A Feature Selection Approach for Terrestrial Hyperspectral Image Analysis

Kyle Loggenberg, Nitesh Poona

Department of Geography & Environmental Studies, Stellenbosch University, South Africa,
kyleloggenberg@sun.ac.za

Abstract

Feature selection techniques are often employed for reducing data dimensionality, improving computational efficiency, and most importantly for selecting a subset of the most important features for model building. The present study explored the utility of a Filter-Wrapper (FW) approach for feature selection using terrestrial hyperspectral remote sensing imagery. The efficacy of the FW approach was evaluated in conjunction with the Random Forest (RF) and Extreme Gradient Boosting (XGBoost) classifiers, to discriminate between water-stressed and non-stressed Shiraz vines. The proposed FW approach yielded a test accuracy of 80.0% (KHAT = 0.6) for both RF and XGBoost, outperforming the more traditional Kruskal-Wallis (KW) filter by more than 20%. The FW approach was also less computationally expensive when compared with the more commonly used Sequential Floating Forward Selection (SFFS) wrapper. Additionally, we examined the effect of hyperparameter optimisation on classification accuracy and computational expense. The results showed that RF marginally outperformed XGBoost when using all wavebands (p = 176) and optimised hyperparameter values. RF yielded a test accuracy of 83.3% (KHAT = 0.67), whereas XGBoost yielded a test accuracy of 81.7% (KHAT = 0.63). Our results further show that optimising hyperparameter values yielded an overall increase in test accuracy, ranging from 0.8% to 5.0%, for both RF and XGBoost. Overall, the results highlight the effect of feature selection and optimisation on the performance of machine learning ensembles for modelling vineyard water stress.

1. Introduction

Hyperspectral remote sensing (HRS) provides a wealth of spectral information (Li *et al.*, 2018) by recording narrow-band reflectance across the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) regions of the electromagnetic (EM) spectrum. These narrow, contiguous wavebands convey meaningful spectral variations (Santara *et al.*, 2017) that can aid in the discrimination of spectrally similar features (Li *et al.*, 2018). Deployment of HRS has traditionally been facilitated through aerial and satellite platforms. However, in the past decade, there has been a

growing interest in the use of terrestrial platforms in HRS applications (Hartzell, Glennie and Khan, 2017).

Hyperspectral imaging coupled with terrestrial remote sensing (i.e. proximal remote sensing) offers a solution for near real-time, on-the-go monitoring of vineyards (Mulla, 2013). Terrestrial HRS produces high spatial resolution imagery that is ideal for managing the narrow spacing (typically from 1.4 to 2.1 m) and overlapping nature of vines (Matese and Di Gennaro, 2015). Furthermore, terrestrial HRS platforms are easily deployable and less restricted by atmospheric effects, compared with aerial and satellite platforms. To date, a limited number of studies have employed terrestrial HRS in precision viticulture (Mohite *et al.*, 2017; Kyle Loggenberg *et al.*, 2018).

Decision tree-based machine-learning techniques have been widely used for the classification of high dimensional data (Georganos *et al.*, 2018; Wietecha *et al.*, 2019). The popularity of these learners are likely due to their high classification performance, robustness to noise and overfitting, interpretability and scalability (Belgiu and Drăguţ 2016; Poona, van Niekerk, and Ismail 2016). Their ability to handle high dimensionality (p) coupled with a small sample size (n)—the $n << p$ problem—makes tree-based ensemble learners particularly useful for analysing HRS data (Poona and Ismail 2014; Loggenberg *et al.* 2018).

The Random Forest (RF) tree-based ensemble (Breiman, 2001) is one of the most popular algorithms utilised for the classification of HRS data. RF is a non-parametric classifier that is robust to outliers and noise (Breiman, 2001). However, it is the algorithm's resistance to overfitting that makes RF ideal for the classification of HRS datasets (Abdel-Rahman *et al.*, 2014). RF prevents overfitting by growing uncorrelated trees using randomly selected subsets of the input data (Breiman, 2001; Abdel-Rahman *et al.*, 2014). Recent studies have successfully demonstrated the utility of RF for the classification of HRS data in precision viticulture. For example, Poblete-Echeverría *et al.* (2017) reported an average classification accuracy of 94.0% (Kappa = 0.91) when discriminating vine canopies from soil profiles. Knauer *et al.* (2017) also reported a high classification accuracy (87.0%) when detecting Powdery Mildew on grapes.

Another tree-based ensemble learner that has gained noteworthy recognition in the literature is Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016). XGBoost is a regularised tree boosting ensemble learner that builds on the Gradient Boosting Machine (GBM) proposed by Friedman (2001). XGBoost has demonstrated considerable promise in precision viticulture applications. For example, Mohite *et al.* (2017) reported XGBoost classification accuracies ranging from 81.6% to 87.6% for detecting pesticide residue on vineyard grapes. Loggenberg *et al.* (2018) employed XGBoost to model water stress in a Shiraz vineyard using HRS data, reporting classification accuracies ranging from 78.0% to 90.0%, with KHAT values ranging from 0.53 to 0.6. Both Mohite *et al.* (2017) and Loggenberg *et al.* (2018) compared the utility of XGBoost and RF and found that RF often outperformed XGBoost.

Most studies to date have employed machine learning algorithms, such as RF, using default hyperparameter values (for example, see Lagrange, Fauvel & Grizonnet 2017; Poona, van Niekerk & Ismail 2016; Taşkın, Hüseyin & Bruzzone 2017). However, Xia *et al.* (2017) assert that machine learning algorithms can be highly sensitive to hyperparameter settings, which can greatly affect the algorithm's performance. Furthermore, the hyperparameter values for a given algorithm are dataset-dependent and are consequently rarely optimal across applications (Rodrìguez, Kuncheva and Alonso, 2006). Hyperparameter optimisation thus forms an integral addition to machine learning classification frameworks (Martinez-de-Pison *et al.*, 2017), providing an efficient automated method that can greatly lessen the burden of manual hyperparameter tuning (Xia *et al.*, 2017). Poona & Ismail (2014) highlighted that using optimised RF hyperparameter values leads to improved classification accuracies, compared with using default hyperparameter values.

Feature selection techniques are often coupled with ensemble learners to reduces data dimensionality and computational complexity (Fu *et al.* 2017; Pedergnana, Marpu & Mura 2013; Vélez Rivera *et al.* 2014). Dimensionality is reduced by removing redundant and/or irrelevant wavebands (Taşkın, Hüseyin and Bruzzone, 2017), thereby lessening computational complexity without decreasing classification competency (Chandrashekar and Sahin, 2014). Notably, varied results have been reported regarding the effects of feature selection on model performance. For example, Vélez Rivera *et al.* (2014) reported a 6.5% decrease in accuracy when classifying mechanical damage in mango fruits using NIR hyperspectral imagery. Li *et al.* (2011) and Pedergnana, Marpu & Mura (2013) also reported decreased classification accuracies when employing feature selection techniques on hyperspectral datasets. Contrary to these findings, several studies have reported improved model performance with feature selection (for example, see Belgiu *et al.* 2014; Poona, van Niekerk & Ismail 2016).

Feature selection techniques are broadly categorised into filter and wrapper approaches (Chandrashekar and Sahin, 2014; Fu *et al.*, 2017). The filter approach, which functions independently from the classifier (Fu *et al.* 2017), is used as a pre-processing step to rank wavebands based on their relative importance (Lagrange, Fauvel & Grizonnet 2017). The highest ranked wavebands, i.e. wavebands that contain the most useful information about a given class, are then selected based on a predefined threshold value and used as input for classification (Chandrashekar and Sahin, 2014). In comparison, the wrapper approach evaluates the suitability of waveband subsets based on the performance of a given classifier (Jović, Brkić and Bogunović, 2015). Wrappers employ searching strategies to automatically select subsets, with the optimal subset selected based on predictive competency (Chandrashekar and Sahin, 2014). Wrapper approaches are thus computationally more expensive compared with filter approaches. However, wrappers have been shown to produce subsets that yield improved classification results (Jović, Brkić and Bogunović, 2015). For a detailed review of popular feature selection methods, see Chandrashekar & Sahin (2014).

Numerous studies (for example Jović, Brkić and Bogunović, 2015; Ghareb, Bakar and Hamdan, 2016; Martinez-de-Pison *et al.*, 2017) have applied hybrid approaches for feature selection. However, in this study, we present an alternative approach that fuses (i.e. embeds) a filter method with the RF and XGBoost learners. Thus, the overarching aim of this study was to assess the efficacy of an embedded filter-wrapper approach for the discrimination of water-stressed vines using terrestrial HRS. The specific objectives are to (i) examine the effect of RF and XGBoost hyperparameter value optimisation on classification accuracy and computational cost, (ii) evaluate RF and XGBoost performance when coupled with a filter and a wrapper, and (iii) explore the utility of a filter applied within a wrapper paradigm (where the predictive competency of filter-selected wavebands is determined by the respective learner). The remainder of the paper is structured as follows: the dataset and corresponding analysis are introduced and detailed in Section 2, the experimental results and discussions are presented in Section 3, and concluding remarks provided in Section 4.

2. Materials and Methods

2.1. Experimental design

The study utilised terrestrial imaging spectrometer data to discriminate water-stressed from non-stressed Shiraz vines. The data was acquired on the Welgevallen experimental farm in Stellenbosch, South Africa (33°56'38.5"S, 18°52'06.8"E). The SIMERA HX MkII hyperspectral line scanner was used to capture terrestrial imagery from a side-on view of the vine canopy (Loggenberg *et al.* 2018). The imagery consisted of 176 wavebands ranging from 473 nm to 708 nm, with a bandwidth range of 0.9 nm to 2 nm. Imagery was captured on February 24, 2017, between 10:00 and 12:00, and pre-processed using Environment for Visualising Images (ENVI) version 5.3.1 software. All further processing was completed in the R statistical software environment (R Development Core Team, 2017) on a contemporary machine running Windows-64 OS, with an i7-4770 CPU @ 3.40GHz and 8 GB RAM. For a detailed account of the study area, data collection, and pre-processing methods, see Loggenberg *et al.* (2018).

2.2. Statistical analysis

Classification models were developed for the RF and XGBoost learners using both default and optimised hyperparameter values. The two tree-based classifiers evaluated the performance of the Kruskal-Wallis (KW) filter (Kruskal and Wallis, 1952), Sequential Floating Forward Selection (SFFS) wrapper (Pudil, Novovičová and Kittler, 1994), and the proposed Filter-Wrapper (FW) (Bischl *et al.*, 2016) feature selection approaches. The predictive competencies of each feature selection approach were compared with using all wavebands, i.e. $p = 176$. The classifiers were trained using $n = 60$ leaf spectra samples extracted from the terrestrial imagery, with 30 samples collected for each of the stressed and non-stressed classes. The data analysis workflow is represented in Figure 1.

2.2.1. *Random forest (RF)*

RF grows a multitude of classification trees (*ntree*), by randomly selecting training samples with replacement (Breiman, 2001). Each tree is maximally grown, i.e. without pruning, on 2/3 of the original data (i.e. bagged sample), using a random subset of wavebands (*mtry*) to determine the split at each tree node (Breiman, 2001). The remaining 1/3 of the data (i.e. out-of-bag sample) is used to calculate the out-of-bag (OOB) error, which is used by RF as an internal measure of accuracy (Breiman, 2001). RF produces a complex forest of trees that have high variance and low bias, and applies majority voting across all the trees in the forest to determine class membership (Breiman, 2001; Belgiu and Drăguţ, 2016). A detailed review on RF can be found in Breiman (2001) and Belgiu & Drăguţ (2016). RF was implemented using the "randomForest" package (Liaw and Wiener, 2002). The default hyperparameter values used for RF were *ntree* = 500 and *mtry* = $\sqrt{p}$, where p is the number of wavebands.

2.2.2. *Extreme gradient boosting (XGBoost)*

Similar to RF, XGBoost transforms weak classification trees into an ensemble of strong predictive capacity. However, unlike RF, where trees are grown independently (Breiman, 2001), XGBoost builds trees that learn from the previously grown tree (Chen and Guestrin, 2016). XGBoost trains a model in an additive manner and finds the best parameters for the given model by defining an objective function (Chen and Guestrin, 2016). The objective function (Equation 1) contains a user-defined predictive term, which measures the predictive competency of the model, and a regularisation term, which controls overfitting and reduces model complexity (Chen and Guestrin, 2016):

$$obj(\theta) = \varphi(\theta) + \gamma(\theta) \qquad [1]$$

where φ is the training loss function, and γ is the regularisation term. The XGBoost classifier was employed using default hyperparameter values (Table 1) and implemented using the "xgboost" package (Chen *et al.*, 2017). For a more detailed account on XGBoost and its hyperparameter settings, see Chen & Guestrin (2016); Xia *et al.* (2017); and Loggenberg *et al.* (2018).
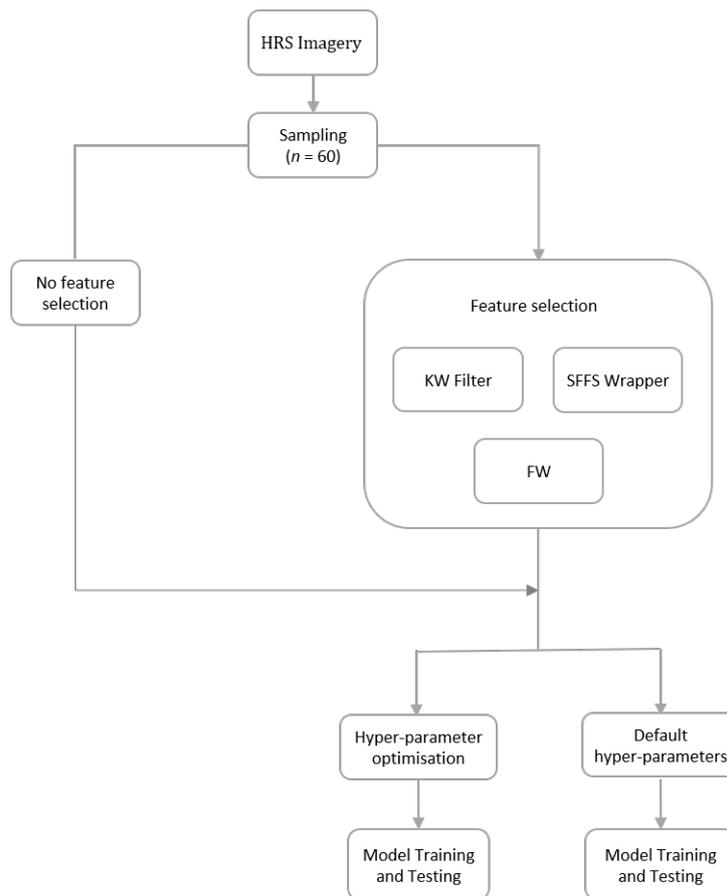
Figure 1: Statistical analysis workflow.

2.3.   Hyperparameter optimisation

In an attempt to obtain the best classification results, hyperparameter values for all RF and XGBoost models were optimised using the grid search optimisation method. For RF, *ntree* and *mtry* values were optimised following the recommendation of Abdel-Rahman *et al.* (2014) and Poona *et al.* (2016). *Ntree* values up to 2500 were evaluated using intervals of 500, with the *mtry* value set to varying multiplicative factors of $\sqrt{p}$ (i.e. 1/4, 1/3, 1/2, 1, 1.5, 2, 2.5, 3). Six hyperparameter values were optimised for XGBoost (see Table 1), based on work done by Xia *et al.* (2017) and Loggenberg *et al.* (2018). The optimal RF and XGBoost hyperparameter values were selected based on model performance, i.e. hyperparameter values that yielded the lowest OOB and classification error, respectively. The Grid Search optimisation method was employed utilising the "caret" package (Kuhn *et al.*, 2017).

Table 1: Optimisation ranges tested for XGBoost hyperparameters

Hyperparameter	Default value	Range	Interval
max_depth	6	3-10	2
subsample	1	0.5-1	0.2
eta	0.3	0.01-0.2	0.2
nrounds	500 (current study)	20-100	20
		50-250	50
		500-2500	500
min_child_weight	1	0.2-1	0.2
colsample_bytree	1	0.5-1	0.2

2.4. Waveband selection

The KW filter, SFFS wrapper, and FW feature selection approaches were all implemented using the "mlr" package (Bischl *et al.*, 2016). The KW filter generated a single waveband subset that served as input to classification for both RF and XGBoost. The SFFS and FW approaches—using default classifier hyperparameters—generated individual subsets for RF and XGBoost, respectively.

2.4.1. Filter

The KW filter was employed to rank wavebands based on their respective importance. The KW filter was utilised based on the recommendation of Vora & Yang (2017). The authors reported the effectiveness of the KW filter when applied to high dimensional datasets in a classification framework. KW is a non-parametric filter that is applied to a labelled dataset with x number of classes (Vora and Yang, 2017). Wavebands are partitioned based on their class membership, and a KW statistic subsequently calculated (Vora and Yang, 2017). The KW statistic is defined as (Zhao *et al.* 2010):

$$K = (N-1)\frac{\sum_{r=1}^{x} i_r(\varpi_r - \varpi)^2}{\sum_{r=1}^{x}\sum_{k=1}^{i_r}(\omega_{rk} - \varpi)^2}$$ [2]

where N is the total number of wavebands across all the classes, ϖ is the average rank of each waveband, i_r is the number of wavebands in class r, ω_r is the average rank of wavebands in class r, and ω_{rk} is the rank of waveband k in class r. The top 10% ($p = 18$) of the ranked wavebands were then selected and used as input for classification following Abdel-Rahman *et al.* (2014) and Loggenberg *et al.* (2018).

2.4.2. Wrapper

Sequential wrapper techniques, such as Sequential Forward Selection (SFS), are frequently utilised for dimensionality reduction (Taşkın, Hüseyin and Bruzzone, 2017) as they are computationally less expensive than exhaustive search methods (Fu *et al.*, 2017). A more flexible variant of SFS, known as SFFS, was employed in the present study as it has been shown to outperform SFS (Taşkın, Hüseyin and Bruzzone, 2017). SFFS builds on its predecessor by integrating feature (i.e. waveband) removal (Chandrashekar and Sahin, 2014), which facilitates the prevention of feature nesting (Pudil, Novovičová and Kittler, 1994). Similar to SFS, SFFS adds wavebands one at a time to an empty subset (k). The predictive prowess of the subset is then evaluated based on a user-defined objective function (Chandrashekar and Sahin, 2014). The subset is then deemed optimal (d) or not; if deemed not optimal, the SFFS algorithm then randomly removes a waveband and the new subset is re-evaluated (see Figure 2). This process is iterated until the algorithm converges, i.e. classification accuracy does not improve more than a given threshold (*alpha*).

SFFS was implemented with an *alpha* value of 0.02 (Bischl *et al.*, 2016). Monte-Carlo Cross Validation (MCCV) (Xu and Liang, 2001) splits the dataset into 2/3 training, with the remaining 1/3 used to test model performance. MCCV was iterated 10-fold and the subset that yielded the lowest Mean Misclassification Error (MMCE) was selected as the optimum subset of wavebands.
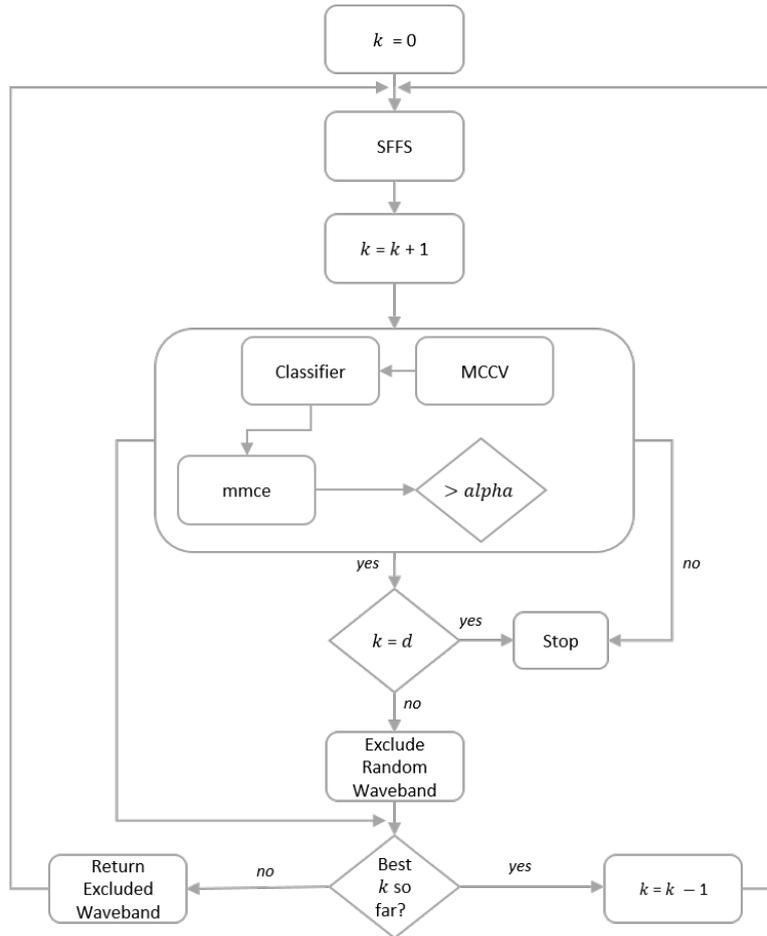
Figure 2: SFFS Wrapper workflow (adapted from Chandrashekar & Sahin 2014).

2.4.3. *Filter-Wrapper (FW)*

The aim of the FW approach was two-fold: to improve the classification accuracy achieved using the KW filter, and to lessen the computational strain often associated with wrapper methods (such as SFFS). Implementation of the FW approach followed a simple workflow (see Figure 3), based on the following steps:

Step 1: Wavebands are ranked using the KW filter, with a subset selected based on the optimal threshold value. The threshold value is optimised using grid search with values ranging from 0.02 to 0.2; equivalent to testing subsets of the top 2% to 20% of the ranked wavebands.

Step 2: The classifier (RF, XGBoost) evaluates the predictive competency of the selected subset using a 10-fold MCCV, similar to the SFFS wrapper approach.

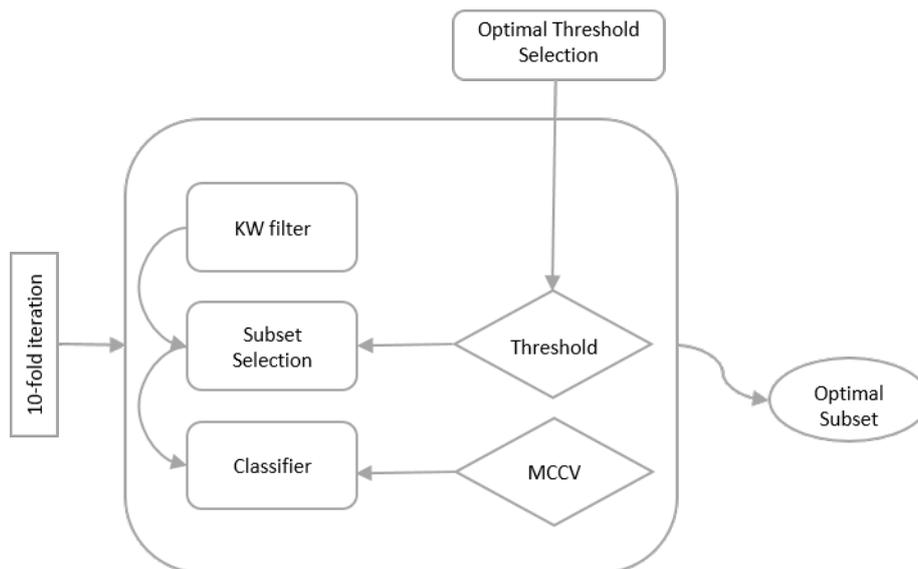Step 3: Steps 1 and 2 are iterated 10-fold and the optimum subset selected based on the lowest MMCE.

Figure 3: Filter-Wrapper workflow.

2.5. Accuracy assessment

An independent test set (n = 60), collected for stressed (n = 30) and non-stressed (n = 30) vines, was used to evaluate model performance. The performance of all models was compared based on their measured mean accuracies (Kohavi and Provost, 1998), computed using a confusion matrix, and KHAT statistic (Congalton and Green, 2009), calculated using kappa analysis:

$$\widehat{K} = \frac{p_a - p_c}{1 - p_c}$$ [3]

where p_a describes the actual agreement and p_c describes the chance agreement. All classification models were iterated 10-fold to ensure model robustness and to prevent overfitting.

3. Results and Discussion

3.1. RF and XGBoost optimisation

Table 2 indicates the optimised hyperparameter values for RF and XGBoost. Using all wavebands (p = 176) as input—optimised hyperparameter values for RF were *ntree* = 1 500 and *mtry* = 4— yielded the lowest OOB error of 6.7%. Notably, for all RF models the optimised *mtry* values were smaller than the default values (see Table 2 (A)). Although smaller *mtry* values lead to decreased computational expense, as fewer wavebands are considered for node splitting, Goldstein *et al.* (2010) assert that a smaller *mtry* can lead to biased RF models and decreased accuracies. However, several authors (for example Abdel-Rahman *et al.* 2015; Adam *et al.* 2017) have shown that smaller *mtry* values lead to improved classification performance.

Table 2: Optimised hyperparameter values using grid search.

(A) RF

Model	Default		Optimised		
	ntree	*mtry*	*ntree*	*mtry*	**OOB Error**
All wavebands (*p* = 176)	500	13	1500	4	6.70%
KW (*p* = 18)		4	500	2	28.30%
FW (*p* = 35)		6	500	3	11.70%
SFFS (*p* = 4)		2	500	1	8.30%

(B) XGBoost

	Model	*max _depth*	*subsample*	*eta*	*nrounds*	*min_child _weight*	*colsample _bytree*	**Classification Error**
	Default	6	1	0.3	500	1	1	-
Optimised	**All wavebands** (*p* = 176)	3	0.7	0.11	250	0.6	0.7	8.30%
	KW (*p* = 18)	5	0.5	0.07	250	1	0.7	27.80%
	FW (*p* = 18)	3	0.7	0.15	100	0.4	0.7	11.20%
	SFFS (*p* = 3)	3	0.9	0.15	80	0.8	0.7	9.30%

An *ntree* value of 500 was determined as optimal for all three feature selection approaches, i.e. the KW filter, SFFS wrapper, and FW approach. A default *ntree* value of 500 was also found to be optimal in previous studies by Abdel-Rahman *et al.* (2014); Abdel-Rahman *et al.* (2015); and Poona *et al.* (2016). The higher *ntree* value (1 500) obtained using all wavebands may be attributed to the dataset comprising weak predictors, resulting in a model requiring a larger number of trees (Goldstein *et al.*, 2010).

For XGBoost, using all wavebands as input yielded the lowest classification error of 8.3%. All XGBoost models comprised a lesser number of trees (*nrounds*) compared with RF (see Table 2 (B)). A similar finding was reported by Georganos *et al.* (2018) where RF required more trees (*ntrees = 2000*) than XGBoost (*nrounds* = 600) when optimising models for land cover classification using high dimensional datasets. Xia *et al.* (2017) noted that an inherent trade-off exists between the number of trees (*nrounds*), tree complexity (*max_depth*), and the learning rate (*eta*). The authors further assert that generally, for a given learning rate, a smaller *max_depth* value leads to a greater number of trees. In this study, learning rates ranging from 0.07 to 0.15 were observed. These relatively small learning rate values indicate that the optimised models would be more robust to overfitting (Xia *et al.*, 2017). However, the slower learning rate (i.e. lower *eta* value) would increase computational expense (Xia *et al.*, 2017).

3.2. Optimal waveband selection

The wavebands selected by the KW filter, SFFS wrapper, and FW approach are shown in Table 3. It is evident from Table 3 that the location of the selected wavebands differs for the three feature selection approaches. For example, the KW filter selected wavebands exclusively in the blue region (473.92-491.95 nm) of the EM spectrum. These wavebands were also present in the subset derived using the FW approach (see Table 3). The blue wavebands selected by both the KW and FW approaches were closely related in terms of wavelengths. This could indicate that the KW filter and the FW approach may not be optimal for reducing the multicollinearity within the present dataset.

Table 3: RF and XGBoost important wavebands as determined by the KW, FW, and SFFS feature selection approaches. Common wavebands are highlighted in bold.

	Classifier	
	RF (nm)	**XGBoost (nm)**
KW	473.92, 474.74, 475.58, 476.41, 478.09, 478.94, 479.78, 480.63, 483.20, 484.06, 484.92, 485.79, 487.53, 488.41, 489.29, 490.17, 491.06, 491.95	
FW	473.92, **474.74**, **475.58**, **476.41**, 477.25, **478.09**, **478.94**, **479.78**, 480.63, 481.48, **482.34**, 483.20, **484.06**, 484.92, **485.79**, 486.66, **487.53**, **488.41**, **489.29**, **490.17**, **491.06**, **491.95**, 492.84, 493.74, 494.64, 495.54, 497.36, 504.76, 577.17, **578.48**, 579.79, **581.11**, **582.43**, 583.77, 585.12	**474.74**, **475.58**, **476.41**, **478.09**, **478.94**, **479.78**, **482.34**, **484.06**, **485.79**, **487.53**, **488.41**, **489.29**, **490.17**, **491.06**, **491.95**, **578.48**, **581.11**, **582.43**
SFFS	475.58, 488.41, 578.48, 644.22	496.45, 521.32, 585.12

In comparison, the FW approach selected wavebands across the blue and green regions for both RF (473.92-585.12 nm) and XGBoost (474.74-582.43 nm). These selected wavebands correspond favourably to wavebands reported by Pôças *et al.* (2015) and Loggenberg *et al.* (2018), highlighting the feasibility of employing narrow wavebands in the VIS to model vineyard water stress. Notably, the FW approach selected a greater number of wavebands for RF ($p = 35$) compared with XGBoost ($p = 18$). Similar results were found by Georganos *et al.* (2018). However, the additional features in the FW-RF subset may indicate RF's resistance to the presence of redundant and/or irrelevant wavebands. Moreover, as shown in Table 3, the FW approach selected wavebands common to both RF and XGBoost. These common wavebands ($p = 18$) were located across the blue (474.74-491.95 nm) and green (578.48-582.43 nm) regions of the EM spectrum. Loggenberg *et al.* (2018) reported similar wavebands when employing internal measures of variable importance to reduce dimensionality. These findings suggest that these wavebands ($p = 18$) may be the most important for water stress modelling in a Shiraz vineyard. However, this requires further investigation.

Of the three feature selection approaches evaluated in this study, the SFFS wrapper approach yielded the smallest subsets for RF ($p = 4$) and XGBoost ($p = 3$). For XGBoost, wavebands were located in the blue (496.45 nm) and green (521.32 nm and 585.12 nm), and in the blue (475.58 nm

and 488.41 nm), green (578.48 nm), and red (644.22 nm) for RF. Loggenberg *et al.* (2018) reported the use of blue (474.74 nm and 497.36 nm) and green (521.32 nm, 578.48 nm and 585.12 nm) wavebands as an indicator of vineyard water status, which correspond to similar wavebands present in the SFFS-XGBoost and SFFS-RF subsets. These wavebands may be significant in modelling crop water stress, as the blue and green regions are highly sensitive to plant pigment (i.e. carotenoid and chlorophyll pigments) absorption (Zygielbaum *et al.*, 2009; Pôças *et al.*, 2015). Vegetative water stress is often expressed as an increase in blue and green reflectance (Zygielbaum *et al.*, 2009).

3.3.   RF and XGBoost classification

Table 4 shows the classification results for RF and XGBoost. For all models, the optimised hyperparameter values yielded improved classification accuracies, ranging from 0.8% to 5.5%. Using all wavebands (p= 176) as input yielded the best-performing models overall, producing a test accuracy of 83.3% (KHAT = 0.67) for RF and 81.7% (KHAT = 0.63) for XGBoost. These results compare favourably with work done by Vélez Rivera *et al.* (2014) and Abdel-Rahman *et al.* (2015), who found machine learning classifiers to perform best when using all wavebands. These results could indicate that both the RF and XGBoost ensembles are insensitive to the curse of dimensionality. Furthermore, the results indicate that RF outperformed XGBoost when using all wavebands. Similar findings were reported by Loggenberg *et al.* (2018).

Subsets generated using the KW filter yielded the lowest overall classification accuracies. Test accuracies for all models were found to be less than 60.0%. The decreased accuracies may be attributed to the selected wavebands (Table 3), with the subset comprised entirely of wavebands located in the blue region of the EM spectrum. The blue region is most often used in combination with longer wavelengths (Maimaitiyiming *et al.* 2017; Pôças *et al.* 2015) for vineyard water stress modelling.

In comparison, the waveband subsets selected by the SFFS wrapper and FW approach yielded higher accuracies, producing a test accuracy of 80.0% (KHAT = 0.60) for both RF and XGBoost. When compared with using all wavebands, the SFFS wrapper subsets resulted in reduced test accuracies for XGBoost (1.7%) and RF (3.3%). However, the SFFS wrapper subsets achieved these results utilising only 2% (approximately 98% reduction in dimensionality) of the original waveband dataset. Moreover, the FW approach presented here provides flexibility as it can be employed across classifiers and combined with different waveband ranking, i.e. filter, approaches.

Table 4: RF and XGBoost classification results. Results for the best-performing and worst-performing models are highlighted in bold.

Feature Selection	Parameters	Dataset	RF		XGB	
			Accuracy (%)	KHAT	Accuracy (%)	KHAT
All Wavebands (*p* = 176)	Default	Train	90.0	0.80	86.7	0.73
		Test	80.0	0.60	78.3	0.57
	Optimised	Train	**93.3**	**0.87**	**91.7**	**0.83**
		Test	**83.3**	**0.67**	**81.7**	**0.63**
KW (*p* = 18)	Default	Train	68.3	0.37	66.7	0.33
		Test	56.7	0.13	53.3	0.07
	Optimised	Train	**71.7**	**0.43**	**72.2**	**0.43**
		Test	**57.5**	**0.15**	**58.3**	**0.17**
FW *p* = 35 (RF) & *p* = 18 (XGBoost)	Default	Train	86.7	0.73	86.7	0.73
		Test	78.7	0.57	80.0	0.60
	Optimised	Train	88.3	0.77	88.8	0.77
		Test	80.0	0.60	80.0	0.60
SFFS (*p* = 4 (RF) & *p* = 3 (XGBoost))	Default	Train	90.0	0.80	88.3	0.77
		Test	80.0	0.60	80.0	0.60
	Optimised	Train	91.7	0.83	90.7	0.80
		Test	80.0	0.60	80.0	0.60

3.4.   Comparison of computational expense

Additionally, the study recorded the computational expense of the three feature selection approaches (Table 5). The KW filter took 2.59 seconds to run. However, the waveband subset derived using the KW filter yielded the lowest classification accuracies. The SFFS wrapper subsets yielded the highest classification accuracies but required 1 402.33 seconds (approximately 23 minutes) to run for RF and 5 820.24 seconds (approximately 1.6 hours) for XGBoost. In comparison, the FW approach required less processing time, approximately 2 minutes for RF and 7 minutes for XGBoost, to produce subsets of equivalent predictive competency.

It is evident from these results that algorithm optimisation does improve classification accuracy. However, it can be argued that the implementation of hyperparameter optimisation is application-specific, i.e. marginal increases in accuracy may not always warrant the added computational expense. Additionally, the choice of classifier should be considered when employing optimisation. In this study, RF optimisation resulted in marginal increases in accuracy, with minimal computational cost; 267.61 seconds using all wavebands (*p*=176). In comparison, XGBoost optimisation was computationally more expensive; 19 646.84 seconds using all wavebands. Chen & Guestrin (2016) and Xia *et al.* (2017) recommend the optimisation of XGBoost hyperparameter values. However, the

results of the present study indicate that the gain in predictive competency does not justify the greater increase in computational expense.

Table 5: RF and XGBoost computational expense for feature selection and hyperparameter optimisation.

		Processing Time (s)			
		KW	FW	SFFS	All Wavebands
Optimisation	RF	27.8	14.75	31.36	267.61
	XGBoost	17 803.06	17 742.21	18 898.73	19 646.84
Feature Selection	RF	2.59	87.55	1 420.33	-
	XGBoost		402.66	5 820.24	

The longer processing time observed for XGBoost optimisation and feature selection may be attributed to the classifier's use of the greedy search algorithm. XGBoost utilises an exact greedy search algorithm (Chen and Guestrin, 2016; Xia *et al.*, 2017) for finding the optimal tree structure. Xia *et al.* (2017) noted that this method is computationally expensive when employed on high-dimensional datasets, such as hyperspectral data. In contrast, RF uses random coefficients, determined through bagging with replacement and randomisation (Breiman, 2001), to find the optimal split for each tree, which is computationally more efficient. Moreover, as asserted by Xia *et al.* (2017), the longer processing times for the XGBoost algorithm may be explained by the slower learning rates (i.e. low *eta* values) utilised in the present study.

Consequently, this study shows that there is an inherent trade-off between dimensionality reduction, classification accuracy and computational expense. The FW approach presented in this study was the most successful in lessening this trade-off. High classification accuracies were obtained, using only 10% to 20% of the original waveband dataset (equivalent to an 80% to 90% reduction in dimensionality), at minimal computational expense. Although the FW approach warrants further investigation, it has demonstrated its potential operational capability.

4. Conclusion

This study investigated the efficacy of the KW filter, SFFS wrapper, and FW approach for waveband selection of terrestrial hyperspectral imaging data. The predictive competency of the generated subsets was evaluated using the RF and XGBoost machine learning classifiers. The classifiers were employed using both default and optimised hyperparameter values. Based on the findings of this study, the following conclusions are drawn:

(1)   The proposed FW approach to feature selection demonstrated considerable promise in both predictive competency and as a means to lessen computational expense.

(2)   The performance of a filter can be improved by implementing the proposed FW approach.

(3)   Both the RF and XGBoost ensemble learners have shown to be insensitive to the curse of dimensionality.

(4)   The VIS region of the EM spectrum shows promise in discriminating between water-stressed and non-stressed Shiraz vines.

(5)   Optimising RF and XGBoost hyperparameter values do lead to increased classification accuracies. However, careful consideration should be given to the choice of classifier and the application.

5. Acknowledgments

6. References

Abdel-Rahman, E.M., Makori, D.M., Landmann, T., Piiroinen, R., Gasim, S., Pellikka, P. & Raina, S.K., 2015. The utility of AISA eagle hyperspectral data and random forest classifier for flower mapping. *Remote Sensing*, 7(10), pp. 13298–13318.

Abdel-Rahman, E.M., Mutanga, O., Adam, E. & Ismail, R., 2014. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, pp. 48–59.

Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E.M. & Mutanga, O., 2017. Detecting the early stage of Phaeosphaeria leaf spot infestations in maize crop using in situ hyperspectral data and guided regularized random forest algorithm. *Journal of Spectroscopy*, 2017, pp. 1–8.

Belgiu, M. & Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24–31.

Belgiu, M., Tomljenovic, I., Lampoltshammer, T.J., Blaschke, T. & Höfle, B., 2014. Ontology-based classification of building types detected from airborne laser scanning data. *Remote Sensing*, 6(2), pp.1347–1366.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G. & Jones, Z., 2016. mlr: Machine Learning in R. *Journal of Machine Learning Research*, 17(170), pp. 1–5.

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5–32.

Chandrashekar, G. & Sahin, F., 2014. A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), pp. 16–28.

Chen, T. & Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. San Francisco,CA, USA, ACM, pp. 785–794.

Chen, T., He, T., Benesty, M., Khotilovich, V. & Tang, Y., 2017. xgboost: Extreme Gradient Boosting.

Congalton, R.G. & Green, K., 2009. *Assessing the accuracy of remotely sensed data: principles and practices*. 2nd ed. Boco Raton, FL, USA: CRC press.

Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232.

Fu, Y., Zhao, C., Wang, J., Jia, X., Yang, G., Song, X. & Feng, H., 2017. An improved combination of spectral and spatial features for vegetation classification in hyperspectral images. *Remote Sensing*, 9(3), pp. 1–16.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S. & Wolff, E., 2018. Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience and Remote Sensing*, 55(2), pp. 221–242.

Ghareb, A.S., Bakar, A.A. & Hamdan, A.R., 2016. Hybrid feature selection based on enhanced genetic algorithm for text categorization. *Expert Systems with Applications*, 49, pp.31–47.

Goldstein, B.A., Hubbard, A.E., Cutler, A. & Barcellos, L.F., 2010. An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. *BMC genetics*, 11(49), pp. 1–13.

Hartzell, P., Glennie, C. & Khan, S., 2017. Terrestrial hyperspectral image shadow restoration through lidar fusion. *Remote Sensing*, 9(5).

Jović, A., Brkić, K. & Bogunović, N., 2015. A review of feature selection methods with applications. In *Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2015)*. Opatija, Croatia: MIPRO Croatian Society: Rijeka, Croatia, pp.1200–1205.

Knauer, U., Matros, A., Petrovic, T., Zanker, T., Scott, E.S. & Seiffert, U., 2017. Improved classification accuracy of powdery mildew infection levels of wine grapes by spatial-spectral analysis of hyperspectral images. *Plant Methods*, pp. 1–15.

Kohavi, R. & Provost, F., 1998. Glossary of terms. *Machine Learning*, 30, pp. 271–274.

Kruskal, W.H. & Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), pp. 583–621.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Candan, C., Tang, Y. & Hunt, T., 2017. *caret: Classification and Regression Training*.

Lagrange, A., Fauvel, M. & Grizonnet, M., 2017. Large-Scale Feature Selection With Gaussian Mixture Models for the Classification of High Dimensional Remote Sensing Images. *IEEE Transactions on Computational Imaging*, 3(2), pp. 230–242.

Li, S., Wu, H., Wan, D. & Zhu, J., 2011. An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems*, 24(1), pp. 40–48.

Li, W., Feng, F., Li, H. & Du, Q., 2018. Discriminant Analysis-Based Dimension Reduction for Hyperspectral Image Classification: A Survey of the Most Recent Advances and an Experimental Comparison of Different Techniques. *IEEE Geoscience and Remote Sensing Magazine*, 6(1), pp. 15–34.

Liaw, A. & Wiener, M., 2002. Classification and Regression by randomForest. *R news*, 2(December), pp. 18–22.

Loggenberg, Kyle, Strever, A., Greyling, B. & Poona, N., 2018. Modelling water stress in a Shiraz vineyard using hyperspectral imaging and machine learning. *Remote Sensing*, 10(2).

Maimaitiyiming, M., Ghulam, A., Bozzolo, A., Wilkins, J.L. & Kwasniewski, M.T., 2017. Early Detection of Plant Physiological Responses to Different Levels of Water Stress Using Reflectance Spectroscopy. *Remote Sensing*, 9(7).

Martinez-de-Pison, F.J., Gonzalez-Sendino, R., Aldama, A., Ferreiro, J. & Fraile, E., 2017. Hybrid Methodology Based on Bayesian Optimization and GA-PARSIMONY for Searching Parsimony Models by Combining Hyperparameter Optimization and Feature Selection. In F. J. de Pisón, R. Urraca, H.

Quintián, & E. Corchado, eds. *In Lecture Notes in Computer Science, Proceedings of the 12th International Conference on Hybrid Artificial Intelligent Systems (HAIS 2017)*. La Rioja, Spain: Springer, Cham, pp. 52–62.

Matese, A. & Di Gennaro, S.F., 2015. Technology in precision viticulture: A state of the art review. *International Journal of Wine Research*, 7(1), pp. 69–81.

Mohite, J., Karale, Y., Pappula, S., Shabeer T. P., A., Sawant, S.D. & Hingmire, S. 2017. Detection of pesticide (Cyantraniliprole) residue on grapes using hyperspectral sensing. In M. S. Kim, K. l Chao, B. A. Chin, & B. K. Cho, eds. *In Sensing for Agriculture and Food Quality and Safety IX, Proceedings of the SPIE Commercial+ Scientific Sensing and Imaging Conference*. Anaheim, CA, USA: International Society for Optics and Photonics: Bellingham,WA, USA.

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), pp. 358–371.

Pedergnana, M., Marpu, P.R. & Mura, M.D., 2013. A Novel Technique for Optimal Feature Selection in Attribute Profiles Based on Genetic Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6), pp. 3514–3528.

Poblete-Echeverría, C., Olmedo, G.F., Ingram, B. & Bardeen, M. 2017. Detection and segmentation of vine canopy in ultra-high spatial resolution RGB imagery obtained from Unmanned Aerial Vehicle (UAV): A case study in a commercial vineyard. *Remote Sensing*, 9(3).

Pôças, I., Rodrigues, A., Gonçalves, S., Costa, P.M., Gonçalves, I., Pereira, L.S. & Cunha, M., 2015. Predicting grapevine water status based on hyperspectral reflectance vegetation indices. *Remote Sensing*, 7(12) , pp. 16460–16479.

Poona, N., van Niekerk, A. & Ismail, R., 2016. Investigating the utility of oblique tree-based ensembles for the classification of hyperspectral data. *Sensors*, 16(11).

Poona, N.K. & Ismail, R., 2014. Using Boruta-selected spectroscopic wavebands for the asymptomatic detection of fusarium circinatum stress. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(9), pp. 3764–3772.

Poona, N.K., Van Niekerk, A., Nadel, R.L. & Ismail, R., 2016. Random Forest (RF) Wrappers for Waveband Selection and Classification of Hyperspectral Data. *Applied Spectroscopy*, 70(2), pp. 322–333.

Pudil, P., Novovičová, J. & Kittler, J., 1994. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11), pp. 1119–1125.

R Development Core Team, R., 2017. R: A Language and Environment for Statistical Computing. In R Foundation for Statistical Computing: Vienna, Austria. https://www.r-project.org/.

Rodrìguez, J.J., Kuncheva, L.I. & Alonso, C.J., 2006. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), pp. 1619–1630.

Santara, A., Mani, K., Hatwar, P., Singh, A., Garg, A., Padia, K. & Mitra, P. 2017. BASS Net: Band-Adaptive Spectral-Spatial Feature Learning Neural Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9): 5293–5301.

Taşkın, G., Hüseyin, K. & Bruzzone, L., 2017. Feature Selection Based on High Dimensional Model Representation for Hyperspectral Images. *IEEE Transactions on Image Processing*, 26(6), pp. 2918–2928.

Vélez Rivera, N., Gómez-Sanchis, J., Chanona-Pérez, J., Carrasco, J.J., Millán-Giraldo, M., Lorente, D., Cubero, S. & Blasco, J., 2014. Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning. *Biosystems Engineering*, 122, pp. 91–98.

Vora, S. & Yang, H., 2017. A comprehensive study of eleven feature selection algorithms and their impact on text classification. In *Proceedings of the 5th Computing Conference*. Hilton Kensington, London, UK: IEEE: 440–449.

Wietecha, M., Jełowicki, Ł., Mitelsztedt, K., Miścicki, S. & Stereńczak, K., 2019. The capability of species-related forest stand characteristics determination with the use of hyperspectral data. *Remote Sensing of Environment*, 231(111232).

Xia, Y., Liu, C., Li, Y.Y. & Liu, N., 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78, pp. 225–241.

Xu, Q.-S. & Liang, Y.-Z., 2001. Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56, pp. 1–11.

Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A. & Liu, H., 2010. Advancing Feature Selection Research. *ASU Feature Selection Repository*, pp. 1–28.

Zygielbaum, A.I., Gitelson, A.A., Arkebauer, T.J. & Rundquist, D.C., 2009. Non-destructive detection of water stress and estimation of relative water content in maize. *Geophysical Research Letters*, 36(12).