WILEY | Hindawi

Research Article

Automatic Mushroom Species Classification Model for Foodborne Disease Prevention Based on Vision Transformer

Boyuan Wang [iD] [1,2]

[1]*Faculty of Information Technology, Macau University of Science and Technology, Taipa, Macau, China*
[2]*Zhongshan Center for Disease Control and Prevention, Zhongshan, Guangdong, China*

Correspondence should be addressed to Boyuan Wang; 2109853pia30001@student.must.edu.mo

Mushrooms are the fleshy, spore-bearing structure of certain fungi, produced by a group of mycelia and buried in a substratum. Mushrooms are classified as edible, medicinal, and poisonous. However, many poisoning incidents occur yearly by consuming wild mushrooms. Thousands of poisoning incidents are reported each year globally, and 80% of these are from unidentified species of mushrooms. Mushroom poisoning is one of the most serious food safety issues worldwide. Motivated by this problem, this study uses an open-source mushroom dataset and employs several data augmentation approaches to decrease the probability of model overfitting. We propose a novel deep learning pipeline (ViT-Mushroom) for mushroom classification using the Vision Transformer large network (ViT-L/32). We compared the performance of our method against that of a convolutional neural network (CNN). We visualized the high-dimensional outputs of the ViT-L/32 model to achieve the interpretability of ViT-L/32 using the t-distributed stochastic neighbor embedding (t-SNE) method. The results show that ViT-L/32 is the best on the testing dataset, with an accuracy score of 95.97%. These results surpass previous approaches in reducing intraclass variability and generating well-separated feature embeddings. The proposed method is a promising deep learning model capable of automatically classifying mushroom species, helping wild mushroom consumers avoid eating toxic mushrooms, safeguarding food safety, and preventing public health incidents of food poisoning. The results will offer valuable resources for food scientists, nutritionists, and the public health sector regarding the safety and quality of mushrooms.

1. Introduction

Fungi represent a highly diversified component in the ecological systems with a major connection with living organisms [1–4]. Despite recent breakthroughs in fungal taxonomic identification, only 5% of the 3,8 million fungal species have been discovered [5]. Of the fungal species, *Morchella*, *Tuber melanosporum*, and *Cantharellus cibarius* are in the macro-fungi group and produce mushrooms that have distinctive fruiting bodies from an underground mycelium [6]. These fungal species are not autotrophs because of the deficiency of chlorophyll. However, their enzymes can break down complex substrates to obtain the nutrients needed for growth [1, 2, 6]. Thus, mushrooms are classified based on their edible, medicinal, or poisonous properties [7]. Mushrooms are also divided into temperate, tropical, or subtropical fungi. Edible mushrooms are a diverse and important group of fungi. A previous study found that more than 3000 varieties of mushrooms were edible, accounting for 20% of all mushroom species taxa reported in worldwide sources [8]. The most common edible mushrooms are White Button, Shitake, Portobello, Oyster, Enoki, Cremini, Lion's Mane, Turkey Tail, Hen of the Woods, Beech, Chanterelle, Porcini, etc. People consume these edible mushrooms for their nutritional and medicinal value. However, many poisoning incidents occur yearly by consuming wild mushrooms [9, 10]. Poisonous mushroom species can cause health complications when ingested, such as liver failure, acute gastroenteritis, dizziness, respiratory distress, renal failure, erythromelalgia, and rhabdomyolysis [11, 12]. Thus, mushroom collectors often confuse edible and nonedible wild mushrooms due to their similarities. Sometimes,

victims do not exhibit the symptoms of poisoning immediately after ingestion as the symptoms often appear after 48 h [13]. The severity of the symptoms varies from case to case. In cases of fatal outcome, poisonous mushrooms can lead to death, and the median time to death was 6.1 days (2.7–13.9 days) after ingestion [14]. Some of the most poisonous mushroom species include *Amanita phalloides* (the death cap), *Amanita virosa* (the destroying angel), *Amanita muscaria* (the Fly Agaric), and *Cortinarius rubellus* [15]. Globally, thousands of poisoning incidents from mushrooms are reported every year, and 80% of them are from unidentified species of mushrooms [16]. In China, 480 distinct types of toxic mushrooms cause seven different clinical syndromes, including acute renal failure, rhabdomyolysis, acute liver failure, gastroenteritis, psychoneurological illness, hemolysis, and photosensitive dermatitis [7]. Particularly, the liver suffers from irreparable damage when consuming toxic species [17]. Mushroom poisoning is the major cause of oral poisoning deaths in China, with a significant risk to farmers owing to its typical temporal aggregation (from summer to autumn) and high mortality rate (approximately 20%) [10]. According to the China Center for Disease Control and Prevention, mushroom poisoning incidents are reported every month, especially from summer to autumn, with a peak in July. In 2020, a total of 676 independent mushroom poisoning incidents were reported in 1719 patients, and 25 deaths were investigated in 24 provincial-level administrative divisions [9].

Experts use the traditional method for classifying and identifying mushrooms based on their morphology. The mushroom structures vary from species to species. However, their overall structure comprises cap, flesh, gills, and stalk. Some may have rings and receptacles. Generally, cap characteristics, such as shape, size, color, and surface covering, are used to identify mushrooms. Color, texture, thickness, and emulsion are used to identify mushrooms' flesh characteristics. Differences between the gill characteristics of mushrooms are based on the bearing, color, density, length, and injury discoloration. Stalk characteristics, such as length, size, shape, texture, color, and coverings, play a dominant role in identifying mushrooms. Experts use ring characteristics, such as color, texture, shape, and growth position, to identify mushrooms. Shape, size, color, and cracking situation are the differences remarked in a receptacle of mycorrhizal [18, 19]. These morphological characteristics are critical tools for distinguishing different species of mushrooms [20]. However, due to a lack of knowledge, skills, and guidance from mushroom experts, many locals face the risk of consuming toxic mushrooms as these mushrooms are morphologically similar to edible mushrooms [3].

Some studies have attempted to learn the mushrooms' characteristics through artificial intelligence and have developed models to assist consumers in identifying different species of mushrooms and in preventing mushrooms poisoning. These studies can be divided into two main learning approaches. One approach is to manually extract mushroom features and classify the input features by using machine learning models such as support vector machines (SVMs) [21], logistic regression [22], and random forest [23]. Another approach involves extracting features automatically from mushroom images using deep learning models (e.g., CNN) [24].

Many studies have used machine learning to classify mushrooms. For example, Ottom et al. [21] collected the mushroom image from a public dataset to classify mushrooms using different machine learning algorithms, such as neural networks (NNs), SVMs, decision trees, and k-nearest neighbors (kNN). Of these methods, kNN achieves the best result for classifying mushroom images with 94% accuracy using features extracted from the images and dimensions of mushroom species. Wagner et al. [25] have established the largest and most comprehensive dataset available for predicting the edible group of mushrooms. They evaluate several different machine learning models, such as naive Bayes, logistic regression, linear discriminant analysis, and random forests (RF). Of these models, RF provides the best results with fivefold Cross-Validation accuracy and an F2-score of 1.0 ($\mu = 1$, $\sigma = 0$), respectively. Tongcham et al. [26] proposed a machine learning algorithm to classify the oyster mushroom spawn. They measured the performance of five machine learning classifiers, and 4-fold cross-validation demonstrated that the deep neural network classifier has a higher accuracy of 98.8% with a residual variance of 2.5%.

Despite the advances of machine learning for classifying and recognizing mushroom classification, machine learning algorithms have some limitations. For example, they require manual feature extraction as input data. Moreover, machine learning has low efficiency and accuracy when using large mushroom samples. Machine learning cannot accurately measure various metrics and cannot automate the full process of recognition. Deep learning (DL) was proposed [27] to solve the problem of automatic feature extraction and image classification, such as CNN [28], recurrent neural network (RNN) [29], and generative adversarial network (GAN) [30]. However, limited studies used deep learning to examine automatic mushroom recognition. Previous studies have focused on using CNN models for mushroom classification by establishing basic architectures or using transfer learning with pretrained architectures. Sajedi et al. [31] used a four-layer basic CNN to automatically identify *mucilaginous taxa*. The initial stage in this approach is to extract image features using a CNN, and these characteristics are input into machine algorithm classifiers such as SVM, XGBoost, and Extreme Learning Machine (MLP). The CNN-MLP model outperformed the others with 80.7% accuracy, 100% precision, and 100% recall, which was approximately 5% better than SVM and XGBoost. Devika et al. [32] suggested a mushroom classification using deep convolutional neural networks (DCNN) model of four convolutional layers and one fully connected layer. On the test set, the DCNN model was pitted against the network structures sNet, LeNet, AlxNet, and cNET. The DCNN shows an accuracy of 93% better than the mushroom classifier. Wang et al. [33] suggested a bilinear convolutional neural network (B-CNN) based on an attention mechanism for the *Amanita* classification. After training, the B-CNN model achieves the accuracy of 95.2% in the test set, which

helps solve the problem associated with the image classification of genus *Amanita* in the wild complex environment. Preechasuk et al. [24] established a basic CNN architecture to classify multiple types of mushrooms. The experimental dataset includes 8556 mushroom images classified into 45 types, of which 35 are edible mushrooms, and the other 10 are poisonous. The suggested method presents results of 78%, 73%, and 74% in terms of average precision, average recall, and average F1-score, respectively. Zahan et al. [4] applied deep learning models such as Inception-V3, VGG-16, and Resnet50 to identify mushroom species on a dataset of 8190 mushroom images. They used the contrast-limited adaptive histogram equalization with the Inceptionv3 network and obtained accuracy of 88.4% on the test set.

Currently, few studies identify mushrooms using deep learning models, and no interpretability studies classify mushrooms using deep learning models. To address these issues, we conduct this study with the following major contributions:

(1) This study proposes a novel deep learning pipeline (ViT-Mushroom) based on the ViT-L/32 network for mushroom classification, which is more suitable for the dataset after fine-tuning. A thorough search of the literature shows that this is the first study to classify mushrooms using a transformer-based model.

(2) Additionally, we visualize the high-dimensional outputs of the ViT-L/32 model to analyze clustering the feature space based on t-SNE and compare the learned features using the CNN models.

2. Datasets

The mushroom dataset used in our experiments was mainly obtained from Kaggle, and the original source of the images was mainly from https://www.mushroom.world, which includes *Agaricus*, *Amanita*, *Boletus*, *Cortinarius*, *Entoloma*, *Exidia*, *Hygrocybe*, *Lactarius*, *Pluteus*, *Russula*, and *Suillus* for a total of 11 different species of mushrooms.

We uploaded the processed data to the Kaggle platform as a public database and available at https://www.kaggle.com/mustai/mushroom-12-9528. The data and labels were examined by the Nordic Association of Mycologists. The dataset consists of 9528 mushroom images, of which 80% were used for training and validation, and the other 20% were used for model testing, as shown in Table 1.

TABLE 1: Training, validation, and test sets for the mushroom classification dataset.

Class	Photo	Dataset			
		Training	Validation	Testing	Total
Agaricus		681	97	195	973
Amanita		523	75	150	748
Boletus		751	107	215	1073
Cortinarius		585	83	168	836
Entoloma		587	84	168	839
Exidia		304	44	87	435
Hygrocybe		520	75	149	744
Lactarius		760	109	218	1087
Pluteus		308	44	89	441
Russula		751	107	215	1073
Suillus		462	66	133	661
Total		6664	953	1911	9528

3. Methods

3.1. ViT-Mushroom. Figure 1 shows the architecture of the ViT-Mushroom. The backbone of ViT-Mushroom is ViT-L/32, and it uses a transfer learning-based method [34, 35]. After the breakthrough of the Transformer [36] for dealing with natural language processing (NLP) tasks recently, ViT [37] has been implemented as an image recognition method for computer vision applications [38]. It is possible to solve the CNN difficulties that require stacking more layers and expanding the receptive field by employing Multi-Head Attention [37, 39–41]. ViT comprises these components: Linear Projection of Flattened Patches (Embedding layer), Transformer Encoder, and MLP Head.

ViT divides the original image into patches and transforms each patch into a vector to obtain a flattened patch. The shape of the input image is $H \times W \times C$, where C represents the number of input image channels, and H and W represent the height and width of the original image. ViT obtains N image patches by segmenting the original image with a $P \times P$ patch. ViT converts the image of $H \times W \times C$ into a sequence of $N \times (P^2 \times C)$. The sequence contains a total of N image patches, and the dimension of each image patch is $P^2 \times C$. Finally, the image patches are flattened and mapped to D dimensions using a linear projection with position-encoded vectors, analogous to the Word Vectors in NLP. The input sequence z of ViT is formulated as
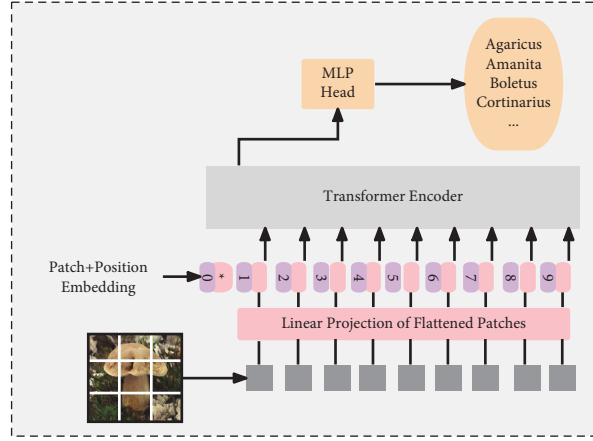
FIGURE 1: The ViT-Mushroom pipeline.

$$z_0 = \left[x_{\text{class}}; x_p^1 \mathbf{E}; x_p^2 \mathbf{E}; \ldots; x_p^N \mathbf{E}\right] + \mathbf{E}_{\text{pos}}, \mathbf{E} \in R^{(p^2 c) \times D}, \mathbf{E}_{\text{pos}} \in R^{(N+1) \times D}, \tag{1}$$

where x denotes an image patch, and the equations for ViT are presented in the following formulas:

$$\hat{x}^i = \text{MSA}\left(\text{LN}\left(x^{i-1}\right)\right) + x^{i-1}, \tag{2}$$

$$x^i = \text{MLA}\left(\text{LN}\left(\hat{x}^i\right)\right) + \hat{x}^i, \tag{3}$$

$$y = \text{LN}\left(\hat{x}^i\right), \tag{4}$$

where y is the output of ViT. ViT is mainly composed of Multi-Head Attention (MSA) and MLP (two fully connected layers and a Gaussian error linear unit activation function), with LayerNorm and residual connections added in front of MSA and MLP, as shown in Figure 2.

3.2. Transfer Learning.
In deep learning, labeled image data are scanty, and the Calibration effort is extremely expensive [42]. Meanwhile, transfer learning has attempted to overcome the problem of insufficient labeled training data. This makes transfer learning become a research hotspot in deep learning to transfer knowledge to a different but relevant second task when solving the first task. With this process, training a new deep network for task 2 will be unnecessary [34, 43]. Pan and Yang [43] put forward a formal definition of the concepts of domain and task. Let X denote an input space; $X = \{x_1, \ldots, x_m\} \in X$ and Y denotes a label space; $Y = \{y_1, \ldots, y_m\} \in Y$, and (x_i, y_i) denote the training pair. Let $\mathcal{D} = \langle X, P(X) \rangle$ denote a special domain, where $P(X)$ is a marginal probability distribution. $T = \langle Y, P(Y|X) \rangle$ denotes a task, where $P(Y|X)$ is a conditional probability distribution, in which the task is learned from training pairs. Given source domain, $D_S = \langle X_S, P(X_S) \rangle$ learning the task $T_S = \langle Y_S, P(Y_S|X_S) \rangle$, target domain $D_T = \langle X_T, P(X_T) \rangle$, and learning task $T_T = \langle Y_S, P(Y_T|X_T) \rangle$ [34, 44]. Transfer learning improves the learning of the target predictive function $P(Y_T|X_T)$ in which T_T uses the knowledge in D_S and T_S. In this issue, the first source domain on the ViT L/32

backbone has been trained in ImageNet-21k [45, 46]. The goal is to assist the network extract the crucial but generic feature representations to categorize images. After that, the original ViT L/32 classifier head was replaced with a new head specifically for mushroom classification.

3.3. t-SNE.
This study also explores the distribution of features generated by the transfer learning model to better understand their class separability [47, 48]. The output of the high-dimensional layers was viewed using dimensionality reduction methods [49]. The t-SNE was presented by Van der Maaten and Hinton [50] in 2008 as a novel method for scaling down high-dimensional data. The t-SNE uses stochastic neighbor embedding to convert high-dimensional Euclidean distances between data points into conditional probabilities. Let X be a vector holding all samples in the dataset and let Y be a target vector representing the low-dimensional representation, as shown in Eq. 5 [49]. The similarity of data point x_j to data point x_i is described using the conditional probability $P_{j|i}$ in the original high-dimensional space, written as a conditional probability [50, 51]:

$$P_{j|i} = \frac{\exp\left(-\left(\|x_i - x_j\|\right)^2 / 2\sigma_i^2\right)}{\sum_{r \neq s} \exp\left(\left(-\|x_r - x_s\|\right)^2 / 2\sigma_i^2\right)}. \tag{5}$$

The probabilities in the original space are expressed as follows:

$$P_{j|i} = \frac{\left(P_{j|i} + P_{i|j}\right)}{2n}. \tag{6}$$

The data size is denoted as the number n. To minimize overpopulation, the t-SNE employs Student's t-distribution with a single degree of freedom [50]. The probability of low-dimensional Q_{ij} is obtained from this distribution, as indicated by the following expression:
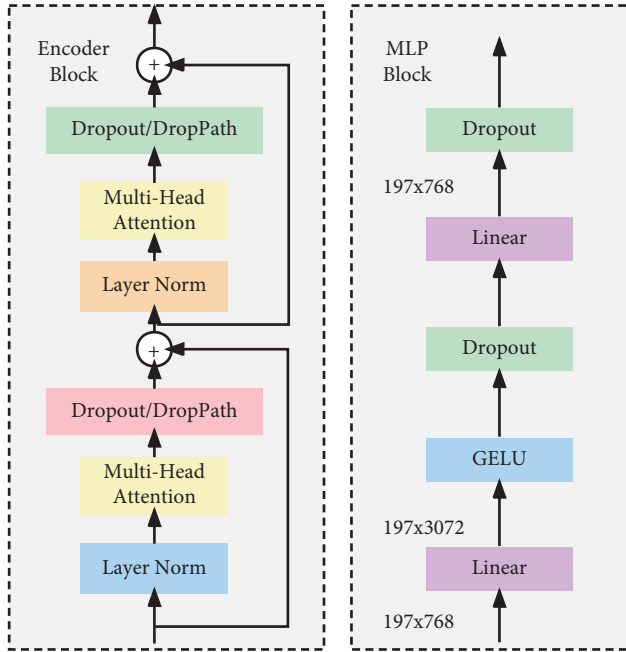
FIGURE 2: Architecture of the encoder block and MLP.

$$Q_{ij} = \frac{\left(1 + \left\| y_i - y_j \right\|^2\right)^{-1}}{\sum_{r \neq s}\left(1 + \left\| y_r - y_s \right\|^2\right)_{-1}}. \tag{7}$$

The goal is to learn the coordination y_i of the low-dimensional space to preserve the distribution of clusters in the low-dimensional embedding space. The t-SNE approach finds the projections of the input data x_i in the lower dimension y_i based on the Kullback–Leibler divergence [52] as well as the loss function and a gradient-based technique:

$$KL(PQ) = \sum_i \sum_j P_{ij} \log \frac{P_{ij}}{Q_{ij}}. \tag{8}$$

4. Experimental Setup

4.1. Augmentation. Data augmentation is applied to the deep learning model to boost the data, prevent overfitting, and develop a more general model. Several augmentation procedures, such as rotation, horizontal flipping, cropping, blurring, salt-and-pepper noise, and Gaussian noise, were used to produce an augmented dataset. Figure 3 shows the examples of each image augmentation method during the mushroom dataset experiment. Finally, the images were normalized using the mean and standard deviation of the ImageNet dataset, and we applied the random order command to disrupt the order of all transformed operations and increase the randomness of the operations.

4.2. Experimental Settings. The pretrained architectures for the classification are (1) ViT-L/32, (2) ResNet-34, (3) VGG-16, (4) Inception-V3, (5) Inception-ResNet-V2, and (6) Xception. These transformer-based and CNN-based pretrained models are fine-tuned according to the principles of transfer learning [34, 53, 54], which aims to transfer the knowledge learned to a different but relevant second task when solving the first task [34]. The weights of the pretrained architectures are first pretrained on ImageNet (I) to obtain a low-level feature extractor, share knowledge among computer vision problems in different fields, and serve as a feature extractor for new image sets. Most of the image data on ImageNet (I) belong to fields such as fish, birds, and objects. Conversely, our targets are mushroom images, and some trained images must fine-tune the pretraining models in the training dataset. Therefore, we fine-tune all pretraining networks, with the full connection layer in the original model. Then, we change the full connection layer to a custom layer and modify the fully connected layer according to the number of classifications.

In our experiment, all models are trained using the Adam optimizer with up to 30 epochs. The training batch size value and the test batch size are set to 16 and 8, respectively. The initial learning rate value is set to 3e-5. All models were built using Python language. The experiments were performed on the GPU NVIDIA CUDA version 11.0 on a Tesla P100-16 GB. In addition, the models applied in this experiment are from the PyTorch 1.9.1 (https://pytorch.org/) and the PyTorch Image Model Library (https://fastai.github.io/timmdocs/).

5. Results and Discussion

Table 2 presents the experimental findings obtained through the ViT-L/32 and other models. ViT-L/32 outperformed the CNN techniques on the mushroom test set, with an accuracy score of 95.97% and an AUC of 99.01%. Xception is the best performing CNN model for mushroom classification, with an accuracy score of 92.95% (approximately 3% lower than ViT-L/32.) and an AUC of 97.82% (approximately 1% lower than ViT-L/32). Xception is the only CNN model with an accuracy of above 90%. Of the CNN models, the VGG-16 produces the worst performance with an accuracy score of 81.31% and an AUC of 92.95%. The VGG-16's worst performance is associated with its structure and lack of new techniques, such as a residual network and an attention mechanism. Moreover, its connection structure is simpler and ineffective for mushroom classification.

Thus, we compared the model's precision, sensitivity (recall), F1 scores of ViT-L/32, and Xception. The average performance measures of macro average and weighted average revealed that ViT-L/32 outperforms Xception in terms of accuracy, sensitivity (recall), and F1 scores, which thereby obtained the best performance.

Table 3 shows the classification performance of ViT-L/32 for each mushroom species. The results suggest that ViT-L/32 score is high in each mushroom species' categorization. For *Exidia* species, ViT-L/32 had a higher F1-score of 99.43%, which outperformed the other six additional mushroom species with F1 scores above 95.00%. *Pluteus* (88.30%) and *Entoloma* (93.05%) were the only two species that performed poorly on the ViT-L/32 model.
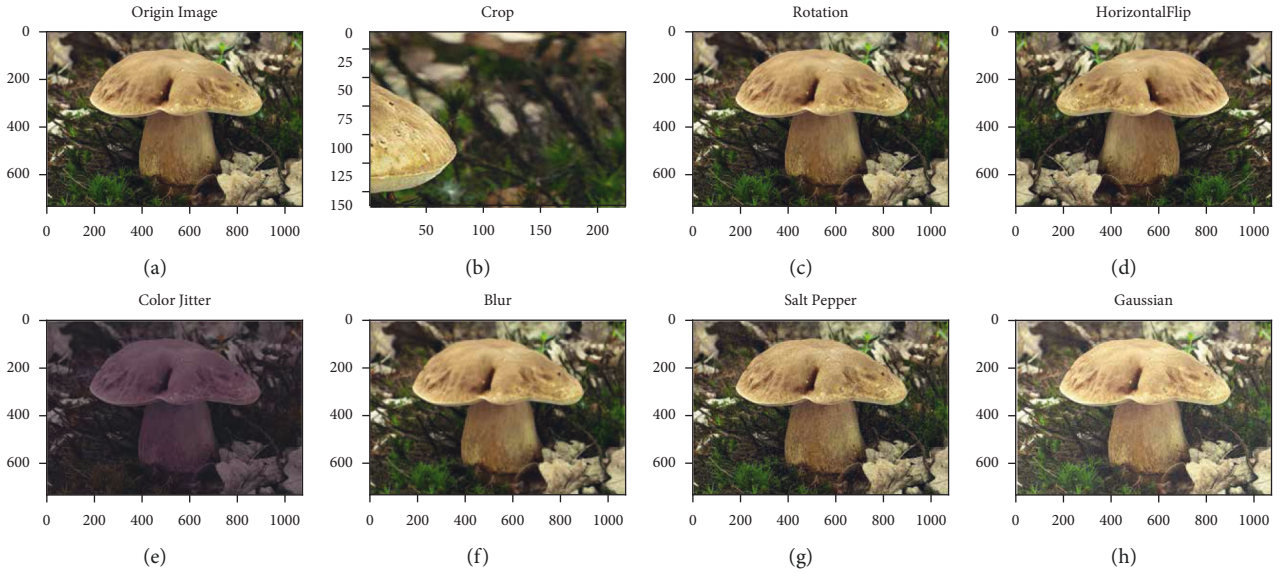
FIGURE 3: Examples of mushroom image augmentation: (a) original image, (b) random crop, (c) random rotation, (d) random horizontal flip, (e) random color jitter, (f) random blur, (g) random salt pepper noise, (h) random Gaussian noise.

TABLE 2: ViT-L/32 vs. SOTA CNN.

Model	Types	Evaluation metrics (%)				
		Precision	Sensitivity	F1-score	ACC	AUC
VGG-16	Macro average	81.85	79.71	80.25	81.31	92.95
	Weighted average	81.59	81.31	81.07		
ResNet-34	Macro average	85.64	84.72	84.88	85.79	94.43
	Weighted average	85.85	85.79	85.59		
Inception-V3	Macro average	82.21	81.91	81.94	83.04	92.88
	Weighted average	83.01	83.04	82.90		
Inception-ResNet-V2	Macro average	85.13	84.29	84.54	85.39	94.42
	Weighted average	85.44	85.39	85.25		
Xception	Macro average	92.85	92.61	92.64	92.95	97.82
	Weighted average	93.20	92.95	92.98		
ViT-L/32	Macro average	95.69	95.92	95.77	95.97	99.01
	Weighted average	96.05	95.97	95.99		

TABLE 3: Classification evaluation results of ViT-mushroom.

Class	Evaluation metrics (%)		
	Precision	Sensitivity	F1-score
Agaricus	96.50	98.97	97.72
Amanita	100.00	96.67	98.31
Boletus	99.07	98.60	98.83
Cortinarius	94.01	93.45	93.73
Entoloma	94.48	91.67	93.05
Exidia	98.86	100.00	99.43
Hygrocybe	99.32	97.99	98.65
Lactarius	95.39	94.95	95.17
Pluteus	83.84	93.26	88.30
Russula	94.95	96.28	95.61
Suillus	96.12	93.23	94.66

We further examine the results of the confusion matrix of the six models' classification, as shown in Figure 4. Most models struggle to distinguish between these three types of mushroom species: (1) *Entoloma* and *Pluteus*, (2) *Lactarius* and *Russula*, and (3) *Cortinarius* and *Suillus*, with several misclassifications. VGG-16 misdiagnosed 30 *Pluteus* photos
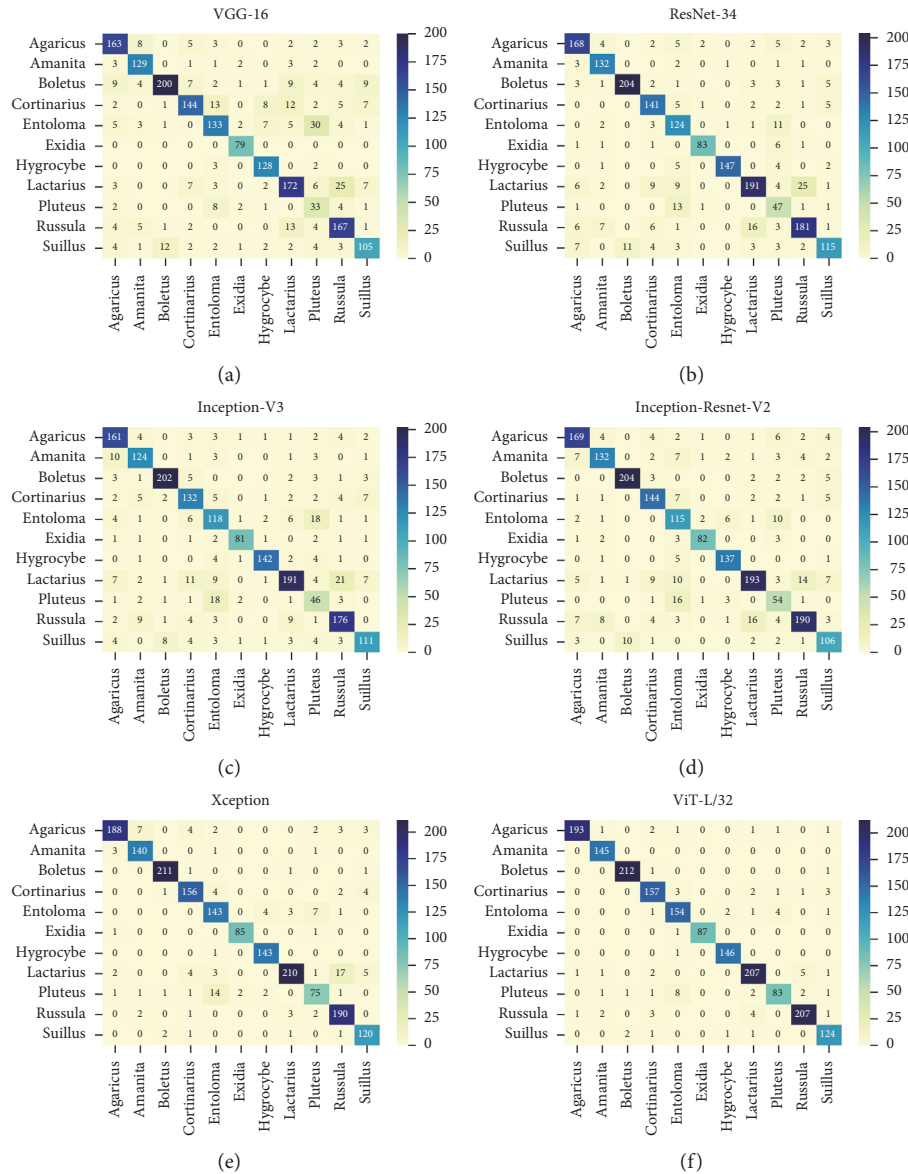
FIGURE 4: Confusion matrix of the classification results of all models: (a) VGG-16, (b) ResNet-34, (c) Inception-V3, (d) Inception-ResNet-V2, (e) Xception, (f) ViT-L/32.

as *Entoloma*, 25 *Russula* images as *Lactarius*, 13 *Lactarius* images as *Russula*, and 7 *Suillus* images as *Cortinarius*. Of the top overall performer among CNNs, Xception was misdiagnosed as follows: 4 *Suillus* as *Cortinarius*, 14 *Pluteus* as *Entoloma*, 17 *Russula* as *Lactarius*, and 3 *Lactarius* as *Russula*.

ViT-L/32 had the fewest classification errors among the CNN models, and it was the best at identifying mushrooms. Moreover, ViT-L/32 has a good classification accuracy for the following mushroom groups: (2) *Lactarius* vs. *Russula* and (3) *Cortinarius* vs. *Suillus*. Only five *Lactarius* photographs were misdiagnosed as *Russula*, whereas no *Russula* images were misidentified as *Lactarius*. Only one *Suillus* photo has been mistaken for *Cortinarius*.

ViT-L/32 seems to be less effective in classifying (1) *Entoloma* vs. *Pluteus*. However, it still outperforms CNN in

total accuracy. ViT-L/32 incorrectly classifies eight *Pluteus* photos as *Entoloma* and four *Entoloma* images as *Pluteus*.

Table 4 compares the performance of the proposed method with methods presented by other published studies, revealing that our approach outperforms the other five approaches, regarding the accuracy rate.

The classification relies heavily on the visual characteristics used for categorization. Since the feature represents the content of an image, its quality has a significant impact on classification performance. We compare the learned features in the CNN and transformer-based models to evaluate how crowded the feature space is. In each model, we extract the output of the last layer of the feature extractor to obtain a multidimensional feature vector. Then, the feature vectors were projected to 2D space using commonly used dimensional reduction methods, such as t-SNE approaches.

TABLE 4: Performance of the proposed method compared with other published methods for mushroom classification.

| Authors	Methods description	Accuracy (%)
Maurya et al. [22]	Classification of mushrooms using texture features based on an SVM classifier	76.60
Sajedi et al. [31]	Four-layer CNN model with MLP classifier	80.70
Zahan et al. [4]	Inception-V3 deep learning network and contrast-limited adaptive histogram equalization	88.40
Kiss et al. [55]	Transfer learning, noisy student, and EfficientNet-B5 model	92.60
Devika et al. [32]	DCNN model with four convolutional layers and one fully connected layer	93.00
Ours	Data augmentation, transfer learning and vit-l/32 network	95.97

(a)

(b)

(c)

(d)
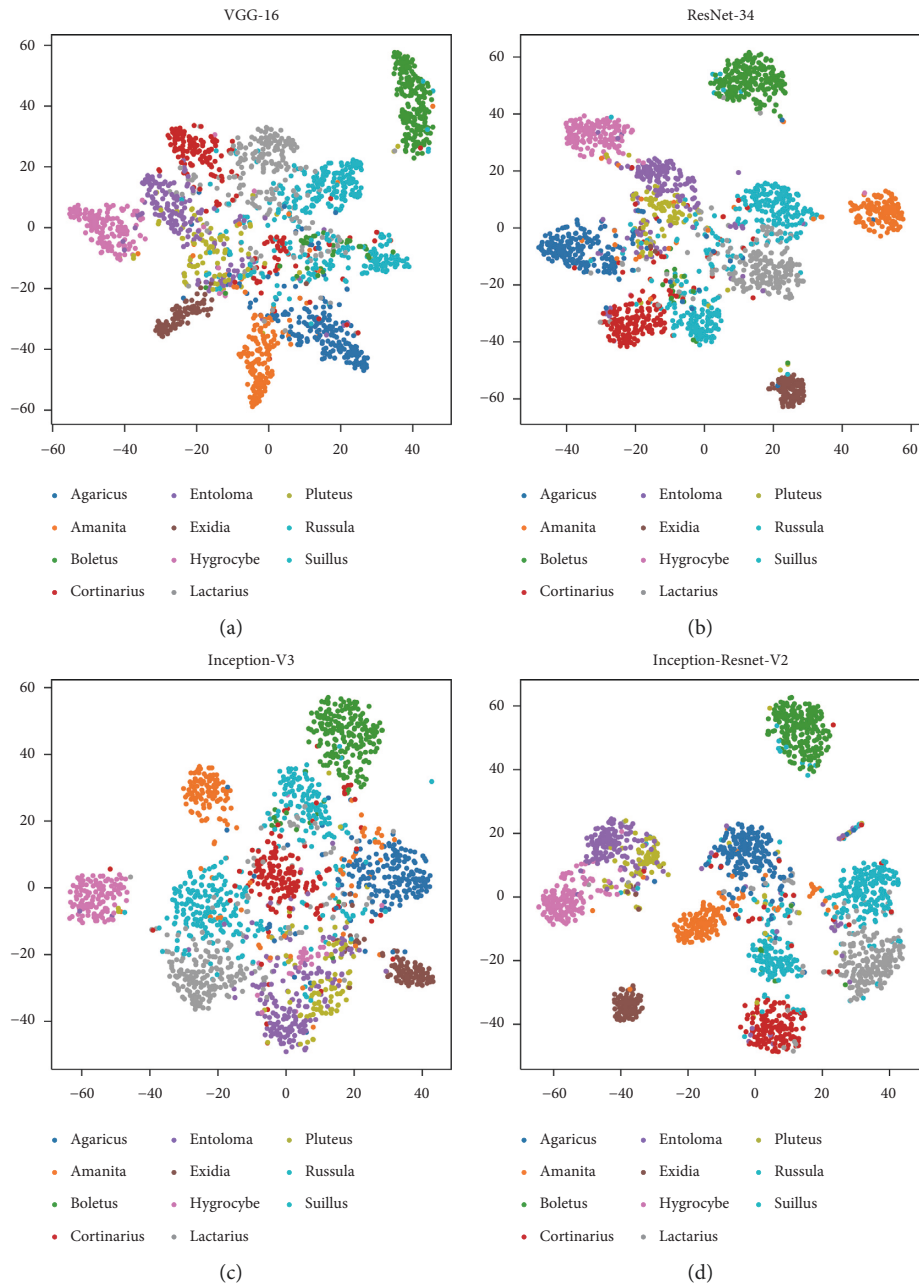
FIGURE 5: Continued.
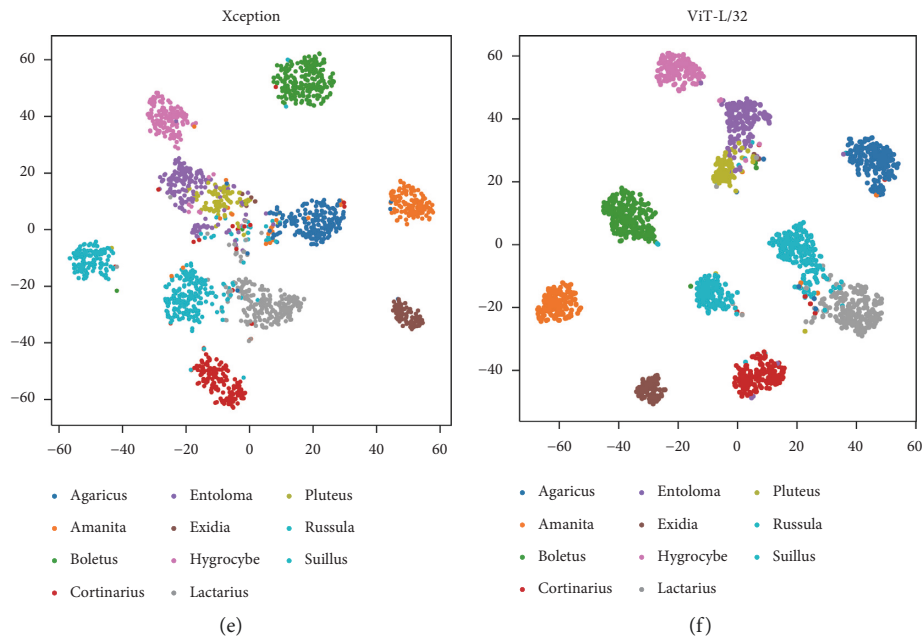
Figure 5: The t-SNE visualization. (a) VGG-16, (b) ResNet-34, (c) Inception-V3, (d) Inception-ResNet-V2, (e) Xception, (f) ViT-L/32.

The findings of t-SNE are depicted in Figure 5. The various colors in the scatterplot signify the images of various mushroom classes in each subgraph. The following conclusions are drawn from the t-SNE feature distribution maps. Compared with other techniques, the t-SNE results of ViT-L/32 are well-plotted in a relatively compact space and exhibit the clearest separation of each class, indicating that ViT-L/32 may minimize intraclass variances and provide well-separated feature embeddings.

6. Conclusion

We used five models based on convolutional neural network architecture (VGG, ResNet, Inception, Inception-ResNet-V2, and Xception) and the ViT-L/32 model based on a transformer architecture to train and classify 11 different types of mushrooms. To select the most suitable deep learning model for mushroom classification, the accuracy of these six classification models was compared. The results show that the ViT-L/32 model outperforms the other five CNN models in all evaluation metrics, and it has the clearest boundaries for the scatterplots in various classes of its high-dimensional output mapping of t-SNE. ViT-L/32 is considered a promising model for the automatic classification of toxic and edible mushrooms. This model can also assist wild mushroom consumers in avoiding eating toxic mushrooms, safeguarding food safety, and helping the public health sector prevent incidents of foodborne diseases. The results will offer valuable resources for food scientists, nutritionists, and the public health sector regarding the safety and quality of mushrooms. In the future, we will investigate ViT network-based mushroom target detection and image segmentation tasks. Moreover, we will compare the performance of ViT with other target detection and segmentation models of mushrooms in future work.

Data Availability

The datasets used during the current study are available at Kaggle, https://www.kaggle.com/mustai/mushroom-12-9528.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

Authors' Contributions

BOYUAN WANG was born in Beijing, China, in 1985. He received his first M.E. degree in software engineering from Beijing Jiaotong University, Beijing, China and the second M.E. degree in E-media from Group T-International University College Leuven, Leuven, Belgium. He is currently a deputy secretary-general of the Spatial Statistics Branch of the Chinese Association for Applied Statistics (CAAS). He is also an engineer at the Centers for Disease Control and Prevention in Zhongshan City, Guangdong Province, China. He is currently pursuing a Ph.D. degree in artificial intelligence from Macau University of Science and Technology, Taipa, Macau. His current research interests include deep learning, spatial statistics, geographic information systems, and their applications. He has published eight papers in Chinese core journals and three SCI papers as a coauthor.

Acknowledgments

References

[1] D. S. Hibbett, M. Binder, J. F. Bischoff et al., "A higher-level phylogenetic classification of the Fungi," *Mycological Research*, vol. 111, no. 5, pp. 509–547, 2007.

[2] N. Liu, "Perspectives into the value of genera, families and orders in classification," *Mycosphere*, vol. 7, no. 11, pp. 1649–1668, 2016.

[3] N. Zahan, M. Z. Hasan, M. A. Malek, and S. S. Reya, "A deep learning-based approach for edible, inedible and poisonous mushroom classification," in *Proceedings of the International Conference on Information and Communication Technology for Sustainable Development(ICICT4SD)*, Dhaka, Bangladesh, 2021.

[4] Y. Zhang, M. Mo, L. Yang et al., "Exploring the species diversity of edible mushrooms in yunnan, southwestern China, by DNA barcoding," *Journal of Fungi*, vol. 7, no. 4, p. 310, 2021.

[5] D. L. Hawksworth, R. Lücking, and J. Heitman, "Fungal diversity revisited: 2.2 to 3.8 million species," *Microbiology Spectrum*, vol. 5, no. 4, 2017.

[6] M.-Q. He, R.-L. Zhao, K. D. Hyde et al., "Notes, outline and divergence times of Basidiomycota," *Fungal Diversity*, vol. 99, no. 1, pp. 105–367, 2019.

[7] F. Wu, L.-W. Zhou, Z.-L. Yang, T. Bau, T. H. Li, and Y. C. Dai, "Resource diversity of Chinese macrofungi: edible, medicinal and poisonous species," *Fungal Diversity*, vol. 98, no. 1, pp. 1–76, 2019.

[8] J. Pérez-Moreno and M. Martínez-Reyes, "Edible ectomycorrhizal mushrooms: biofactories for sustainable development," in *Biosystems Engineering: Biofactories for Food Production in the Century XXI*, R. Guevara-Gonzalez and I. Torres-Pacheco, Eds., pp. 151–233, Springer International Publishing, Berlin, Germany, 2014.

[9] H. Li, H. Zhang, Y. Zhang et al., "Mushroom poisoning outbreaks — China, 2019," *China CDC Weekly*, vol. 2, pp. 19–24, 2020.

[10] H. Li, H. Zhang, Y. Zhang et al., "Mushroom poisoning outbreaks — China, 2020," *China CDC Weekly*, vol. 3, pp. 41–45, 2021.

[11] J.-Z. Chen, W.-S. Fu, F. Xu et al., "Acute mushroom poisoning of Amanita pseudosychnopyramis: a case report from Fujian, China with exact species identification and descriptive study," *Toxicon*, vol. 212, pp. 55–61, 2022.

[12] W.-S. Jo, M. A. Hossain, and S.-C. Park, "Toxicological Profiles of Poisonous, Edible, and Medicinal Mushrooms," *Mycobiology*, vol. 42, no. 3, pp. 215–220, 2014.

[13] M. Beaumier, J.-P. Rioult, M. Georges, I. Brocheriou, T. Lobbedez, and A. Lanot, "Mushroom poisoning presenting with acute kidney injury and elevated transaminases," *Kidney International Reports*, vol. 4, no. 6, pp. 877–881, 2019.

[14] B. Le Daré, P.-J. Ferron, and T. Gicquel, "Toxic effects of amanitins: repurposing toxicities toward new therapeutics," *Toxins*, vol. 13, no. 6, p. 417, 2021.

[15] R. Wennig, F. Eyer, A. Schaper, and H. Andresen-Streichert, "Mushroom Poisoning," *Deutsches Arzteblatt international*, vol. 117, no. 42, pp. 701–708, 2020.

[16] C. S. Bever, C. A. Adams, R. M. Hnasko, L. W. Cheng, and L. H. Stanker, "Lateral flow immunoassay (LFIA) for the detection of lethal amatoxins from mushrooms," *PLoS One*, vol. 15, no. 4, Article ID e0231781, 2020.

[17] H. Karagöz, a. erden, and K. Esmeray, "Acute liver failure caused by mushroom poisoning: a case report and review of the literature," *International Medical Case Reports Journal*, vol. 6, pp. 85–90, 2013.

[18] E. Alkronz, K. Moghayer, M. Meimeh, M. Gazzaz, B. S. Abu-Nasser, and S. S. Abu-Naser, "Prediction of whether mushroom is edible or poisonous using back-propagation neural network," *International Journal of Corpus Linguistics*, vol. 3, pp. 1–8, 2019.

[19] N. Chitayae and A. Sunyoto, "Performance comparison of mushroom types classification using K-nearest neighbor method and decision tree method," in *Proceedings of the 3rd International Conference on Information and Communications Technology (ICOIACT)*, 2020.

[20] H. Yin, W. Yi, and D. Hu, "Computer vision and machine learning applied in the mushroom industry: a critical review," *Computers and Electronics in Agriculture*, vol. 198, Article ID 107015, 2022.

[21] M. A. Ottom and N. A. Alawad, "Classification of mushroom fungi using machine learning techniques," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 5, pp. 2378–2385, 2019.

[22] P. Maurya and N. P. Singh, "Mushroom classification using feature-based machine learning approach," in *Proceedings of the 3rd International Conference on Computer Vision and Image Processing*, Singapore, Asia, 2020.

[23] H. Rahman, M. O. Faruq, T. B. Abdul Hai et al., "IoT enabled mushroom farm automation with Machine Learning to classify toxic mushrooms in Bangladesh," *Journal of Agriculture and Food Research*, vol. 7, Article ID 100267, 2022.

[24] J. Preechasuk, O. Chaowalit, F. Pensiri, and P. Visutsak, "Image analysis of mushroom types classification by convolution neural networks," in *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*, Kobe, Japan, 2019.

[25] D. Wagner, D. Heider, and G. Hattab, "Mushroom data creation, curation, and simulation to support classification tasks," *Scientific Reports*, vol. 11, no. 1, p. 8134, 2021.

[26] P. Tongcham, P. Supa, P. Pornwongthong, and P. Prasitmeeboon, "Mushroom spawn quality classification with machine learning," *Computers and Electronics in Agriculture*, vol. 179, Article ID 105865, 2020.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[28] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: a review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, 2017.

[29] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Computer Speech & Language*, vol. 30, no. 1, pp. 61–98, 2015.

[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[31] H. Sajedi, F. Mohammadipanah, and A. Pashaei, "Automated identification of myxobacterial genera using convolutional neural network," *Scientific Reports*, vol. 9, no. 1, Article ID 18238, 2019.

[32] G. Devika and A. G. Karegowda, "Identification of edible and non-edible mushroom through convolution neural network," in *Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021)*, Berlin, Germany, 2021.

[33] P. Wang, J. Liu, L. Xu et al., "Classification of Amanita species based on bilinear networks with attention mechanism," *Agriculture*, vol. 11, no. 5, p. 393, 2021.

[34] F. Zhuang, Z. Qi, K. Duan et al., "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[35] Y. Brima, M. Atemkeng, S. Tankio Djiokap, J. Ebiele, and F. Tchakounte, "Transfer learning for the detection and diagnosis of types of pneumonia including pneumonia induced by COVID-19 from chest X-ray images," *Diagnostics*, vol. 11, no. 8, p. 1480, 2021.

[36] Y. Dong, J.-B. Cordonnier, and A. Loukas, "Attention is not all you need: pure attention loses rank doubly exponentially with depth," *Proceedings of Machine Learning Research*, vol. 139, pp. 2793–2803, 2021.

[37] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: transformers for image recognition at scale," *International Conference on Learning Representations*, , 2021, https://arxiv.org/abs/2010.11929.

[38] B. Wang, Q. Xie, J. Pei, P. Tiwari, Z. Li, and J. fu, "Pre-trained Language Models in Biomedical Domain: A Systematic Survey," *CoRR*, vol. 2110, 2021 https://arxiv.org/abs/2110.05006, Article ID 05006.

[39] B. Heo, S. Yun, D. Han, S. Chun, J. Choe, and S. J. Oh, "Rethinking Spatial Dimensions of Vision Transformers," IEEE/CVF International Conference on Computer Vision (ICCV), 2021, https://arxiv.org/abs/2103.16302.

[40] Z. Liu, Y. Lin, Y. Cao et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," *IEEE/CVF International Conference on Computer Vision (ICCV)*, https://arxiv.org/abs/2103.14030, 2021.

[41] H. Fan, B. Xiong, K. Mangalam et al., *Multiscale Vision Transformers*, IEEE/CVF International Conference on Computer Vision (ICCV), , 2021.

[42] R. B. Altman, "Artificial intelligence (AI) systems for interpreting complex medical datasets," *Clinical Pharmacology & Therapeutics*, vol. 101, no. 5, pp. 585-586, 2017.

[43] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[44] X. Sun, J. Shi, L. Liu et al., "Transferring deep knowledge for object recognition in Low-quality underwater videos," *Neurocomputing*, vol. 275, pp. 897–908, 2018.

[45] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[46] H. Benbrahim and A. Behloul, "Fine-tuned xception for image classification on tiny ImageNet," in *Proceedings of the International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, El Oued, Algeria, 2021.

[47] M. Shorfuzzaman, M. Masud, H. Alhumyani, D. Anand, and A. Singh, "Artificial neural network-based deep learning model for COVID-19 patient detection using X-ray chest images," *Journal of Healthcare Engineering*, vol. 202116 pages, Article ID 100340, 2021.

[48] B. Melit Devassy, S. George, and P. Nussbaum, "Unsupervised clustering of hyperspectral paper data using t-SNE," *Journal of Imaging*, vol. 6, no. 5, 2020.

[49] M. Shorfuzzaman, "IoT-enabled Stacked Ensemble of Deep Neural Networks for the Diagnosis of COVID-19 Using Chest CT Scans," *Computing*, https://www.scienceopen.com/document?vid=cff05885-2021-4369-b6bb-a5b98a94dbcc, 2021.

[50] L. van der Maaten and G. Hinton, "Viualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[51] B. Melit Devassy and S. George, "Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE," *Forensic Science International*, vol. 311, Article ID 110194, 2020.

[52] S. Kullback, *Information Theory and Statistics*, Wiley, Hoboken, NY, USA, 1959.

[53] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt, and J. Carretero, "Detection and analysis of COVID-19 in medical images using deep learning techniques," *Scientific Reports*, vol. 11, no. 1, 2021.

[54] A. Bhatt, A. Ganatra, and K. Kotecha, "COVID-19 pulmonary consolidations detection in chest X-ray using progressive resizing and transfer learning techniques," *Heliyon*, vol. 7, no. 6, Article ID e07211, 2021.

[55] N. Kiss and L. Czùni, "Mushroom image classification with CNNs: a case-study of different learning strategies," in *Proceedings of the 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Zagreb, Croatia, 2021.