Quantifying Root Colonization in Arbuscular Mycorrhizas by Image Segmentation and Machine Learning

Ivan Arcangelo Sciascia
University of Turin
Crosino Andrea
University of Turin
Mara Novero
University of Turin
Mara Politi
University of Turin
Andrea Genre (andrea.genre@unito.it)
University of Turin

Research Article

Keywords: biostatistics, image segmentation, microscopy, arbuscular mycorrhiza

Posted Date: June 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-538682/v1

Quantifying root colonization in arbuscular mycorrhizas by image segmentation and machine learning

Sciascia I. [1], Crosino A. [1], Novero M.[1], Politi M.[1], Genre A.[1*]

1. Department of Life Sciences and Systems Biology, University of Turin. Viale P.A. Mattioli 25, 10125 Turin, Italy.

*Corresponding author: Andrea Genre, Department of Life Sciences and Systems Biology, University of Turin. Viale P.A. Mattioli 25, 10125 Turin, Italy. andrea.genre@unito.it

Abstract

Motivation

Arbuscular mycorrhizas are the most widespread plant symbioses and involve the majority of crop plants. The beneficial interaction between plant roots and a group of soil fungi (Glomeromycotina) grants the green host a preferential access to soil mineral nutrients and water, supporting plant health, biomass production and resistance to both abiotic and biotic stresses. The nutritional exchanges at the core of this symbiosis take place inside the living root cells, which are diffusely colonized by specialized fungal structures called arbuscules. For this reason, the vast majority of studies investigating arbuscular mycorrhizas and their applications in agriculture require a precise quantification of the intensity of root colonization. To this aim, several manual methods have been used for decades to estimate the extension of intraradical fungal structures, mostly based on optical microscopy observations and individual assessment of fungal abundance in the root tissues.

Results

Here we propose a novel semi-automated approach to quantify AM colonization based on digital image analysis and compare two methods based on image thresholding and machine learning. Our results indicate in machine learning a very promising tool for accelerating, simplifying and standardizing this critical type of analysis, with a direct potential interest for applicative and basic research.

Contact

ivan.sciascia@unito.it; andrea.genre@unito.it

Key words: biostatistics, image segmentation, microscopy, arbuscular mycorrhiza

1. Introduction

Arburscural mycorrhizas (AM) are widespread plant endosymbioses that develop between Glomeromycotina fungi and the roots of the majority of plants species, including most crops. The symbiosis benefits extend to both partners by improving plant mineral absorption, tolerance to biotic and abiotic stresses and overall fitness, while rewarding fungal symbionts with carbon compounds derived from the photosynthetic process, such as sugars and lipids [1]. Such exchanges of nutrients represent the functional core of the symbiosis and mainly take place in highly branched fungal structures - called arbuscules - that are hosted within the living cortical cells of the host root, as described in Figure 1 [2].

The central ecological role of AM in the functioning of low-input ecosystems and the ability of most crop plants to develop this symbiosis has focused a growing number of investigations on the use of AM in sustainable agricultural practices. A critical factor in all such studies is represented by the precise quantification of root colonization by AM fungi, with particular attention to arbuscule abundance [3]. To this aim, molecular analyses (based on the quantification of fungal sequences in total root DNA or arbuscule-specific markers in total root RNA extracts) are outnumbered by the direct quantification of intraradical fungal structures through optical microscopy imaging and manual analyses [4, 5]).

In one of the most commonly used methods, root samples are stained with lactic blue or alternative dyes to label intraradical fungal structures; roots are then cut into 1cm-long segments, mounted on microscope slides and carefully observed under an optical microscope to classify each segment based on visual criteria such as the extension of intraradical hyphae and the abundance of arbuscules in the colonized areas [6, 7]. Such methods are extremely time consuming, based on the ability of trained operators and subject to errors.

In an attempt to improve speed, repeatability and reliability of this type of analysis, we developed a semi-automated approach based on digital imaging and different types of post processing. Firstly,

an automatic method was developed to discriminate between mycorrhizal and non-mycorrhizal root images. Secondly, we designed a semi-automated algorithm to generate quantitative indexes of root colonization deriving from either image thresholding using ImageJ, or based on machine learning analyses [8, 9] using the commercial software Zeiss Intellesis [10]. Our results indicate machine learning as the most effective approach, with interesting applicative perspectives as an alternative to manual quantification.

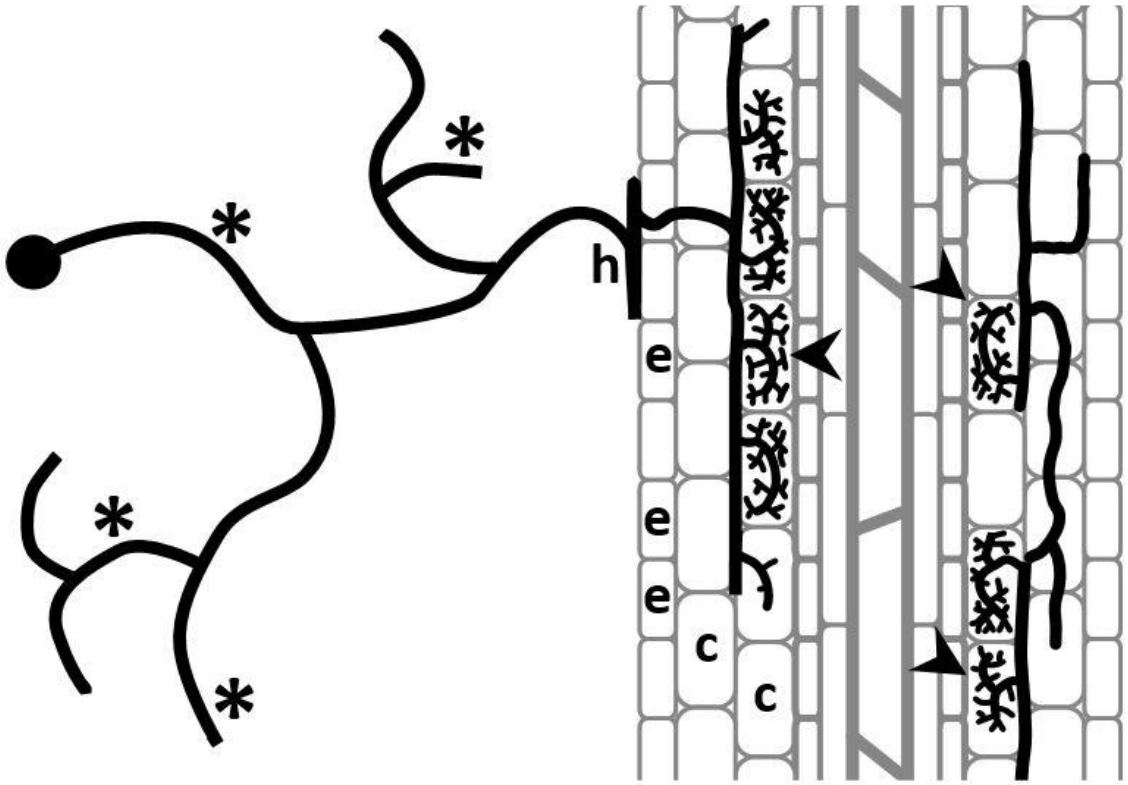


Figure 1. Schematic representation of a host root (grey) colonized by an arbuscular mycorrhizal fungus (black). The extraradical mycelium (*) explores the soil surrounding the root, while intraradical structures produced from the hyphopodium (h) penetrate root epidermal cells (e), colonizing single cortical cells (c), where they eventually develop into branched arbuscules (arrowhead), the sites of nutrient exchanges between symbionts.

2. System and methods

All images used in this research (Supplementary file 1) were acquired at a resolution of 1280*720 pixels, using one of the following brightfield optical microscopes: Nikon Eclipse Ci-L mounting 4x / 0.10 WD30, 10x / 0.25 WD10.5 and 40x / 0.65 WD0.57 objectives; Leica DM500 mounting a PLAN 4x/0.10 objective. The intraradical fungal structures are stained in dark blue while the plant tissues and cells are completely transparent or light blue, this allows to distinguish the fungal structures from the plant structures.

The segmentation analysis based on digital microscopy images on gray scale pixels was carried out using ImageJ software (Rueden *et al*., 2017) and was based on two phases analysis:

- a dataset of 143 images from non-mycorrhizal and 60 images from mycorrhizal roots of *Medicago truncatula* colonized by the AM fungus *Funneliformis mosseae*, previously classified manually according to a standard protocol for quantifying AM colonization [6] (Supplementary file 1), has been used for an explorative analysis to assign images to mycorrhizal/non mycorrhizal classes.
- a second dataset consisting of 180 images was used for both the thresholding (using ImageJ) and machine learning (using Intellesis) approaches. All images had previously been classified manually into 6 classes of 30 images, ranging from 1 (non mycorrhizal) to 6 (maximum root colonization), adapting to the same standard protocol [6].

  The digital images in grayscale and the relative thresholding are described in Supplementary file 2.

Algorithm

Binary segmentation analysis.

In the first phase of the analysis, images were loaded into ImageJ [11] and transformed into 8-bit digital images before further processing. A simple segmentation technique based on pixel intensity thresholding was then applied in order to distinguish darker areas (likely corresponding to stained fungal structures) from lighter areas (not colonized tissues). In more detail, the pixel brightness range [0-100] was chosen as representative of colonized areas, whereas the [101-255] range corresponded to uncolonized tissues (Figure 2). Each image was classified based on the results of the segmentation analyses and correlated with the manual classification with binary logistic regression. The study variable (y) is the classification with binary variable 0 (not mycorrhized) or 1 (mycorrhized), obtained by manual classification. The explanatory variable (x) is the expression of the pixel area determined as a consequence of the chosen threshold and it is released by the ImageJ macro in a value report.

This whole set of image processing steps has been automated, taking advantage of the integrated ImageJ macros feature. (Supplementary file 3).

The results demonstrated a good discrimination capacity of the binary regression model with the table of manual classifications, and by the ROC curve with an area under the curve of 0.870 (Supplementary file 4).

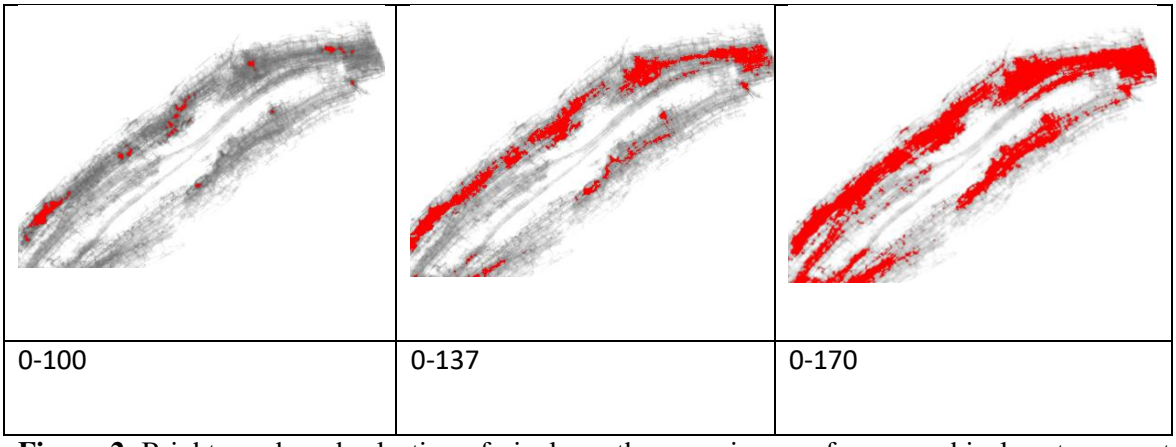


Figure 2. Brightness-based selection of pixels on the same image of a mycorrhizal root segment. Pixels are selected (in red) based on arbitrary brightness thresholds: 100, 137, 170 in a range from 0 (black) to 255 (white) using Fiji/ImageJ.

129 Statistical analyzes were carried out with the SPSS software (IBM Statistics for Windows version
130 26.0).
131
132 *Thresholding analysis.*
133 Also in this case digital images were loaded into ImageJ and transformed into 8-bit digital images.
134 Two ImageJ macros were designed to process the images (Supplementary file 3). The first image
135 segmentation macro was designed to discriminate the entire root section area from the globally
136 lighter background; in this case the pixel intensity threshold was set to 155.
137 The second macro produced image segmentation based on pixel intensity, with a threshold
138 empirically set at 65 (in the 0-255 range) to discriminate between darker pixels (corresponding to
139 fungal structures) and lighter (uncolonized) areas.
140 Both results are described in Figure 3.
141 After applying both segmentation macros to all our images, a set of quantitative values was
142 obtained corresponding to the supposed colonized area (darkest pixels) and the total root section
143 area (as isolated from the image background).
144
145
146
147

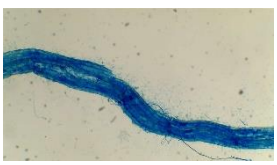
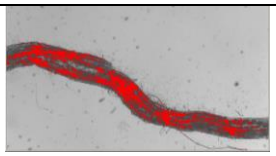
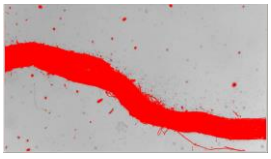
Threshold	Original digital 8 bit image	Segmentation
65		
155		

Figure 3. Image thresholding in ImageJ. The first threshold set at 65 identifies the colonized area, while the second threshold at 155 outlines with a good approximation the total area of the root section.

The ratio between these two values was used to obtain the *m* index (for mycorrhization) for each image in the dataset:

$$m = \frac{mycorrhized\ area}{total\ area} * 100 \quad [1]$$

This m-index was therefore used in the subsequent statistical analyses as an independent variable in the predictive model for the quantification of root colonization (Table 1).

Thresholding index (*m*)

manual_classification	Number of images	Mean	Std. Deviation	Median	Range
1	30	1,011369	1,1947128	0,607633	5,5826
2	30	5,878176	4,4051314	4,631726	16,4411
3	30	14,396212	10,1959085	10,330562	38,9856
4	30	15,174801	8,3606544	14,529747	28,6474
5	30	29,751744	13,2256586	31,321755	51,3397
6	30	29,970579	11,6878341	28,027334	45,3563
Total	180	16,030480	14,2051105	12,332234	60,3251

Table 1a. *m*-index, descriptive statistics

Classification	**Qualitative description**	**Confidence interval (α=0,05)**
1	Not mycorrhized (mycorrhization 0%)	$m = 1{,}011369 \pm 1{,}96 * \frac{1{,}1947128}{\sqrt{30}}$
2	Slightly mycorrhized (few mycorrhization)	$m = 5{,}878176 \pm 1{,}96 * \frac{4{,}4051314}{\sqrt{30}}$
3	Moderately mycorrhized (mycorrhization < 10%)	$m = 14{,}396212 \pm 1{,}96 * \frac{10{,}1959085}{\sqrt{30}}$
4	Averagely mycorrhized (mycorrhization from 11 to 50%)	$m = 15{,}174801 \pm 1{,}96 * \frac{8{,}3606544}{\sqrt{30}}$
5	Highly mycorrhized (mycorrhization from 51 to 90%)	$m = 29{,}751744 \pm 1{,}96 * \frac{13{,}2256586}{\sqrt{30}}$
6	Strongly mycorrhized (mycorrhization > 90%)	$m = 29{,}970579 \pm 1{,}96 * \frac{11{,}6878341}{\sqrt{30}}$

Table 1b. Confidence interval for m-index related to the multinomial classification

In the Table 1a we show the descriptive statistics of the calculated *m*-index based on ImageJ macros relative for each manual classified classes in the range 1 – 6. Confidence intervals and qualitative descriptions for each class are shown in Table 1b.

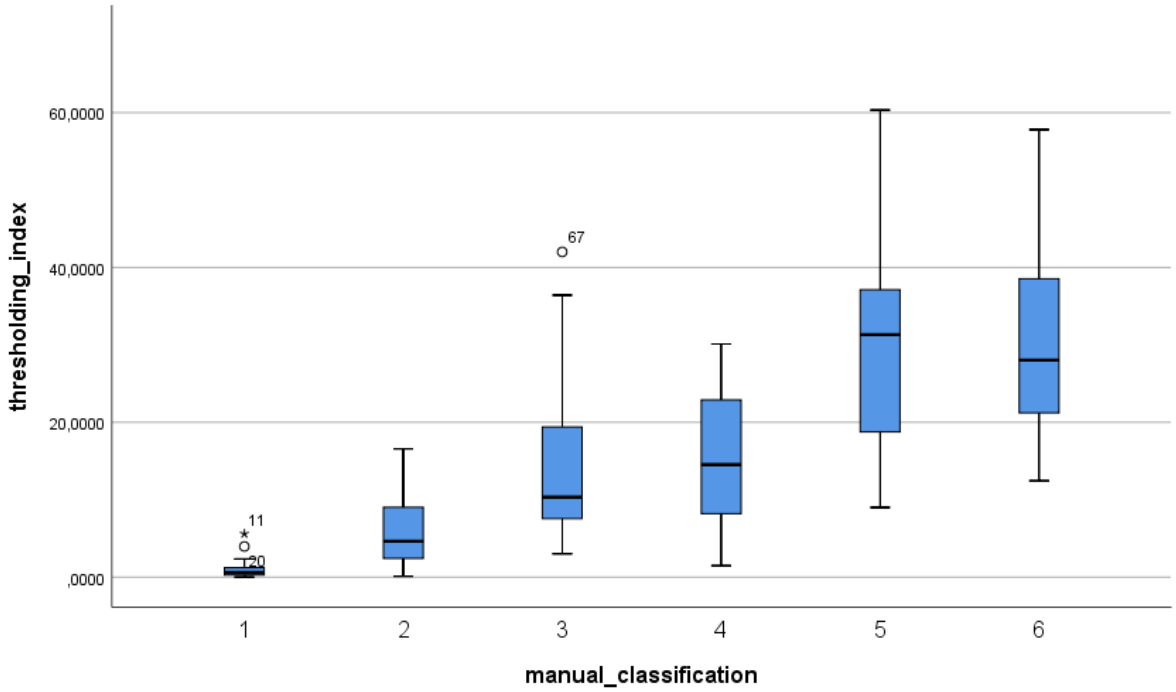


Figure 4. *m*-index, box plot

As shown in Figure 4, the average value of the *m* index increases for each class considered, indicating a good correlation between the study variable y (manual classification) and the independent variable x (*m*-index).

Indeed, a strong correlation was detected between the two variables, with a significant Pearson coefficient at the 0.01 two tailed level with a value of 0.748. Furthermore, ANOVA analysis (Table 2) and post hoc test of multiple comparisons performed with Bonferroni correction (Table 3) confirmed the statistical significance of the differences between classes. In particular, Bonferroni post hoc test demonstrated significant differences in 12 pairwise comparisons (on a total of 15) with exceptions for class 1 vs 2, class 3 vs 4 and class 5 vs 6.

ANOVA

m-index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	21439,364	5	4287,873	50,823	,000
Within Groups	14680,181	174	84,369		
Total	36119,544	179			

Table2. ANOVA for the six groups of digital images index t

(I) manual_classification	(J) manual_classification	Mean Difference (I-J)	Std. Error	Sig.
1	2	-4,8668071	2,3716219	,625
	3	-13,3848433*	2,3716219	,000
	4	-14,1634322*	2,3716219	,000
	5	-28,7403757*	2,3716219	,000
	6	-28,9592100*	2,3716219	,000
2	1	4,8668071	2,3716219	,625
	3	-8,5180362*	2,3716219	,006
	4	-9,2966252*	2,3716219	,002
	5	-23,8735687*	2,3716219	,000
	6	-24,0924030*	2,3716219	,000
3	1	13,3848433*	2,3716219	,000
	2	8,5180362*	2,3716219	,006
	4	-,7785890	2,3716219	1,000
	5	-15,3555325*	2,3716219	,000
	6	-15,5743668*	2,3716219	,000
4	1	14,1634322*	2,3716219	,000
	2	9,2966252*	2,3716219	,002
	3	,7785890	2,3716219	1,000
	5	-14,5769435*	2,3716219	,000
	6	-14,7957778*	2,3716219	,000
5	1	28,7403757*	2,3716219	,000
	2	23,8735687*	2,3716219	,000
	3	15,3555325*	2,3716219	,000

	4	14,5769435*	2,3716219	,000
	6	-,2188343	2,3716219	1,000
6	1	28,9592100*	2,3716219	,000
	2	24,0924030*	2,3716219	,000
	3	15,5743668*	2,3716219	,000
	4	14,7957778*	2,3716219	,000
	5	,2188343	2,3716219	1,000

Table 3. Post hoc comparison Bonferroni. * indicates that the mean difference is significant at the 0.05 level

In order to build a prediction model of the level of mycorrhization, we used regression analysis with linear, quadratic or cubic models (Figure 5 and Table 4). Our results show that the most satisfactory R^2 (0.687) is obtained with the cubic model expressed in the form:

$$y_{index\,t} = a + b_1 x + b_2 x^2 + b_3 x^3 \qquad [2]$$

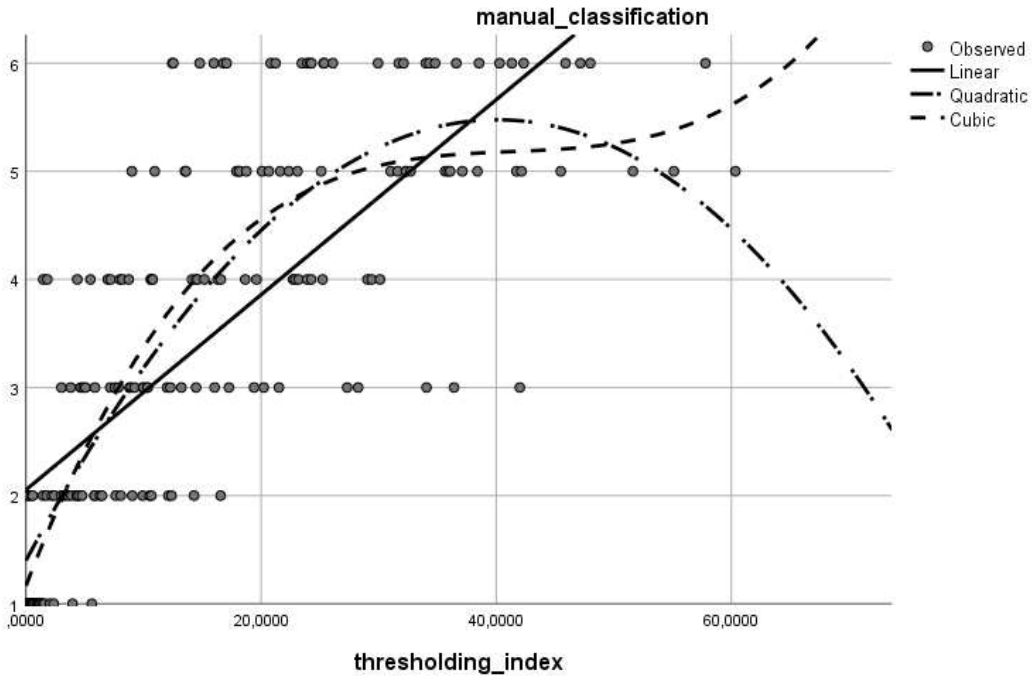


Figure 5. Curve fit with linear, quadratic and cubic regression models t index

Dependent Variable: manual_classification

	Model Summary					Parameter Estimates			
Equation	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	0,560	226,301	1	178	0,000	2,054	0,090		
Quadratic	0,673	181,823	2	177	0,000	1,401	0,203	-0,003	
Cubic	0,687	128,714	3	176	0,000	1,168	0,282	-0,007	0,000054

The independent variable is thresholding_index.

Table 4. Estimation of the parameters of the predictive model with *m*-index

189

190

Machine learning analysis

As a last and most advanced approach to image analysis, we decided to test a machine learning system using the Zeiss Zen Intellesis application (Carl Zeiss Microscopy GmbH Jena, Germany https://www.zeiss.com/microscopy/int/products/microscope-software/zen-intellesis-image-segmentation-by-deep-learning.html).

Machine learning is a branch of artificial intelligence that solves tasks using algorithms that are capable of learning from experience (training), without being explicitly programmed for a specific task (https://github.com/zeiss-microscopy/OAD/tree/master/Machine_Learning).

Zeiss Intellesis can apply a number of machine learning tools called *basic features*. In this study, we used *basic feature 25,* which achieves digital image segmentation by applying Gaussian, Sobel of Gaussian, Gabor and Hessian filters to the region surrounding classified pixels, to generate a final feature vector with 25 dimensions describing each pixel (https://github.com/zeiss-microscopy/OAD/blob/master/Machine_Learning/Feature_Extractors/feature_extractors.md).

This technique involves a training phase for the construction of an image segmentation model and the subsequent application of the model to the image dataset. The training phase consisted in training the software with a few initial images of roots with a high level of colonization. On each different image we focused on the discrimination between colonized areas, non-colonized areas and image background (Figure 6).

Once the training was completed, the model was applied to the entire dataset of 180 images. Intellesis outputs different quantitative parameters of the areas in square pixels as in the case of the thresholding technique explained above.

Also in this case we built a relationship index between segmented areas.

Original digital image	Segmentation

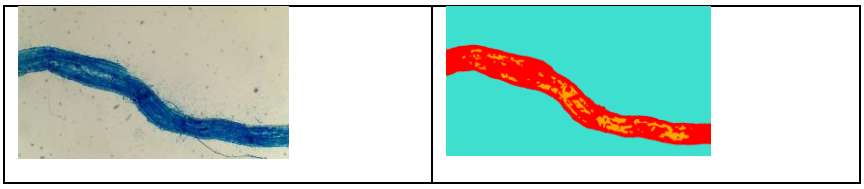


Figure 6. Segmentation in Zeiss Intellisis with machine learning. Light blue: background, yellow: mycorrhized area, red: non mycorrhized area

In this case, the *ml*-index for each image in the dataset is:

$$ml = \frac{colonized\ area}{colonized\ area + non\ colonized\ area} * 100 \qquad [3]$$

The *ml*-index was therefore used in the subsequent analyses for the construction of a prediction model to evaluate the level of mycorrhization.

In the Table 5a we show the descriptive statistics of the calculated *ml*-index, based on Intellesis machine learning, relative for each manual classification (1 – 6). Confidence intervals and qualitative descriptions for each class are shown in Table 5b.

Manual classification	N	Mean	Std. Deviation	Range
1	30	6,425657	6,2031851	18,6092
2	30	7,889029	8,6432691	41,8196
3	30	17,301326	11,7548653	40,6675
4	30	26,794820	10,5940464	49,6036
5	30	44,150405	11,3738218	37,4350
6	30	48,430012	8,8434900	33,8745
Total	180	25,165208	19,0615764	65,7459

Table 5a. *ml*-index, descriptive statistics

Classification	Qualitative description	Confidence interval (α=0,05)
1	Not mycorrhized (mycorrhization 0%)	$ml = 6{,}425657 \pm 1{,}96 * \frac{6{,}2031851}{\sqrt{30}}$
2	Slightly mycorrhized (few mycorrhization)	$ml = 7{,}889029 \pm 1{,}96 * \frac{8{,}6432691}{\sqrt{30}}$
3	Moderately mycorrhized (mycorrhization < 10%)	$ml = 17{,}301326 \pm 1{,}96 * \frac{11{,}7548653}{\sqrt{30}}$
4	Mycorrhized (mycorrhization from 11 to 50%)	$ml = 26{,}794820 \pm 1{,}96 * \frac{10{,}5940464}{\sqrt{30}}$
5	Highly mycorrhized (mycorrhization from 51 to 90%)	$ml = 44{,}150405 \pm 1{,}96 * \frac{11{,}3738218}{\sqrt{30}}$
6	Strongly mycorrhized (mycorrhization > 90%)	$ml = 48{,}430012 \pm 1{,}96 * \frac{8{,}8434900}{\sqrt{30}}$

Table 5b. Confidence interval for ml-index related to the multinomial classification

And the box plot of the descriptive statistics is described in Figure 7.

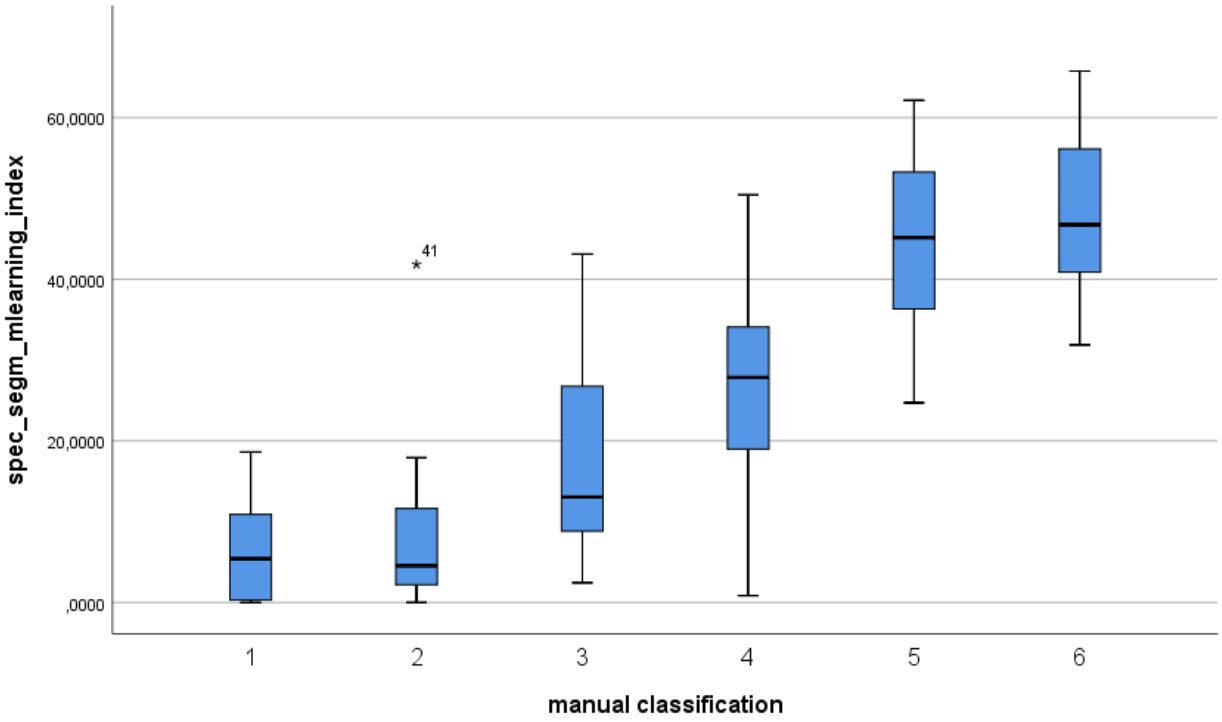


Figure 7. *ml*-index, box plot

The average value of the *ml*-index increases between classes, suggesting a good correlation with the manual classification. This was confirmed by the correlation analysis with a significant Pearson coefficient at the 0.01 two tailed level of 0.843.

Also in this case ANOVA analysis (Table 6) confirmed the statistical significance of the differences between classes. Post hoc test of multiple comparisons performed with Bonferroni correction (Table 7) further demonstrated significant differences between 13 comparisons (on a total of 15), except for class 1 vs 2 and class 5 vs 6.

ANOVA

Segmentation *ml*-index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	48474,666	5	9694,933	101,843	,000
Within Groups	16563,855	174	95,195		
Total	65038,521	179			

242 **Table 6.** ANOVA for the six groups of digital images *ml*-index

243

244

Multiple Comparisons

Dependent Variable: spec_segm_mlearning_index

Bonferroni

		Mean Difference			95% Confidence Interval	
(I) T_manual_class	(J) T_manual_class	(I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
1	2	-1,4633715	2,5191873	1,000	-8,961158	6,034415
	3	-10,8756684*	2,5191873	,000	-18,373455	-3,377882
	4	-20,3691622*	2,5191873	,000	-27,866949	-12,871376
	5	-37,7247477*	2,5191873	,000	-45,222534	-30,226961
	6	-42,0043545*	2,5191873	,000	-49,502141	-34,506568
2	1	1,4633715	2,5191873	1,000	-6,034415	8,961158
	3	-9,4122969*	2,5191873	,004	-16,910084	-1,914510
	4	-18,9057907*	2,5191873	,000	-26,403577	-11,408004
	5	-36,2613762*	2,5191873	,000	-43,759163	-28,763590
	6	-40,5409830*	2,5191873	,000	-48,038770	-33,043196
3	1	10,8756684*	2,5191873	,000	3,377882	18,373455
	2	9,4122969*	2,5191873	,004	1,914510	16,910084
	4	-9,4934938*	2,5191873	,003	-16,991280	-1,995707
	5	-26,8490792*	2,5191873	,000	-34,346866	-19,351293
	6	-31,1286860*	2,5191873	,000	-38,626473	-23,630899
4	1	20,3691622*	2,5191873	,000	12,871376	27,866949
	2	18,9057907*	2,5191873	,000	11,408004	26,403577
	3	9,4934938*	2,5191873	,003	1,995707	16,991280

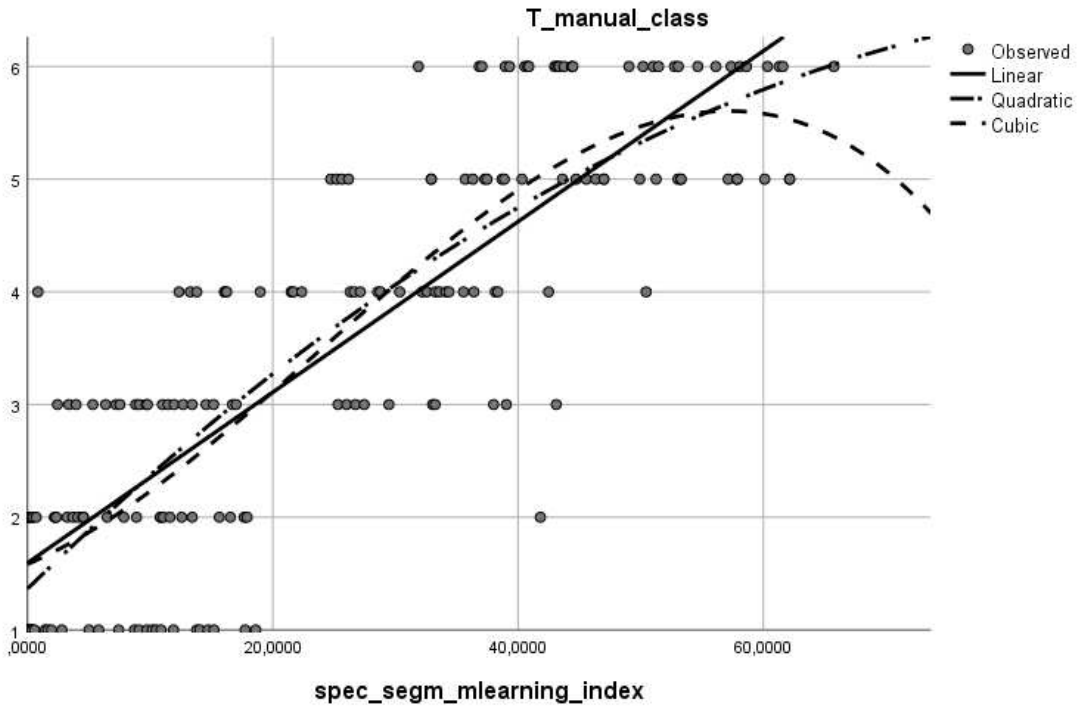
	5	-17,3555855*	2,5191873	,000	-24,853372	-9,857799
	6	-21,6351923*	2,5191873	,000	-29,132979	-14,137406
5	1	37,7247477*	2,5191873	,000	30,226961	45,222534
	2	36,2613762*	2,5191873	,000	28,763590	43,759163
	3	26,8490792*	2,5191873	,000	19,351293	34,346866
	4	17,3555855*	2,5191873	,000	9,857799	24,853372
	6	-4,2796068	2,5191873	1,000	-11,777393	3,218180
6	1	42,0043545*	2,5191873	,000	34,506568	49,502141
	2	40,5409830*	2,5191873	,000	33,043196	48,038770
	3	31,1286860*	2,5191873	,000	23,630899	38,626473
	4	21,6351923*	2,5191873	,000	14,137406	29,132979
	5	4,2796068	2,5191873	1,000	-3,218180	11,777393

*. The mean difference is significant at the 0.05 level.

Table 7. Post hoc Bonferroni comparison. * indicates that the mean difference is significant at the 0.05 level

A prediction model for the level of mycorrhization was then built using regression with linear, quadratic, cubic models (Figure 8). The most satisfactory R^2 (0.728) was obtained for the cubic model expressed in the form:

$$y_{index\ ml} = a + b_1x + b_2x^2 + b_3x^3 \qquad [4]$$



253

254 **Figure 8.** Curve fit with linear, quadratic and cubic regression models *ml*-index

In the Table 8 below, R squared for ml-models are resumed.

Model Summary and Parameter Estimates

Dependent Variable: T_manual_class

	Model Summary					Parameter Estimates			
Equation	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	0,710	436,258	1	178	,000	1,595	0,076		
Quadratic	0,720	227,308	2	177	,000	1,365	0,106	-0,001	
Cubic	0,728	156,956	3	176	,000	1,583	0,047	0,002	-0,00002856

The independent variable is spec_segm_mlearning_index.

255 **Table 8.** Estimation of the parameters of the predictive model with *ml*-index

256

257 We summarize in Table 9 the results of the statistical analysis performed comparing thresholding

258 and machine learning training using the best predictive cubic models.

259

Automatic methods of mycorrhization prediction

	Thresholding	Machine learning training
Pearson correlation index	0,748	0,824

R^2		
Linear	0,56	0,710
Quadratic	0,673	0,720
Cubic	0,687	0,728
Bonferroni	12/15	13/15
Statistically significant comparisons on total comparisons		

Table 9. Methods comparison

The best prediction model is the cubic model with an R^2 reaching 0.728 compared to the 0.687 achieved by the thresholding.

Discussion

The degree of root colonization is a major feature in most studies of AM biology and field applications. Assessing the extent of fungal development inside the host root system provides a direct indication of symbiosis development and functioning. Indeed, quantitative estimates of AM colonization are a pre-requisite for all studies of symbiotic promotion of plant nutrition and growth [12].

Two major approaches are commonly used to quantify fungal presence in root samples: a quantitative analysis of fungal marker genes, providing a direct indication of fungal DNA abundance in total extracted DNA, and microscopic investigation of a representative sample of root segments, assessing the extension of colonized areas, the spread of intraradical hyphae and the relative abundance of arbuscules or vesicles in the whole root system [13]. Overall, the molecular approach provides a relatively fast and reliable determination of fungal presence but cannot address important questions, such as the relative abundance of different intraradical structures (e.g. arbuscules and hyphae), which limits its interest when studying symbiosis functionality, unless used in combination with functional markers, such as plant P transporters that are only expressed in arbusculated cells [3].

By contrast, morphological methods, albeit time consuming, provide more direct information on AM functionality and remain of very common use. In particular, those described by Trouvelot and

283 Brundrett [6, 14]. The two methods are based on non-vital staining of fungal cell wall in mycorrhizal
284 roots, followed by the manual ranking of a representative sample of root segments and a statistical
285 analysis of the results, to extrapolate whole root system estimates.
286 Even if manual ranking is based on rather objective traits, morphological methods perform best
287 when the same expert operator analyses all the samples that need to be compared [15]. Another critical
288 limitation is the high amount of time required for manual analyses of sufficiently large samples to
289 support solid statistics and conclusions. In the present study, we used a manual method based on
290 Trouvelot et al (1986) [6] as a benchmark to compare the reliability of three semi-automated methods
291 based on image analysis.
292 A preliminary binary segmentation approach was initially used, to provide a raw distinction
293 between mycorrhizal and non-mycorrhizal root images. The good correlation of this binary
294 segmentation method with the manual classification opened the way to the subsequent thresholding
295 analysis. This approach used the gradient of pixel brightness (inversely related to lactic blue
296 staining) as an indicator of intraradical fungal development. The thresholding model generated 6
297 classes of mycorrhization, with a good correlation with manual scores of the corresponding images
298 and can therefore be considered as a reliable and cheap method for a rapid screening of root
299 samples and their classification in a range of mycorrhization intensity values. A few critical aspects
300 should anyway be considered. One major limitation of the thresholding method is the variability of
301 brightness range between images: different dyes, optical setups, root translucence and the presence
302 of additional microorganisms (such as bacteria, algae, endophytic fungi, invertebrates) especially in
303 field samples, can generate very diverse patterns that are impossible to discriminate only based on
304 pixel brightness. In addition, the method is strongly affected by image background noise and the use
305 of images acquired at different magnification levels. Lastly, the segmentation process can only be
306 set *ex ante*, during macro editing, without any subsequent correction by the operator.
307 The machine learning method, based on the Zeiss Zen Intellesis application, resulted to be the most
308 efficient. In this case, individual fungal structures such as intracellular hyphae and arbuscules were

manually selected during the training phase to generate a model that the software then applied to the whole data set. This method identified 6 classes of mycorrhization intensity, achieving the best correlation (Pearson correlation coefficient 0.824) with manual analysis.

Importantly, the training process was relatively quick (it required 50 minutes overall) and resulted to be effective even when using a limited number of images (10 images). Lastly, the use of machine-learning identification of shapes and objects allowed a reliable discrimination between equally stained structures such as extraradical hyphae and arbuscules, a capability that our image thresholding methods could never achieve.

A critical factor emerging from the machine learning-based method is the need to train the software by an expert operator. Nevertheless, the training phase generates a reference file containing all the image analysis data that can subsequently be shared with other researchers. This opens a new perspective where the expertise of a few operators could be made available to the whole community of researchers by simply sharing a file. Furthermore, the training file could continuously be implemented and stored online, thus becoming a common resource and a unifying instrument for the quantification of root colonization intensity in AM.

The current model accuracy should be further implemented to achieve the level of detail of manual methods. Our method can currently only assign a mycorrhization class to each image, without reliably discerning between arbuscules and other intraradical structures such as vesicles or hyphal coils. Such limitations could be overcome by improving the number of images used for training or modifying the image analysis algorithm. The latter option is anyway currently not possible, due to the commercial nature of the software.

In conclusion, it should be stressed that this is one of the first attempt to use a machine learning approach for the evaluation of the level of root colonization and the current results represent a promising base for future improvements.

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

The data-sets generated and analysed during the current study are available in the Figshare repository:
Binary segmentation non myc DOI https://doi.org/10.6084/m9.figshare.14679642
Thresholding and machine learning segmentation - mycorrhized roots DOI
https://doi.org/10.6084/m9.figshare.14679729
Thresholding and machine learning segmentation – non mycorrhized roots DOI
https://doi.org/10.6084/m9.figshare.14679684

Competing interests

The authors declare that they have no competing interests.

Funding

Ministero dell'Istruzione, dell'Università e della Ricerca: PhD fellowship to AC Università degli Studi di Torino

Authors' contributions

IS designed the image analysis approach, performed image analyses and statistical analyses, and wrote the text; AC contributed to design the image analysis approach, performed image analyses and wrote the text; MN contributed to image analyses and writing of the text; MP contributed to image analyses; AG contributed to the experimental design, and wrote the text. All authors have read and approved the final manuscript.

Acknowledgements

We are grateful to Carl Zeiss Microscopy GmbH for the licensing agreement of the Zeiss Zen Intellesis software and to Alessandro Cometta and Francesco Biancardi (Carl Zeiss Spa, Milan, Italy) for fruitful discussion. This work was supported by MIUR (PhD fellowship to AC).

References

1. Rich, M. K., Nouri, E., Courty, P. E. & Reinhardt, D. Diet of Arbuscular Mycorrhizal Fungi: Bread and Butter? *Trends in Plant Science* vol. 22 652–660 (2017).

2. Gutjahr, C. & Parniske, M. Cell and Developmental Biology of Arbuscular Mycorrhiza Symbiosis. *Annual Review of Cell and Developmental Biology* **29**, 593–617 (2013).

3. Ferrol,N. and Lanfranco,L. (2020) Arbuscular mycorrhizal fungi: methods and protocols. New York, Springer US.

4. McGONIGLE, T. P., MILLER, M. H., EVANS, D. G., FAIRCHILD, G. L. & SWAN, J. A. A new method which gives an objective measure of colonization of roots by vesicular-arbuscular mycorrhizal fungi. *New Phytologist* **115**, 495–501 (1990).

5. Novero, M. *et al.* Dual requirement of the LjSym4 gene for mycorrhizal development in epidermal and cortical cells of Lotus japonicus roots. *New Phytologist* **154**, 741–749 (2002).

6. Trouvelot,A. et al. (1986) Mesure du taux de mycorhization VA d'un système radiculaire. Recherche de méthodes d'estimation ayant une signification fonctionnelle. In: Gianinazzi-Pearson V, Gianinazzi S (Eds) Physiological and Genetical Aspects of Mycorrhizae INRA Press, Paris, 217-221.

7. Vierheilig, H. & Piché, Y. A modified procedure for staining arbuscular mycorrhizal fungi in roots. *Zeitschrift für Pflanzenernährung und Bodenkunde* **161**, 601–602 (1998).

8. Arganda-Carreras, I. *et al.* Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification. *Bioinformatics* **33**, 2424–2426 (2017).

9. Evangelisti, E. et al (2021) Artificial intelligence enables the identification and quantification of arbuscular mycorrhizal fungi in plant roots. Biorxiv doi: https://doi.org/10.1101/2021.03.05.434067

10. Volkenandt, T., Freitag, S. & Rauscher, M. Machine Learning Powered Image Segmentation. *Microscopy and Microanalysis* **24**, 520–521 (2018).

11. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).

12. Lekberg, Y. & Koide, R. T. Is plant performance limited by abundance of arbuscular mycorrhizal fungi? A meta-analysis of studies published between 1988 and 2003. *New Phytologist* **168**, 189–204 (2005).

13. Vierheilig, H., Schweiger, P. & Brundrett, M. An overview of methods for the detection and observation of arbuscular mycorrhizal fungi in roots+. *Physiologia Plantarum* **0**, 051021083431001-??? (2005).

14. Brundrett,M.C. et al. (1996) Roots and VAM fungi. In Working with mycorrhizas in forestry and agriculture, 184-187. Aust. Centre Int. Agricult. Res., Canberra, Australia.

15. Sun, X. G. & Tang, M. Comparison of four routinely used methods for assessing root colonization by arbuscular mycorrhizal fungi. *Botany* **90**, 1073–1083 (2012).

Figures

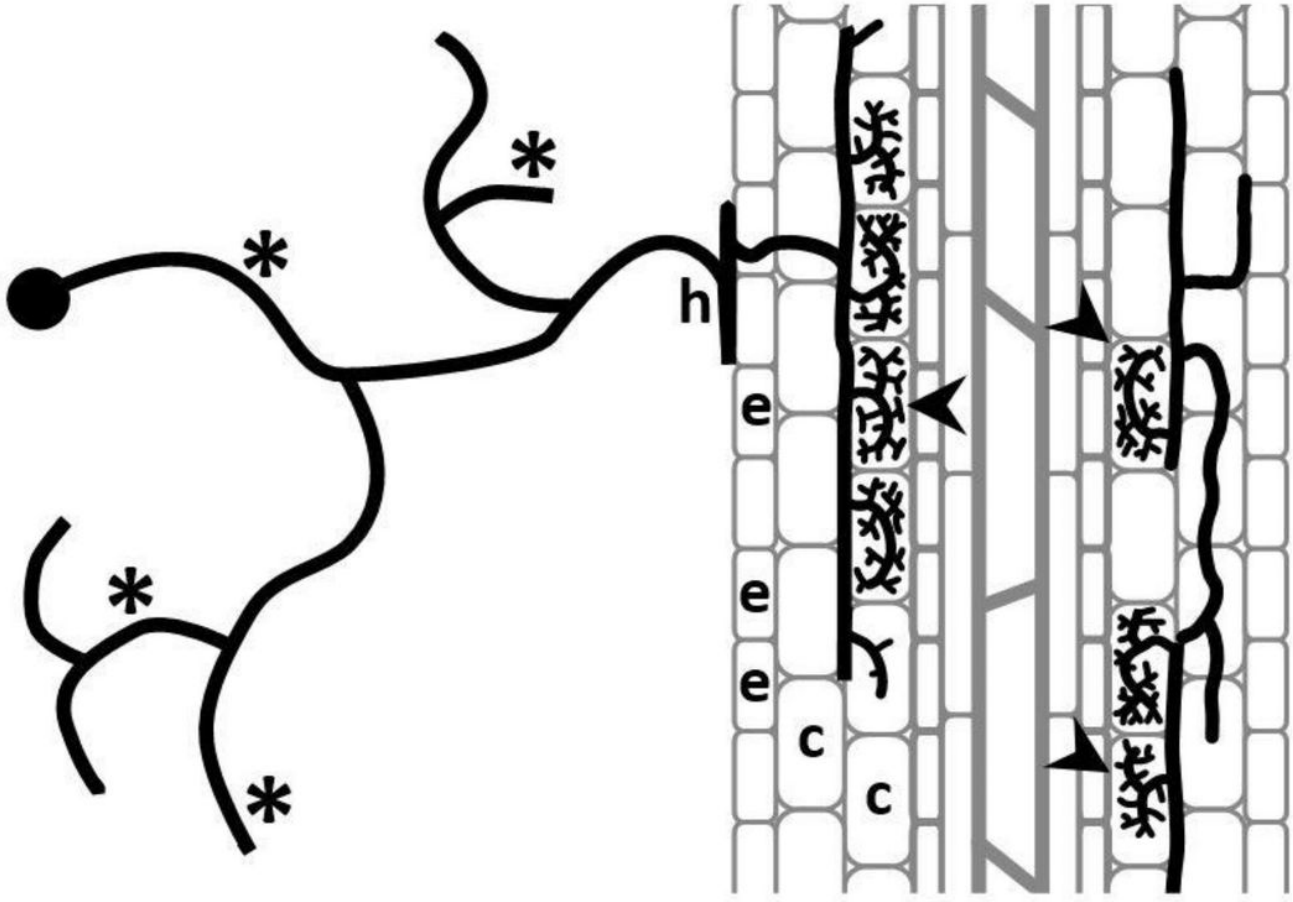


Figure 1

Schematic representation of a host root (grey) colonized by an arbuscular mycorrhizal fungus (black). The extraradical mycelium (*) explores the soil surrounding the root, while intraradical structures produced from the hyphopodium (h) penetrate root epidermal cells (e), colonizing single cortical cells (c), where they eventually develop into branched arbuscules (arrowhead), the sites of nutrient exchanges between symbionts.

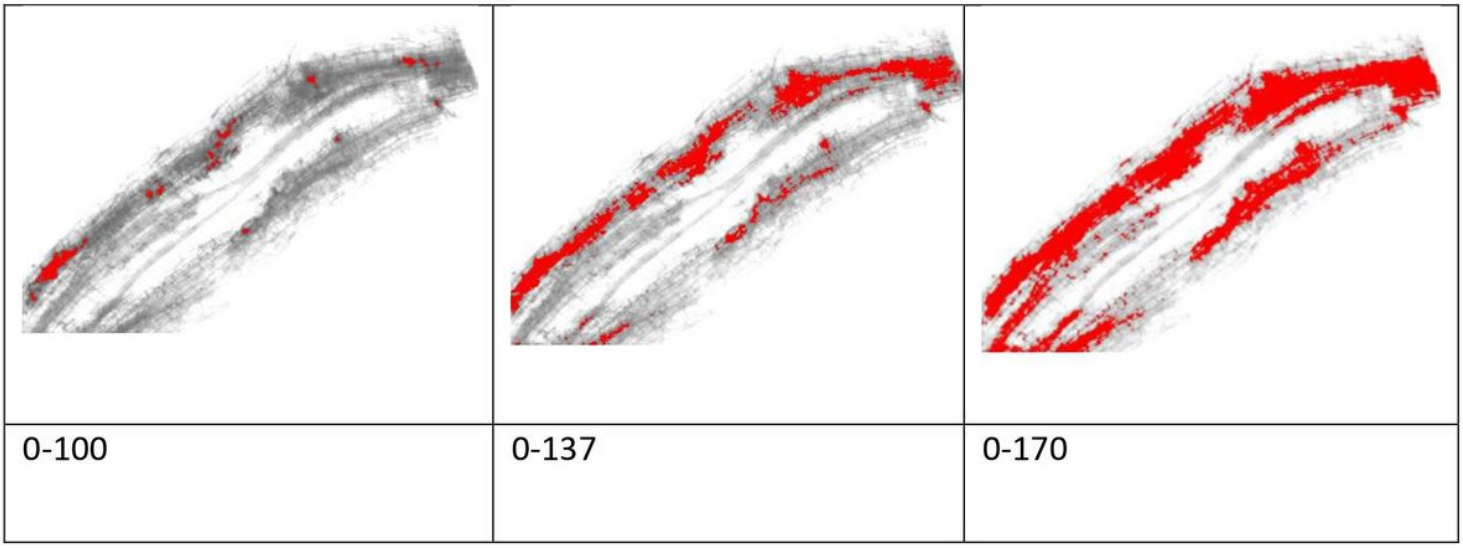


Figure 2

Brightness-based selection of pixels on the same image of a mycorrhizal root segment. Pixels are selected (in red) based on arbitrary brightness thresholds: 100, 137, 170 in a range from 0 (black) to 255 (white) using Fiji/ImageJ.

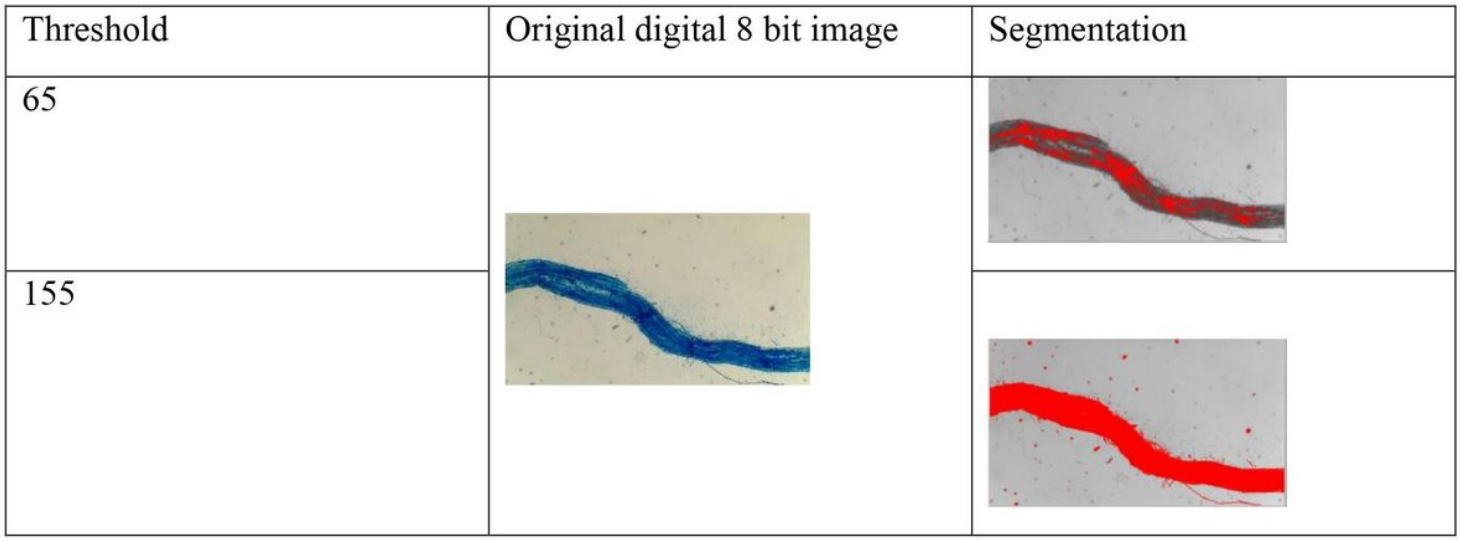


Figure 3

Image thresholding in ImageJ. The first threshold set at 65 identifies the colonized area, while the second threshold at 155 outlines with a good approximation the total area of the root section.

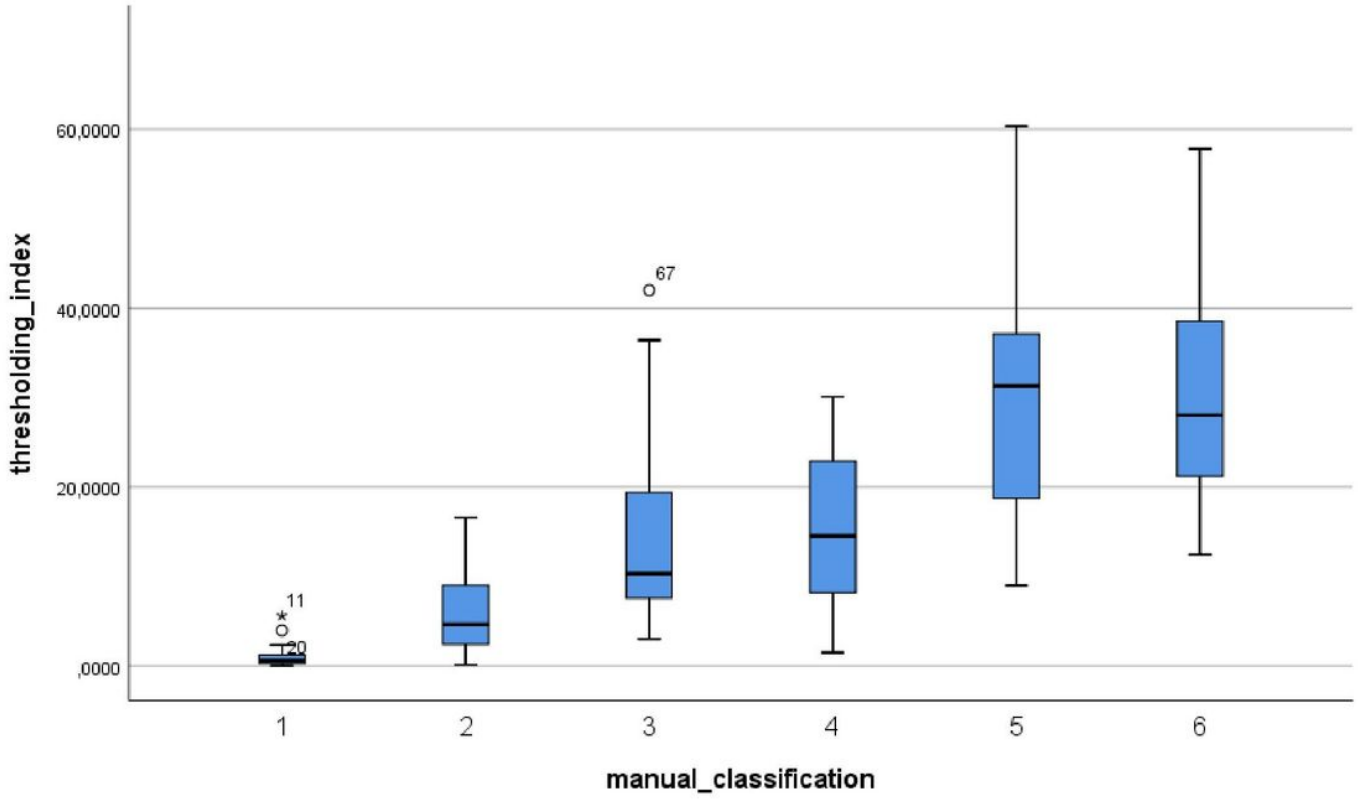


Figure 4

m-index, box plot

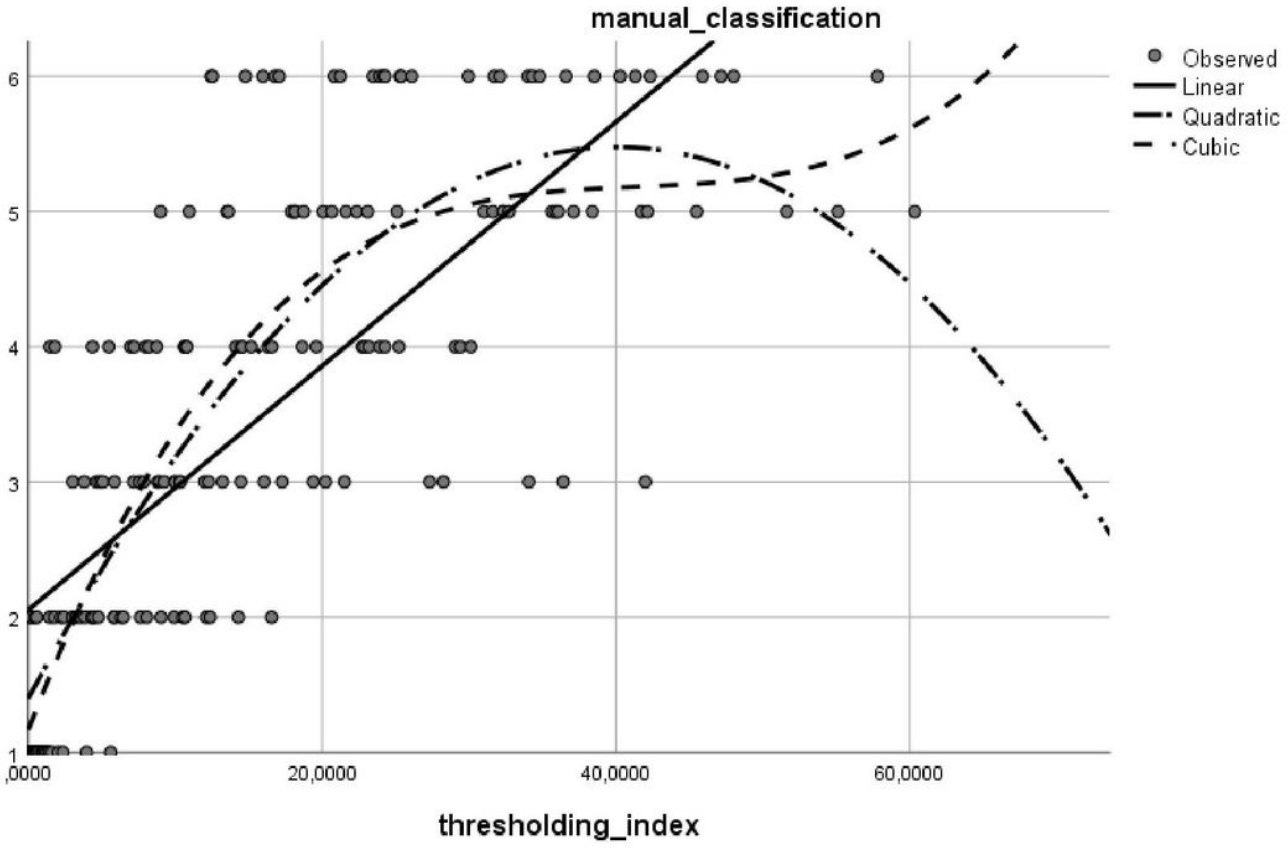


Figure 5

Curve fit with linear, quadratic and cubic regression models t index

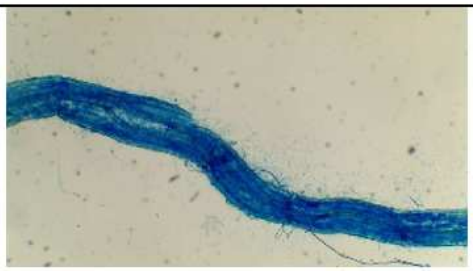
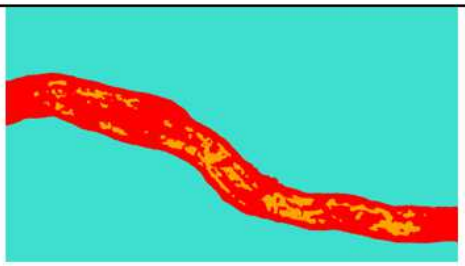
Original digital image	Segmentation
	

Figure 6

Segmentation in Zeiss Intellisis with machine learning. Light blue: background, yellow: mycorrhized area, red: non mycorrhized area

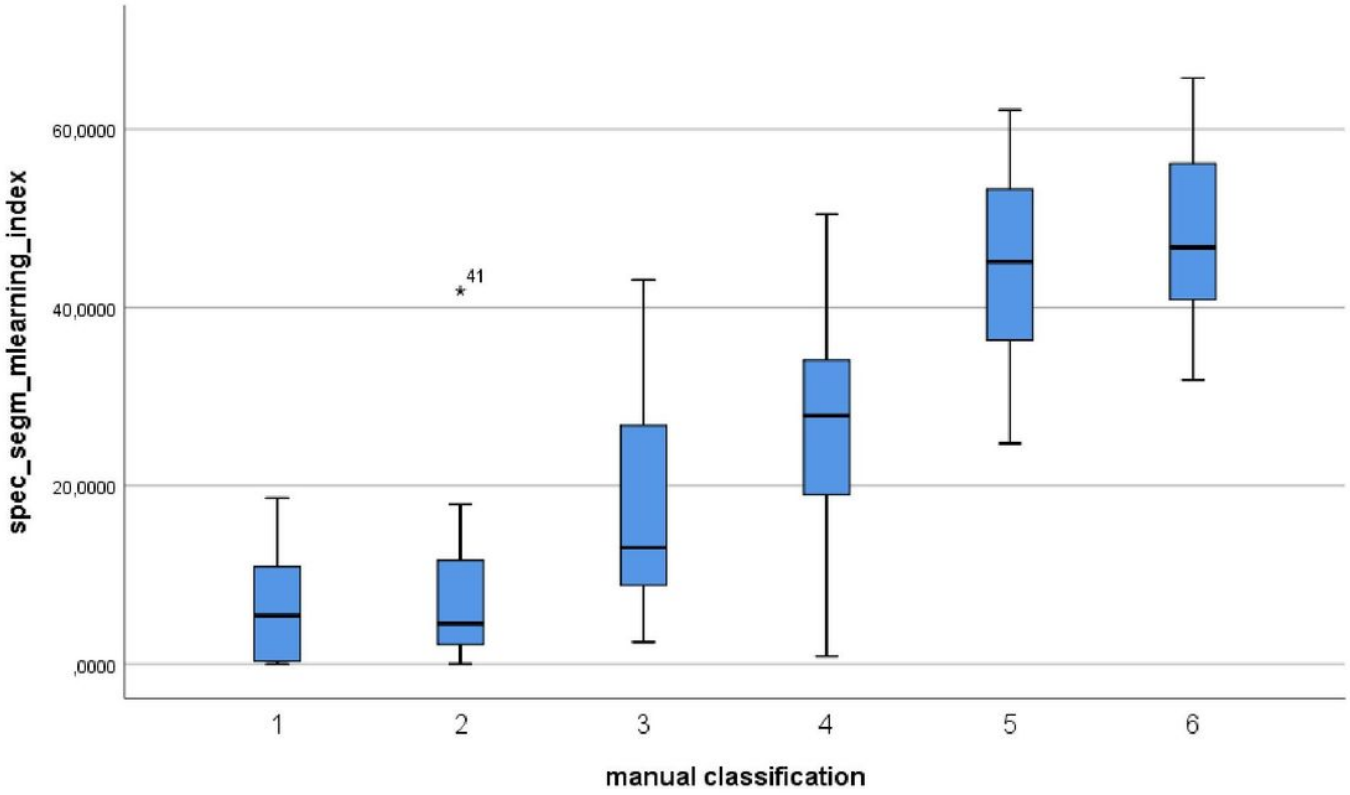


Figure 7

ml-index, box plot

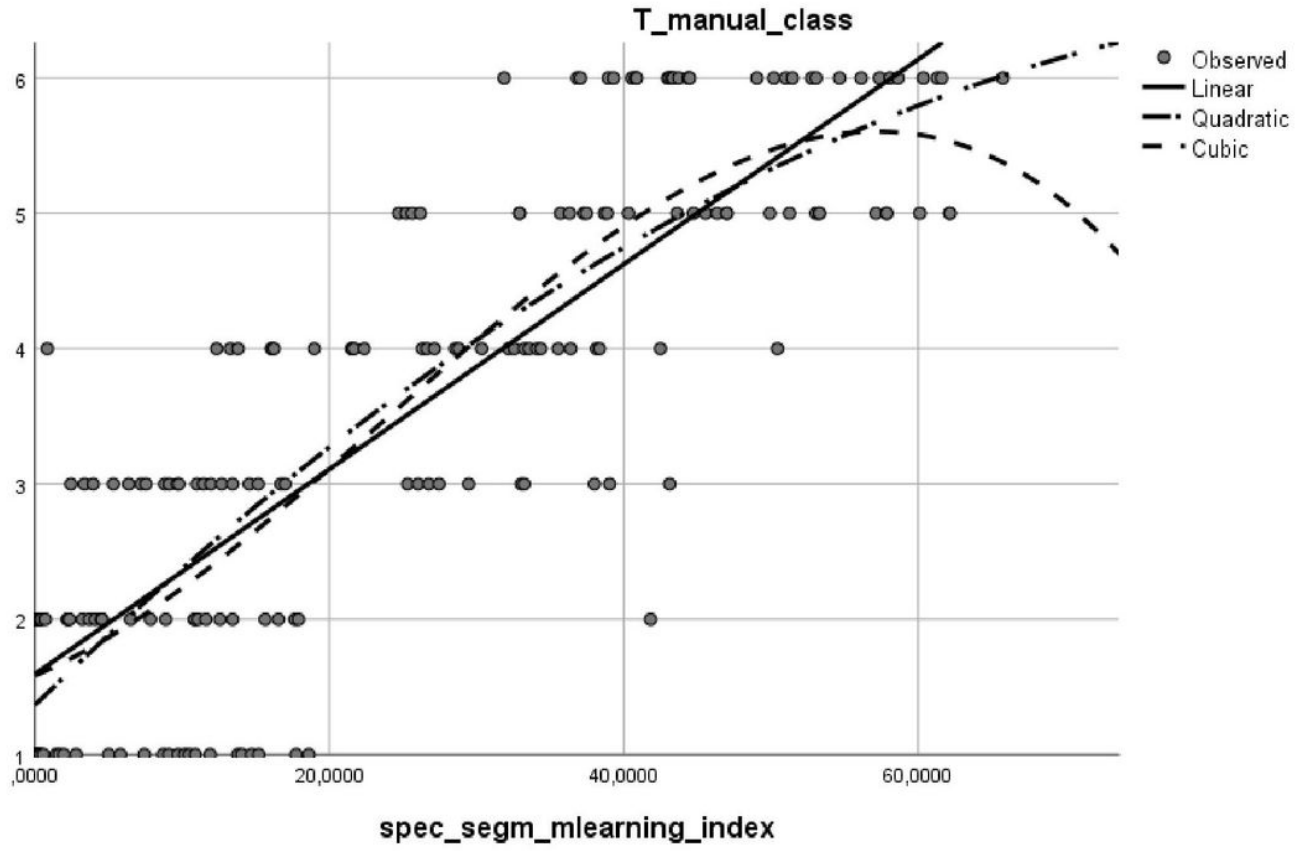


Figure 8

Curve fit with linear, quadratic and cubic regression models ml-index

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryfilesQuantifying.docx