OPEN ACCESS

sensors

ISSN 1424-8220

www.mdpi.org/sensors *Research Paper, Review, Communication (Type of Paper)*

Combining a Standardized Growth Class Assessment, UAV Sensor Data, GIS Processing, and Machine Learning Classification to Derive a Correlation with the Vigour and Canopy Volume of Grapevines

Supplementary materials:

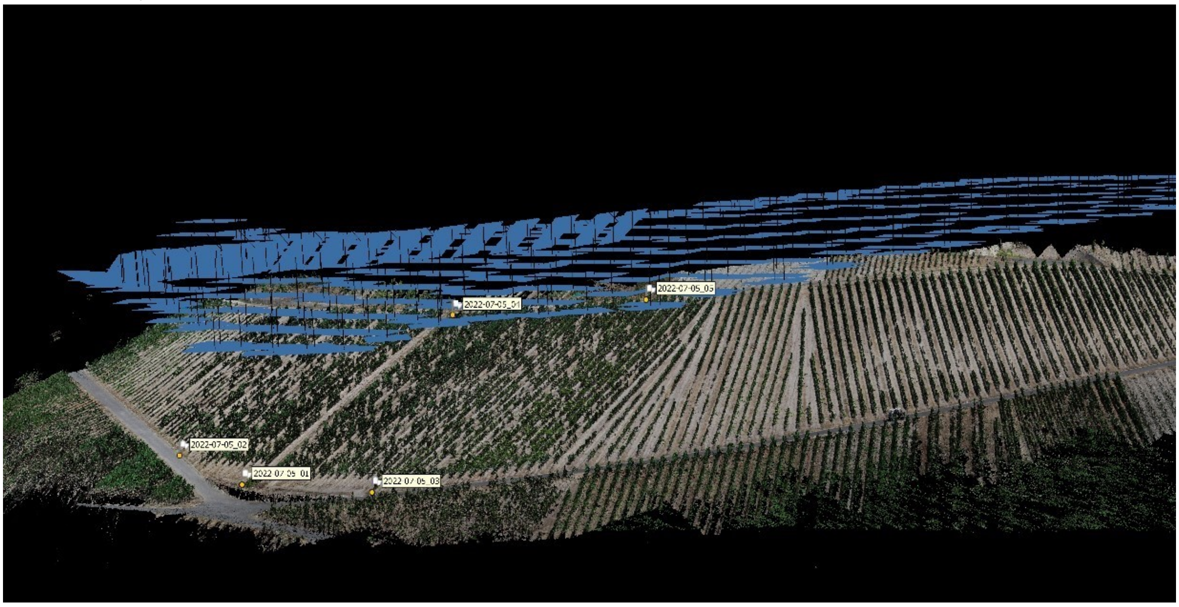

Figure 1: Generated 3D orthomosaick of the investigation area during aerotriangulation process in the software Agisoft Metashape Professional (version 1.7.0).

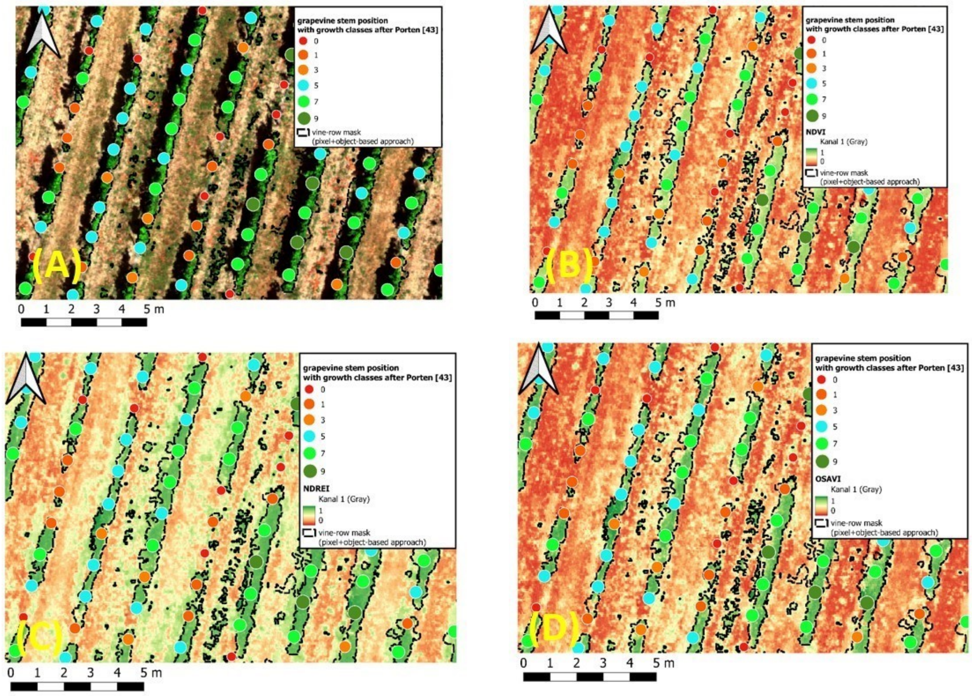

Figure 2. Color-coded growth class categorized grapevine stems according to the assesment after Porten [43], overlying (A) True-color orthomosaic with vine-row mask outline (B) NDVI with vine-row mask outline (C) NDREI with vine-row mask outline (D) OSAVI with vine-row mask outline. Layouts created with QGIS 3.22.

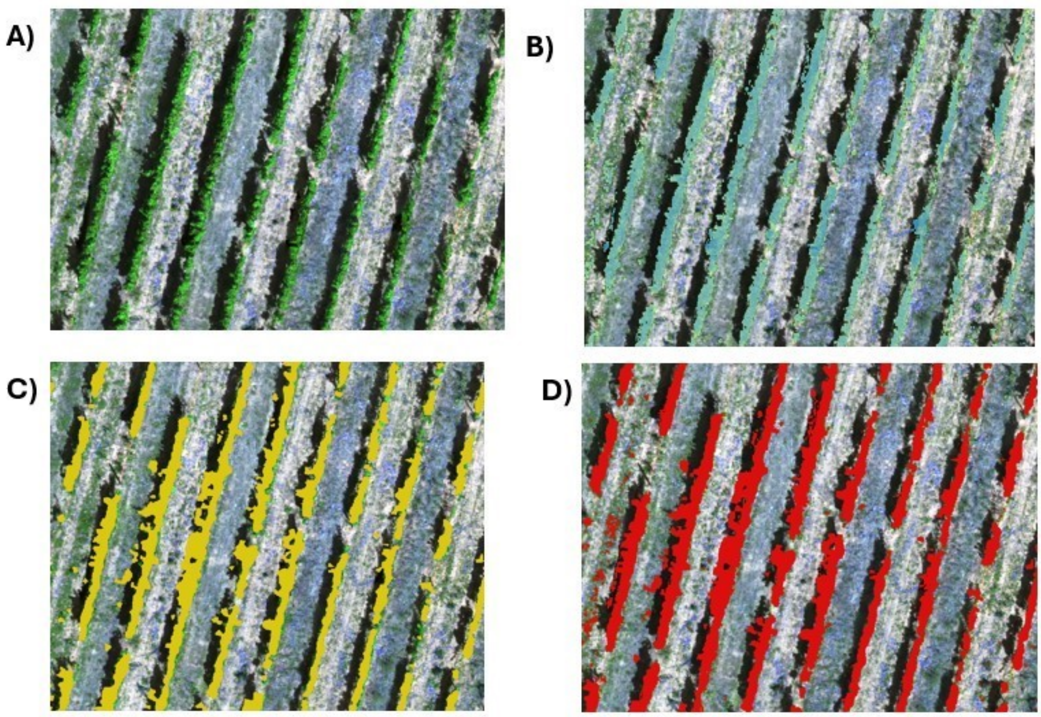

Figure 3. Multispectral Orthomosaick A), various extraction masks: based on OSAVI with Threshold B), OBIA (C) and a combined mask D) (B+C). Layouts created with QGIS 3.22.

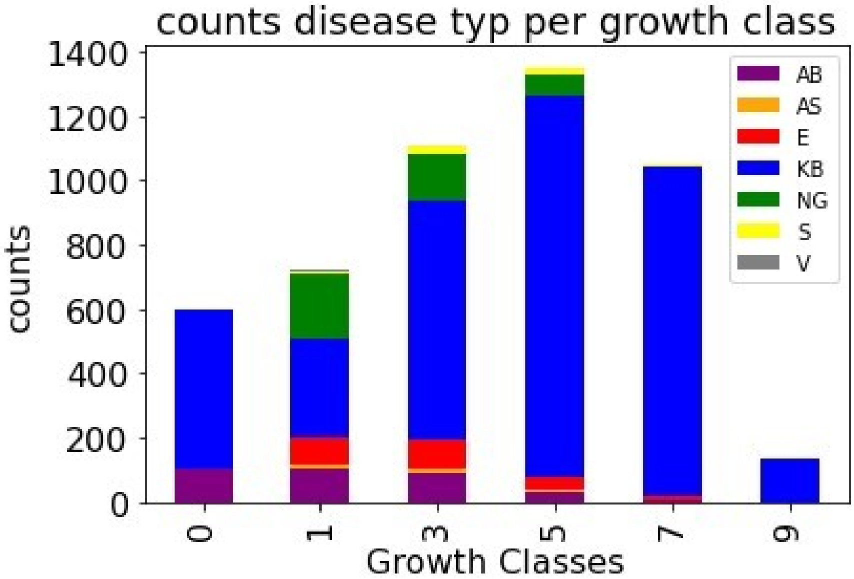

Figure 4. Frequency distribution of the infected and non-infected grapevines grouped by each growth class according to the assesment after Porten [43]. AB- Other infection, AS-, E-ESCA, KB- No infection, NG- New planted, S- Bois Noir, V- Virus infection. The statistics include data from different data acquisition dates. Like that in the main publication these were not further used for machine learning model predictions of the growth classes, respectively vigour.

Table 1: Frequency statistics in % of disease type grouped by growth classes after Porten [43] in percent for ground truth data of numerous dates within the investigation area. Statistics created with RStudio version 4.3.3.

Growth Class	*Other infection (%)*	*E/ S (ESCA+ Bois Noir) (%)*	*E (ESCA)*	*Not Infected*	*N (newly planted)*	*S (Bois Noir)*	*V- Virus*
0	30.56	0.00	0.43	12.72	0.00	0.00	0.00
1	32.10	42.42	34.33	7.89	49.16	11.29	66.67
3	28.09	39.39	38.63	19.05	35.56	33.87	33.33
5	8.64	18.18	18.88	30.44	15.27	40.32	0.00
7	0.62	0.00	7.30	26.25	0.00	14.52	0.00
9	0.00	0.00	0.43	3.52	0.00	0.00	0.00

Table 2. Mean value of infection strength for each growth class after Porten [43]. Statistic created in QGIS 3.22 with statistics by categories tool.

Growth Class	count	unique	min	max	range	sum	mean (IS)
0	576	1	0	0	0	0	0.0
1	711	6	0	4.5	4.5	490.5	0.69
3	1137	6	0	4.5	4.5	517.5	0.46
5	1392	6	0	4	4	210	0.15
7	993	4	0	4	4	75	0.08
9	132	2	0	3	3	9	0.07

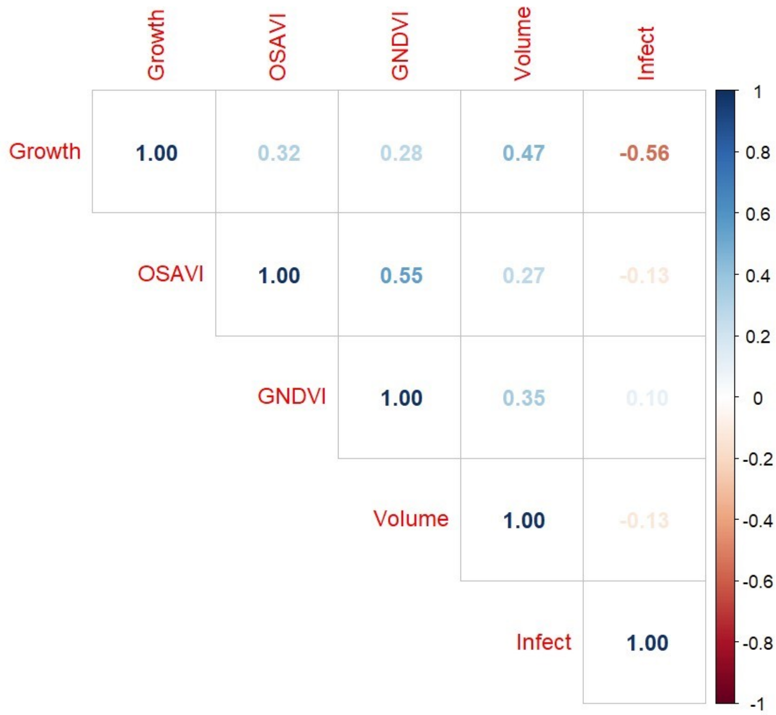

Figure 5. Correlation- matrix of the growth classes (growth) after Porten [43], the VI and the infection strength (infect- IS) with correlation coefficients after Pearson filtered for all vines with at least infection strengthn (IS) of 0.5. It can be seen that in the category of infected vines, is a pronounced negative correlation between infection strength and growth classes after Porten [43]. Plot created with RStudio version 4.3.3.

Table 3: Descriptive Statistics of OSAVI grouped by growth classes after Porten [43]. Statistic created in QGIS 3.22 with statistics by categories tool.

Growth Class	count	MW	Median	Stabw.	Minimum	Maximum
0	192	0.63	0.67	0.15	0.13	0.85
1	237	0.67	0.69	0.10	0.21	0.86
3	379	0.72	0.73	0.08	0.16	0.88
5	464	0.77	0.77	0.04	0.62	0.86
7	331	0.80	0.81	0.04	0.65	0.87
9	44	0.81	0.82	0.03	0.73	0.85

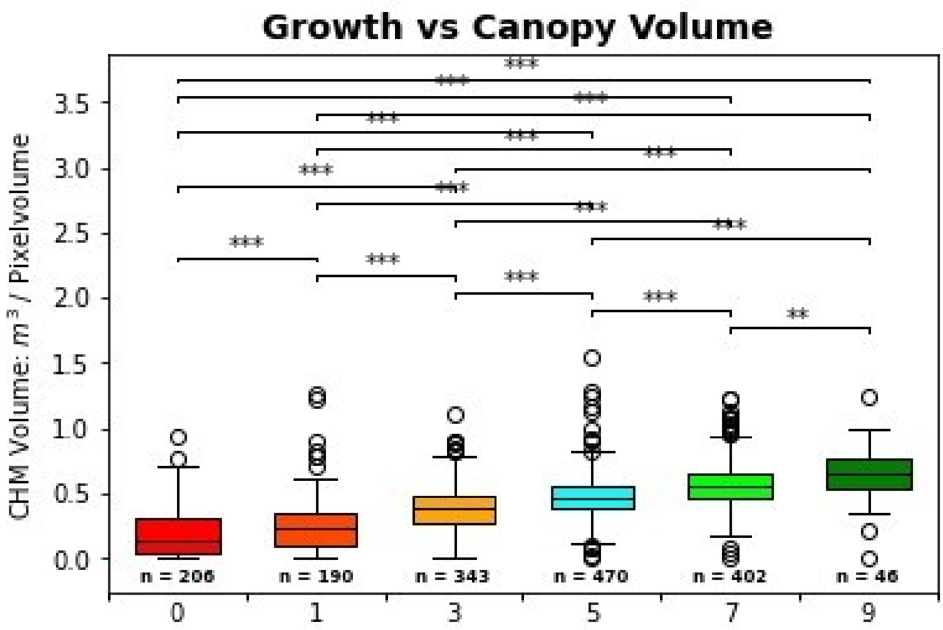

Figure 6. Growth class grouped boxplots after Porten [43] for vine volume estimated from CHM (Canopy Height Model) (CHM volume) with significance stars *) generated from the Mann-Whitney-U-test with p- value significance classes *<0.5, >0.1 (weak significance), **<0.1, >0.01 (intermediate significance), ***<0.001 (strong significance). Plot created in Python 3.9 with matplotlib.

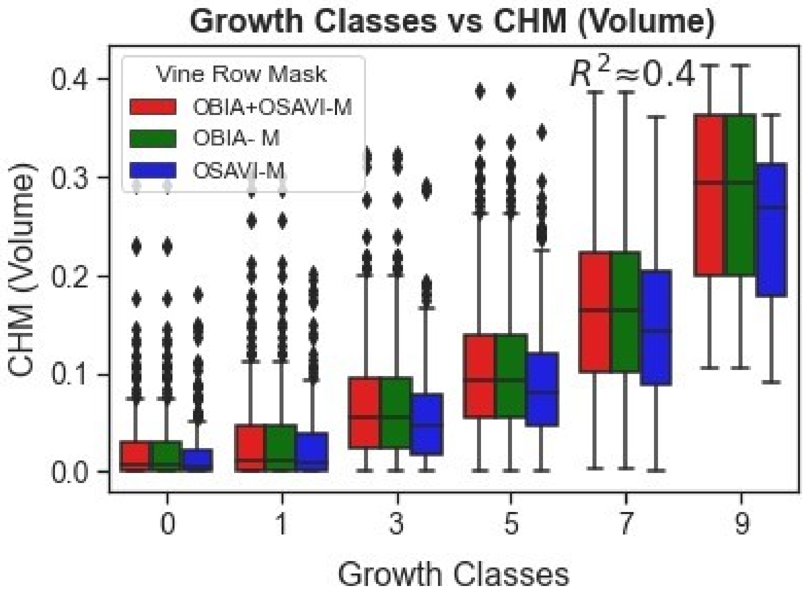

Figure 7. Growth class grouped boxplots after Porten [43] for vine volume estimated from CHM (Volume CHM) for the individual extraction masks (OSAVI-M-mask, OBIA-M-mask and the combined OBIA+OSAVI-M-mask). Plot created in Python matplotlib.

Figure 8. Correlation- matrix of the growth classes after Porten [43] and the spectral features (Vegetation Indices) with correlation coefficients after Pearson. Plot created with RStudio version 4.3.3.

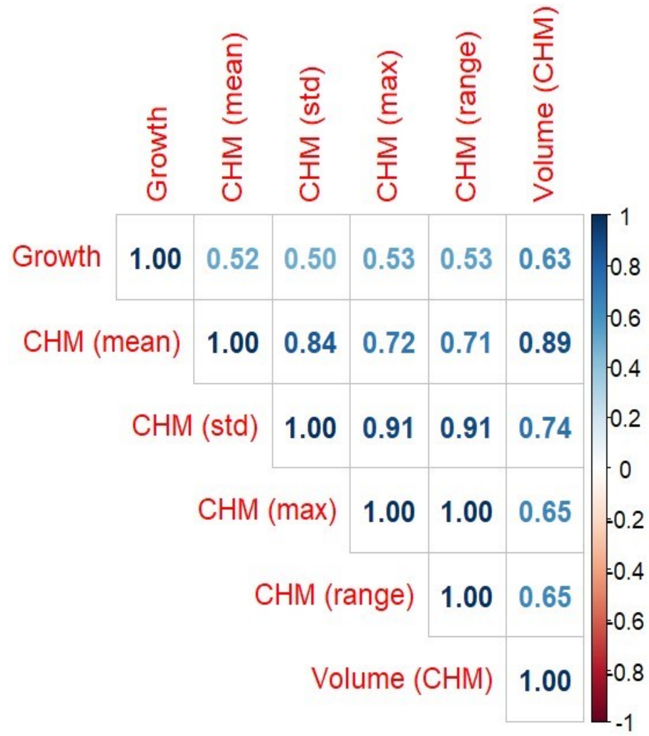

Figure 9. Correlation- matrix of the growth classes after Porten [43] and structural features with correlation coefficients after Pearson. Plot created with RStudio version 4.3.3.

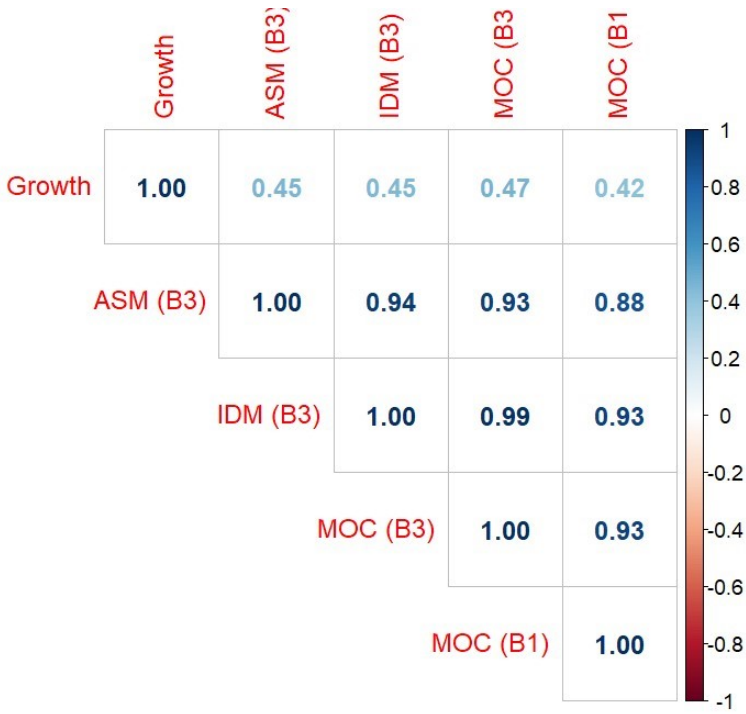

Figure 10. Correlation- matrix of the growth classes after Porten [43] and texture features (ASM (B3), MOC (B3), MOC (B1) with correlation coefficients after Pearson. Plot created with RStudio version 4.3.3.

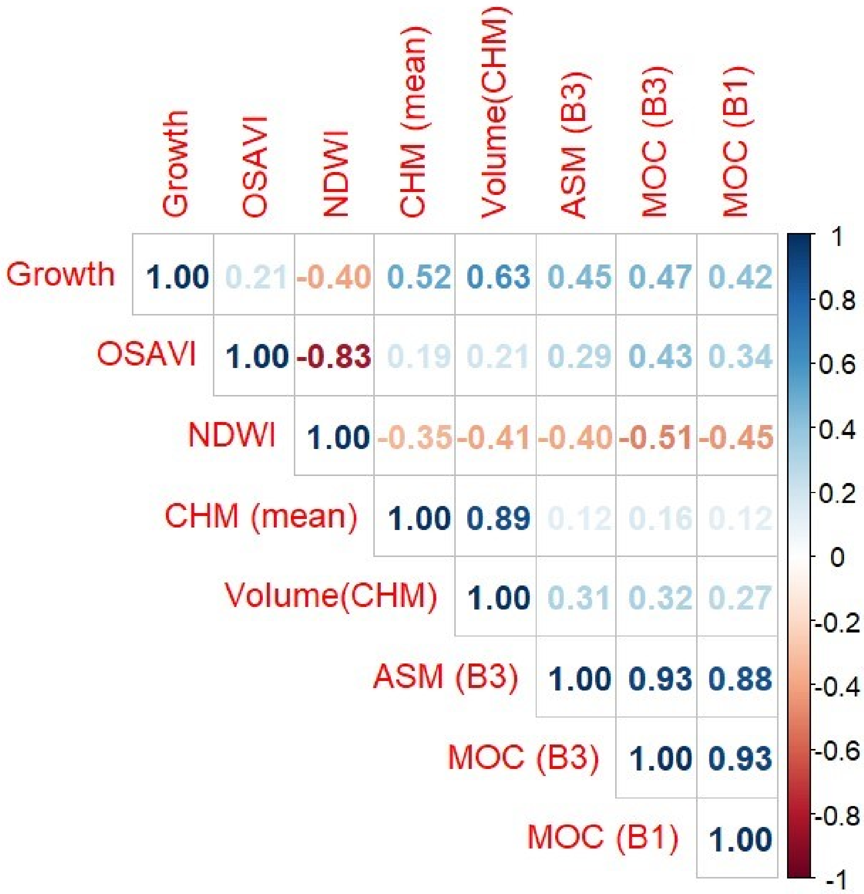

Figure 11. Correlation- matrix of the growth classes after Porten [43] and the spectral- (Vegetation Indices), structural- (CHM mean, Volume CHM) and textural features (ASM (B3), MOC (B3), MOC (B1) with correlation coefficients after Pearson. Plot created with RStudio version 4.3.3.

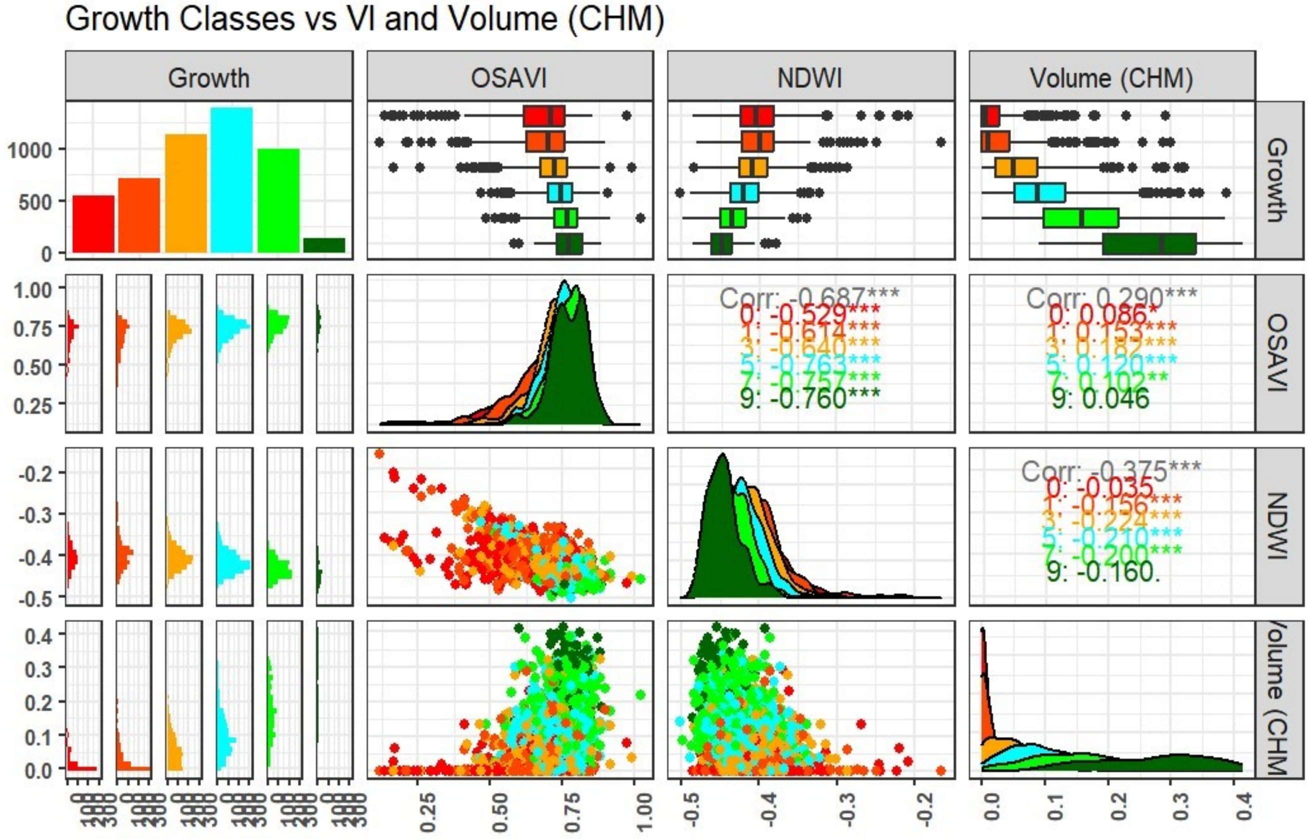

Figure 12. Summary of uni-, and bivariate statistics of the growth classes after Porten [43], Vegetation Indices (VI) and the calculated volume from the CHM (Canopy Height Model) grouped after the growth classes, from different ground truth labeling dates within the investigation area. Plot created with RStudio version 4.3.3.

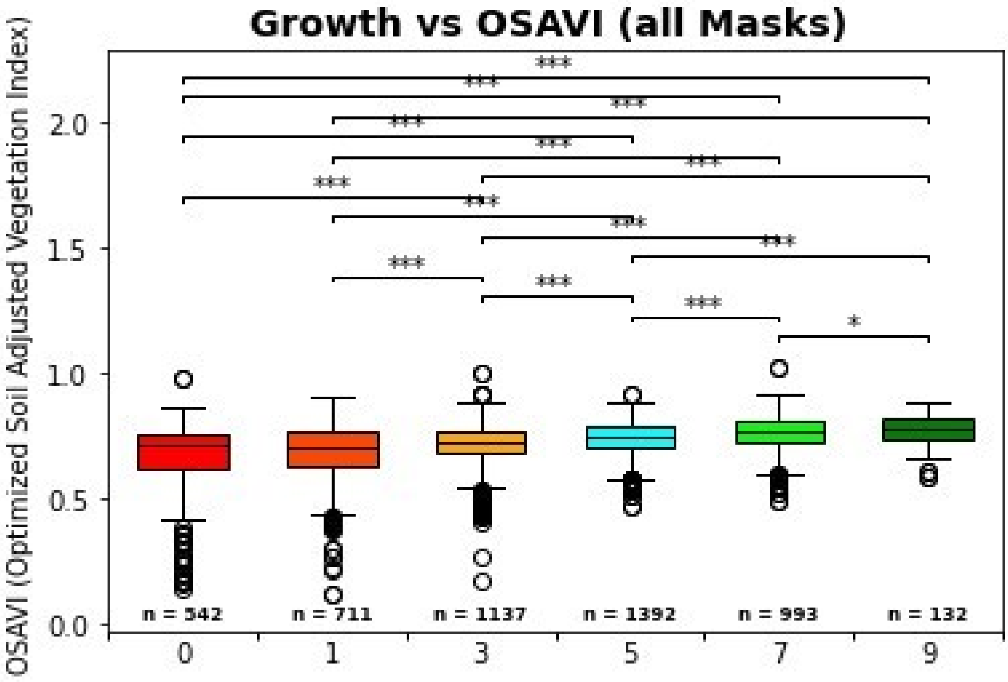

Figure 13. Growth Classes after Porten [43] against OSAVI with significance stars (*) generated from the Mann-Whitney-U-test with significance classes *<0.5, >0.1 (weak sig.), **<0.1, >0.01 (intermediate sig.), ***<0.001 (strong sig.). Plot created with matplotlib (Python 3.9).

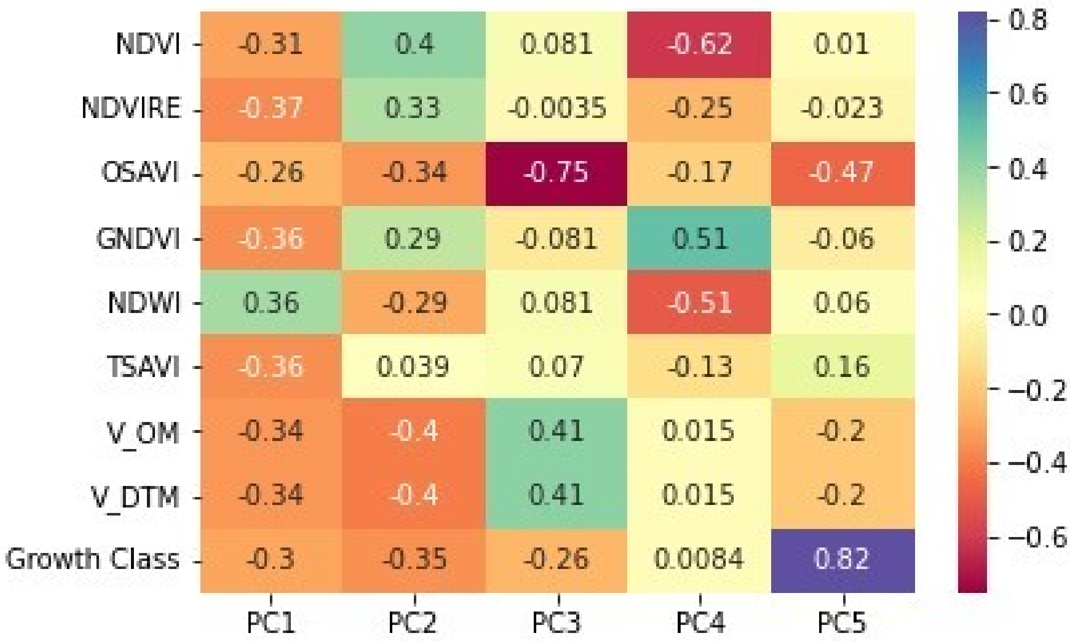

Figure 14. Correlation heatmap between chosen spectral,- and strutural features against first five PC, for geoprocessing results including different data acquisition dates. PCA+plot created with matplotlib (Python 3.9).

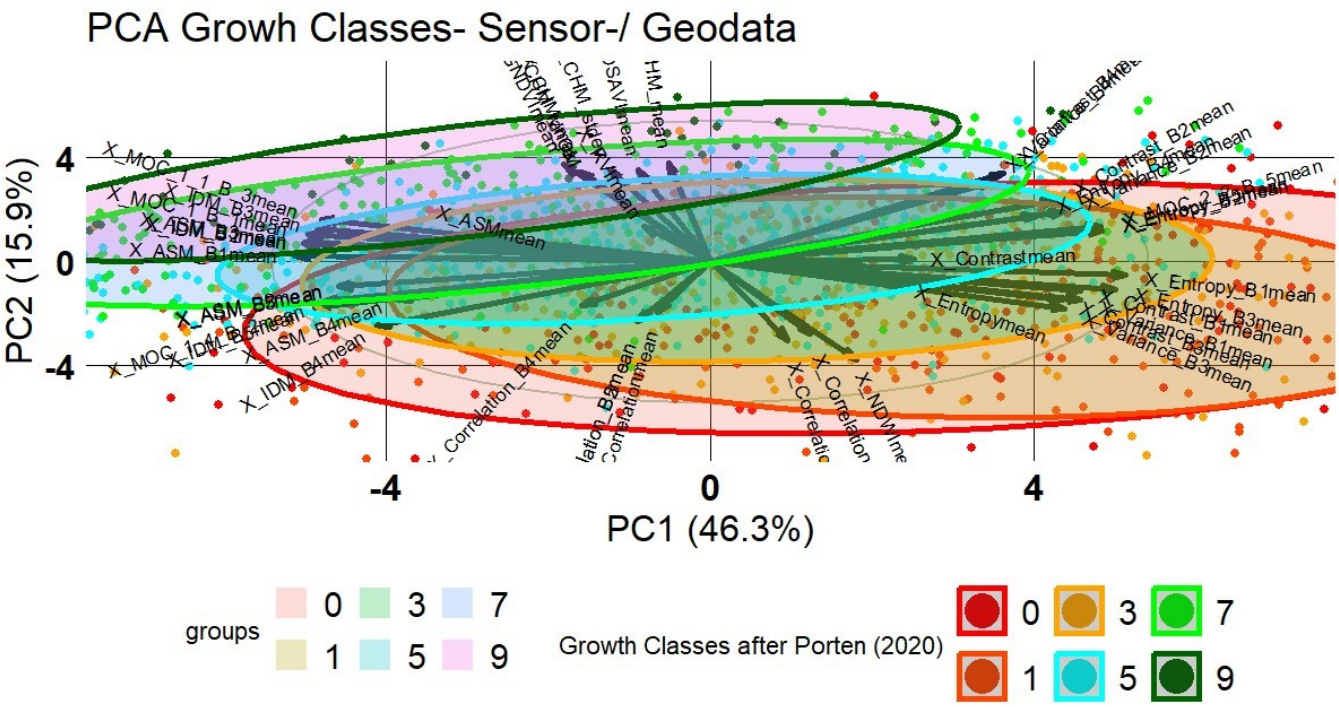

Figure 15. PCA (Principal Component Analysis) - Biplot with Principal Components PC1 and PC2, labeled features (with ellipses- colour-coded after growthclass assesment) and individual samples color-coded also after the growth classification assesment of Porten [43]. The Biplot shows the complex major influencing factors and their correlation to the input features of the data set (Biplot and PCA created with RStudio version 4.3.3).

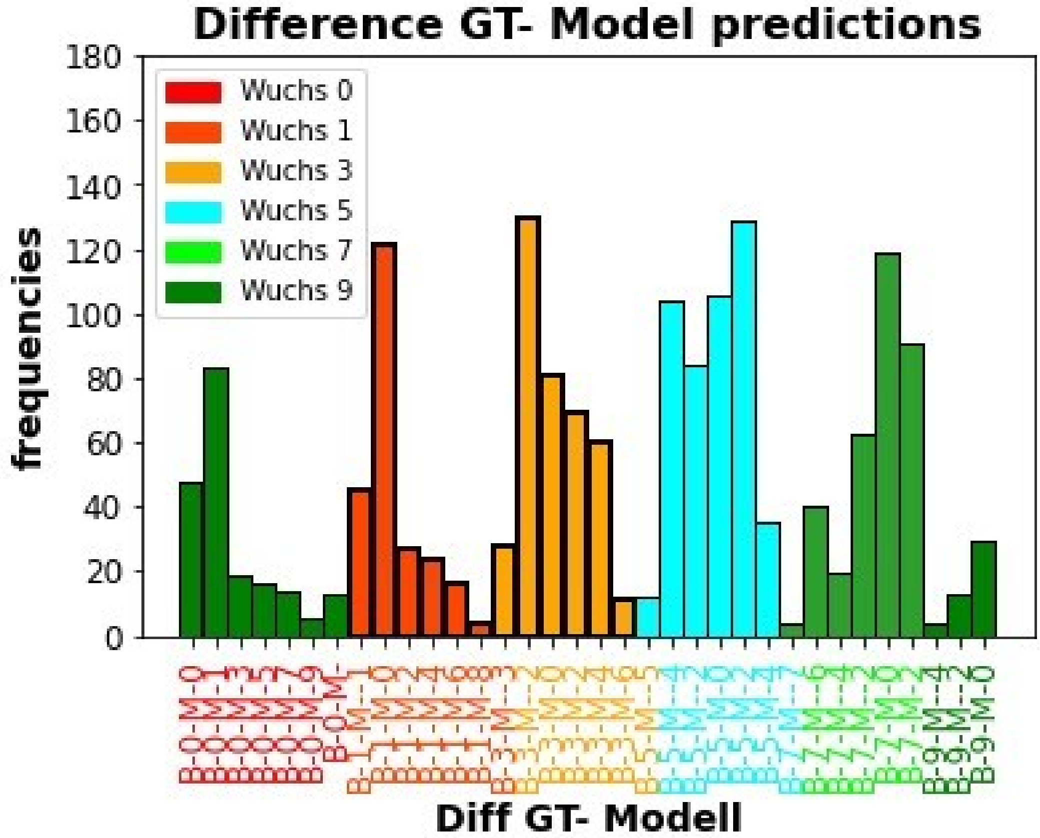

Figure 16. Frequencies of the GT (Ground Truth) and the predicted values for each Growth Class (B) (Wuchs 0- Growth Class 0, Wuchs 1- Growth Class 1, Wuchs 3- Growth Class 3, Wuchs 5- Growth Class 5, Wuchs 7- Growth Class 7, Wuchs 9- Growth Class 9) to model (M) difference pairs. It can be seen that exact matches and lower differences between GT and Model prediction are the most frequent. Plot created with matplotlib (Python 3.9). The statistics were generated for the sum of different ground-truth and sensor- data collections within the investigation area.

Description of Geoprocessing and Classification workflow software/ programming languages and packages (QGIS.3.22, SAGA GIS 7.8.2, Python 3.9, RStudio 4.3.3)

1. Grapevine positions were set manually based on the spatial highly-resolved orthomosaic from unfoliaged season + the ground- truth growth classes were joined per ID join onto the ID, with the same sampling date
2. Photogrammetry for calculation of Multispectral Orthomosaic, Digital Terrain Model, and Digital Surface Model (Agisoft Metashape professional), see also manuscript description
3. Geoprocessing in QGIS/ PyQGIS (version 3.22)

3.1. Create zonal statistics sampling area (derived from average width (1.00 m) of vine- row and average space between grapevine stems (1.20) in the same row) - the aspect of the vine- row was determined from the aspect of the Digital Terrain Model (DTM- calculated in QGIS 3.22) and subsequently aggregated for each grapevine with zonal statistics. The aspect values in the generated column were used as parameter input within the tool rectangles, diamonds to adjust the zonal statistics sampling area, according to the aspect of the vineyard respectively the fall line of the vine rows

3.2. Calculate spectral features (Vegetation Indices according to literature in the table) (see Table 4 and formula 1-6)

Table 4. Calculated VI for vine row extraction and as classification model input parameters with relevant literature references

ID	Spectral Indices	Source Reference	Abbreviation
1	Normalized Difference Vegetation Index	[72]	NDVI
2	Normalized Difference Red Edge Index	[73]	NDREI
3	Optimized Soil Adjusted Vegetation Index	[74]	OSAVI
4	Green Normalized Difference Vegetation Index	[75]	GNDVI
5	Transformed Soil Adjusted Vegetation Index	[76]	TSAVI
6	Normalized Difference Water Index	[77]	NDWI

[redacted]

VI formula:

1)

2)

3)

4)

5)

6)

3.3. Calculate structural features (CHM= DSM- DTM) /QGIS/ PyQGIS). See structural features in Table 5.

Table 5. Utilized structural features used as input for the classification model of the different model types.

ID F.I.	Height and Volume Measures	Name
1	Mean Height	CHM_{mean}
2	Median Height	CHM_{median}
3	Minimum Height	CHM_{min}
4	Maximum Height	CHM_{max}
5	Standard Deviation Height	CHM_{std}
6	Variance	CHM_{var}
7	Aggregated Pixel Volume	CHM Volume

3.3.4. Calculate (CHM- pixel volume) (CHM*count*pixel (area) in vine- row extracted area) (also described in the manuscript/ article)
3.5. Calculate texture features (GLCM- Gray Level Covariance Matrix) - in SAGA GIS (also described in the manuscript/ article) - this was done in the SAGA GUI Manually setted inputs (see also Table 7)

Table 7. The Gray-Level Co-occurrence Matrix (GLCM) textural features and their definitions used in this study to predict the growth classes after [43].

S.N.	Texture Measure	Formula
1.	Mean (ME)	$ME = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} kP(x,z)$
2.	Variance (VAR)	$VAR = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} (x-\mu)^2 P(x,z)$
3.	Homogeneity (HOM)	$HOM = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} \frac{1}{1+(x+z)^2} P(x,z)$
4.	Contrast (CON)	$CON = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} (x-z)^2 P(x,z)$
5.	Dissimilarity (DIS)	$DIS = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} P(x,z)\|x-z\|$
6.	Entropy (ENT)	$ENT = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} P(x,z)\log(P(x,z))$
7.	Angular Second Moment (ASM)	$ASM = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} (P(x,z))^2$
8.	Correlation (COR)	$COR = \sum_{x=0}^{N-1}\sum_{z=0}^{N-1} P(x,z)\left[\frac{(x-ME)(z-ME)}{\sqrt{VA_x VA_z}}\right]$

Note: $P(x,z) = \frac{V(x,z)}{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} V}$; where VA (x, z) represents the value in the row at the cell x and column z within the moving window. N represents the number of rows or columns in the window.

Correlation Analysis:

Correlate the texture features with the growth classes after Porten [43] and select those texture features with (Gray Level Covariance Matrix) a correlation coefficient (after Pearson) equal or higher than 0. 4). These features together with the chosen vegetation indices and structural features (selected based on literature sources and with the highest correlation after Pearson) were used as spectral, structural and texture features, subsequently extracted with the vine- row mask, spatially aggregtated with zonal statistics and were the basis for the machine learning classification process. For the spectral and structural features those based on literature research were chosen.

Generation of vine- row masks/ mask combination

Pixel- based (OSAVI >0.7, Otsu's Threshold))

Object- based image segmentation (SAGA- GIS- van Neumann Neighbourhood) with OSAVI as input and specific settings (settings described in the article/ manuscript)

→merge both masks to one masks with the QGIS 3.22 raster calculator

Extraction of the features based on the zonal statistics mask and intersection with the vine- row mask with QGIS raster calculator. The extraced features were attached as new columns to the geopackage

→ the shapefile and/ or geopackage was converted to csv and subsequently used as input for the classification routine (see classification script in Python of the following description)

-QGIS calculation of the overlap of the digitized vine- row ground-truth with the generated vine- row mask

$area_{intersection}/area_{union}$

Machine Learning Classification

Python/ Spyder (as development environment- the also github link at the end of the publication)

-Machine- learning classification script with visualization of the model results

-loading of all relevant packages and libraries (see list below)

- The csv which contained all selected features (spectral, structural and texture extracted + aggregated features for each vinegrape stem position) was

  loaded with Pandas, than filtered according to the different feature groups and classes
- -pre- processing of the data with standard scaler and filtering of Nan (Null) values
- The features were selected for model fitting to the target classes
- Defining of parameters for the hyperparameter tuning repeated- k-fold-cross validation of the classifiers SVM and RFC and setting of the pipeline (classifiers SVM and RFC were selected based on extensive literature research)

  including train-test split ratio etc.
- Definition of the for loop with cross val predict, cross val score, cross validate of the classifiers (RFC, SVM) with the

  grid search hyperparameter tuning repeated k-fold cross validation
- The machine learning classification with the best overall accuracy (SVM 7- Support Vector Machines mit input feature group 7 (all feature types)) was selected and for these output, the confusion matrix was

  calculated with User and Produser accuracy for the class specific growth class evaluation
- Moreover the classification results were exported from pandas dataframes to csv file format
- Some of the classification results were stacked together and visualized in boxplot format with significance stars

  based on the Mann- Whitney- U-Test (see also grouped boxplot figures within the manuscript/ publication)