

# DEEP LEARNING APPROACH TO MUSHROOM SPECIES CLASSIFICATION

**Yanni Alan Alevras**

Student# 1009330706

yanni.alevras@mail.utoronto.ca

**Nicholas Biancolin**

Student# 1009197726

n.biancolin@mail.utoronto.ca

**Eric Liu**

Student# 1009098450

ey.liu@mail.utoronto.ca

**Jason Ruixuan Zhang**

Student# 1008997631

jasonrx.zhang@mail.utoronto.ca

## 1 INTRODUCTION

Fungi identification is an increasingly critical task, with implications in food security, industrial use, conservation efforts, and biosafety. However, visual and image classification of fungi is a difficult task due to the wide variety of species (Lücking et al., 2020).

Nearly 87% of Ontario’s land is Crown Land, filled with expansive forests and a diverse array of plant species. Among these, mushrooms stand out as both common and challenging to identify, due to their wide variety and subtle differences between species. Thus, identification is particularly important because some species are edible and others are highly poisonous, which could pose a risk to foragers and nature enthusiasts.

Our project developed a deep learning model based on a convolutional architecture to accurately identify macrofungi (fungi with large bodies) based on their genus. Deep learning formed a suitable choice for this task, as it constitutes a powerful and accurate method for image recognition and classification tasks, including in ecological settings (Schneider et al., 2019).

We used the MIND.Funga dataset described in Drechsler-Santos et al. (2023), which has approximately 17 000 images of more than 500 species of fungi spanning 216 genera. This dataset is well-suited for our project, as it is built primarily for use in deep classification models. Images are also curated to be of a high quality and are labelled by species.

Our code can be found at this GitHub repository. It contains Jupyter notebooks with pre-processing, baseline, main model, and evaluation code.

## 2 BACKGROUND & RELATED WORK

## 3 DATA PROCESSING

The original dataset contains images of 509 classes. Many species have a small number of images associated with them, insufficient for training a deep learning model. To mitigate this, we grouped images into larger “buckets” based on their genus to reduce the number of classes to 216 classes, which increases the number of images per class. Genera form an ideal way to group images together, as species of the same genus share physical characteristics and the same root name (first word in the species name) (Hollister et al., 2023).

Once this combination is done, we take the top 10 genera with the most images to form an intermediate dataset of 6131 images. We then randomly split the dataset into training, validation, and testing sets with a 75%, 15%, and 10% split, respectively. This split ensures the model has enough data to train on, while ensuring validation and testing better reflects its performance.

Finally, we apply data augmentation techniques described by Shorten & Khoshgoftaar (2019) solely to the training set, to avoid poisoning the validation and testing sets. An example of this applied to one image is described in Figure 1. We applied flips, rotations (90, 180, 270 degrees), Gaussian noise, and random erasing (of black rectangular regions). These manipulations were chosen as they would not alter the image’s quality or classification. As mushrooms are identifiable based on visual traits, preserving key details is necessary to distinguish between genera. As our model is agnostic to colour, colour space transformations were not applied.

Our final dataset contains 16 962 training images, 3379 validation images, and 2324 testing images. To generate a usable set, we save a comma-separated value (CSV) file for each set, containing the image path and corresponding label. We also serialize the dataset into a binary Pickle object.

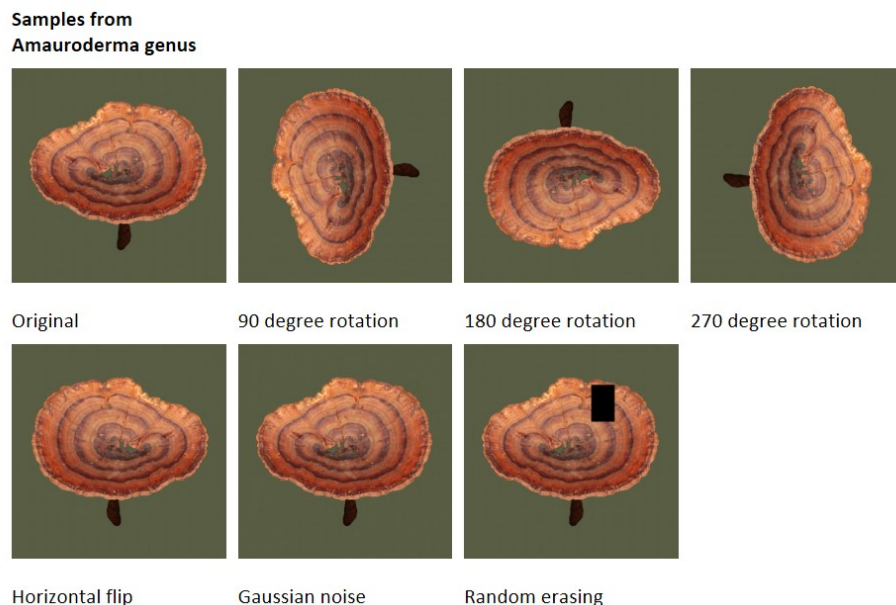


Figure 1: Data augmentation techniques applied to a sample image

## 4 ARCHITECTURE

Our primary model relies on two distinct sections: a non-tunable transfer learning section, and a tunable convolutional and fully-connected section. Our data pre-processing and model training pipeline is described in Figure 2.

We also use cross entropy loss, as it is a standard loss function for multi-class classification tasks. We also use a stochastic gradient optimizer, with a momentum of 0.8, weight decay of 0.001, and learning rate of 0.001. Our training flow uses 100 epochs with a batch size of 64.

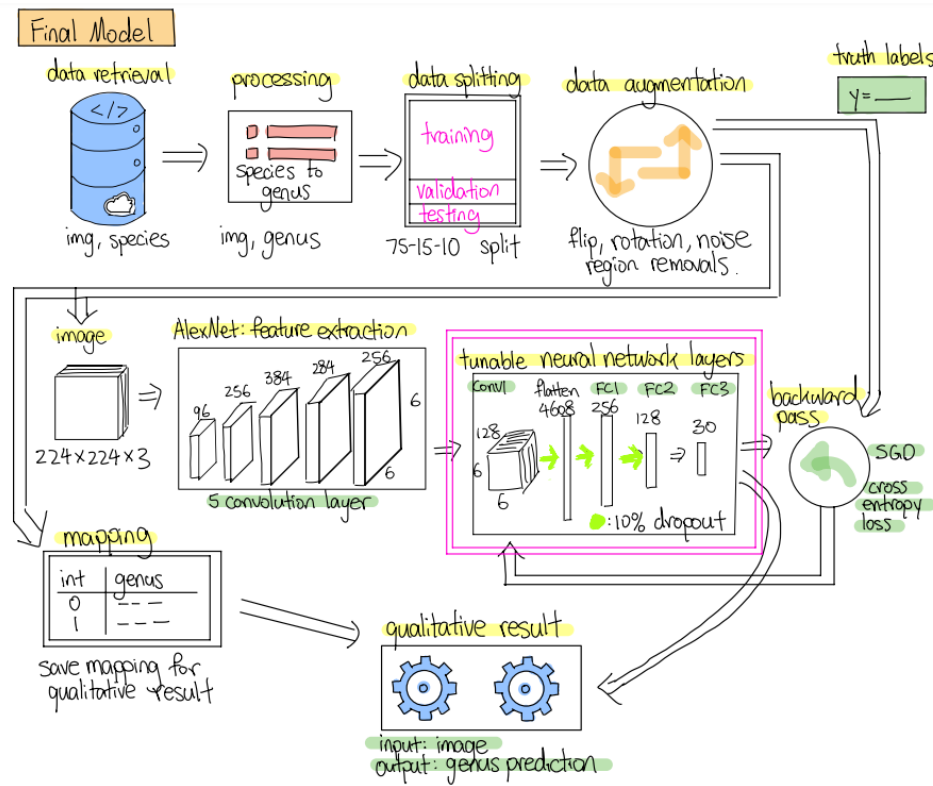


Figure 2: Primary model processing and training pipeline

#### 4.1 NON-TUNABLE SECTION

We use AlexNet for high-level feature extraction, sourced from PyTorch’s pre-trained models. AlexNet is a well-known convolutional architecture for image classification tasks (Krizhevsky et al., 2012). Since the model needs to differentiate between mushrooms with similar appearances, AlexNet excels in extracting the fine features that set them apart.

Its architecture is described in Figure 3. It processes a  $3 \times 224 \times 224$  input image, and the feature extraction outputs a  $256 \times 6 \times 6$  tensor. There are five convolutional layers and three pooling layers (Krizhevsky et al., 2012).

### AlexNet Architecture

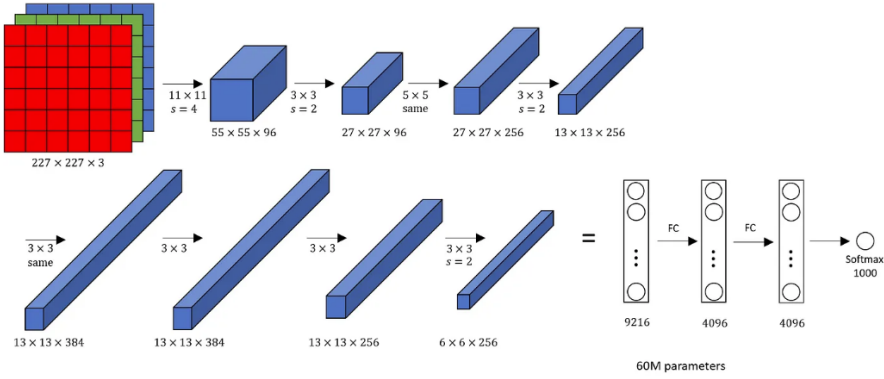


Figure 3: AlexNet architecture

#### 4.2 TUNABLE SECTION

To make the model specific to our team’s project, we add an additional convolutional layer, which outputs a  $128 \times 6 \times 6$  tensor. After this layer, the output is flattened and passed through three fully-connected layers with ReLU activation functions. The fully-connected layers have 256, 128, and 10 neurons, respectively, which matches the number of output classes in our dataset.

In total, there are  $5 + 3$  layers in the non-tunable section, and  $1 + 3$  layers in the tunable section, making our model architecture 12 layers in total.

### 5 BASELINE MODEL

Our baseline model is a random forests classifier trained with scikit-learn on the same dataset as our deep learning model. Decision tree classifiers, like random forests, are generally well-suited for multiclass problems (Gall et al., 2012). We follow similar processing steps as the primary model, including a train-test split of 75%-25%, feature extraction, training, and evaluation. We do not apply data augmentation, as this benefits CNNs more.

We apply feature extraction on grayscale images with histogram of oriented gradients (HOG), an edge extraction descriptor algorithm (N. Dalal & B. Triggs, 2005-06-20/2005-06-25) and computed with scikit-image. As random forests does not automatically perform feature extraction (compared to deep convolutional networks), this step is done to provide the model with a more effective set of features to learn from. We compute HOG features on both the training and testing set with default parameters, in line with the process outlined by Dutta (2024). We also use default hyperparameters for training, with the exception of the number of estimators, set to 500, which suggests is an optimal number of trees for increasing accuracy.

Our final baseline model achieves an accuracy of 61%. We achieve an average precision of 0.7, recall of 0.51, and F1 score of 0.54.

### 6 QUANTITATIVE RESULTS



Figure 4: Train vs validation loss: final model

The team achieved a testing accuracy of 81.4% with our final model. More than a 16% increase from the progress report.

During model training, the group prioritized validation loss over training loss, as validation loss serves a more unbiased measure of the model's performance, bringing better generalization. After training with a unique set of hyperparameters, we select epochs with low validation loss, saving the models then run the testing set to determine a testing accuracy. Finally, the group selected the model with the highest testing accuracy.

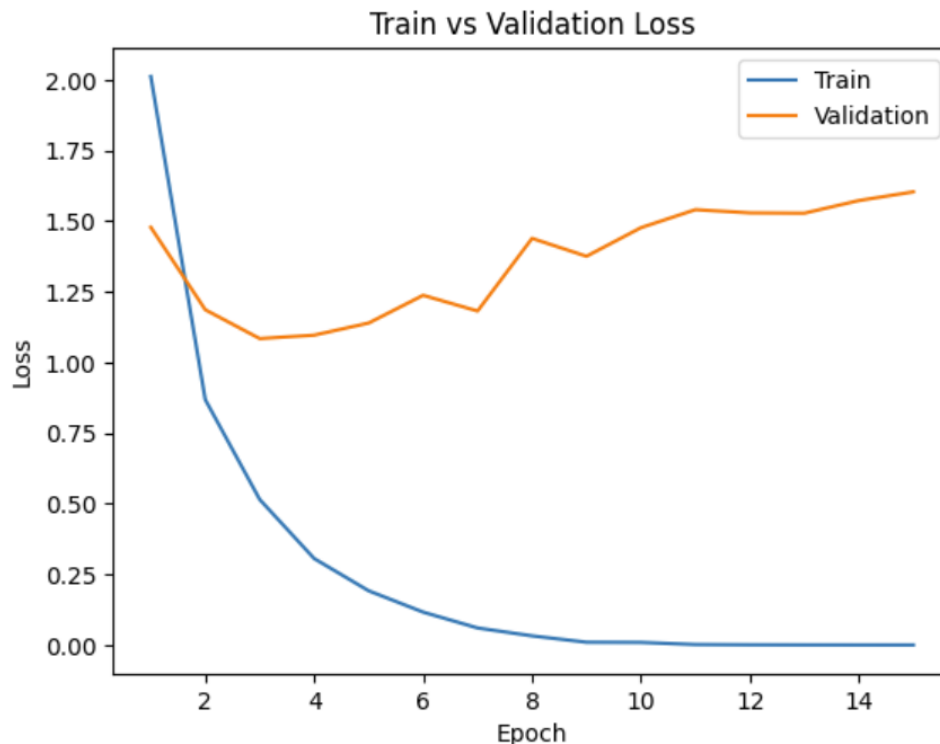


Figure 5: Train vs validation loss: progress report

From the above graph, the validation curve does not overfit immediately as the one seen in the progress report. This improvement is due data augmentation, weight decay, dropout, and hyperparameter tuning discussed above and in previous sections.

Our final hyperparameters were:

- Learning rate: 0.0006
- Number of epochs: 101, chosen epoch for model: 90
- Batch size: 64
- Weight decay: 0.003

## 7 QUALITATIVE RESULTS

### 8 EVALUATION ON NEW DATA

We also evaluated our model on data collected from outside the MIND.Funga dataset. We collected 26 images of our dataset’s genera, with about 2-3 images per genus. This was collected from Wikimedia Commons, an online repository of permissively-licensed images. We took special care to ensure that the images were not included in the original dataset. As MIND.Funga is primarily made up of images from the public and academics, we are confident that no images from Wikimedia Commons were included in the original dataset (Drechsler-Santos et al., 2023).

Notably, our baseline model performs significantly worse than our deep learning model, with correct predictions on only 8% of new images. It overpredicts the genus *Hygrocybe*, the second most frequent genus in the dataset, on 61% of images. This may be a result of the baseline model’s lack of robustness with non-curated data, and largely suggests the rigidity of non-deep learning models without excessive hyperparameter tuning in handling image classification tasks.

## 9 DISCUSSION

## 10 ETHICAL CONSIDERATIONS

Many machine learning models rely on the “unauthorized uses of copyrighted” data for training purposes. This broadly results in infringement of the intellectual property rights and personal interests of the data owners (Sobel, 2021). The dataset used in this project, MIND.Funga, is publicly licensed under a permissive copyleft license (CC BY NC 3.0) (Drechsler-Santos et al., 2023). Images were sourced from contributions from the public and academics. However, the organizing project does not disclose that public contributions will be released under this license (Drechsler-Santos et al., 2023), suggesting that the dataset may contain images that were not intended to be shared under this license as a result of inadequate informed consent.

The dataset also mostly contains images of tropical fungi from Brazil, with a small subset (approximately 500 out of 14 000) of the globally recorded number of fungi species (Lücking et al., 2020). This results in a model that is primarily representative of Brazilian fungi species. As some fungi species can be poisonous, any practical real-world applications of our model may result in misclassifications, which may result in harm to users that rely on the model. This is a significant ethical concern, as the model may be used by individuals who are not experts in mycology, and may not be able to identify the potential risks of consuming certain fungi species.

## REFERENCES

- Elisandro Ricardo Drechsler-Santos, Fernanda Karstedt, Thiago Zimmermann Loureiro Chaves, Mahatmā Titton, Kelmer Martins da Cunha, Eloisa Leopoldo, Genivaldo Alves-Silva, Diogo Rezende, Thiago Kossmann, Emerson Luiz Gumboski, and Aldo von Wangenheim. MIND.Funga, July 2023.
- Srijan Dutta. Random Forest for Image Classification Using OpenCV. *GeekforGeeks*, 2024.
- Juergen Gall, Nima Razavi, and Luc Van Gool. An Introduction to Random Forests for Multi-class Object Detection. In Frank Dellaert, Jan-Michael Frahm, Marc Pollefeys, Laura Leal-Taixé, and Bodo Rosenhahn (eds.), *Outdoor and Large-Scale Real-World Scene Analysis*, pp. 243–263, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-34091-8.
- Jack D. Hollister, Xiaohao Cai, Tammy Horton, Benjamin W. Price, Karolina M. Zarzeczny, and Phillip B. Fenberg. Using computer vision to identify limpets from their shells: A case study using four species from the Baja California peninsula. *Frontiers in Marine Science*, 10, 2023. ISSN 2296-7745.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- Robert Lücking, M. Catherine Aime, Barbara Robbertse, Andrew N. Miller, Hiran A. Ariyawansa, Takayuki Aoki, Gianluigi Cardinali, Pedro W. Crous, Irina S. Druzhinina, David M. Geiser, David L. Hawksworth, Kevin D. Hyde, Laszlo Irinyi, Rajesh Jeewon, Peter R. Johnston, Paul M. Kirk, Elaine Malosso, Tom W. May, Wieland Meyer, Maarja Öpik, Vincent Robert, Marc Stadler, Marco Thines, Duong Vu, Andrey M. Yurkov, Ning Zhang, and Conrad L. Schoch. Unambiguous identification of fungi: Where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus*, 11(1):14, July 2020. ISSN 2210-6359. doi: 10.1186/s43008-020-00033-z.
- N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, volume 1, pp. 886–893 vol. 1, 2005-06-20/2005-06-25. ISBN 1063-6919. doi: 10.1109/CVPR.2005.177.
- Stefan Schneider, Graham W. Taylor, Stefan Linquist, and Stefan C. Kremer. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, 10(4):461–470, April 2019. ISSN 2041-210X. doi: 10.1111/2041-210X.13133.

Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1):60, July 2019. ISSN 2196-1115. doi: 10.1186/s40537-019-0197-0.

Benjamin Sobel. Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning. In Jyh-An Lee, Reto Hilty, and Kung-Chung Liu (eds.), *Artificial Intelligence and Intellectual Property*, pp. 0. Oxford University Press, February 2021. ISBN 978-0-19-887094-4. doi: 10.1093/oso/9780198870944.003.0011.