

Poison-Aware Open-Set Fungi Classification: Reducing the Risk of Poisonous Confusion

Stefan Wolf^{1,2,3}, Philipp Thelen^{2,1,3} and Jürgen Beyerer^{2,1,3}

¹*Vision and Fusion Lab, Karlsruhe Institute of Technology KIT, Vincenz-Priessnitz-Straße 3, 76131 Karlsruhe, Germany*

²*Fraunhofer IOSB, Institute of Optronics, System Technologies and Image Exploitation, Fraunhoferstrasse 1, 76131 Karlsruhe, Germany*

³*Fraunhofer Center for Machine Learning*

Abstract

The FungiCLEF 2024 challenge aims to foster research in the field of application-oriented fine-grained open-set classification. Particularly, it sets the challenge to optimize fungi species classification while recognizing unknown species with the evaluation of multiple metrics targeting the problems of actual use-cases, e.g., the risk of a highly detrimental confusion of a poisonous species for an edible species. To develop a well-performing approach, we focus on reducing this particular risk by introducing multiple improvements. The major improvements are a poisonous reranking which prevents predicting an edible species while a significant chance of the sample being poisonous exists and a genus loss which provides additional training information improving the regularization of the feature space. The advancements provide a large improvement in terms of poisonous confusion but also in terms of overall classification accuracy. With this approach, we achieved the 1st place in the challenge's main metric. Code is available at <https://huggingface.co/stefanwolf/fungi2024>.

Keywords

Fungi classification, Open-set classification, FungiCLEF, Entropy

1. Introduction

Fine-grained open-set classification is an important topic in the biology context in order to find samples of rare species and to provide inexperienced citizen scientist with a support to identify species of plants and animals. Particularly, fungi species classification has an additional use case: identifying poisonous species to reduce the risk of accidental eating poisonous fungi. Thus, the FungiCLEF 2024 challenge [1], part of the LifeCLEF 2024 lab [2], sets up the task of open-set fungi species classification with an emphasis on correctly identifying fungi species while preventing confusing poisonous species for edible species.

While the task of open-set fungi classification has been intensively investigated in the recent iterations of the FungiCLEF challenge [3, 4], optimizing the poisonous confusion error has only been lightly approached rendering the room for improvements quite large. Thus, we focus on the poisonous confusion error and can achieve significant gains with multiple advancements. These advancements are

- a poisonous reranking which predicts the highest ranking poisonous species if its confidence is only lower by a certain factor than the overall highest ranking species.
- a genus loss that regularizes the feature space by incorporating the genus label in training.
- a second open-set threshold to reduce the risk of misclassifying a poisonous sample as an unknown species.
- a two-stage metadata integration that enhances the overall classification accuracy.

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

✉ stefan.wolf@kit.edu (S. Wolf); philipp.thelen@iosb.fraunhofer.de (P. Thelen); juergen.beyerer@iosb.fraunhofer.de

(J. Beyerer)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

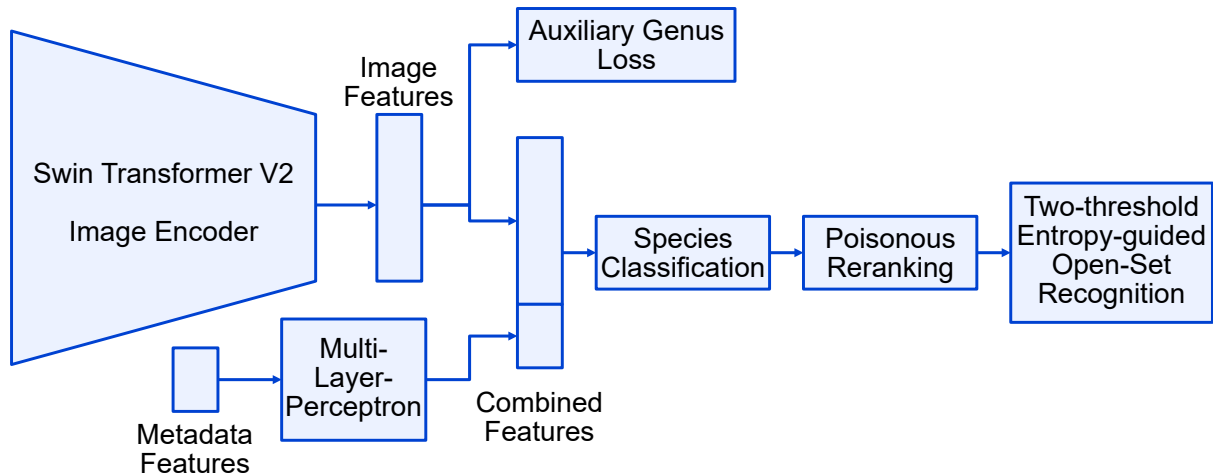


Figure 1: Overview of our approach. We employ a Swin Transformer V2 Base backbone to extract image features. The image features are passed to an auxiliary genus loss during training. The metadata features are processed by a multi-layer-perceptron and thereafter, they are combined with the image features and fed into the species classifier. Afterwards, our poisonous reranking and the two-threshold entropy-guided open-set recognition is applied.

2. Related work

A wide range of approaches have been proposed targeting fine-grained fungi classification in wildlife images. Sulc et. [5] employ an ensemble of CNNs to classify images of fungi. Picek et al. [6] propose a simple but effective probabilistic strategy to exploit metadata in order to improve the accuracy of fine-grained fungi classification. Kiss and Czùni [7] provide a study about a broad-range of design choices optimizing mushroom type classification accuracy. The 2022 [3] and the 2023 [4] iterations of the FungiCLEF challenge summarize a variety of approaches with the 2022 iteration being focused on improving open-set fungi classification and the 2023 iteration emphasizing the importance of choosing metrics based on use cases to focus research on relevant aspects, e.g., focusing on reducing the confusion of poisonous species with edible species.

3. FungiCLEF 2024 challenge

The 2024 iteration of the FungiCLEF challenge [1] aims to stimulate research about efficient open-set fungi classification. The target is to distinguish 1,604 fungi species by using an observation consisting of one or multiple images with different perspectives and additional metadata information such as habitat, substrate, time and location. Apart from distinguishing the known species, the submitted approach needs to be able to solve an open-set scenario, i.e., it needs to recognize whether a sample is of a species not part of the training samples. The provided training data consists of the Danish Fungi 2020 dataset [6], the validation data consists of the test set of FungiCLEF’s 2022 iteration [3] and the test data consists of new data for the 2024 iteration. The evaluation is done based on three metrics:

- Track 1: Classification error – standard classification with "unknown" category.
- Track 2: Poisonous confusion error – cost for confusing edible species for poisonous and vice versa (with $100\times$ weight for confusing edible species for poisonous).
- Track 3: User-focused error – user-focused loss composed of both the classification error and the poisonous/edible confusion.

4. Approach

Our challenge submission is based on the approach by Wolf and Beyerer [8]. To simplify the training and due to insignificant impact, we refrain from applying the resampling-based class balancing. We extend the approach by applying several improvements as described in this section, i.e., a two-stage integration of metadata, a genus loss, a poisonous reranking and a two-threshold open-set recognition strategy. The overall architecture of our approach is shown in Figure 1.

4.1. Model architecture

We employ a Swin Transformer V2 Base [9] as an image feature extractor backbone. Additionally, we use the metadata information provided by the Danish Fungi dataset [6] to improve the classification accuracy. The metadata is encoded similarly to the approach by Ren et al. [10] We encode the month m and day d of each observation as a vector $(\sin(\frac{2\pi m}{12}), \cos(\frac{2\pi m}{12}), \sin(\frac{2\pi d}{31}), \cos(\frac{2\pi d}{31}))^\top$. The geographical locations country code, substrate and habitat are encoded as one-hot vectors. All metadata vectors are concatenated and fed into two fully connected layers with an output size of 64 and each being followed by a ReLU activation and a layer norm [11]. The resulting metadata feature vector and the image feature vector are concatenated and fed into a final linear classification layer followed by a softmax activation. We apply an auxiliary second classification head for predicting the genus of a sample during training which is fed the image features.

4.2. Training process

For training the model, we apply two losses: a classification loss on the species level as commonly used and an auxiliary classification loss on the genus level. Both are label-smooth losses [12] with a smoothing value of 0.9. To prevent a degradation of the image features when training with metadata, we use a two-stage training with the first stage only training the image classification stream of the network and the second stage training the complete network including the metadata feature extractor with a newly initialized species classifier.

4.3. Inference

During inference, we extract the feature vectors of all images of an observation and concatenate the observation-wise mean of the image features with the metadata features before feeding the result into the species classification head. Based on the resulting softmaxed confidence scores, we apply our poisonous reranking which reranks the poisonous species with the highest confidence to the top of the species ranking if its confidence is higher than the actual top-1 species' confidence divided by a poisonous reranking factor α . After the poisonous reranking, we apply the entropy-based open-set thresholding based on the approach by Ren et al. [10] If the entropy of the output confidences is above a certain threshold τ , we predict the observation to be out-of-distribution. We extend this approach by employing two thresholds τ_e and τ_p , which are applied if the predicted species is edible or poisonous, respectively. The threshold τ_p is selected higher than the threshold τ_e to reduce the risk of misclassifying a poisonous species as an out-of-distribution species. Both, the poisonous reranking and the second threshold, are improvements targeting the challenge's Track 2 metric which is measuring the poisonous confusion with a significantly higher weight for mispredicting poisonous species than edible species.

5. Evaluation

5.1. Datasets

We use the official FungiCLEF 2024 [1] datasets. The Danish Fungi 2020 [6] dataset is used for training. All metrics reported in this study are based on the official validation set which is the test set of FungiCLEF's 2022 iteration [3]. The test set for the official ranking is a set of images which has not

Table 1

Impact of the auxiliary genus loss. Our auxiliary genus loss shows a small improvement for the classic classification metrics F1 score and Track 1 (classification error). However, particularly the poison-focused Track 2 metric is drastically improved due to better feature separation on the genus level which is important for distinguishing poisonous from edible species. Note: the images of each observation are combined by a post-classification mean fusion.

Genus loss	F1 score	Track 1	Track 2	Track 3
No	49.1	0.413	0.220	0.633
Yes	49.7	0.410	0.181	0.591

been disclosed publicly prior to the end of the challenge. Only the results of the public part of the test set was publicly visible with the results on the private part only being disclosed after the challenge deadline.

5.2. Implementation and baseline

We use the MMPreTrain [13] classification framework based on PyTorch [14] for the training and the inference of the models. All models are pre-trained on the ImageNet-21k dataset [15] and trained for 24 epochs with an AdamW optimizer [16], a base learning rate of $6.25 \cdot 10^{-5}$, a learning rate warm-up for 2100 iterations and a cosine learning rate decay. We train with a total batch size of 128. The metadata training is performed for two epochs in a second stage with a frozen image encoder. Our image pre-processing pipeline for training includes a random crop of an image area between 8% and 100%, a resize to 384×384 pixels, a random horizontal image flip, RandAugment [17] and random erasing [18]. We use 8 Nvidia A100 GPUs for training. The pre-processing pipeline for the inference includes an image scaling with 438 pixels output size on the shorter edge and a center crop of size 384×384 pixels.

5.3. Genus loss

We compare a training with and without our genus loss in Table 1. All metrics are improved by the application of the genus loss. Particularly, the Track 2 error focused on identifying poisonous samples as such improves significantly with a drop of 0.22 to 0.18. Nonetheless, also the classification-focused F1 score and Track 1 error show an improvement. The strong increase in terms of identifying poisonous species is likely due to most genus containing only edible species. Thus, considering the genus level in training the feature space results in a denser feature representation of these poisonous-wise uniform genus. Therefore, the chance of misclassification of a species of a uniformly edible genus with a non-edible species is heavily reduced. The improvements in terms of classification accuracy are probably induced by species with a low number of samples. The risk of misclassifying them with species from other genus due to a lack of variance in the data is reduced when also training the feature space on the genus level.

5.4. Poisonous reranking

We apply the poisonous reranking additional to the genus loss and evaluate it with different factors. It reranks poisonous species to the top-1 if their confidence is higher than the actual top-1’s confidence divided by an $\alpha > 1$ in order to prevent costly confusions of poisonous fungi with edible fungi. The results are shown in Table 2. As expected, it drastically reduces the Track 2 error from 0.181 to 0.076 due to a lower number of samples misclassified as an edible species. While the Track 1 error is slightly increased from 0.41 to 0.416 due to some so far correctly classified samples from edible species now being misclassified as a poisonous species, the increase is small compared to the drop in terms of Track 2 error resulting in a significant drop in the overall Track 3 error from 0.591 to 0.492. Out of the evaluated values of α , the Track 2 and 3 errors are reduced until an α of 10 while a value of 20 is leading to

Table 2

Impact of the poisonous reranking with different values of α . The results show the positive impact of reducing the risk of misclassifying a poisonous species for an edible species with significantly lowered Track 2 and 3 errors outweighing the small increase in terms of Track 1 error due to an overall increased species misclassification. Note: the images of each observation are combined by a post-classification mean fusion.

Poisonous reranking factor α	F1 score	Track 1	Track 2	Track 3
-	49.7	0.410	0.181	0.591
2	49.7	0.411	0.144	0.555
5	49.3	0.413	0.106	0.518
10	49.1	0.416	0.076	0.492
20	48.8	0.421	0.079	0.500

Table 3

Impact of the two-threshold entropy-based open-set recognition. For the softmax-score-based recognition, we report results for thresholds optimized separately for each metric while we report results for a pair of thresholds for the entropy-based strategy which is manually optimized for Track 3. Thus, the softmax-score-based approach shows the best results for all metrics but the combined Track 3. For the Track 3 metric, the entropy-based approach outperforms the softmax score.

Open-set recognition	F1 score	Track 1	Track 2	Track 3
-	49.9	0.412	0.082	0.494
Softmax score	52.1	0.364	0.070	0.469
Entropy-based with two thresholds	50.3	0.376	0.074	0.449

Table 4

Impact of the metadata integration. Exploiting the metadata gives a significant improvement for all metrics.

Metadata	F1 score	Track 1	Track 2	Track 3
No	50.3	0.378	0.068	0.445
Yes	56.4	0.329	0.051	0.380

an increase in all error metrics. Particularly, also the Track 2 error increases showing that even the comparatively lowly weighted case of mispredicting an edible species for a poisonous species now playing a significant role.

5.5. Open-set recognition

We employ the entropy-based open-set recognition by Ren et al. [10] and extend it by a second threshold for poisonous species. If the predicted species is poisonous, a higher entropy is needed to classify the sample as out-of-distribution since out-of-distribution samples are considered edible by the Track 2 metric and thus, misclassifying a poisonous sample as out-of-distribution increases the Track 2 error heavily. We compare it to applying no open-set recognition and applying a simple softmax-based thresholding. The results including genus loss and poisonous reranking as baseline are shown in Table 3 and indicate an improvement in all metrics for both open-set recognition methods. The entropy-based approach with two thresholds provide an additional improvement over the softmax-based thresholding.

5.6. Metadata

We integrate metadata information in the inference process by feeding the encoded metadata through two fully connected layers and concatenating the resulting vector to the feature vector of the image encoder before the final linear classification layer. The impact of this metadata exploitation strategy is shown in Table 4 including all previously mentioned improvements. The results show a significant improvement across all evaluated metrics.

Table 5

Overview of the top-5 challenge submissions. The best performing submission according to the challenge’s main metric Track 3 is selected. Performance is measured on the private test set.

Team	F1 score	Track 1	Track 2	Track 3
IES	54.3	0.311	0.090	0.401
jack-etheredge	54.9	0.244	0.163	0.407
upupup	53.6	0.390	0.072	0.462
chirmy	51.8	0.269	0.415	0.684
TingTing1999	51.4	0.275	0.438	0.713

5.7. Final model

The final best-performing model includes all proposed improvements with the following adjustments:

1. a poisonous reranking factor α of 13, the overall best performing value on the public test set.
2. an open-set entropy thresholds of $\tau_e = 2.5$ for edible species and $\tau_p = 7$ for poisonous species.
3. including the validation set in training with unknown samples being assigned a vector with each element having the same value as target label similar to the approach by Ren et al. [10] and the true genus label for the genus loss.

5.8. Challenge results

The final private test set results of the top-5 challenge participants are shown in Table 5. We ranked first with the lowest error in the main metric Track 3 due to a high emphasis on optimizing Track 2 while not lacking too far behind in terms of Track 1. Particularly, we achieve the first place with an efficient solution that consists of only a single model. While the runner-up team achieved a better Track 1 error, the Track 2 error is almost twice as high. In contrast, the third-placed team achieved an even better Track 2 metric than our approach. However, this achievement comes at a large Track 1 error outweighing the advantage.

6. Conclusion

In this study, we described our top-ranking approach for the FungiCLEF 2024 challenge. With a high emphasis on reducing the risk of confusing poisonous species for edible species, we propose several advancements which improve the respective error drastically while also improving the overall classification accuracy. Particularly, we introduced a poisonous reranking, a genus loss, two-threshold open-set recognition and an efficient two-stage metadata exploitation strategy.

Acknowledgments

This work was supported by the Helmholtz Association’s Initiative and Networking Fund on the HAICORE@FZJ partition.

References

- [1] L. Picek, M. Sulc, J. Matas, Overview of FungiCLEF 2024: Revisiting fungi species recognition beyond 0-1 cost, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.
- [2] A. Joly, L. Picek, S. Kahl, H. Goëau, V. Espitalier, C. Botella, B. Deneu, D. Marcos, J. Estopinan, C. Leblanc, T. Larcher, M. Šulc, M. Hruz, M. Servajean, et al., Overview of lifeclef 2024: Challenges

- on species distribution prediction and identification, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2024.
- [3] L. Pícek, M. Šulc, J. Heilmann-Clausen, J. Matas, Overview of FungiCLEF 2022: Fungi recognition as an open set classification problem, in: Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, 2022.
 - [4] L. Pícek, M. Šulc, R. Chamidullin, J. Matas, Overview of fungiclef 2023: Fungi recognition beyond 1/0 cost, in: CLEF 2023-Conference and Labs of the Evaluation Forum, 2023.
 - [5] M. Sulc, L. Pícek, J. Matas, T. Jeppesen, J. Heilmann-Clausen, Fungi recognition: A practical use case, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 2316–2324.
 - [6] L. Pícek, M. Šulc, J. Matas, T. S. Jeppesen, J. Heilmann-Clausen, T. Læssøe, T. Frøslev, Danish fungi 2020-not just another image recognition dataset, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 1525–1535.
 - [7] N. Kiss, L. Czùni, Mushroom image classification with cnns: A case-study of different learning strategies, in: 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE, 2021, pp. 165–170.
 - [8] S. Wolf, J. Beyerer, Optimizing fine-grained fungi classification for diverse application-oriented open-set metrics, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), 2023.
 - [9] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, B. Guo, Swin transformer v2: Scaling up capacity and resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 12009–12019.
 - [10] H. Ren, H. Jiang, W. Luo, M. Meng, T. Zhang, Entropy-guided open-set fine-grained fungi recognition, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), 2023.
 - [11] J. L. Ba, J. R. Kiros, G. E. Hinton, Layer normalization, arXiv preprint arXiv:1607.06450 (2016).
 - [12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.
 - [13] MMPreTrain Contributors, Openmmlab’s pre-training toolbox and benchmark, <https://github.com/open-mmlab/mmpretrain>, 2023.
 - [14] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, *Advances in neural information processing systems* 32 (2019).
 - [15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.
 - [16] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 (2017).
 - [17] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, Randaugment: Practical automated data augmentation with a reduced search space, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.
 - [18] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, in: Proceedings of the AAAI conference on artificial intelligence, volume 34, 2020.