

Monitoring root rot in flat-leaf parsley via machine vision by unsupervised multivariate analysis of morphometric and spectral parameters

Avinash Agarwal^{1*}, Filipe de Jesus Colwell², Julian Bello Rodriguez¹, Sarah Sommer¹, Viviana Andrea Correa Galvis², Tom Hill³, Neil Boonham¹, Ankush Prashar¹

¹School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK

²Crop Science R&D Division Infarm - Indoor Urban Farming B.V., The Netherlands

³Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

*Corresponding author: avinash.agarwal.1612@gmail.com

Submitted to: *European Journal of Plant Pathology*

SUPPLEMENTARY MATERIAL

Supplementary tables

Table S1. Primer sequences for strain identification

Primer name	Primer sequence	Reference
ITS1 (ITS)	5' TCCGTAGGTGAACCTGCGG 3'	White et al., 1990
ITS4 (ITS)	5' TCCTCCGCTTATTGATATGC 3'	White et al., 1990
FM66 (COX II)	5' TAGGATTTCAGATCCTGC 3'	Martin, 2000
FM58 (COX II)	5' CCACAAATTTCACTACATTGA 3'	Martin, 2000

Table S2. Correlation coefficients (*r*) of all parameters for selecting variables (*n* = 512).

	DB	Ht	Ht_m	LAn	TLA	LAI	PLA	Linc	LPD	GLI	Hue	NDVI	NPCI	PSRI	R	G	B	NIR	R/G	G/R	R+G+B	R+G-B	GMR	AGRI
Ht	0.944																							
Ht_m	0.933	0.988																						
LAn	0.034	0.058	0.029																					
TLA	0.789	0.584	0.583	0.037																				
LAI	0.789	0.584	0.583	0.037	1.000																			
PLA	0.770	0.567	0.562	0.160	0.990	0.990																		
Linc	-0.026	-0.045	-0.019	-0.994	-0.042	-0.042	-0.164																	
LPD	0.436	0.635	0.631	-0.051	0.067	0.067	0.041	0.064																
GLI	0.765	0.761	0.746	-0.003	0.631	0.631	0.608	0.011	0.378															
Hue	0.404	0.404	0.379	0.068	0.355	0.355	0.356	-0.060	0.219	0.784														
NDVI	0.838	0.801	0.793	0.094	0.753	0.753	0.738	-0.088	0.346	0.847	0.459													
NPCI	0.538	0.583	0.591	-0.160	0.331	0.331	0.284	0.162	0.306	0.590	0.051	0.589												
PSRI	-0.691	-0.650	-0.631	-0.184	-0.656	-0.656	-0.661	0.176	-0.276	-0.787	-0.706	-0.856	-0.196											
R	-0.831	-0.811	-0.801	-0.051	-0.731	-0.731	-0.712	0.042	-0.398	-0.884	-0.527	-0.978	-0.572	0.846										
G	-0.774	-0.766	-0.765	-0.034	-0.647	-0.647	-0.626	0.026	-0.394	-0.646	-0.165	-0.925	-0.596	0.707	0.911									
B	-0.812	-0.786	-0.780	-0.033	-0.726	-0.726	-0.704	0.026	-0.368	-0.823	-0.379	-0.958	-0.645	0.726	0.976	0.927								
NIR	-0.670	-0.760	-0.757	0.181	-0.381	-0.381	-0.343	-0.197	-0.603	-0.759	-0.528	-0.676	-0.500	0.623	0.774	0.692	0.703							
R/G	-0.707	-0.689	-0.671	-0.044	-0.620	-0.620	-0.606	0.035	-0.325	-0.972	-0.868	-0.778	-0.420	0.796	0.835	0.542	0.757	0.699						
G/R	0.795	0.775	0.760	0.030	0.677	0.677	0.658	-0.021	0.355	0.977	0.773	0.850	0.508	-0.812	-0.891	-0.652	-0.828	-0.741	-0.979					
R+G+B	-0.821	-0.805	-0.799	-0.040	-0.712	-0.712	-0.691	0.032	-0.397	-0.791	-0.353	-0.973	-0.614	0.777	0.980	0.972	0.984	0.739	0.715	-0.798				
R+G-B	-0.802	-0.796	-0.791	-0.047	-0.671	-0.671	-0.652	0.038	-0.415	-0.729	-0.312	-0.955	-0.558	0.805	0.950	0.986	0.932	0.753	0.642	-0.735	0.981			
GMR	-0.118	-0.141	-0.160	0.025	-0.021	-0.021	-0.009	-0.026	-0.112	0.304	0.713	-0.173	-0.234	-0.077	0.093	0.495	0.182	0.039	-0.452	0.304	0.281	0.380		
AGRI	0.310	0.268	0.249	0.035	0.347	0.348	0.347	-0.032	0.063	0.688	0.872	0.290	0.069	-0.444	-0.365	0.043	-0.277	-0.310	-0.800	0.713	-0.185	-0.078	0.877	
R+G	-0.818	-0.805	-0.800	-0.043	-0.701	-0.701	-0.681	0.034	-0.405	-0.774	-0.341	-0.971	-0.598	0.789	0.974	0.981	0.971	0.747	0.694	-0.781	0.998	0.991	0.315	-0.151

Linear correlation analysis was performed by collating the data obtained from all five intervals. Attributes (ordered as per appearance in table) - DB: Digital biomass; Ht: Mean plant height; Ht_m: Maximum plant height; LAn: Leaf angle; TLA: Total leaf area; LAI: Leaf area index; PLA: Projected leaf area; Linc: Leaf inclination; LPD: Light penetration depth; GLI: Green leaf index; NDVI: Normalised difference vegetation index; NPCI: Normalized pigment chlorophyll ratio index; PSRI: Plant senescence reflectance index; R: Red; G: Green; B: Blue; NIR: Near infrared; GMR: G minus R; AGRI: Augmented green-red index.

Table S3. Feature selection based on correlation matrix

Category	Features selected	Features omitted based on correlation
Morphometric	Digital biomass	-
	Mean plant height	Maximum plant height
	Leaf area index	Total leaf area, Projected leaf area
	Leaf inclination	Leaf angle
	Light penetration depth	-
Spectral	GLI	R/G, G/R
	Hue	-

	NDVI	R, B, R+G+B, R+G-B, R+G
	NPCI	-
	PSRI	-
	G	R+G+B, R+G-B, R+G
	NIR	-
	GMR	-
	AGRI	-

GLI: Green leaf index; NDVI: Normalised difference vegetation index; NPCI: Normalized pigment chlorophyll ratio index; PSRI: Plant senescence reflectance index; R: Red; G: Green; B: Blue; NIR: Near infrared; GMR: G minus R; AGRI: Augmented green-red index.

Table S4. Variance explained by the different PCs for PCA with morphometric features (M_all).

	2 DPI		4 DPI		7 DPI		9 DPI		11 DPI	
PC	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)
PC1	53.3291	53.3291	51.689	51.689	55.6745	55.6745	61.2114	61.2114	59.8226	59.8226
PC2	28.8266	82.1557	26.1339	77.8229	20.9983	76.6728	20.4637	81.6751	20.771	80.5936
PC3	16.1498	98.3055	19.389	97.2119	19.84	96.5128	15.1784	96.8535	15.9089	96.5025
PC4	1.5367	99.8422	2.6682	99.8801	3.3339	99.8467	2.9559	99.8094	3.3074	99.8099

Total number of input features = 5; DPI, days post inoculation; PC, principal component; VE, variance explained; CVE, cumulative variance explained.

Table S5. Variance explained by the different PCs for PCA with spectral features (S_all).

	2 DPI		4 DPI		7 DPI		9 DPI		11 DPI	
PC	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)
PC1	40.5335	40.5335	51.9955	51.9955	67.1678	67.1678	78.9839	78.9839	81.6531	81.6531
PC2	36.5647	77.0982	30.0492	82.0447	20.9077	88.0755	13.148	92.1319	12.8866	94.5397
PC3	14.1576	91.2558	11.0253	93.07	8.0229	96.0984	5.2491	97.381	3.042	97.5817
PC4	5.8231	97.0789	4.4231	97.4931	2.6628	98.7612	1.9323	99.3133	1.5542	99.1359
PC5	2.4222	99.5011	2.2524	99.7455	0.9739	99.7351	-	-	-	-

Total number of input features = 9; DPI, days post inoculation; PC, principal component; VE, variance explained; CVE, cumulative variance explained.

Table S6. Variance explained by the different PCs for PCA with morphometric and spectral features combined (MS_all).

	2 DPI		4 DPI		7 DPI		9 DPI		11 DPI	
PC	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)
PC1	35.1129	35.1129	45.7361	45.7361	58.4421	58.4421	69.4338	69.4338	70.2374	70.2374
PC2	29.6045	64.7174	21.418	67.1541	15.4759	73.918	11.1041	80.5379	11.3619	81.5993
PC3	14.1795	78.8969	9.932	77.0861	9.1713	83.0893	6.8913	87.4292	6.7867	88.386
PC4	09.5083	88.4052	7.923	85.0091	6.4097	89.499	5.1807	92.6099	4.7278	93.1137
PC5	03.7703	92.1755	7.0508	92.06	4.5911	94.0901	2.9705	95.5803	3.0952	96.2089
PC6	03.3961	95.5716	3.9307	95.9906	3.0307	97.1207	2.2295	97.8099	1.4529	97.6618
PC7	02.3397	97.9113	1.9443	97.9349	1.4911	98.6118	1.0247	98.8345	1.0957	98.7575
PC8	01.4059	99.3172	1.124	99.0589	0.7247	99.3366	0.7667	99.6012	0.7594	99.5169

Total number of input features = 14; DPI, days post inoculation; PC, principal component; VE, variance explained; CVE, cumulative variance explained.

Table S7. Variance explained by the different PCs for three best morphometric and spectral features (MS_3-3).

	2 DPI		4 DPI		7 DPI		9 DPI		11 DPI	
PC	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)
PC1	66.2759	66.2759	71.6379	71.6379	84.9924	84.9924	89.368	89.368	87.1771	87.1771
PC2	17.5467	83.8226	12.55	84.1879	7.4189	92.4113	5.6634	95.0314	8.113	95.2901
PC3	10.4774	94.3	10.559	94.7469	5.2515	97.6627	3.1368	98.1682	3.0757	98.3658
PC4	4.7694	99.0694	4.4332	99.1801	1.8037	99.4664	1.4467	99.6149	1.2446	99.6104

Total number of input features = 6; DPI, days post inoculation; PC, principal component; VE, variance explained; CVE, cumulative variance explained.

Table S8. Variance explained by the different PCs for the six best features (MS_top6).

	2 DPI		4 DPI		7 DPI		9 DPI		11 DPI	
PC	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)	VE (%)	CVE (%)
PC1	61.0562	61.0562	70.8015	70.8015	86.5222	86.5222	91.814	91.814	90.3307	90.3307

PC2	17.6113	78.6675	14.2105	85.012	6.7929	93.3151	3.594	95.408	04.7785	95.1092
PC3	15.8809	94.5484	11.0204	96.0324	5.2378	98.5529	3.4452	98.8532	03.7328	98.842
PC4	2.7112	97.2596	2.3695	98.4019	0.6446	99.1975	0.6085	99.4617	0.6586	99.5006
PC5	1.972	99.2315	1.122	99.5239	-	-	-	-	-	-

Total number of input features = 6; DPI, days post inoculation; PC, principal component; VE, variance explained; CVE, cumulative variance explained.

Supplementary figures

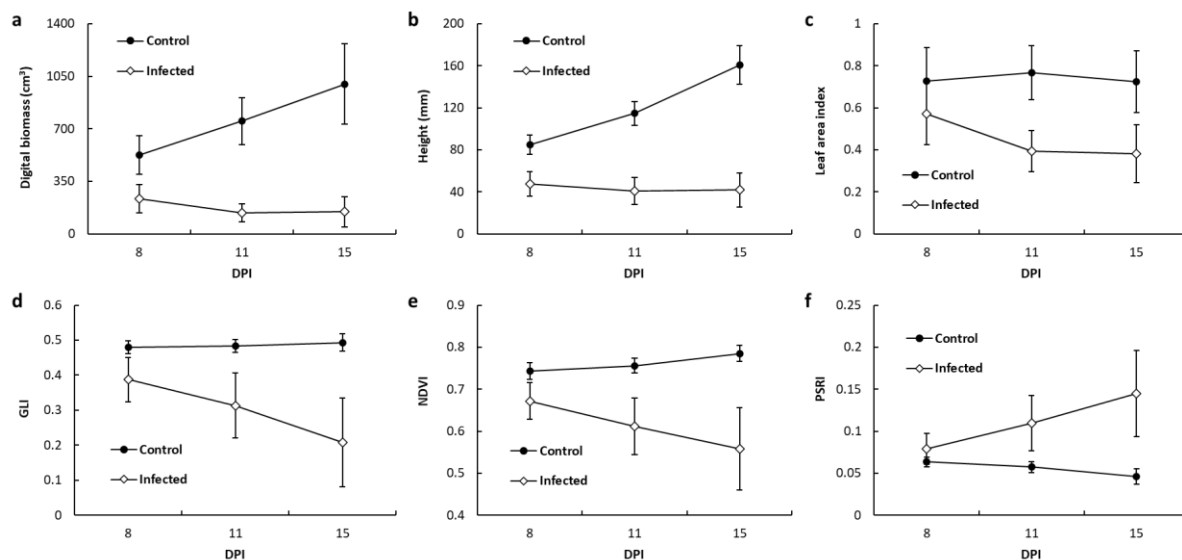


Fig. S1 Indicative morphometric (a–c) and spectral (d–f) attributes of healthy and infected plants in the preliminary trial. Values have been expressed as mean $\pm$ SD ($n = 52$). DPI, days post inoculation; GLI, Green Leaf Index; NDVI, Normalised Difference Vegetation Index; PSRI, Plant Senescence Reflectance Index.

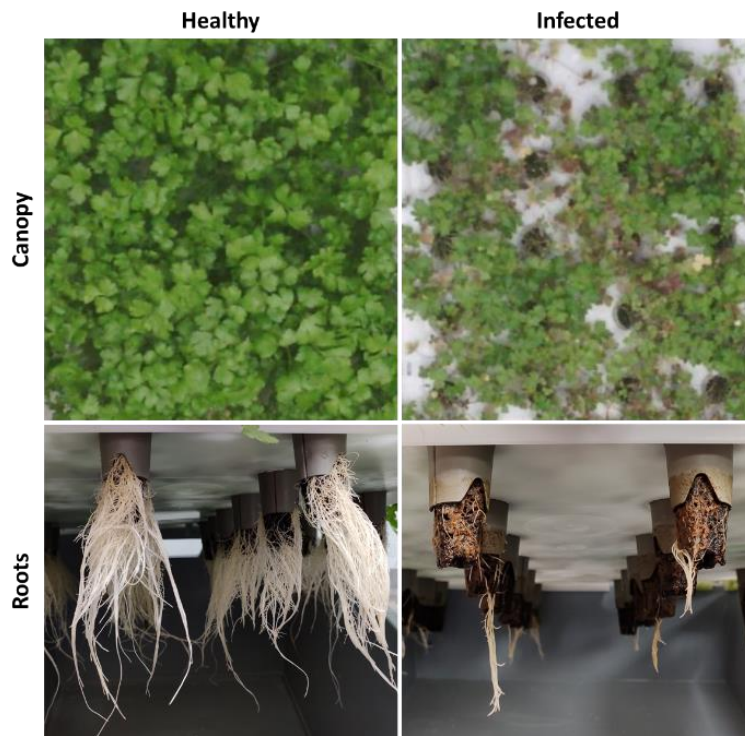


Fig. S2 Canopy and roots of healthy and infected plants at 11 DPI (main trial).

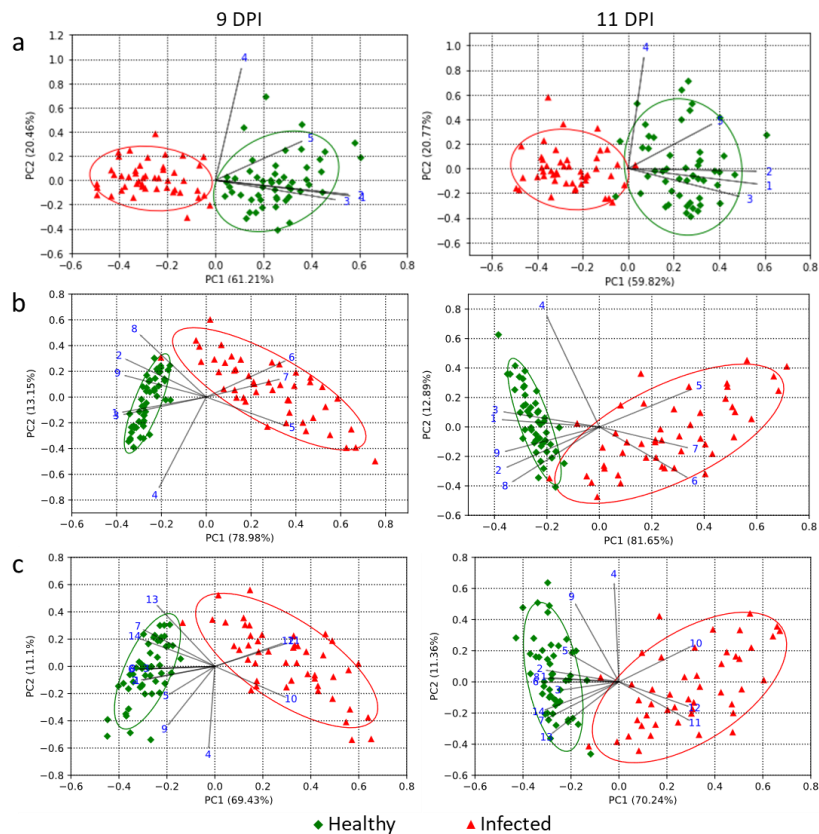


Fig. S3 PCA of morphometric and spectral attributes for healthy (green, diamonds) and infected (red, triangles) plants at 9 and 11 days post inoculation (DPI). Biplots represent the first two principal components (PCs) for the analyses with morphometric (a; five features) and spectral (b; nine features) features, and both datasets combined (c; fourteen features). Values in parentheses indicate the percentage of variance explained by the corresponding PC. Ellipses represent the 95% confidence interval for each category (healthy,

$n = 52$; infected, $n = 48$). Features are in the order: DB, Height, LAI, LInc, LPD (a, morphometric); GLI, Hue, NDVI, NPCI, PSRI, G, NIR, GMR, AGRI (b, spectral); morphometric followed by spectral in the same order (c).

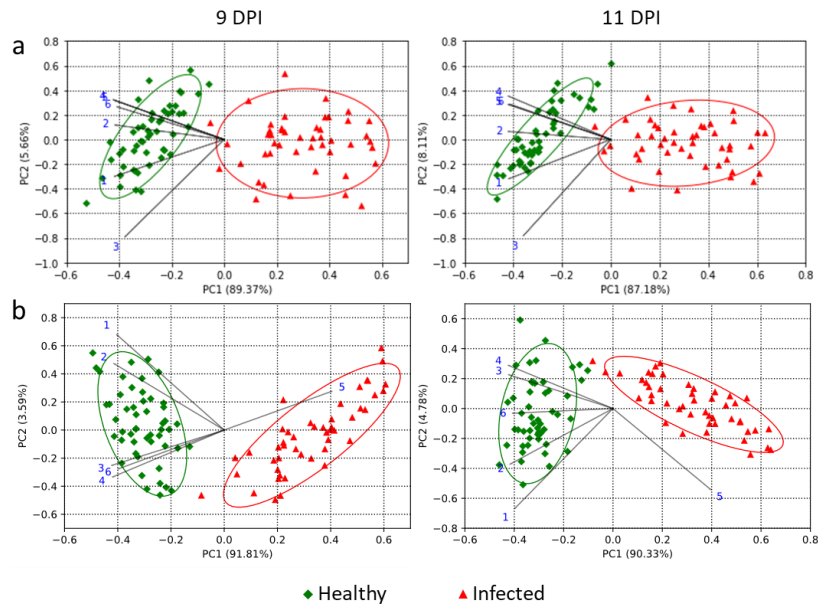


Fig. S4 PCA of morphometric and spectral attributes for healthy (green, diamonds) and infected (red, triangles) plants at 9 and 11 days post inoculation (DPI). Biplots represent the first two principal components (PCs) for the analyses performed by combining three best morphometric and three best spectral features (a; MS_3-3), and top six features (b; MS_top6). Values in parentheses indicate the percentage of variance explained by the corresponding PC. Ellipses represent the 95% confidence interval for each category (healthy, $n = 52$; infected, $n = 48$). Features are in the order: DB, Height, LAI, GLI, NDVI, AGRI (a, MS_3-3); DB, Height, GLI, NDVI, G, AGRI (b, MS_top6).