ghosh, S., Singh, A., Kumar, S.: Multiplier leadership optimization algorithm (MLOA): unconstrained global optimization approach for melanoma classification. Discov. Internet Things **5**(1), 70 (2025). https://doi.org/10.1007/s43926-025-00168-8
54. Ghosh, S., Singh, A., & Chanda, J. (2024). Enhancing land flora quality through a novel PSO-based optimizations approach of YOLO architecture for weed classification. In International Conference on Mobile Radio Communications & 5G Networks (pp. 423–438). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-96-4226-7_32
55. Kumru, E., Ekinci, F., Açici, K., Altindal, Ö.B., Güzel, M.S., Akata, I.: Advanced deep learning approaches for the accurate classification of Phallaceae fungi with explainable AI. Turk. J. Bot. **49**(5), 388–405 (2025). https://doi.org/10.55730/1300-008X.2871
56. Ekinci, F., Ugurlu, G., Ozcan, G.S., Acici, K., Asuroglu, T., Kumru, E., Guzel, M.S., Akata, I.: Classification of mycena and marasmius species using deep learning models: an ecological and taxonomic approach. Sensors **25**(6), 1642 (2025). https://doi.org/10.3390/s25061642
57. Korkmaz, A.F., Ekinci, F., Altaş, Ş, Kumru, E., Güzel, M.S., Akata, I.: A deep learning and explainable ai-based approach for the classification of discomycetes species. Biology **14**(6), 719 (2025). https://doi.org/10.3390/biology14060719
58. Kumru, E., Ugurlu, G., Sevindik, M., Ekinci, F., Güzel, M.S., Acici, K., Akata, I.: Hybrid deep learning framework for high-accuracy classification of morphologically similar puffball species using cnn and transformer architectures. Biology **14**(7), 816 (2025). https://doi.org/10.3390/biology14070816
59. Kumru, E., Korkmaz, A.F., Ekinci, F., Aydoğan, A., Güzel, M.S., Akata, I.: Deep ensemble learning and explainable AI for multiclass classification of earthstar fungal species. Biology **14**(10), 1313 (2025). https://doi.org/10.3390/biology14101313

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Fatih Ekinci received his PhD in Nuclear Physics and is currently affiliated with Ankara University. He conducts multidisciplinary research integrating artificial intelligence with medical and radiation sciences. His work focuses on radiation-based simulation systems, AI-driven medical imaging, biological mycology, clinical decision support systems, and explainable artificial intelligence (XAI) applications.



Mehmet Serdar Güzel is a full-time professor. His research areas include artificial intelligence, computer vision, medical imaging, and robotics. Throughout his academic career, Dr. GÜZEL has contributed to numerous significant publications in these fields. He has also played a crucial role in various national and international projects. Drawing on the knowledge and experience he has gained during his academic career, Dr. GÜZEL has actively participated as a speaker at numerous national and international conferences.



Koray Acici is currently an associate professor at Ankara University. He received a PhD in science from the Başkent University in 2021. His current main research includes artificial intelligence, machine learning, and deep learning.



Abdullah Aydoğan is a researcher working in the fields of deep learning and medical imaging. He has developed deep learning-based models for the taxonomic classification of macrofungal species. He is currently working on the automatic segmentation of brain tumors from magnetic resonance imaging (MRI). His research integrates image processing, data preprocessing, and model optimization techniques to contribute to both biological and biomedical data analysis.



Şehmus Altaş is a computer engineer specializing in artificial intelligence, with experience in computer vision, natural language processing, and machine learning and has contributed to the development and deployment of AI-driven solutions, including model training, data processing, and system integration in real-world applications, and involved in both research-oriented and applied projects, with a focus on building scalable and efficient intelligent systems across domains such as healthcare and interactive technologies.



Ömer Burak Altındal is a computer engineer developing projects in the field of medical image processing. He works on medical data analysis, segmentation, and volumetric image processing using deep learning-based models. He focuses particularly on model optimization, performance improvement, and accuracy enhancement on high-dimensional datasets.



Eda Kumru completed her undergraduate education at Ankara University and is currently pursuing her master's degree at the same institution. Her research interests include mycology, macrofungal diversity, morphological and molecular identification, and artificial intelligence-assisted fungal classification.



Tunc Asuroglu completed his PhD studies in computer engineering at Baskent University, Ankara, Turkey. He was an Assistant Professor with the Department of Computer Engineering, Baskent University. He was also a Guest Researcher with the Faculty of Computer Sciences, Østfold University College, Norway, and a Postdoctoral Research Fellow with the Tampere University, Finland. He is currently a Senior Scientist with VTT Technical Research Centre of Finland and holds the title of Docent with Tampere University. His research interests include the applications of computational intelligence in health informatics and wearable sensor systems. He serves on the editorial boards of various journals.



Ilgaz Akata is a mycologist at the Department of Biology, Faculty of Science, Ankara University. His research focuses on macrofungal diversity, taxonomy, systematics, and applied mycology, and he has made important contributions to fungal research in Türkiye.