A. Evaluation Metrics

The diversity and unique features of the FungiTastic dataset allow for the evaluation of various fundamental computer vision and machine learning problems. We present several distinct benchmarks, each with its own evaluation protocol. This section provides a detailed description of all evaluation metrics for each benchmark.

A.1. Closed set classification

For closed-set classification, the main evaluation metric is F_1^m, i.e., the macro-averaged F_1-score, defined as

$$\mathrm{F}_1^m = \frac{1}{C}\sum_{c=1}^{C} F_c, \quad F_c = \frac{2P_c \cdot R_c}{P_c + R_c}, \tag{1}$$

where P_c and R_c are the recall and precision of class c and C is the total number of classes. Additional metrics of interest are Recall@k, defined as

$$\text{Recall@}k = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(y_i \in q_k(x_i)\right), \tag{2}$$

where N is the total number of samples in the dataset, x_i, y_i are the i-th sample and its label and $q_k(x)$ are the top k predictions for sample x.

A.2. Few-shot classification

The few-shot classification challenge does not have any unknown classes and can be considered as closed-set classification. Unlike other FungiTastic subsets, the few-shot subset does not suffer from high class imbalance and we choose the Top1 accuracy as the main metric. F_1 score and Top3 accuracy are also reported. All metrics are as defined in closed-set classification.

A.3. Open-set classification

The primary metric used for evaluation is the Receiver Operating Characteristic Area Under the Curve (ROCAUC), which measures the ability of the model to distinguish between classes across various threshold values. ROCAUC is defined as the area under the ROC curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at different classification thresholds where

$$\text{TPR} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}, \tag{3}$$

$$\text{FPR} = \frac{\text{False Positives (FP)}}{\text{False Positives (FP)} + \text{True Negatives (TN)}}. \tag{4}$$

In addition to ROCAUC, the True Negative Rate (TNR) at 95% TPR (TNR^{95}) is also reported. The TNR, also known as specificity, is defined as:

$$\text{TNR} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}}. \tag{5}$$

The TNR^{95} metric indicates the specificity achieved when the True Positive Rate (TPR) is fixed at 95%, reflecting the model's ability to minimize false positives while maintaining a high sensitivity.

The F1-score of the unknown-class, F^u, and the F-score over the known classes, F_k, are also of particular interest, with F_k defined as

$$F_K = \frac{1}{|K|}\sum_{c \in K} F_c, \tag{6}$$

where $K = \{1 \dots C\} \setminus \{u\}$ is the set of known classes.

A.4. Classification beyond 0-1 loss function

For the classification beyond 0-1 cost, we follow the definition we set for the annual FungiCLEF competition. A metric of the following general form should be minimized.

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} W(y_i, q_1(x_i)), \tag{7}$$

where N is the total number of samples, (x_i, y_i) are the i-th sample and its label, $q_1(x)$ is the top prediction for sample x and $W \in \mathbb{R}^{C \times C}$ is the cost matrix, C being the total number of classes. For the poisonous/edible species scenario, we define the cost matrix as

$$\mathrm{W}^{p/e}(y, q_1(x)) = \begin{cases} 0 & \text{if } d(y) = d(q_1(x)) \\ c_p & \text{if } d(y) = 1 \text{ and } d(q_1(x)) = 0, \\ c_e & \text{otherwise} \end{cases} \tag{8}$$

where $d(y), y \in \mathrm{C}$ is a binary function that indicates dangerous (poisonous) species ($d(y) = 1$), $c_p = 100$ and $c_e = 1$.

A.5. Segmentation

Provided segmentation masks allow the evaluation of many different segmentation scenarios; here, we highlight two.

Binary segmentation, where the positive class is the foreground (mushroom) and the negative class is the background (the complement). The metric is the intersection-over-union (IoU) averaged over all images ($\mathrm{IoU_B}$), giving each image the same weight

$$\mathrm{IoU_B} = \frac{1}{I}\sum_{i=1}^{I} \frac{|P_i \cap G_i|}{|P_i \cup G_i|}, \tag{9}$$

where I is the total number of images, P_i is the predicted set of foreground pixels for image i, G_i is the ground truth set of foreground pixels for image i, $|P_i \cap G_i|$ is the intersection

(true positives) for image i and $|P_i \cup G_i|$ is the union (true positives + false positives + false negatives) for image i.

For **semantic segmentation**, we adopt the standard mean intersection-over-union (mIoU) metric, where per-class IoUs are averaged, giving each class the same weight

$$\text{mIoU} = \frac{1}{C} \sum_{c=1}^{C} \frac{|P_c \cap G_c|}{|P_c \cup G_c|}, \tag{10}$$

where C is the total number of classes, P_c is the predicted set of pixels for class c, G_c is the ground truth set of pixels for class c, $|P_c \cap G_c|$ is the intersection (TPs) and $|P_c \cup G_c|$ is the union (TPs + FPs + FNs).

B. Supporting Figures and Tables

B.1. Closed-set experiment with higher input size

Following the results provided in the paper, we further experimented with how input size affects classification performance. Switching from 224×224 to 384×384 increased the performance by around five percentage points in all measured metrics and for almost all the architectures. Still, the best-performing model, i.e., BEiT-Base/p16 achieves "*just*" 75% accuracy and less then 50% in terms of F_1^m.

Table 8. **Closed-set fine-grained classification FungiTastic and FungiTastic–M**. A set of selected state-of-the-art CNN- and Transformer-based architectures evaluated on the test sets. All reported metrics show the challenging nature of the dataset.

	FungiTastic–M – 384^2			FungiTastic – 384^2		
Architectures	**Top1**	**Top3**	$\mathbf{F}_1^m$	**Top1**	**Top3**	$\mathbf{F}_1^m$
ResNet-50	66.3	82.9	39.8	66.9	80.9	36.3
ResNeXt-50	67.0	84.0	39.9	68.1	81.9	37.5
EfficientNet-B3	67.4	82.8	40.5	68.2	81.9	37.2
EfficientNet-v2-B3	70.3	85.8	43.9	72.0	84.7	41.0
ConvNeXt-Base	70.2	85.7	43.9	70.7	83.8	39.6
ViT-Base/p16	73.9	87.8	46.3	74.9	86.3	43.9
Swin-Base/p4w12	72.9	87.0	47.1	74.3	86.3	43.3
BEiT-Base/p16	**74.8**	**88.3**	**48.5**	**75.3**	**86.7**	**44.5**

B.2. FungiTastic – Dataset statistics

The FungiTastic dataset offers a rich and diverse collection of observations and metadata. To provide a clearer understanding of its scope, Table 9 presents a statistical overview of its subsets, including the number of observations, associated images, species categories, and metadata availability. Each subset caters to specific benchmarking needs, ensuring comprehensive evaluation scenarios.

Table 9. **FungiTastic dataset splits – statistical overview.** The number of observations, images, and classes for each benchmark and the corresponding dataset. "**Unkn**own classes" are those with no available data in training. DNA stands for DNA-sequenced data.

Dataset	**Subset**	**Observ.**	**Images**	**Classes**	**Unkn.**	Metadata	Masks	Captions
FungiTastic *Closed Set*	Train.	246,884	433,701	2,829	—	✓	–	✓
	Val.	45,613	89,659	2,306	—	✓	–	✓
	Test	48,378	91,832	2,336	—	✓	–	✓
	Test^{DNA}	2,041	5,105	725	—	✓	–	✓
FungiTastic–M *Closed Set*	Train.	25,786	46,842	215	—	✓	✓	✓
	Val.	4,687	9,412	193	—	✓	✓	✓
	Test	5,531	10,738	196	—	✓	✓	✓
	Test^{DNA}	211	642	93	—	✓	✓	✓
FungiTastic–FS *Closed Set*	Train.	4,293	7,819	2,427	—	✓	–	✓
	Val.	1,099	2,285	570	—	✓	–	✓
	Test	999	1,911	567	—	✓	–	✓
FungiTastic *Open Set*	Train.	246,884	433,702	2,829	—	✓	–	✓
	Val.	47,450	96,756	3,360	1,053	✓	–	✓
	Test	50,084	97,551	3,349	1,000	✓	–	✓
Total unique values:		**349,307**	**632,313**	**6,034**	**1,678**			

B.3. Additional figures

To further highlight the unique features of the FungiTastic dataset, we provide additional figures. These include: (i) a time series sample of temperature data illustrating climatic variability over a 3-year period (Figure 8), (ii) examples of detailed text descriptions generated for individual images to aid in species identification (Figure 9), and (iii) visual samples of ground truth segmentations that highlight different fruiting body parts of fungi (Figure 10).

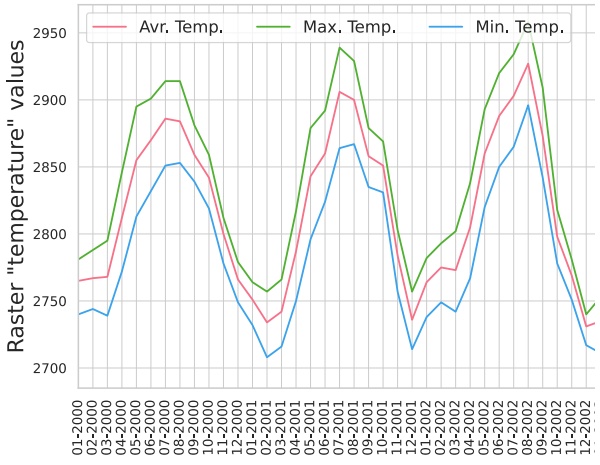

Figure 8. **Sample of available temperature data**. First 3 years (out of 20 available years) of average, maximum, and minimum monthly temperatures for a randomly selected location.

The image shows two mushrooms growing on a dark, decaying log. The mushroom on the left is larger and more prominent. It has a light brown cap with a slightly darker brown rim. The cap's surface appears smooth and slightly shiny. Its stem is thick and light brown, with a hint of green at the base. The smaller mushroom on the right has a similar light brown cap, but its rim is more pronounced and has a white, almost translucent appearance. This gives it a delicate, lacy look. The stem of this mushroom is thinner and lighter in color compared to its larger companion. Both mushrooms have a gilled structure visible underneath their caps. The gills are light brown .....

The image shows a cluster of mushrooms growing on a decaying log. The fungi have distinctive features that are clearly visible: The caps of the mushrooms are dark brown in color, with a slightly lighter shade around the edges. They have a dome-like shape, resembling small umbrellas or bell-shaped structures. The gills of the mushrooms are visible underneath the caps. They appear as thin, white, vertical lines radiating from the stem to the edge of the cap. These gills are a key identifying feature of many mushroom species. The stems of the mushrooms are light brown in color and appear to be relatively thick and sturdy. They support the caps and extend into .....

The image shows four mushrooms arranged in a square formation on a grassy surface. These fungi have distinctive visual features that are clearly visible: The caps of the mushrooms are dark brown in color, with a smooth, slightly shiny appearance. They have a rounded, dome-like shape that's typical of many mushroom species. The stems of the mushrooms are white, providing a stark contrast to the dark caps. They appear to be relatively thick and sturdy, supporting the caps firmly. The texture of the caps seems to be relatively smooth, without any visible ridges or bumps. The stems appear to have a slightly fuzzy or velvety texture. In terms of size, the mushrooms are of .....

The image shows a single mushroom lying on its side in a grassy area. The mushroom has a distinctive shape with a white cap and a long, cylindrical stem. The cap is rounded and slightly curved, while the stem is straight and elongated. The mushroom's surface appears smooth and uniform in color, with a clean white appearance. The cap and stem are of similar width, creating a balanced look. The stem extends from the base of the cap and continues downward, with the tip pointing towards the bottom right corner of the image. The mushroom is covered in dirt, particularly on its stem and parts of the cap. This gives it a freshly harvested appearance and .....

The image shows a large, flat mushroom as the central focus. Its cap is predominantly gray with a slightly mottled appearance, featuring some darker patches. The surface of the cap appears smooth and slightly shiny, reflecting light in certain areas. The mushroom's cap is circular in shape, though it's not perfectly round. It has a slightly irregular edge, giving it a natural, organic look. The cap's size is quite substantial, dominating the frame of the image. On the top of the mushroom cap, there are several small, white, circular structures. These are likely the gills or spore-bearing areas of the mushroom. They stand out in contrast against the gray cap, ...

The image shows two mushrooms growing on a dark, decaying log. The mushroom on the left is larger and more prominent. It has a light brown cap with a slightly darker brown rim. The cap's surface appears smooth and slightly shiny. Its stem is thick and light brown, with a hint of green at the base. The smaller mushroom on the right has a similar light brown cap, but its rim is more pronounced and has a white, almost translucent appearance. This gives it a delicate, lacy look. The stem of this mushroom is thinner and lighter in color compared to its larger companion. Both mushrooms have a gilled structure visible underneath their caps. The gills are light brown .....

Figure 9. **Additional image caption samples.** For each photograph, we provide a Malmo-7B image caption-like text description.

Figure 10. **Additional samples of ground truth fruiting body part segmentation**.