An Interpretable Framework to Characterize Compound Treatments on Filamentous Fungi using Cell Painting and Deep Metric Learning

Supplementary Materials

1 Deep Cosine Metric Learning (DCML)

Algorithm 1: Training DML for one iteration

Sample batch of images $\boldsymbol{I}$, phenotype $\boldsymbol{y}$, and category $\boldsymbol{c}$
$G_{\kappa_j}, G_{\psi_j}, G_{\omega_j}, G_\theta \leftarrow \mathbf{0} \quad \forall j$; `// Initialize gradients`
foreach I, c, y **do**
  $X \leftarrow f_\theta(I)$; `// Pass image in feature extractor`
  $\boldsymbol{h} \leftarrow g_{\psi_c}(X)$; `// Compute embedding given concept`
  Append global intensity statistics to get $\boldsymbol{h}'$
  L_2-normalize embeddings and centroids to get $\tilde{\boldsymbol{h}}'$ and $\tilde{\boldsymbol{\omega}}_c$
  Compute softmax outputs $p(y = k|\tilde{\boldsymbol{h}}'), \forall k$
  Compute cosine softmax loss $\mathcal{L}$
  $G_{\kappa_c} \mathrel{+}= \frac{\partial \mathcal{L}}{\partial \kappa_c}$; `// Accumulate gradients`
  $G_{\psi_c} \mathrel{+}= \frac{\partial \mathcal{L}}{\partial \psi_c}$
  $G_{\omega_c} \mathrel{+}= \frac{\partial \mathcal{L}}{\partial \omega_c}$
  $G_\theta \mathrel{+}= \frac{\partial \mathcal{L}}{\partial \theta}$
Back-propagate all gradients

We show in Fig. 1 a schema of our architecture, and give a training pseudo-code in Alg. 1

2 Fungi-Profiler

We give here additional details related to *Fungi-Profiler*, our handcrafted feature extractor.

2.1 Segmentation

As a first pre-processing step, We aim to segment regions where fungi is located to focus our downstream feature extraction pipeline.

We first apply Adaptive Histogram Equalization [7] on bright-field images to reduce luminosity variations accross different experiments. Next, we detect edges by applying an adaptive thresholding step, and follow with a serie of morphological operations. We show an example in Fig. 2.

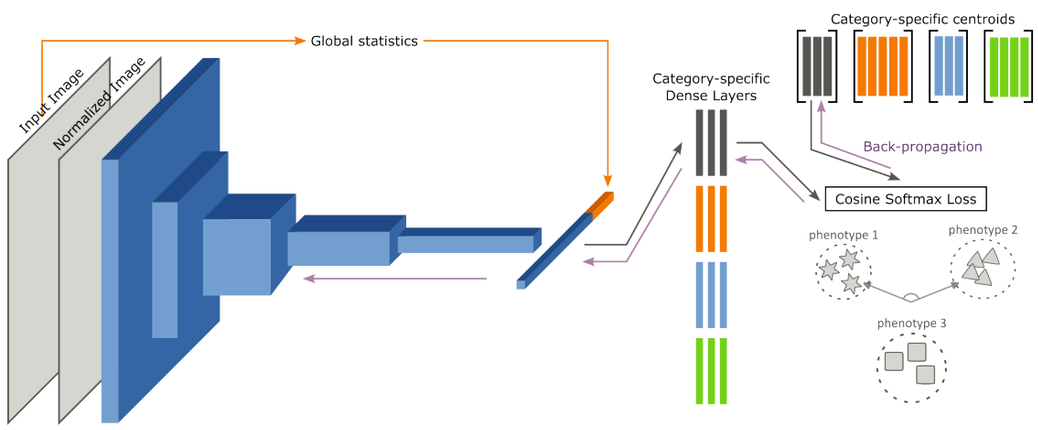

Figure 1: Deep Cosine Metric Learning model. Images are normalized, then fed into a Convolution Neural Network. We concatenate global image statistics computed on the original image to the resulting feature vector and feed it into a serie of dense layers according to the phenotypical category captured by its modality (red, green, blue and grey blocks). Last, we compute the cosine softmax loss on each image of a batch, and aggregate gradients to update the centroids, the category-specific dense layers, and the feature extraction backbone.

	Apical cell wall			Cell wall			Lipid			Morphology			
	DCML	DCML/IR	FPCML/IR	DCML	DCML/IR	FPCML/IR	DCML	DCML/IR	FPCML/IR	DCML	DCML/IR	FPCML/IR	GI
Compound													
AculeacinA	0.36	0.36	0.56	0.14	0.14	0.12	0.12	0.12	0.33	0.13	0.12	0.10	0.59
Boscalid	0.17	0.17	0.37	0.19	0.19	0.44	0.09	0.09	0.09	0.13	0.13	0.06	0.11
Bromuconazole	0.42	0.42	0.06	0.23	0.24	0.46	0.09	0.09	0.13	0.10	0.10	0.02	0.09
Carbendazim	0.03	0.03	0.02	0.33	0.29	0.12	0.01	0.01	0.00	0.02	0.02	0.00	0.17
Carboxin	0.11	0.11	0.31	0.10	0.10	0.33	0.15	0.15	0.17	0.08	0.08	0.14	0.16
Cyprodinil	0.15	0.14	0.57	0.54	0.54	0.24	0.01	0.01	0.01	0.08	0.08	0.04	0.07
Fluopicolide	0.07	0.07	0.04	0.05	0.05	0.18	0.09	0.09	0.15	0.12	0.10	0.25	0.21
Fluquinconazole	0.29	0.27	0.13	0.19	0.31	0.46	0.05	0.05	0.07	0.09	0.09	0.03	0.09
Iprodione	0.21	0.21	0.56	0.17	0.17	0.22	0.28	0.28	0.21	0.20	0.19	0.11	0.09
Itraconazole	0.52	0.52	0.01	0.23	0.23	0.67	0.00	0.00	0.00	0.01	0.01	0.01	0.07
Mancozeb	0.10	0.10	0.37	0.03	0.03	0.29	0.17	0.17	0.41	0.07	0.07	0.06	0.06
Metalaxyl	0.07	0.07	0.06	0.04	0.04	0.18	0.09	0.09	0.17	0.20	0.20	0.08	0.11
Nikkomycin	0.23	0.23	0.03	0.74	0.74	0.55	0.00	0.00	0.00	0.15	0.15	0.60	0.67
Picoxystrobin	0.12	0.12	0.25	0.10	0.10	0.77	0.16	0.15	0.11	0.20	0.20	0.15	0.15
PolyoxinB	0.15	0.15	0.41	0.30	0.30	0.26	0.10	0.10	0.06	0.16	0.16	0.53	0.51
Prochloraz	0.54	0.54	0.02	0.23	0.23	0.66	0.25	0.25	0.25	0.01	0.01	0.00	0.06
Pyrimethanil	0.16	0.16	0.81	0.59	0.59	0.24	0.01	0.01	0.00	0.09	0.09	0.03	0.09
Terbinafine	0.30	0.30	0.22	0.07	0.13	0.39	0.43	0.43	0.41	0.11	0.11	0.09	0.08
All	0.22	**0.22**	0.27	0.24	**0.24**	0.37	0.12	**0.12**	0.14	0.11	**0.11**	0.13	0.19
Std	0.15	0.15	0.24	0.19	0.19	0.19	0.11	0.11	0.13	0.06	0.06	0.17	0.19

Table 1: Mean absolute errors on each phenotypical categories, compound, and methods. DCML: Proposed deep cosine metric learning method, DCML/IR: DCML combined with isotonic regression post-processing. FPCML: Metric learning based on fungi-profiler features. GI: Growth-Inhibition. In bold, we show the best performing method.

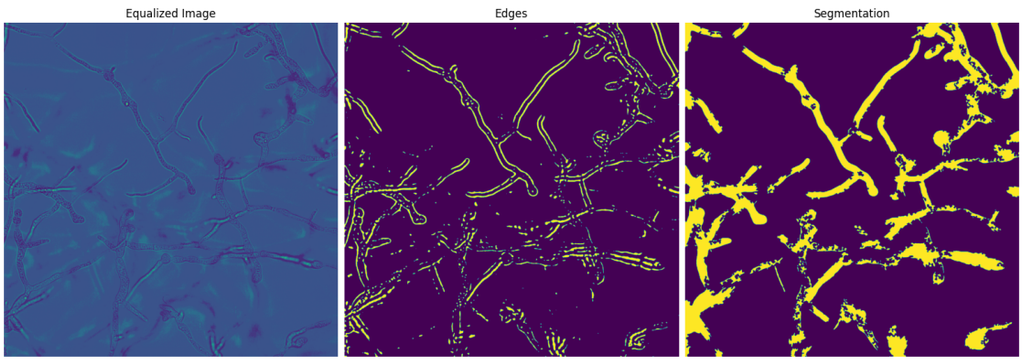

Figure 2: Segmentation step of Fungi-Profiler. (Left) Original image after adaptive histogram equalization. (Center) Edge map obtained through morphological operations. (Right) Segmentation map.

Category	Feature description	Dimension
Intensity	Mean & standard deviation	2
	Median & Mean absolute difference from median	2
	Upper and lower quartiles	2
	Area (Pixel counts)	1
	Min and max intensity	2
Texture	Haralick features: Rotation invariant features in local neighborhood computed at multiple scale [5]	156
	Granularity spectrum using morphological operators [6]	16
Co-Localization	Correlation coefficient	1
	Least-square regression slope coefficient	$M-1$
	Manders overlap coefficient [3]	$M-1$
	Rank-weighted coefficient [8]	$M-1$
	Count of overlapping pixels above threshold	$M-1$
	Costes automated threshold [1]	$M-1$

Table 2: Summary of feature set computed by Fungi-Profiler. M denotes the number of image modalities.

2.2 Illumination Correction

As noted in [9], microscopy images manifest an uneven illumination pattern accross the field of view, where regions close to the border show reduced brightness with respect to the center. This has been shown to have an important impact on the quality of visual features as measured by downstream prediction tasks.

To circumvent this artefact, we implement a simple illumination correction procedure as suggested by [9], where we compute an illumination correction function (ICF) as follows:

All bright-field images acquired during a plate assay are averaged pixel-wise, and then smoothed spatially using a squared median filter of size 25% of the original image size. Last, we apply the ICF to bright-field and fluorescent channels following:

$$I' = \frac{I}{1 + ICF} \tag{1}$$

2.3 Feature Extraction

In this step, we apply to each corrected image (both bright-field and fluorescent channels) the segmentation mask of the corresponding site, so as to set background pixels to 0. Next, we apply a serie of feature extraction routines summarized in Tab. 2 to obtain a vector of dimension 210 for each image. Note that, aside from traditional features computed on each image taken individually, we further leverage cross-channel information by extracting co-localization features on each pair of images of a given stack.

3 Pooled-Adjacent Violators Algorithm (PAVA)

We give in Alg. 2 a pseudo-code of the PAVA algorithm [2] by which we impose a monotonicity constraint on our dose-response estimations.

4 Testing Dataset

We show in Tab. 3 all compounds used in our experiments, along with their modes of action and biological targets.

Algorithm 2: Pooled-Adjacent Violators Algorithm (PAVA)

input : $\boldsymbol{x} \in \mathbb{R}^N$: An ordered set of Design-points
$\boldsymbol{r} \in \mathbb{R}^N$: A set of observations taken at $\boldsymbol{x}$
output: $\hat{\boldsymbol{r}} \in \mathbb{R}^N$: A set of monotous-increasing values that best approximate $\boldsymbol{r}$ in the least-squares sense

$\forall j \in 1, \ldots, N, \quad \hat{r}_j \leftarrow r_j$; `// Initialize all observations as valid`
$\mathbb{C} \leftarrow \emptyset$; `// For book-keeping contiguous violations`
`// Find all violations`
while $\mathbb{V} \leftarrow \{j : j < m, \hat{r}_j > \hat{r}_{j+1}\} \neq \emptyset$ **do**
  $l \leftarrow \min(\mathbb{V})$; `// Select first occurence`
  if $l \in \mathbb{C}$ **then**
    $\mathbb{C} \leftarrow \mathbb{C} \cup \{l+1\}$; `// last correction did not break violation`
  else
    $\mathbb{C} \leftarrow \{l, l+1\}$; `// pool adjacent observation`
  $\forall j \in \mathbb{C}, \quad \hat{r}_j \leftarrow \mathbb{E}_{k \in \mathbb{C}}[r_k]$; `// replace values with expectation`

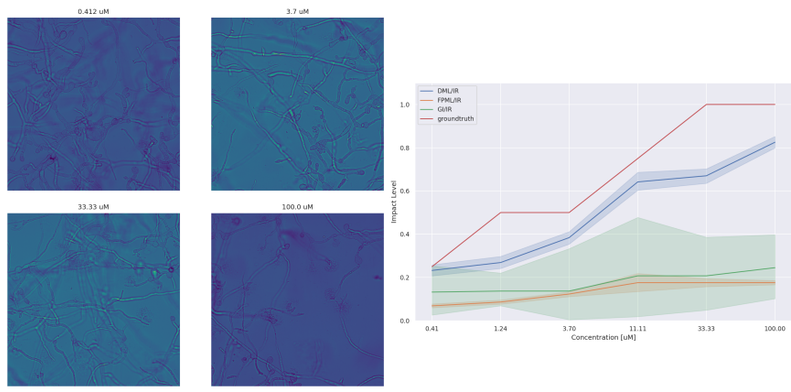

Figure 3: Example of a subtle phenotypical variation where spores turn bigger as concentration increases, while fungal mass is largely unaffected. Methods that relie on fungal mass (FPCML, GI) fail to capture this transformation, in contrast with DCML, which relies on a CNN feature extraction backbone.

5 Experiments

5.1 Dose-response Estimation

We give in Tab. 1 the mean absolute errors of our methods on all tested compounds. In Fig. 3, we show a set of images along with estimated dose-response curves where a phenotype is largely independent on the fungal mass. In particular, we emphasize a limitation of methods **FPCML** and **GI** in capturing subtle variations in morphology, while our best method **DCML**, as it relies on a CNN backbone, behaves as expected.

References

[1] Sylvain V Costes, Dirk Daelemans, Edward H Cho, Zachary Dobbin, George Pavlakis, and Stephen Lockett. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical journal*, 86(6):3993–4003, 2004. 4

[2] Jan De Leeuw, Kurt Hornik, and Patrick Mair. Isotone optimization in r: pool-adjacent-violators algorithm (pava) and active set methods. *Journal of statistical software*, 32(5):1–24, 2009. 3

[3] Kenneth W Dunn, Malgorzata M Kamocka, and John H McDonald. A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology*, 300(4):C723–C742, 2011. 4

[4] Frederick M Fishel and MM Dewdney. Fungicide resistance action committee's (frac) classification scheme of fungicides according to mode of action. *PI94. University of Florida. 7p*, 2012. 6

[5] Tommy Löfstedt, Patrik Brynolfsson, Thomas Asklund, Tufve Nyholm, and Anders Garpebring. Gray-level invariant haralick texture features. *PLOS ONE*, 14(2):1–18, 02 2019. 4

[6] Petros Maragos. Pattern spectrum and multiscale shape representation. *IEEE Transactions on pattern analysis and machine intelligence*, 11(7):701–716, 1989. 4

[7] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987. 3

[8] Vasanth R Singan and Jeremy C Simpson. Implementation of the rank-weighted co-localization (rwc) algorithm in multiple image analysis platforms for quantitative analysis of microscopy images. *Source code for biology and medicine*, 11:1–3, 2016. 4

[9] S. SINGH, M.-A. BRAY, T.R. JONES, and A.E. CARPENTER. Pipeline for illumination correction of images for high-throughput microscopy. *Journal of Microscopy*, 256(3):231–236, 2014. 3

Compound	MoA group	MoA sub-group	Target
Metalaxyl	Nucleic Acid Metabolism	RNA polymerase I	-
Cyprodinil	Amino Acid and Protein Synthesis	Methionine Biosynthesis	-
Pyrimethanil	Amino Acid and Protein Synthesis	Methionine Biosynthesis	-
Carbendazim	Cytoskeleton and Motor Proteins	tubulin polymerization	-
Fluopicolide	Cytoskeleton and Motor Proteins	Delocalisation of spectrin-like proteins	-
Fenpiclonil	Signal Transduction	Osmotic Signal Transduction	MAP / histidine- kinase (os-2, HOG1)
Fludioxonil	Signal Transduction	Osmotic Signal Transduction	MAP / histidine- kinase (os-2, HOG1)
Picoxystrobin	Respiration	Complex III cytochrome bc1 (ubiquinol oxidase) at Qo site (cyt b gene)	-
Boscalid	Respiration	Complex II: succinate-dehydrogenase	-
Iprodione	Signal Transduction	Osmotic signal transduction	MAP / histidine kinase (os-1, Daf1)
Carboxin	Respiration	Complex II: succinate-dehydrogenase	-
Fluazinam	Respiration	Uncouplers of oxidative Phosphorylation	-
Nikkomycin	Cell wall	Chitin synthase	-
PolyoxinB	Cell wall	Chitin synthase	-
AculeacinA	Cell wall	1,3-beta-D-glucan synthase	-
Bromuconazole	Sterol Biosynthesis in Membranes	C14-demethylase in sterol biosynthesis	-
Itraconazole	Sterol Biosynthesis in Membranes	C14-demethylase in sterol biosynthesis	-
Fluquinconazole	Sterol Biosynthesis in Membranes	C14-demethylase in sterol biosynthesis	-
Prochloraz	Sterol Biosynthesis in Membranes	C14-demethylase in sterol biosynthesis	-
Terbinafine	Sterol Biosynthesis in Membranes	squalene epoxidase in sterol biosynthesis	-
Mancozeb	Chemicals with Multi-Site Activity	-	-
Zineb	Chemicals with Multi-Site Activity	-	-

Table 3: Summary of testing dataset. We show compound names, MoA group, MoA sub-group, and biological target according to the FRAC classification when applicable [4].