Critical Analysis of Machine Learning and Deep Learning Models for Mushroom Classification

MSc Research Project
Masters in Artificial Intelligence

Akashdip Neogi
Student ID: x23218461

School of Computing
National College of Ireland

Supervisor: Dr. Muslim Jameel Syed

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Akashdip Neogi
Student ID:	x23218461
Programme:	Masters in Artificial Intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr. Muslim Jameel Syed
Submission Due Date:	12/12/2024
Project Title:	Critical Analysis of Machine Learning and Deep Learning Models for Mushroom Classification
Word Count:	6170
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Akashdip Neogi
Date:	9th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Critical Analysis of Machine Learning and Deep Learning Models for Mushroom Classification

Akashdip Neogi
x23218461

Abstract

Accurate mushroom classification is crucial to prevent the consumption of toxic species. This study compares machine learning and deep learning models, including Random Forest, MobileNetv2m ResNet50, ResNet101, VGG16 and custom CNNs, for classifying edible and poisonous mushrooms. **MobileNetV2** had achieved the highest accuracy of (**82.14**%), followed by **ResNet50** (**75**%), while traditional methods underperformed. The findings emphasize the effectiveness of deep learning, particularly with transfer learning and hyperparameter optimization. Practical applications include real-time identification tools for foragers and researchers. Despite limitations such as a small dataset, this research provides a strong foundation for future work on more diverse datasets and multi-modal approaches, advancing sage mushroom classifications.

1 Introduction

1.1 Mushroom importance and nutritional value

Mushrooms have been long valued for their culinary, medicinal, and ecological significance. They were cultivated approximately 6 million metric tonnes in 2010; among them, china was the world's leading producer. Many countries often consume mushrooms as a delicacy, especially in central and eastern Europe (Kalač; 2012). The consumption of mushrooms in countries like China is very high compared to other countries. Mushrooms are rich in nutrients such as digestible proteins, carbohydrates, fibre and certain vitamins, as well as minerals and antioxidants (Wang et al.; 2014).

1.2 Mushroom foraging and about it

Since mushrooms offer such great nutritional and medicinal benefits, foraging for wild mushrooms is one such practice that is deeply rooted in various cultures (Procházka et al.; 2023). Mushroom picking in general is the activity of picking wild mushrooms to eat and is a practice that has recently come back to life at least in part due to concerns over sustainability, food trends, and experiencing nature. Yet not only the same give a chance to taste various delicious and healthy mushrooms but also has a wide amount of health advantages. Engaging in mushroom foraging encourages outdoor activities, improves blood pressure, improves bone density, and serves as an excellent cardiovascular workout (Areo; 2024).

1.3 Challenges in Mushroom Classification

One of the main problems is the ability to identify the edible and poisonous species, that are dangerous for people's health of foragers and consumers. The main difficulty is attributed to several causes. Firstly, most deadly mushrooms in density, form and colour resemble the edible ones. For instance, one of the deadly mushrooms named **Cortinarious rebellus** can be easily mistaken for the edible varieties due to its common physical characteristics. Secondly, other factors like the type of soil, climate and regions influence the structural change of the mushroom's appearance and at times the colour, size and shape and thus the identification is almost impossible. Furthermore, different species of mushrooms also change their morphological appearance radically as they grow and develop from one stage to the other, a particular species containing nutrients good for the body at one stage may contain poisons that are bad for the body when fully grown from another.

Field guides and identification applications are scarce, depending on the country or region, and may not include all the known species of the area, or consider all possible variations in morphological features, which can lead to mistakes. The health authorities have also discouraged data science and policy use from these tools owing to these limitations. First, some of the mushrooms that can be edible have toxic look-alikes that can only be differentiated by experts. (Kaaronen; 2020).

1.4 Use of ML techniques

More recently as deep learning has improved new avenues for automating and improving the accuracy of these classifications have emerged. In this paper specific reference will be made to Convolutional Neural Networks (CNNs) which are very instrumental in image recognition and mapping in several applications, including in health imaging and even environmental imaging. As a result of their work, it applies CNNs created a new systematic way that offers a high level of differentiation and makes it possible to classify mushrooms as edible or poisonous when using photographic data. (He et al.; 2019).

Mascarenhas and Agarwal (2021) have explored and experimented with several architectures of Convolutional Neural Networks (CNNs) namely VGG16, ResNet, and MobileNet, and have found out that they are renowned for their effectiveness in handling complex image classification tasks. In that research, they have found out that each model offers their unique advantages in terms of computational efficiency, and the ability to generalize across different imaging conditions. VGG16 is known for its simplicity and depth, ResNet for its ability to train intense networks through the use of residual connections, and MobileNet for its efficiency in environments with limited computational resources.

1.5 Research Objectives and Goals

1.5.1 Research Objectives

The research objectives of this research were as follows:-

- **RQ1**: **Which model performs the best at classifying edible and poisonous mushrooms?**

In this study, we have critically analyzed the performance of one machine learning namely **Random Forest** and 6 deep learning models, namely **Resnet101**, **Resnet50**, **VGG16**, **MobileNetv2**, **Custom CNN model implemented using TensorFlow and PyTorch**. These models were applied to a dataset specially curated for this study, comprising 326 images of mushroom species under diverse environmental conditions. The goal of RQ1 is to evaluate the classification performance of these models and to determine the most effective one for mushroom classification.

- **RQ2**: **How can the accuracy and precision of these models be improved for real-world applications in mycology and public safety?**

  Building on RQ1, this research further investigates ways to optimize the accuracy of the model, this includes implementing hyperparameter tuning, applying advanced data augmentation techniques, and experimenting with modifications to the model architectures. By doing so, we aim to adapt the models for practical applications, ensuring they are robust and reliable for identifying edible and poisonous mushrooms in real-world scenarios.

1.5.2 Implications and Goals

In answering these research questions, we aim to fill this gap of mushroom classification using different machine learning as well as deep learning models in order to overcome the complex aspects of automated image classification in biology. This work has important implications for improving the forager risk factor, easing identification, and developing a reusable solution for other botanical categorization problems.

1.6 Paper Structure

This paper is structured as follows: after it, the authors provide the related work section, dataset description, models that were employed, training process, as well as evaluation measures. Next, there is the design specification section in which techniques and the architecture underlying the implementation have been described. Next, we have the implementation section which covers the actual technique/architecture that has been used in this study. The performance analysis section of the evaluation provides a detailed performance of each of the models presented. In the conclusion, we have also reviewed the interpretation of these results based on the theoretical and empirical writings identified earlier, and the conclusion also describes possible future research and uses of our results.

2 Related Work

Using machine learning, especially convolutions of a neural network type for the classification of edible and poisonous versions of the fungi has received much attention in the recent past. This review of the literature reviews the main paper, highlighting certain studies, which use different types of CNN architectures to overcome the difficulties flowing from the mushroom classification. In other words, it is essential to discuss the effects of mushrooms on human health, as a component of the exploration. Mushrooms are amongst the oldest vegetables used by human beings in their diets and system of traditional medicine, with numerous nutritional and medicinal values. Temperature regulation is also a factor, as the author of the research, mushrooms give a special flavour to dishes and they must be an important part of the human diet since they contain several useful nutrients.

Some of their nutritional profile includes:-

- **Macro nutrients**: Mushrooms have got low calories and fats; something that makes them ideal for those who want to shed a few pounds. It gives a rather limited quantity of proteins that include the amino acids that are needed by the body for various processes.
- **Vitamins**: These contain B vitamins, such as riboflavin, niacin and pantothenic acid each of which is involved in energy production and synthesis of red blood cells.
- **Minerals**: L-arginine found in mushrooms provides energy for muscles and cells, while selenium, potassium and zinc act on the digestive and immune systems, transmitting nerve impulses and enzymes.

2.1 Challenges in Mushroom Classification

Earlier, several researchers tried to classify mushrooms by applying several different techniques for the classification of mushrooms, but the results were not satisfactory. Humans have been classifying poisonous and edible mushrooms by considering their shape, colour, odour, and outer skin layer by experience for several years (Fukuwatari et al.; 2001).

The identification process relied heavily on observable trails such as colour, shape, gill structure and spore print. However, these characteristics can be highly variable and are subject to interpretation, necessitating significant expertise for accurate assessment. For instance, the genus ***Pleurotus*** encompasses species with overlapping characteristics, thus complicating the taxonomic classification (Albertó et al.; 2002).

Most cichlid species are cryptic and can therefore be considered as cryptic species, Overall, the cichlid taxocoenosis can be characterized by a high morphological convergence associated with the formation of the syntopic species.

There are supposed to be some diminutive taxa that are in genetic terms relatively different, but in the visual sense, they are mostly the same. On the other hand, different kinds of organisms can also show that they have evolved convergently and may possess similar adaptation features. These all question the identification methods in this area, as morphologically based approaches cannot even distinguish between the species in question (Lücking et al.; 2020).

In addition to learning how to categorize a myriad of subspecies, students will also learn how the environment has an impact on the morphology of animals.

This makes the identification challenge, complicated by other factors like the geographical locations of the mushrooms and climate influences like the soil quality that results in differences in size and shape of the mushrooms. Such variability also affects identification as an individual of a particular species will have different morphology depending on the condition under which it was exposed to (Yan et al.; 2021).

2.1.1 Phenotypic plasticity and Development States

Mushrooms pass through many forms in their life cycle and thus have many appearances distinct from one another. This phenotypical flexibility results in cases of confusion, since young examples of toxic species may look like typical examples of the edible species, and vice versa. This rearing observation therefore demands a clear understanding of the structural form of the species in the development phases.

First, there is a limitation in the applicability of Field Guides and related identification tools.

Emphasizing some drawbacks inherent in the Field and smart applications like Mushroom identification, it is also possible to conclude about the general hazards connected with the misidentification of mushrooms.

2.1.2 Field Guides: Limitation and Cultural Bias

Traditional foraging field guides have turned out to be very robust and serve a great purpose to the mushroom foragers by describing and even illustrating the given species. However, they possess inherent limitations such as:-

- **Cultural Bias**: Research indicated that the field guides may reflect their cultural biases, thus potentially leading to inconsistent or inaccurate information regarding the edibility of certain species (Rubel and Arora; 2008).
- **Static Information**: Printed guides cannot be taken into account for the dynamic nature of the fungal taxonomy, thus resulting in outdated or incomplete data.
- **Regional Limitations**: Many guides focus on specific geographic areas, thus limiting their usefulness in diverse or unfamiliar regions.

2.1.3 Mushroom Identification Apps: Accuracy Concerns

The advent use of smartphone applications aimed at identifying mushrooms has introduced new challenges

- **Inaccurate Identification**: Studies have found out that AI-powered mushroom identification apps are correct only about 50% of the time, with some instances of misidentifying deadly mushrooms as safe to eat.
- **Overreliance on Technology**: Users may develop a false sense of security, which leads to risky foraging behaviours without proper verification from reliable sources.

2.2 Health Risks Associated with Misidentification

The study published by Schenk-Jaeger et al. (2012) and their team in the year of 2012 analyzed the mushroom exposures in Switzerland reported by the Swiss Toxicological Information Centre between January 1995 to 2009 has found that the misidentification of amateur foragers accounted for more than 90% of the mushroom toxicity cases. From their extensive research it was also found that the impact of the consumption of poisonous mushrooms causes various symptoms and signs in our body, such as gastroenteritis or other chronic illnesses, it can also have a severe impact on the body such as liver failure which is irreversible or can be as fatal as organ failure, which can only be fixed by performing organ transplantation process.

The primary cause of mushroom poisoning is the accidental consumption of toxic species mistaken for edible ones. For instance, the death cap ***Amanita phalloides*** is responsible for the majority of fatal mushroom poisoning that happens worldwide (Ye and Liu; 2018).

The rise in mushroom foraging has been accompanied by an increase in reliance on identification applications, which have often provided inaccurate information. Health authorities have warned against the use of these applications due to their potential for errors, which emphasizes the importance of traditional knowledge and caution in mushroom foraging.

2.3 Advancement in Mushroom Classification

Recent advancements in machine learning have significantly enhanced mushroom classification, addressing the challenges that are inherent in traditional methods. Several studies have explored various machine-learning techniques to improve the accuracy and efficiency of mushroom classification.

2.3.1 Application of Machine Learning Techniques

Research conducted by Tarawneh et al. (2023) in the year 2023 investigated the use of machine learning algorithms such as Naïve Bayes, Decision Tree, and Support Vector Machine (SVM) for classifying mushrooms as edible or poisonous. Their study proposed an integrated model that combines the decision of the most accurate techniques, thus achieving an accuracy of 94%.

2.3.2 Deep Learning Approaches

Recent studies have explored the application of machine-learning techniques to improve mushroom classification accuracy. Such as a study by Liao et al. (2022) in the year of 2022 introduced "MushroomNet", a model based on the MobileNetV3 combined with an attention mechanism. This approach achieved an accuracy of 83.9% on a public dataset and 77.4% on a local dataset, thus demonstrating the potential of this lightweight network in mushroom recognition tasks.

Research conducted by Zhang et al. (2022) in the year 2022 applied deep convolutional neural networks (CNNs) to classify poisonous and edible mushrooms that are found in China. Their model analyzed hundreds of smartphone images, providing a decision-making basis to reduce the morbidity and mortality that have been caused by consuming poisonous mushrooms.

2.3.3 Integration of Genetic Information

Xiao et al. (2022) conducted research in 2022 and proposed a multi-branching recognition framework that integrates genetic information with image recognition. This approach bridged the gap between macroscopic biological information and microscopic molecular data, achieving over 90% accuracy in species identification and offering a novel method for intelligent biometrics.

The research conducted by Gautam et al. (2022) in the year 2022, developed molecular techniques such as DNA barcoding to address these limitations, offering more precise identification by analyzing genetic material. The methods can overcome the challenges posed by morphological variability and cryptic species, thus providing a more reliable approach to mushroom identification.

2.3.4 Lightweight Models for Mobile Applications

Research published by Peng et al. (2023) in the year 2023, has developed an improved MobileViT deep learning model for wild mushroom classification. By combining convolutional networks with attention mechanisms, their model achieved accuracies of 96.21% and 91.83% on two test datasets, respectively, thus highlighting the effectiveness of lightweight models that are suitable for deployment on mobile devices.

2.3.5 Comparison of different machine learning models for mushroom classification

Research conducted by Y and Chandrasekhar (2020), in the year 2020 has compared the performance of two machine learning models that are MobileNetV2 and VGG16 in predicting the six medicinal mushroom species. In that study, it was found that both the architecture were effective, with MobileNetV2 offering a balance between the accuracy and the computational efficiency.

Further, a comparative analysis was done by Bongat et al. (2023), in 2023, their research used transfer learning methods using two CNN models namely ResNetRS50 and EfficientNetV2B0, which were trained on ImageNet and were used for feature extraction. From their research, they found out that the EfficientNetV2B0 emerged as the most effective model with a validation accuracy of it being 90%. And the average testing accuracy was 92%.

3 Methodology

The methodology involved systematic data collection, preprocessing, model training, and evaluation to identify the most effective model for the task.

3.1 Dataset

The dataset utilized in this study was a meticulously curated collection of mushroom images from an online website that is Ultimate Mushroom *1160 Mushroom Identifications Await Your Discovery — ultimate-mushroom.com* (n.d.) and Mushroom Observer *Mushroom Observer — mushroomobserver.org* (n.d.), comprising ten species from both edible and poisonous categories. The dataset provides a comprehensive basis for analyzing and classifying mushroom species into edible and poisonous categories.

3.2 Data Preprocessing

Data preprocessing is the first and the most crucial step in data analysis, where the data are obtained from the dataset, cleaned and preprocessed which could be used as an input for the model (Stojnev and Ilic; 2020).

This step in cleaning, transforming, and organizing raw data to ensure it is ready for analysis and model building. By ensuring that the dataset is clean, consistent and well-structured. This step lays the foundation for successful machine-learning tasks.

3.3 Machine Learning and Deep Learning Models

3.3.1 VGG16

VGG16 is a deep learning architecture, that is, it is a convolutional neural network trained Simonyan and Zisserman (2015) at University of Oxford, in the year 2014. This model is described as simple and highly symmetrical, it consists of 16 layers having the 3×3 convolutional filters. However, it can be made clear that VGG16 has found great application in most image classification problems due to its simplicity and efficiency bearing in mind that this is an older model.

3.3.2 ResNet101, ResNet50

ResNet is a hyper-parameter tuned version of Densenet, and thus we also call it Resnet101 and ResNet50 are professional deep convolutional neural network models that were developed in 2015 by He et al. (2015) and other authors of Microsoft Research. Resnet101 and Resnet50 are residual networks containing a hundred and one layers making it ResNet101 and containing fifty layers hence its name ResNET50.

These models employ this architecture to enhance the training of very deep architectures hence reducing the vanishing gradient problem in order to enhance the network's ability to learn complex features from complex data sets.

That is, the Resnet50 model is deeper and more complex than the VGG16. It has outstanding feature extraction functions. Due to the large number of layers coupled with more computation and longer train ties, that requires a great amount of hardware. It can also work for vanishing gradient issues which are typical for deep networks. Unlike the ResNet101, this model has fewer layers, but it works with a favourable depth-complexity ratio. With the help of residual connections, ResNet50 allows dragging at the imperfection of other deeply trained networks making it possible to learn details.

3.3.3 MobileNetV2

MobileNet is a mini convolutional neural networks Model that is specially designed for devices with low computational power, It has depthwise separable convolution that reduces the computational cost while still achieving reasonable accuracy. This architecture was launched by Google in the year of 2018 and it is still in the same space as the MobileNetV1 (Sandler et al.; 2018).

This model is particularly useful, because of its high measurability, and the capacity to work effectively even if the resources are limited. However, due to its small structure, this model is oriented to the solution of comparatively simple problems and its application for analyzing data with more complex patterns could be not very effective. In this study, MobileNet was used due to its relatively low training time and overall moderate mean Average Precision.

3.3.4 Random Forest

Random Forest also known as Random Decision Forest was developed by Ho (1995) in 1995, the researcher is Tin Kam Ho who is a computer scientist, who worked for IBM research in northern California. This is a breeding clustering learning method that builds multiple decision trees at the training stage and outputs the mean of the results. This is somewhat of a very strong approach to overfitting and also can work with different kinds of data.

3.3.5 Hyperparameter tuning

Hyperparameter tuning is another crucial factor whereby the hyperparameter that is in the model is modified to enhance the model performance on data (Yu and Zhu; 2020). Hyper-parameter is a type of parameter that cannot be trained within the course of the improvement of the machine learning model but they can be incorporated into the construction of the model as a part of its architecture or into the efficiency and accuracy of the learning and optimization process of the model, respectively.

3.4 Training Accuracy

Training accuracy in this study concluded that deep learning yielded the best performance in helping architects to successfully train neural networks. Of the four measures of accuracy: training accuracy, cross validation accuracy, prediction accuracy, and overall accuracy, it is possible to explain training accuracy as follows: Training accuracy is the level of accuracy exhibited by training data that are used in training the model. It depicts the capacity of the model in terms of how well it has learned patterns from the train data. The formula for Training Accuracy is as follows:-

$$\text{Training Accuracy} = \frac{\text{Number of Correct Predictions on the Training Data}}{\text{Total number of training instances}}$$

3.5 Confusion Matrix

Confusion matrix is a tabular representation that is used to evaluate the performance of the machine learning model on the test dataset. It provides a detailed breakdown of the model's predictions, categorizing them into accurate and inaccurate instances across different classes. This matrix is particularly used for classification tasks, where the goal is to assign a categorical label to each of the input instances.
With the help of the confusion matrix, various performance metrics are derived such as:-

3.5.1 Accuracy

Accuracy represents the proportion of correctly predicted instances compared to the total number of instances in the dataset.
The formula for accuracy is as follows:-

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

3.5.2 Precision

Precision quantifies the accuracy of the model's positive predictions by calculating the proportion of the true positive instances among all the instances predicted as positive. A high precision score reflects that the model makes fewer false positive errors.
The formula for precision is as follows:-

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

3.5.3 Recall

Recall also referred to as sensitivity or the true positive rate is a measure of the proportion of actual positive instances that the model correctly identifies. It indicates how well the model captures all of the positive instances in the dataset.
This formula for recall is as follows:-

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negative}}$$

3.5.4 F1 Score

F1-score is a key metric for assessing the overall performance of a classification model, particularly in situations where there is an imbalance between classes. It combines two metrics that are precision and recall together into a single score, providing a balanced measure of the model's accuracy on positive predictions.
The formula for calculating F1-Score is as follows:-

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

4 Design Specification

Figure 1, shows the Research Framework that has been used in this study.

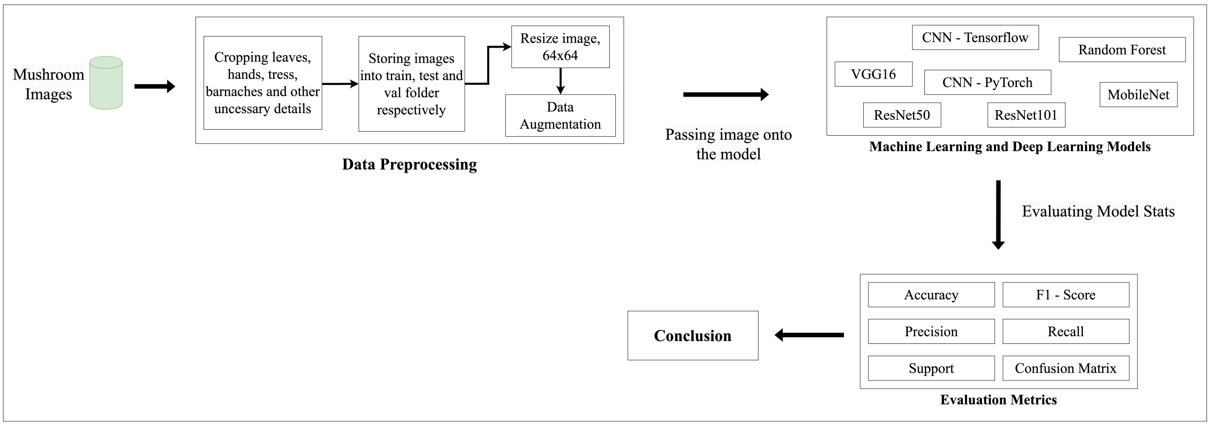

Figure 1: Research Framework

4.1 Preprocessing Techniques

In the first step, preprocessing of the data was performed, to eliminate the noise and irrelevant features such as grass, leaves, and hands, each image was manually cropped, ensuring only the mushroom remained in focus. This step was completed using a free open-source tool named **GIMP** (*GIMP — gimp.org*; n.d.). In this step the images of the mushroom were manually cropped to remove the extraneous elements such as grass, leaves, and hands, to focus solely on the mushrooms. Later the images were stored in the train, test and val folders respectively. The pixel values were normalized to the [0,1] range to accelerate the model convergence the to improve the training stability.

The input data consisted of the image files from the data which were adjusted to a uniform size of 64×64 pixels so that it doesn't take too much computational power to run all the models. To increase the robustness of the model image augmentation techniques were used in which the images were rotated randomly up to 20 degrees, to simulate the real-world variations. To make the data even more diverse it was flipped in both horizontal and vertical ways. For better accuracy of the model, the brightness and contrast were also adjusted to handle varying lighting conditions.

4.2 Machine Learning and Deep Learning Architectures

The number of epochs for the models was set to 30, to have a base for comparison.

4.2.1 Custom CNN Models

Two custom Convolutional Neural Network models were developed using both TensorFlow and PyTorch frameworks.

The model which was designed using the TensorFlow library consisted of **11 layers** and total as the table Table 4.2.1 shows the detailed architecture of the model.

Layer Name	**No of layers present**
Convolutional layer	**3**
Pooling layer	**3**
Dense layer	**3**
Dropout layer	**1**
Flatten layer	**1**

Table 1: Model Architecture, CNN TensorFlow

The model which was designed using the PyTorch Model consisted of **21 layers** in total as the table Table 4.2.1 shows the detailed architecture of the model.

Layer Name	No of layers present
Convolutional layer	**10**
Batch normalization layer	**1**
Pooling layer	**5**
Dense layer	**4**
Flatten layer	**1**

Table 2: Model Architecture, CNN Pytorch

4.2.2 Pre-Trained Deep Learning Models

The use of pre-trained models was done which involved the use of 4 models namely, ResNet101, ResNet50, VGG16 and MobileNetV2.

Resnet101 consisted of a total of **101 layers** in total and the table Table 4.2.2 shows the detailed architecture of the model.

Layer Name	No of layers present
Convolutional layer	**1**
Residual layer	**99**
Fully connected layer	**1**

Table 3: Model Architecture, Resnet101

Res50, on the other hand, is comprised of **50 layers** in total and the table Table 4.2.2 shows the detailed architecture of the model.

Layer Name	No of layers present
Convolutional layer	**1**
Residual layer	**48**
Fully connected layer	**1**

Table 4: Model Architecture, Resnet50

For the VGG16, the base model architecture which consisted of a total of **29** layers, in which the VGG16 model had **16** layers and **13** custom layers were added to the model for better feature extraction, and the table Table 4.2.2 shows the detailed architecture of the model.

Model Name	Layer Name	No of Layers Present
VGG16 (Base Model)	**Convolutional Layer**	**13**
VGG16 (Base Model)	**Fully Connected Layer**	**3**
Custom Layers	**Conv2D**	**2**
Custom Layers	**BatchNormalization**	**2**
Custom Layers	**MaxPooling2D**	**1**
Custom Layers	**GlobalAveragePooling2D**	**1**
Custom Layers	**Fully Connected Layer**	**4 (3 Dense + 1 Output)**
Custom Layers	**Dropout**	**3**

Table 5: Model Architecture, VGG16

This model followed a straightforward architecture with consecutive convolutional layers followed by a pooling layer, converging into a connected layer for classification.

MobileNetV2 is a lightweight model, which consists of **55** layers in total, and the table Table 4.2.2 shows the detailed architecture of the model.

Layer Name	No of layers present
Convolutional layer	**1**
Residual Blocks	**51**
Final Convolutional layer	**1**
Global Average Pooling layer	**1**
Fully Connected layer	**1**

Table 6: Model Architecture, MobileNetV2

4.2.3 Machine Learning Model

Random forest which is a machine-learning model was used, it consisted of **several decision trees** together, which are typically built using **CART (Classification Regression Tree)**.

This model uses the bagging technique to train the individual decision trees. At each split of the decision tree, a random subset of features is considered rather than using all the available features. The final predictions are made through **majority voting**, where each tree casts a vote for the predicted class, thus the final output is generated which is the average of the predictions from all the trees.

4.3 Equipment and Setup

The models were trained and tested on an Apple Macbook Pro. The system had an Apple M3 Pro chip that included an 11-core processing unit (CPU) with a maximum clock speed of up to 3.6 GHz. It also has an integrated 18-core graphics processing unit (GPU), which operates at a clock speed of up to 1.2 GHz. The system used was equipped with 18 GB of unified memory, which efficiently handles the loading of large datasets, execution of complex operations and simultaneously running of multiple processes without significant slowdowns.

5 Implementation

The implementation of this research focused on the comparative analysis of machine learning and deep learning models for the classification of edible and poisonous mushrooms. A structured and methodical approach was followed to ensure the models' successful development, training, and evaluation. This section outlines the key processes, tools, and outputs generated during the implementation phase. Figure 1, shows the implementation of the architecture.

5.1 Dataset Preparation

The dataset consists of images of 10 species of edible and poisonous classes, and it consists of a total of 235 images out of which there were 194 images in the train directory, 28 images in the val directory and 13 images in the test directory. It has 2 classes of mushrooms that are edible and poisonous. The data is divided the data into train and val folders, and in each of them, folders consist of images of mushrooms. The dataset was split into 80% for training and 20% for testing. The test folder consisted of images of mushrooms from both the edible and poisonous classes which consisted of images, that were not shown while training the model to check the accuracy of the model based on new images from the same class. The images were resized to 64×64 to standardize the model training process for all the models. Figure 2, you can see an example of an image contained in the dataset. In Table 5.1 is the name of the author and the respective images taken from the Ultimate Mushroom webpage.

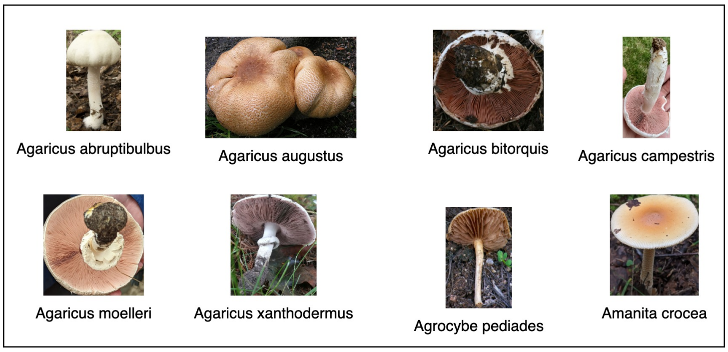

Figure 2: Some sample mushroom image data contained in the dataset

Image Name	**Author Name**
Agaricus abruptibulbus	Σ64
Agaricus augustus	Dick Culbert
Agaricus bitorquis	Nathan Wilson
Agaricus campestris	Rick Farwell
Agaricus moelleri	Zonda Grattus
Agaricus xanthodermus	Tomasz Sobczak
Agrocybe pediades	leschampignons
Amanita crocea	Alan rockefeller

Table 7: Image Name and Author Name

5.2 Model Development

For this study, we have implemented in total of seven models, comprising of both machine learning and deep learning approaches. For all the models accuracy, confusion matrix, precision, recall and f1-score were printed using the **scikit-learn** python library.

Below are the model's architecture explanations of all the models used in this research.

5.2.1 ResNet50 and ResNet101

For both deep learning models, ResNet101 and ResNet50 initially, the models were loaded with their pre-trained weights from the **ImageNet** dataset to leverage the knowledge already embedded in them. Subsequently, the classifier layers of both architectures were modified to align with the number of classes in the mushroom dataset, ensuring compatibility with the specific classification task. For both models, the **CrossEntropyLoss** function was set as the loss metric, and the optimizer was set to **Stochastic Gradient Descent (SGD)** to facilitate effective optimization during training. Both models were implemented using the **PyTorch** python library.

The ResNet50 model was performed using the **Optuna** framework to identify the optimal set of hyperparameters. The learning rate was searched over a logarithmic scale between $\mathbf{1}\times10^{-5}$ and $\mathbf{1}\times10^{-2}$, while the **momentum** was varied between two values, **0.7** and **0.99**. The **batch size** was evaluated across three options: **16**, **32**, and **64**.

On the other hand, the in ResNet101 model, the **learning rate** was fixed at **0.001**, and the **momentum** parameter was set to **0.9**, which helped maintain stability and improve convergence during training.

5.2.2 VGG16

This deep CNN model was implemented using the **TensorFlow** python library. The model was loaded using the **imagenet** weights.

Hyperparameter tuning was conducted on this enhanced VGG16 model to further optimize its performance. The choice of optimizer was narrowed down to **SGD (Stochastic Gradient Descent)** or **Adam**, and the loss function was set to **categorical crossentropy**, given the multi-class nature of the problem.

5.2.3 MobileNetV2

This lightweight model was implemented using the **TensorFlow** python library with its pretrained weights, setting the pretrained parameter to **True**. The classifier layer was modified to accommodate the binary classification task, corresponding to the two classes of edible and poisonous mushrooms. The **Binary Cross-Entrpy** as loss function was utilized for training, along with the optimizer set to **Adam** and the learning rate was set to **0.001** for efficient convergence.

5.2.4 Custom CNN models

Hyperparameter tuning for both the custom CNN models was performed as follows.

Hyperparameter tuning for the TensorFlow model was performed using the **Keras** and **Keras Tuner** libraries, leveraging the **HyperModel** class to optimize the architecture. The **kernel size** for **convolutional layers** was chosen between **3 and 5**, with **ReLU** set as the **activation function**. The **number of filters** for the **convolutional layers** was tuned as follows: **32–128**, **64–256**, and **128–512**, all in **increments** of **16**. For the **dense layer**, the **number of neurons** was tuned between **128 and 512** in **increments** of **32**. The **Adam optimizer** was used, with the **learning rate** logarithmically selected between **1e-4** and **1e-2**.

Hyperparameter tuning for the PyTorch model was performed using the **ParameterGrid utility** from the **sklearn.model_selection** library in Python to systematically generate all possible combinations of parameter values. The **learning rate** was tested at **0.0001**, **0.001**, and **0.01**, while the **dropout rate** were varied at **0.3**, **0.5**, and **0.7** to mitigate overfitting. Additionally, **batch size** of **16**, **32**, and **64** were evaluated to identify the optimal configuration. The model's **loss function** was set to **CrossEntropyLoss**, and the hyperparameter tuning aimed to maximize validation accuracy.

5.2.5 Random Forest

In this research, the model was implemented using the **scikit-learn** library. To ensure consistency, the **random state** value was set to **42**.

6 Evaluation

In this section, the performance of various machine learning and deep learning models was systematically, evaluated based on the validation accuracy, confusion matrices and training accuracy curves. The models are discussed in ascending order of their accuracy, highlighting their strengths and weaknesses.

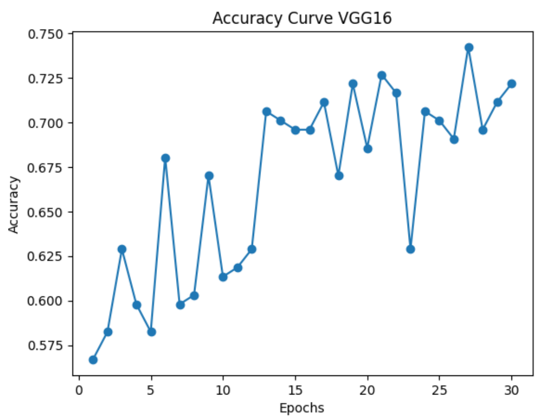

Figure 3: VGG16, Training Accuracy Curve

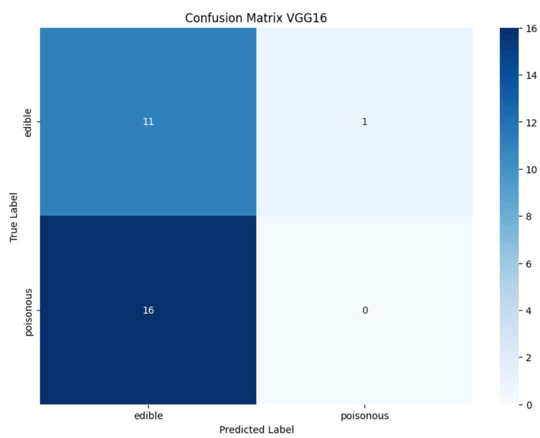

Figure 4: VGG16, Confusion Matrix

VGG16 had achieved the lowest accuracy of **39.28%**, with a training accuracy of

71.14% which can be inferred from this figure 3. Despite the rapid increase in training accuracy, the validation accuracy remained much lower, which indicated a potential overfitting issue. The confusion matrix, which can be referred to in this figure 4 highlighted the misclassification across all classes, which further reinforces its inability to capture complex image features. This poor performance, suggests that this model, despite its historical success in simpler datasets, was not suitable for the complexities of this mushroom classification task.

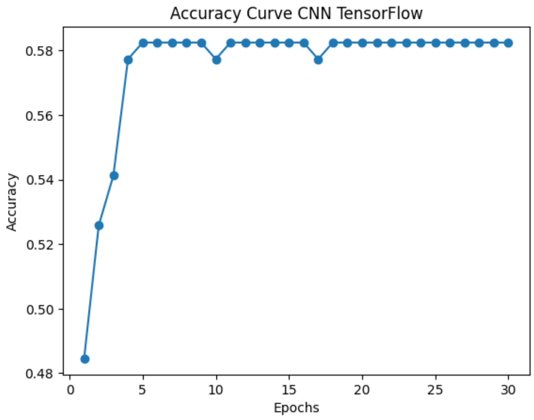

Figure 5: CNN TensorFlow, Training Accuracy Curve

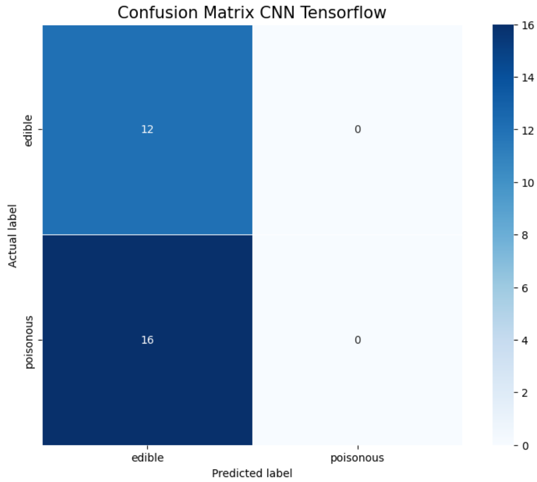

Figure 6: CNN TensorFlow, Confusion Matrix

The custom CNN Model which was designed using the **TensorFlow**, showed slightly improved performance and had achieved **42.85%** of accuracy, and **56.22%** in training accuracy. The model struggled to converge during the training, as evidenced by the flat accuracy curve, which can be seen in figure 5. Statistical analysis of the confusion matrix also revealed poor class separation, with the majority of predictions skewed towards the specific classes, which can be seen in figure 6. These findings suggest the need for more sophisticated architectural improvements or additional regularization techniques.

Metric	Score
Accuracy	0.5
Precision	0.625
Recall	0.5
F1 Score	0.433333

Figure 7: Random Forest, Model Stats

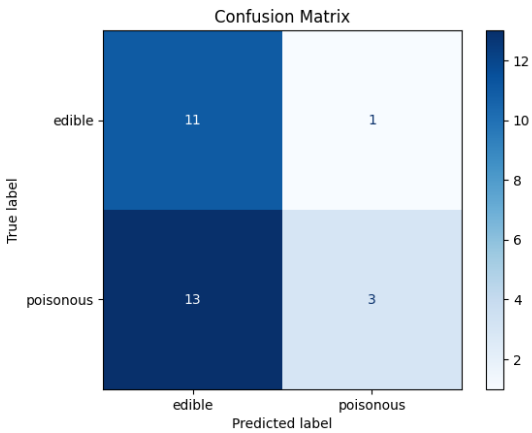

Figure 8: Random Forest, Confusion Matrix

Random Forest was the only machine learning model that was implemented in this research, and it had achieved an accuracy of **46.43%** that can be seen from figure 7. Although it had performed slightly better than the Tensorflow-based CNN, it fell short because this model is inherently limited in its ability to capture hierarchical features from the image data, which explains its lower performance. From the confusion matrix shown in figure 8it can be seen that it shows consistent misclassification across both the classes. However, this model's simplicity and interpretability were useful for feature selection tasks.

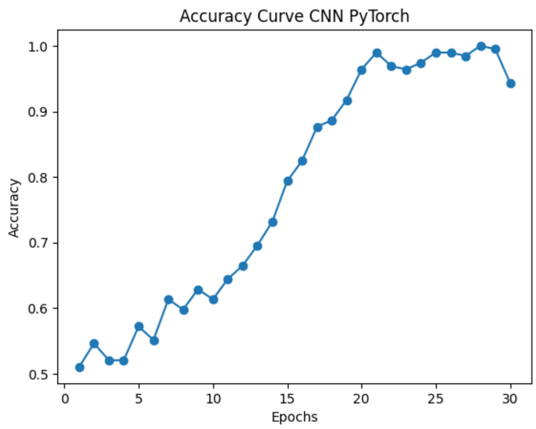

Figure 9: CNN PyTorch, Training Accuracy Curve

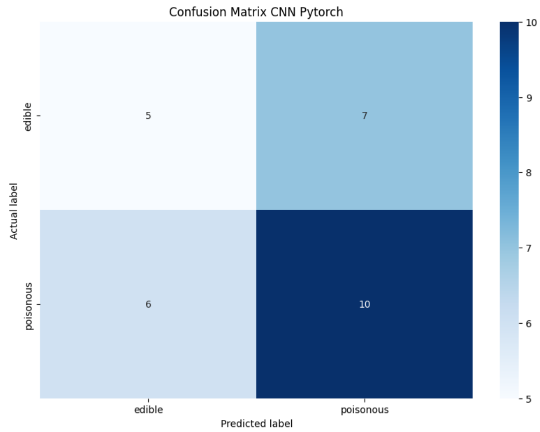

Figure 10: CNN PyTorch, Confusion Matrix

The **PyTorch-based** custom CNN model had achieved an accuracy of **53.57%**, and a training accuracy of **93.33%**, which can be also seen in this figure 9. These results show that the model was able to learn from the training data effectively but struggled to generalize to the test set, as revealed by the confusion matrix, in figure 10. This result emphasizes the importance of regularization techniques such as dropout and data augmentation to improve generalization. While the model demonstrated its potential, its performance lagged behind the pre-trained architectures, underscoring the value of transfer learning in complex classification tasks.

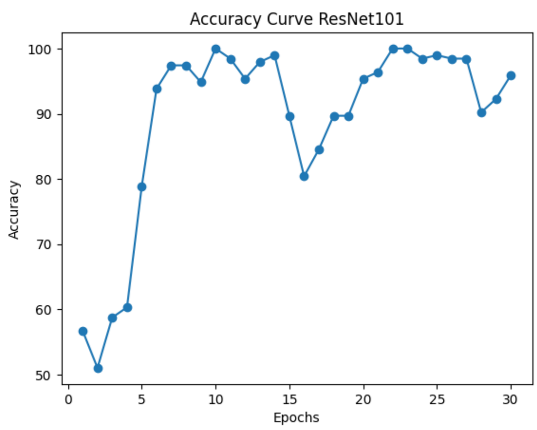

Figure 11: Resnet101, Training Accuracy Curve

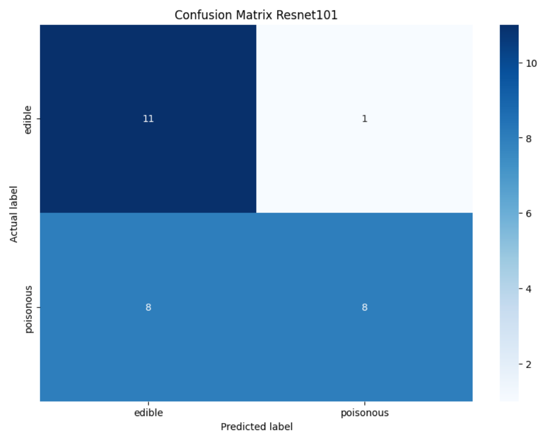

Figure 12: Resnet101, Confusion Matrix

In the case of the **Resnet101** model the achieved **67.86%** for the accuracy and **95%** for training accuracy, which can be seen in figure 11. Its deeper architecture enabled this model to capture complex features effectively. From the confusion matrix, from figure 12 it can be inferred that the model had done moderate misclassification, particularly in edge cases. Although ResNe101 had outperformed the custom CNNs and Random Forest, its accuracy is slightly lower than the ResNet50 model which suggests that the deeper model may be more prone to overfitting on relatively small datasets.

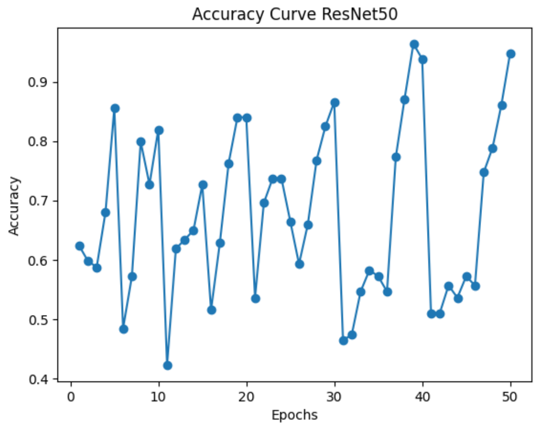

Figure 13: Resnet50, Training Accuracy Curve

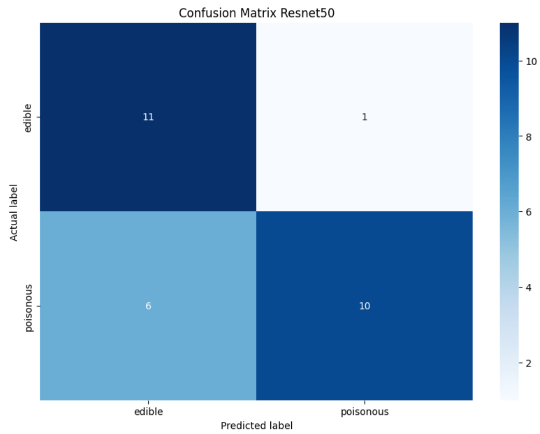

Figure 14: Resnet50, Confusion Matrix

The **Resnet50** has outperformed the **ResNet101**, by achieving an accuracy of **75**% with a training accuracy score of **95**%, which can be seen from figure 13. Due to the model's residual connections, it effectively mitigated the vanishing gradient problem and enabled a robust feature extraction. From the confusion matrix, figure 14, it can be seen that it has fewer misclassifications compared to the models discussed earlier, indicating better generalization. This result validates the effectiveness of this model in handling complex image classification tasks, particularly when paired with careful hyperparameter tuning,

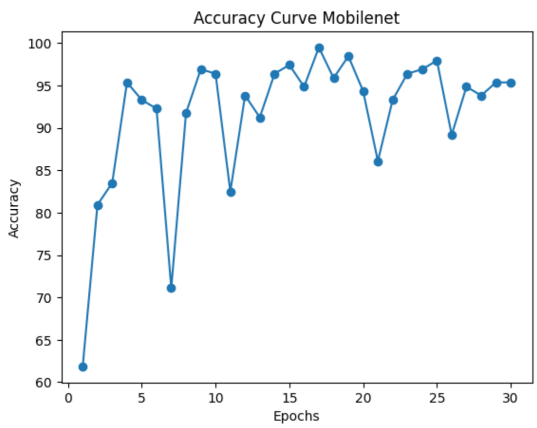

Figure 15: MobileNet, Training Accuracy Curve

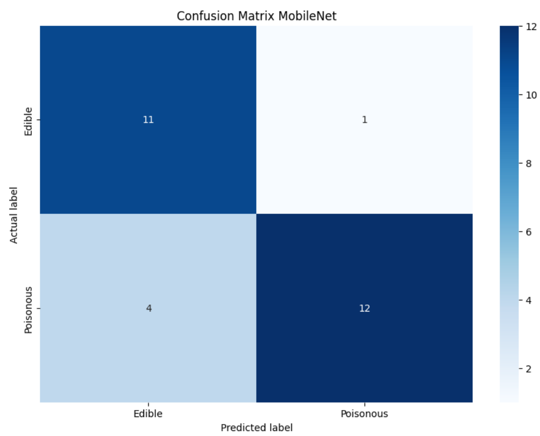

Figure 16: MobileNet, Confusion Matrix

The **MobileNetV2** model had achieved the highest accuracy of **82.14**%, with a training accuracy of **95.36**% which can be seen in figure 15, making it the best performing model in this research. Its lightweight architecture combined with its ability to extract the features efficiently contributed to its superior performance. From the confusion matrix, figure 16, it can be seen that this model has minimal misclassification, and the model demonstrated exceptional generalization capabilities. These results emphasize the practicality of this model for real-world applications, especially in resource-constrained environments where computational efficiency is crucial.

7 Conclusion and Future Work

The purpose of this study was to compare machine learning and deep learning models in the classification of edible and poisonous mushrooms. This study aimed at assessing the performance of different models such as Random Forest, MobileNetv2, ResNet50, ResNet101, and VGG16 and also customized CNN models and the study was able to fulfil the research objectives by presenting statistically viable results on their efficiency.

This study also found that the deep learning models showed the best performances compared to other machine learning models, such as Random Forest to be 75% with the accuracies of 82.14%, for MobileNet V2 and ResNet50, respectively. The MobileNetV2 because of its lightweight accuracy and strong generalization, makes it the most suitable one that should be used in real-world applications. Outcomers such as Random Forest as well as previous architectures like VGG16 perform worse, suggesting that advanced deep learning is useful. This calls for the use of enhanced deep learning approaches and transfer learning such as shown in the research.

The consequences can be measured as the difference between a deadly outcome of consuming toxic mushrooms or a successful identification of edible mushrooms. Through this work, it positions itself for the future application of deep learning models into usable products such as smartphone/mobile applications, or physical computer systems to aid foragers and researchers.

However, the study had its weaknesses which include having a small number of images and using images that have been manually cropped meaning they do not represent close to real environments. The same can be done in future work and concerns the collection of larger, diverse datasets and the use of more sophisticated data augmentation methods. Expanding upon the fine-tuning of VL models, as well as introducing the integration of image data with chemical or environmental metadata are also areas of enhancement.

The study also shows how the knowledge can be commercialized in the form of real time identification tools for mushroom foragers, scientists as well as farmers. In this regard, by employing deep learning methods, this study enables more accurate and safe mushroom distinction and defines promising research and innovation avenues.

References

1160 Mushroom Identifications Await Your Discovery — ultimate-mushroom.com (n.d.). https://ultimate-mushroom.com. [Accessed 01-12-2024].

Albertó, E. O., Petersen, R. H., Hughes, K. W. and Lechner, B. (2002). Miscellaneous notes on pleurotus, *Persoonia - Molecular Phylogeny and Evolution of Fungi* **18**(1): 55–69.

Areo, G. (2024). Mushroom foraging in woodland areas: A fusion of recreation and sustainable tourism.

Bongat, K. P. L., Garcia, L. A. P. and Caya, M. V. C. (2023). Fine-grained classification of edible mushrooms using convolutional neural network, *2023 IEEE 5th Eurasia Conference on IOT, Communication and Engineering (ECICE)*, pp. 537–542.

Fukuwatari, T., Sugimoto, E., Yokoyama, K. and Shibata, K. (2001). [establishment of animal model for elucidating the mechanism of intoxication by the poisonous mushroom clitocybe acromelalga], *Shokuhin eiseigaku zasshi. Journal of the Food Hygienic Society of Japan* **42**(3): 185—189.
URL: *https://www.jstage.jst.go.jp/article/shokueishi/42/3/42$_{31}$85/$_p$df*

Gautam, A. K., Verma, R. K., Avasthi, S., Sushma, Bohra, Y., Devadatha, B., Niranjan, M. and Suwannarach, N. (2022). Current insight into traditional and modern methods in fungal diversity estimates, *Journal of Fungi* **8**(3).
URL: *https://www.mdpi.com/2309-608X/8/3/226*

GIMP — gimp.org (n.d.). `https://www.gimp.org`. [Accessed 05-12-2024].

He, K., Zhang, X., Ren, S. and Sun, J. (2015). Deep residual learning for image recognition.
URL: *https://arxiv.org/abs/1512.03385*

He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J. and Li, M. (2019). Bag of tricks for image classification with convolutional neural networks, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 558–567.

Ho, T. K. (1995). Random decision forests, *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Vol. 1, pp. 278–282 vol.1.

Kaaronen, R. O. (2020). Mycological rationality: Heuristics, perception and decision-making in mushroom foraging, *Judgment and Decision Making* **15**(5): 630–647.

Kalač, P. (2012). A review of chemical composition and nutritional value of wild-growing and cultivated mushrooms, *Journal of the Science of Food and Agriculture* **93**(2): 209–218.

Liao, W., Xiao, J., Zhao, C., Han, Y., Geng, Z., Wang, J. and Yang, Y. (2022). Mushroom image recognition and distance generation based on attention-mechanism model and genetic information.
URL: *https://arxiv.org/abs/2206.13383*

Lücking, R., Catherine, A. M., Robbertse, B., Miller, A. N., Ariyawansa, H. A., Aoki, T., Cardinali, G., Crous, P. W., Druzhinina, I. S., Geiser, D. M., Hawksworth, D. L., Hyde, K. D., Irinyi, L., Jeewon, R., Johnston, P. R., Kirk, P. M., Malosso, E., May, T. W., Meyer, W. and Öpik, M. (2020). Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal dna barcoding?, *IMA Fungus* **11**(1): 14.
URL: *https://doi.org/10.1186/s43008020000033z*

Mascarenhas, S. and Agarwal, M. (2021). A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification, *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, Vol. 1, pp. 96–99.

Mushroom Observer — mushroomobserver.org (n.d.). `https://mushroomobserver.org`. [Accessed 01-12-2024].

Peng, Y., Xu, Y., Shi, J. and Jiang, S. (2023). Wild mushroom classification based on improved mobilevit deep learning, *Applied Sciences* **13**(8).
URL: *https://www.mdpi.com/2076-3417/13/8/4680*

Procházka, P., Soukupová, J., Tomšík, K., Mullen, K. J. and Čábelková, I. (2023). Climatic factors affecting wild mushroom foraging in central europe, *Forests* **14**(2).
URL: *https://www.mdpi.com/1999-4907/14/2/382*

Rubel, W. and Arora, D. (2008). A study of cultural bias in field guide determinations of mushroom edibility using the iconic mushroom, amanita muscaria, as an example, *Economic Botany* **62**(3): 223–243.
URL: *https://doi.org/10.1007/s1223100890409*

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520.

Schenk-Jaeger, K. M., Rauber-Lüthy, C., Bodmer, M., Kupferschmidt, H., Kullak-Ublick, G. A. and Ceschi, A. (2012). Mushroom poisoning: A study on circumstances of exposure and patterns of toxicity, *European Journal of Internal Medicine* **23**(4): e85–e91.
URL: *https://www.sciencedirect.com/science/article/pii/S0953620512000830*

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.
URL: *https://arxiv.org/abs/1409.1556*

Stojnev, D. and Ilic, A. (2020). Pre-processing image data for deep learning, *International Scientific Conference On Information Technology And Data Related Research*.

Tarawneh, O., Tarawneh, M., Sharrab, Y. and Husni, M. (2023). Mushroom classification using machine-learning techniques, *AIP Conference Proceedings* **2979**(1): 030003.
URL: *https://doi.org/10.1063/5.0174721*

Wang, X.-M., Zhang, J., Wu, L.-H., Zhao, Y.-L., Li, T., Li, J.-Q., Wang, Y.-Z. and Liu, H.-G. (2014). A mini-review of chemical composition and nutritional value of edible wild-grown mushroom from china, *Food Chemistry* **151**: 279–285.
URL: *https://www.sciencedirect.com/science/article/pii/S0308814613016968*

Xiao, J., Liao, W., Zhang, M., Wang, J., Wang, J. and Yang, Y. (2022). Taxonomy and evolution predicting using deep learning in images.
URL: *https://arxiv.org/abs/2206.14011*

Y, R. K. and Chandrasekhar, D. V. (2020). Machine learning methods to classify mushrooms for edibility-a review, *International Journal for Modern Trends in Science and Technology* **06**(09): 54–58.
URL: *https://www.ijmtst.com/volume6/issue09/9.IJMTST0609023.pdf*

Yan, Z., Liu, H., Li, J. and Wang, Y. (2021). Application of identification and evaluation techniques for edible mushrooms: A review, *Critical reviews in analytical chemistry* **53**: 1–21.

Ye, Y. and Liu, Z. (2018). Management of amanita phalloides poisoning: A literature review and update, *Journal of Critical Care* **46**: 17–22.
URL: *https://www.sciencedirect.com/science/article/pii/S0883944118301862*

Yu, T. and Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications.
URL: *https://arxiv.org/abs/2003.05689*

Zhang, B., Zhao, Y. and Li, Z. (2022). Using deep convolutional neural networks to classify poisonous and edible mushrooms found in china.
URL: *https://arxiv.org/abs/2210.10351*