

# COMPARATIVE ANALYSIS OF DEEP LEARNING TECHNIQUES AND ATTENTION MECHANISMS FOR MUSHROOM DISEASE DETECTION

Monika Singla<sup>1\*</sup>, Deepali Gupta<sup>2</sup>, Deepam Goyal<sup>3</sup>

<sup>1</sup>Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India  
E-mail - monika1015cse.phd23@chitkara.edu.in

<sup>2</sup>Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India  
E-mail - deepali.gupta@chitkara.edu.in

<sup>3</sup>Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India  
E-mail - bkdeepamgoyal@outlook.com

## ABSTRACT

Deep learning (DL) has transformed image-based diagnosis by identifying learning discriminative features from complex agricultural images automatically. However, it remains challenging to select an appropriate deep learning architecture for mushroom disease detection due to the subtle differences in the visual characteristics of mushroom disease categories and the limited availability of annotated mushroom datasets. In this paper, the authors consider the most popular CNN-based architectures including CNN, AlexNet, VGG16, VGG19, ResNet50, DenseNet121, InceptionV3, and Vision Transformer (ViT), alongside recent attention mechanisms such as Channel Attention (CA), Spatial Attention (SA), SE-Net, Convolutional Block Attention Module (CBAM), Self-Attention, and Transformer Attention and provide a detailed comparative analysis of these architectures. All methods are systematically evaluated in terms of their architectural characteristics, feature learning capability, computational complexity, pros and cons, as well as their suitability for mushroom disease classification. The comparative evaluation demonstrates that DenseNet121 is the most appropriate backbone architecture because of its dense feature connectivity, efficient gradient propagation and effective parameter utilization. The analysis presented will furnish a foundation for developing robust, efficient, and interpretable DL methods for automated mushroom disease detection.

**KEYWORDS:** Deep Learning, DenseNet121, Attention Mechanism, CBAM, Grad-CAM++, Mushroom Disease Detection, Explainable AI.

## 1. INTRODUCTION

With its nutritional and medicinal benefits and commercial potential, mushroom cultivation has become one of the most rapidly expanding sectors of modern agriculture. Edible mushrooms are rich in proteins, vitamins, minerals, dietary fiber and bioactive compounds, making them an important component of human nutrition. However, mushroom production is susceptible to many bacterial, fungal and viral diseases, such as green mold, cobweb disease, dry bubble disease, wet bubble disease, and false truffle disease. Early detection of mushroom diseases is crucial to minimize mushroom crop losses, improve disease management, and increase production efficiency [1]-[4].

DL methods have shown impressive results in automatically learning discriminative features in images without the need for manual feature engineering. Consequently, DL has become one of the most promising technologies for the intelligent detection of agricultural diseases [1], [5]-[7]. CNNs are a type of DL architecture that are particularly well-suited for image classification, as they can learn hierarchical representations from images. The feature extraction ability, classification accuracy and model generalization capability of CNNs have been further enhanced by several advanced CNN architectures, such as AlexNet, VGG16, VGG19, ResNet50, and DenseNet121, which have been widely adopted in agricultural image analysis. The dense connectivity of DenseNet121 facilitates feature reuse, improves gradient propagation, and reduces parameter redundancy, making it a highly appealing network for learning with relatively small datasets [8]-[11]. Although, deep convolutional networks achieve high classification accuracy, they often fail to effectively focus on the most informative regions of an image and may still capture irrelevant features, such as background details. To selectively attend to informative features that correspond to disease symptoms, attention mechanisms have therefore been introduced into neural networks. The different attention strategies considered in our work include Channel Attention (CA), which boosts the representation of relevant feature channels associated with a disease, and Spatial Attention (SA), which focuses on infected regions within the image. The aim of this paper is to summarize recent achievements in deep learning architectures and attention mechanisms for mushroom disease classification. The following summaries the main contribution of this study:

- a) A comprehensive comparative analysis of well-known DL architectures for mushroom disease classification.
- b) An in-depth comparative analysis of attention mechanisms, such as CA, SA, SE-Net, CBAM, Self-Attention, Transformer Attention.

- c) Identification of DenseNet121 as the most suitable backbone feature extraction network because of its dense feature connectivity, superior feature reuse and efficient gradient propagation.
- d) Recommendation of the integration of CA with DenseNet121 for improved disease-specific feature learning.

## 2. Related Work

Recently, DL has been shown to be effective in mushroom mushroom and disease classification based on image features [1, 5, 7]. Early CNN architectures, like AlexNet, laid the foundation for modern image classification using deep CNNs [12]. Subsequently, ResNet50 [16] and InceptionV3 [17] further advanced DL by introducing residual connections and multi-scale features, respectively. In the field of agriculture, DenseNet has improved the recognition performance by facilitating feature reuse and efficient gradient propagation, making it well suited for agricultural image analysis [10]. Other transfer learning models have also been successfully adopted for mushroom species identification and disease classification [6] [8-9], [13]. Furthermore, CNNs have been extended with attention mechanisms that highlight relevant features of the image to the disease while ignoring redundant information [14]. More recently, self-attention mechanisms have been incorporated into ViT-based models to capture global contextual relationships [15]. Although significant progress has been made, only a few studies have investigated mushroom disease classification using attention-enhanced DenseNet architectures, indicating the need for further research in this area.

## 3. Deep Learning Architectures for Mushroom Disease Classification

This section discusses the major DL architectures employed for mushroom disease classification, including CNN, AlexNet, VGG16, VGG19, and DesNet201 etc. Their architectures, key characteristics, merits, limitations, and suitability for mushroom disease detection are comparatively reviewed.

### 3.1 Convolutional Neural Networks (CNNs)

For image processing and computer vision applications, one of the most popular DL architectures is the CNN. CNNs are used for processing grid-like data, such as images, through convolution operations. The architecture is able to learn complex visual patterns from images, automatically extracting hierarchical the spatial features without manual feature engineering. A typical CNN architecture is composed of an input layer, convolutional layers, activation layers, pooling layers, fully connected layers, and an output layer as shown in Fig. 1. These convolutional layers learn filters that can identify key features, such as edges, textures, colors, and disease symptoms. Pooling layers reduce the spatial dimensions of feature maps while retaining meaningful information. The extracted features are then fed into FC layers for final classification. A major advantage of CNNs is their ability to perform automatic feature extraction and classification. By learning complex disease features directly from image data, CNNs have been successfully used for mushroom disease detection, mushroom species identification, and the diagnosis of agricultural diseases.

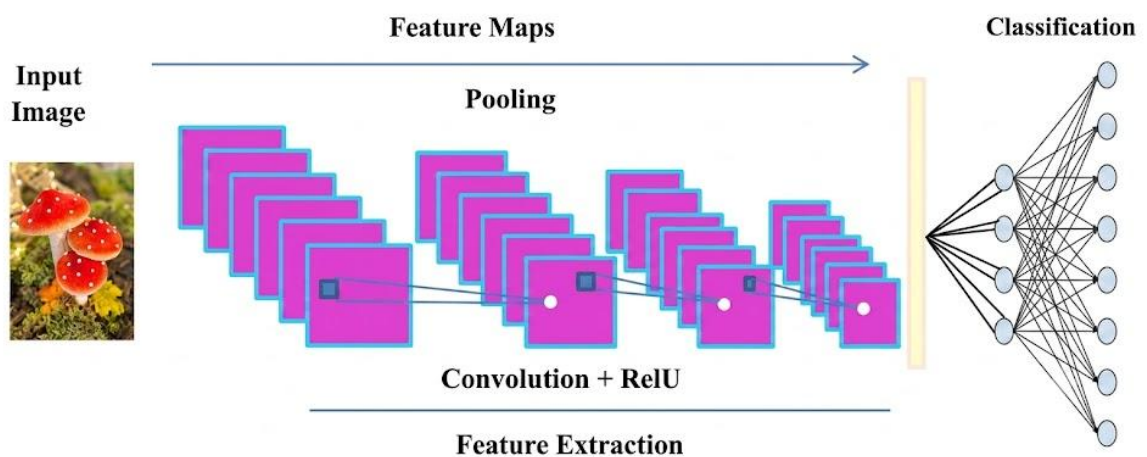


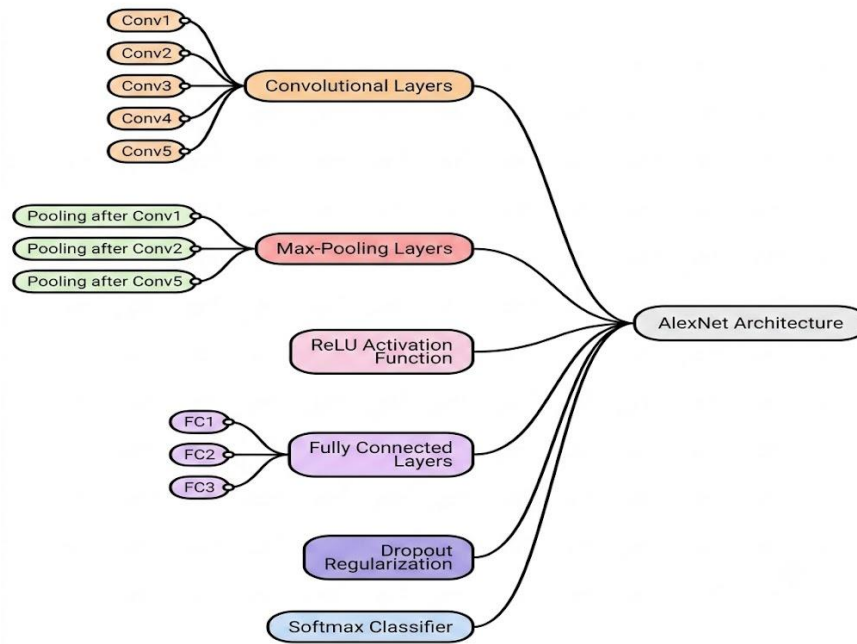
Figure 1: CNN Architecture

### 3.2 AlexNet

Krizhevsky et al. [12] proposed AlexNet, a deep convolutional network. It became well-known for its outstanding performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). AlexNet successfully demonstrated the effectiveness of DL for image classification and laid the groundwork for contemporary CNN architectures. The architecture is composed of 5 convolutional layers and 3 fully connected (FC) layers as shown in Fig. 2. ReLU activation functions are used to speed up the training process, while the dropout mechanism helps prevent overfitting. To reduce the size of feature maps and lower computational complexity, max-pooling layers are incorporated into the network. AlexNet significantly improved image classification accuracy compared with traditional ML models. In agricultural disease detection, AlexNet has been successfully applied to recognizing diseases from visual symptoms and extracting discriminative features from images of both plants and mushrooms.

### 3.3 VGG16

It is a deep CNN architecture, characterized by the consistent use of small 3×3 convolutional filters arranged in a deep



**Figure 2: AlexNet Architecture**

layered stack. The architecture is mainly composed of several convolutional layers, ReLU activation functions, and max-pooling layers followed by three FC layers. The VGG16 model consists of 13 convolutional layers organised into 5 blocks, each followed by a max-pooling layer, and ends with 3 FC layers. The network uses the ReLU activation function to add non-linearity and speed up network convergence. Feature extraction is performed using convolutional kernels, while pooling layers reduce the spatial dimension of the feature maps, thereby decreasing computational complexity and preserving the most informative features. The use of small convolutional filters throughout the network help in enhancing performance and generalization capability.

### 3.4 VGG19

It extends VGG16 by increasing the number of convolutional layers. VGG16 comprises 13 convolutional layers and 3 FC layers, resulting in 16 weight layers, whereas VGG19 has 16 convolutional layers and 3 FC layers, resulting in 19 weight layers. VGG19 can be considered a deeper version of VGG16. The additional convolutional layers enable the network to learn more complex and discriminative feature representations. The increased network depth improves feature extraction capability; however, it also leads to higher computational cost, greater memory requirements, and longer training time than VGG16. Table 1 shows the architecture of VGG19.

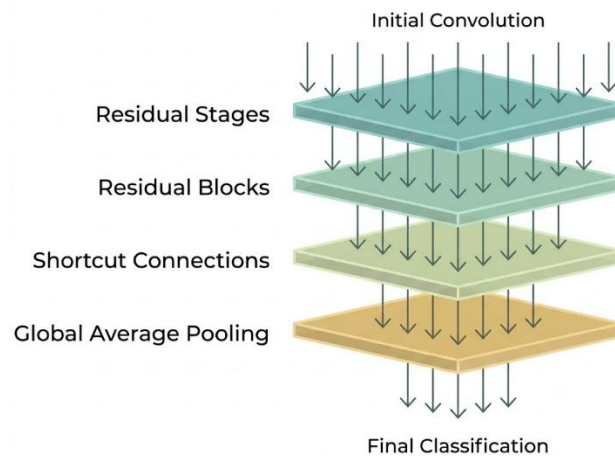
**Table 1: Architecture of VGG19**

| Layer name                         | Shape                                                                            | Output                |
|------------------------------------|----------------------------------------------------------------------------------|-----------------------|
| Conv1_1, Conv1_2                   | $[3 \times 3]:64 \rightarrow [3 \times 3]:64$ , MaxPool $[2 \times 2]$ , $s=2$   | $[112 \times 112]:64$ |
| Conv 2_1, Conv 2_2                 | $[3 \times 3]:128 \rightarrow [3 \times 3]:128$ , MaxPool $[2 \times 2]$ , $s=2$ | $[56 \times 56]:128$  |
| Conv3_1, Conv3_2, Conv3_3, Conv3_4 | $[3 \times 3]:256 \rightarrow [3 \times 3]:256$ , MaxPool $[2 \times 2]$ , $s=2$ | $[28 \times 28]:256$  |
| Conv4_1, Conv4_2, Conv4_3, Conv4_4 | $[3 \times 3]:512 \rightarrow [3 \times 3]:512$ , MaxPool $[2 \times 2]$ , $s=2$ | $[14 \times 14]:512$  |
| Conv5_1, Conv5_2, Conv5_3, Conv5_4 | $[3 \times 3]:512 \rightarrow [3 \times 3]:512$ , MaxPool $[2 \times 2]$ , $s=2$ | $[7 \times 7]:512$    |
| Others                             | FC-4096 $\rightarrow$ FC-4096 $\rightarrow$ Softmax-1000                         | 1000                  |

### 3.5 Residual Network (ResNet50)

It is a deep CNN developed by He et al. [17] to overcome performance degradation and vanishing gradient problems encountered during the training of very deep neural networks. In traditional CNNs, the performance can degrade as the depth of the network increases due to weaker gradients during backpropagation. This limitation can be overcome by ResNet50's shortcut connections, also known as or skip connections, which allow feature information to propagate directly across multiple layers. Such connections facilitate the network to learn residual mappings rather than direct mappings, which enhance the convergence and classification accuracy. Owing to its excellent feature extraction capability and efficient training process, ResNet50 has been widely adopted for image classification, object detection, medical imaging, and agricultural disease recognition applications. The architecture of ResNet50 consists of 50 layers and starts with a  $7 \times 7$  convolutional layers with 64 filters, followed by batch normalization, ReLU activation, and max-pooling layer. The network is then organized into four residual blocks: Conv2\_x (3), Conv3\_x (4), Conv4\_x (6), Conv5\_x (3). The residual blocks are implemented using three consecutive convolutional layers with kernel sizes of  $1 \times 1$ ,  $3 \times 3$ , and  $1 \times 1$ , forming a bottleneck design. The first  $1 \times 1$  convolution reduces the feature dimensionality, the  $3 \times 3$  convolution extracts discriminative spatial features, and the final  $1 \times 1$  convolution restores the feature dimensions. Shortcut connections are added between blocks to learn residual mappings, thereby enhancing gradient flow and facilitating the training of deeper

networks. The output of the last residual block is then subjected to global average pooling to further reduce the feature maps, and the resulting feature vector is fed into a FC layer with a Softmax classifier to obtain the final classification result. The basic architecture of ReNet 50 is shown in Fig. 3.



**Figure 3: Basic Architecture of ResNet50**

### 3.6 DenseNet201

It is considered a strong DL architecture for image classification tasks. It is characterized by its dense connectivity, where each layer is directly connected to every preceding layer within the same dense block. This architecture facilitates efficient information flow and maximum feature reuse across layers, thereby mitigating the vanishing gradient problem. The architecture comprises 201 layers organized into several dense blocks separated by transition layers. Each dense block contains multiple convolutional layers, in which the feature maps from all layers are fed into the next layer. ReLU activation is used to introduce non-linearity, and batch normalization is applied to increase the stability and convergence rate of the model. DenseNet201 is designed for classification problems, with its final layers comprising global mean pooling layer followed by a FC with Softmax activation. Table 2 provides the comparative analysis of DL architectures for mushroom disease detection.

**Table 2. Comparative Analysis of Deep Learning Architectures for Mushroom Disease Detection**

| Model                    | Year | Architecture Characteristics                                                                | Advantages                                                                                  | Limitations                                               | Ref.      |
|--------------------------|------|---------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|-----------------------------------------------------------|-----------|
| CNN                      | 1998 | consists of fully linked layers, pooling, and convolution for automatic feature extraction. | Automatic feature extraction, good classification performance, reduced manual intervention. | Requires large datasets, susceptible to overfitting.      | [6], [13] |
| AlexNet                  | 2012 | 5 convolutional layers and 3 FC layers with ReLU and dropout.                               | Faster training, improved accuracy, effective feature extraction.                           | Large number of parameters, high memory consumption.      | [12]      |
| ResNet50                 | 2016 | Uses residual blocks and shortcut connections to enable deeper networks.                    | Solves vanishing gradient problem, high classification accuracy.                            | Computationally intensive and memory demanding.           | [17]      |
| InceptionV3              | 2016 | Employs parallel convolution filters of multiple sizes for multi-scale feature extraction.  | Captures features at different scales, efficient computation.                               | Complex architecture and training process.                | [16]      |
| DenseNet121              | 2017 | Dense connections between all layers encourage feature reuse and efficient gradient flow.   | Reduced parameter redundancy, improved feature propagation, high accuracy.                  | Higher memory usage due to dense connectivity.            | [9], [14] |
| Vision Transformer (ViT) | 2021 | Utilizes self-attention mechanisms to learn global image relationships.                     | Captures long-range dependencies and global context.                                        | Requires large datasets and high computational resources. | [15]      |

From the above comparison, it is observed that DenseNet201 offers superior feature reusability and more effective gradient propagation due to the dense connectivity, making it more suitable for mushroom disease classification in complex situations. It can learn more discriminative features than traditional CNN architectures since it can preserve and reuse features across layers.

#### 4. Attention Mechanism

Attention mechanisms have proven to be a successful solution for improving the capacity of deep neural networks to learn discriminative features from images by selectively emphasizing the most informative parts of an image while ignoring irrelevant background information. Inspired by the human visual system, attention mechanisms learn to assign higher weights to relevant features and lower weights to less relevant features, and thus achieve better classification results. Symptoms of mushroom diseases often appear as localized spots, lesions, discolorations and texture changes on mushroom caps, stems or gills. Unlike conventional CNNs, which lack an explicit mechanism to prioritize disease-relevant regions, the extracted features maps may include both relevant and irrelevant information, which can lead to reduced performance in correctly identifying infected areas. To overcome this limitation, attention mechanisms focus on disease-relevant features, leading to better feature representation and classification performance.

##### 4.1 Channel Attention Mechanism

Channel Attention aims to extract the relative importance of the feature channels generated by a CNN and to boost the most informative channels for classification. The relative contribution of each channel to disease recognition varies according to the type of information it contains. The CA module dynamically allocates weights to each feature channel, which allows the network to focus on important and relevant features while emphasizing irrelevant or redundant features. For an input feature map,

$$F \in R^{C \times H \times W} \quad (1)$$

Global Average Pooling (GAP) is used to extract channel-wise descriptors where C, H, and W represent the number of channels, height, and width, respectively. The descriptors are fed into a multiplayer perceptron (MLP), followed by a sigmoid activation function to produce channel attention weights.

##### 4.2 Spatial Attention Mechanism

Spatial Attention is an extension of CA, which seeks to identify the optimal informative locations in a spatial feature map. CA assigns importance to the important feature channels, whereas SA identifies the spatial locations of the important features. During mushroom disease detection, the disease-affected area within a mushroom image is generally small, which makes it important to determine the location of the area affected by the disease in order to improve the classification performance. GAP and Global Max Pooling (GMP) are performed along the channel dimension to obtain the spatial attention map. The outputs of these feature maps are then concatenated and fed into a  $7 \times 7$  convolutional layer, followed by a sigmoid activation function. The Spatial Attention Map (SAM) is computed as:

$$M_{s(F)} = \sigma(f_{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (2)$$

where AvgPool( $\cdot$ ) and MaxPool( $\cdot$ ) denote the average and maximum pooling operations along the channel dimension,  $f_{7 \times 7}$  represents a  $7 \times 7$  kernel convolution operation, where  $\sigma$  is the sigmoid activation function.

The attention map generated by the model is used to emphasize disease-relevant regions and suppress irrelevant information in the image to highlight specific characteristics of mushroom diseases, including lesions, discoloration, and abnormal texture patterns, while assisting the network in accurately classifying mushroom diseases.

##### 4.3 Squeeze-and-Excitation Network (SE-Net)

It is a network that uses a squeeze-and-excitation block to adaptively recalibrate channel-wise feature responses and boost feature representation. It captures the contextual information from the input feature map and aggregates it globally using the squeeze operation, and then learns the importance of each channel using the excitation mechanism. The whole channel attention process can be summarized as:

$$Y = X \odot \sigma(SE(X)) \quad (3)$$

where X denotes the input feature map, SE(X) represents the squeeze-and-excitation operation,  $\sigma$  is the sigmoid activation function, and Y is the recalibrated output feature map. SE-Net assigns higher weights to informative channels and suppresses less informative ones, thereby improving feature representation capability. However, it only models channel attention and does not explicitly model the spatial relationships within the feature map.

##### 4.4 Convolutional Block Attention Module (CBAM)

The CBAM improves feature representation by sequentially applying CA and then SA to the input feature map. The CA module first selects the feature channels with the highest information content, and the SA module then identifies the most relevant regions in the feature map. The entire operation of CBAM can be stated as:

$$Y = M_s(M_c(X)) \times X \quad (4)$$

where X represents the input feature map,  $M_c(X)$  and  $M_s$  represent the CA module and SA module, respectively, and Y is the refined output feature map. CBAM has been considered because it effectively combines channel and spatial attention to enhance feature representation while maintaining relatively low computational complexity.

#### CONCLUSION

This study compared the most recent deep learning architectures and attention mechanisms proposed for mushroom disease classification. The analysis demonstrated the strengths of each architecture in terms of feature extraction, computational efficiency, and learning capability. DenseNet121 emerged as the the best performing model in terms of feature reuse among the models evaluated. It has low parameter redundancy and enables efficient gradient propagation, making it particularly suitable for identifying disease symptoms. Similarly, the Convolutional Block Attention Module (CBAM) was identified as effective attention mechanism because it jointly utilizes channel and spatial information while

maintaining relatively low computational complexity. Considering the findings of this work, it provides a solid theoretical basis for selecting appropriate DL architecture and attention mechanisms for mushroom disease classification. They also offer valuable guidance for future research aimed at developing precise, efficient, and interpretable mushroom diagnosis systems.

To further improve the identification of disease-related characteristics from mushroom images, a possible framework for the future could integrate DenseNet201 with CBAM. The CBAM module could emphasize informative feature channels and disease-relevant spatial regions while suppressing redundant information. The framework could enable the network to learn more discriminative disease characteristics and enhance feature representation. It could highlight the important characteristics and remove unnecessary data, potentially improving the accuracy of mushroom disease classification.

Additionally, the use of DenseNet201 in conjunction with CBAM might improve the robustness of classification process across different imaging conditions. In addition, the enhanced feature representation might aid in better distinguishing visually similar disease classes. This approach could improve the reliability and efficiency of automated mushroom disease diagnosis. Various mushroom datasets could be used in future studies to validate this framework and compare its performance with that of other state-of-the-art DL models. Overall, the integration of DenseNet201 with CBAM represents a promising direction for future research in automated mushroom disease classification

## REFERENCES

1. Yin H, Yi W, Hu D. Computer vision and machine learning applied in the mushroom industry: A critical review. *Computers and Electronics in Agriculture*. 2022;198:107015.
2. Sharma SR, Kumar S. Diseases and Competitor Moulds of Mushrooms and Their Management. National Research Centre for Mushroom (ICAR); Technical Bulletin; 2007.
3. Ghimire A, Pandey KR, Joshi YR, Subedi S. Major fungal contaminants of mushrooms and their management. *International Journal of Applied Sciences and Biotechnology*. 2021;9(2):80–93.
4. Carrasco J, Navarro MJ, Gea FJ. Cobweb, a serious pathology in mushroom crops: A review. *Spanish Journal of Agricultural Research*. 2017;15(2):e10R01.
5. Zahan N, Hasan MZ, Uddin MS, Hossain S, Islam SF. A deep learning-based approach for mushroom diseases classification. In: *Application of Machine Learning in Agriculture*. Elsevier; 2022. p. 191–212.
6. Wu L, Chen Y. Mushroom recognition and classification based on convolutional neural network. In: *Proceedings of the IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*; 2022. p. 1430–1433.
7. Yan Z, Liu H, Li J, Wang Y. Application of identification and evaluation techniques for edible mushrooms: A review. *Critical Reviews in Analytical Chemistry*. 2023;53(3):634–654.
8. Ketwongsa W, Boonlue S, Kokaew U. A New Deep Learning Model for the Classification of Poisonous and Edible Mushrooms Based on Improved AlexNet Convolutional Neural Network. *Applied Sciences*. 2022;12(7):3409.
9. Albayrak U, Golcuk A, Aktas S, Coruh U, Tasdemir S, Baykan OK. Classification and analysis of *Agaricus bisporus* diseases with pre-trained deep learning models. *Agronomy*. 2025;15:226.
10. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, and future directions. *Journal of Big Data*. 2021;8:53.
11. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*. 2018;147:70–90.
12. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*. 2012;25:1097–1105.
13. Sutayco MJY, Caya MVC. Identification of medicinal mushrooms using computer vision and convolutional neural network. In: *Proceedings of the 6th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*; 2022. p. 167–171.
14. Hassan E, Abu Ghazalah S, El-Rashidy N, Abd El-Hafeez T, Shams MY. DenseNet model with attention mechanisms for robust date fruit image classification. *International Journal of Computational Intelligence Systems*. 2025;18:228.
15. Wang B. Automatic mushroom species classification model for foodborne disease prevention based on vision transformer. *Journal of Food Quality*. 2022;2022:Article ID 1173102.
16. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016:2818–2826.
17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. p. 770–778.