Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 707–712
DOI: 10.15439/2024F7286
ISSN 2300-5963 ACSIS, Vol. 39

A bottom-up approach to select constrained spectral bands discriminating vine diseases

Shurong Zhang*, Alban Goupil*, Valeriu Vrabie*, Eric Perrin*, and Marie-Laure Panon†
*Université de Reims Champagne-Ardenne, CReSTIC, Reims, France
†Comité Champagne, 5 Rue Henri Martin, 51200 Épernay, France
shurong.zhang@univ-reims.fr; marie-laure.panon@civc.fr

***Abstract*—The detection and control of diseases constitute a primary objective of French viticultural research. In this paper, we present a bottom-up hierarchical approach for selecting spectral bands suitable for class discrimination of spectra acquired by Infrared spectroscopy. Our method entails evaluating neighboring bands using various similarity metrics, applying aggregation criteria, and ultimately identifying a limited number of the most relevant bands for the separation of classes. The bandwidths are limited within a range as is typically required for choosing existing optical filters or specifying colored filter arrays. Our approach facilitates the discovery of distinctive spectral bands associated with a disease of interest, enabling the customization of multispectral cameras to meet specific requirements. It was applied to spectra collected on vine leaves spanning a three-year period with the goal to identify the most discriminant bands for the detection of grapevine yellows. The results show that a limited number of bands are sufficient to identify this class of interest through a classifier based on Linear Discriminant Analysis.**

***Index Terms*—Band selection, hierarchical model, classification, Grapevine Flavescence Dorée**

I. Introduction

FLAVESCENCE Dorée, a serious epidemic disease, is one of the yellow diseases of grapevines. In order to control the risk of its spread, it is necessary to develop an effective and high-throughput detection tool. In this context, we are interested in developing an approach for selecting discriminative spectral bands based on spectra acquired by Near Infrared (NIR) spectroscopy, with a view to specifying a multispectral camera suitable for large-scale acquisitions. These spectral bands must allow optimal separation of the different classes, two in our application: yellow plants and healthy plants. On the other hand, we want to identify a limited number of spectral bands that can have different widths, but with restricted minimum and maximum bandwidths.

We position ourselves in the context of reducing the dimensions of spectral data. This goal can be achieved through band selection techniques that allow the selection of representative bands from an original spectrum. Although there are several selection methods, we will focus on those that are most relevant to our topic.

In [1], unsupervised hierarchical clustering approaches that merge similar wavelengths into bands were proposed. The models named WaluMi for mutual information and WaluDi for Kullback-Leibler divergence use these metrics to assess the similarity between bands. The wavelength representing a band with the highest average similarity to others is selected as the band's representative. These models offer the advantage of merging wavelength similarities into a single band and choosing the optimal wavelength as the representative of this band. The main disadvantage is that only wavelengths are selected, which means that if one wants to specify an optical filter for a multispectral camera, it must have a very narrow band, which is very difficult to achieve in practice. Our approach focuses on aggregating multiple wavelengths into bands, thereby reducing computational complexity while maintaining the performance of identified bands. The rationale behind this choice is rooted in the physical correlation of adjacent wavelengths of NIR spectra, making the computation of similarities restricted to neighboring wavelengths sufficient. Additionally, the bands in our context are inherently adjacent wavelengths with minimal bandwidth. This deliberate restriction not only simplifies the computational requirements but also aligns with the practical constraints of our intended application.

A similar approach was proposed in [2], but using an adaptive hyperbolic distance as similarity measure. This distance avoids obtaining bands with a single or very limited number of wavelengths. Thereby, it partially respects our constraints, however, does not impose a maximum width of the bands. This is problematic in practice, particularly if we use the same sensor on which a colored filter array is placed, the quantity of absorbed light by the sensor being dependent on the bandwidth of each filter. Furthermore, this approach only calculates similarities restricted to neighboring bands.

In [3], the MRMR (Max Relevance and Min Redundancy) algorithm uses statistical selection to independently evaluate the importance of features. Its objective is to select an optimal subset that maximizes the relevance between features and classes, while minimizing the correlation between the selected features, achieved through the application of mutual information. This algorithm is typically applied to identify a set of discriminating wavelengths. In our case, we used this algorithm to identify discriminating bands after several wavelengths have been merged into bands.

In [4] an approach is proposed that uses a Random Forest method based on Recursive Feature Elimination (RFE) criteria. This approach is able to automatically eliminate feature redundancy and generally provides better and more compact subsets.

The project leading to this publication has received funding from the Grand-Est region, France.

The authors used Random Forest to evaluate the importance of features, and we have chosen it to compare it with MRMR in the second step of our method, where bands are selected. The feature selection of the RFE approach is also driven by the classes, but by a classification approach that may be different from the information theory used in MRMR. Note that there are similar approaches to RFE, such as Sequential Feature Selection [5], which greedily incorporates features by either adding or removing them to create a subset. Since fairly similar results were obtained with a Random Forest as an estimator, we only subsequently chose RFE for the sake of simplifying the display of the results.

These methods have been proposed to select wavelengths. Multispectral cameras typically capture a range of wavelengths with a central value and a certain number of wavelengths on either side. Those composed of several sensors use an optical filter for each sensor having fixed bandwidths imposed by the filter manufacturers, typically at least 10 nm. For specifying such a camera, it is possible to integrate the available NIR spectra with windows of desired width, such as Hanning, and then apply one of the previous methods by identifying bands which correspond to these integrations. However, all the filters will have the same bandwidth. For multispectral cameras using a single sensor with a color filter array, it is possible to specify the bandwidths of each filter but which, for practical reasons, must have a specified minimum and maximum limit.

The objective here is to propose an approach allowing aggregating similar wavelengths into bands and then identify a limited number of the most relevant bands for the separation of classes. In the first step, our hierarchical model allows multiple wavelengths to be combined into a band, emulating the physics of optical filters that can have different bandwidths, as may be the case for specifying color filter arrays. In the second step, a few bands are selected, as is necessary when one wishes to specify a suitable multispectral camera. Indeed, for cameras composed of several sensors, increasing the number of bands requires as many sensors and suitable electronics, which leads to cameras with important dimensions and weight and to a significant parallax phenomenon. For cameras using colored filter arrays, increasing the number of bands leads to a decrease in spatial resolution. This work is an extension of our previous work [6] in which we take into account new constraints, minimal and maximal bandwidths. The application is the same: specify a camera adapted to the detection of grapevine yellows by identifying optimal bands capable of effectively distinguishing healthy plants and grapevine yellows.

Since our method consists of two crucial steps: band grouping and band selection, we chose to use the spectral clustering algorithm studied in [7] to compare the performance of the band grouping. Spectral clustering is widely used to group data points based on their similarity. The author explores the concept of representing data as a graph, where connections between points reflect their similarity, and aims to divide the data into clusters such that points within a cluster are similar, while points in different clusters are dissimilar. In our approach, we adapt spectral clustering to group similar adjacent bands by generating an adjacency matrix in which only adjacent bands are connected. This adjacency matrix is then passed to the spectral clustering algorithm with the affinity set to 'precalculated' to use the user-defined adjacency matrix. Once we have the groups of bands, we apply the same selection criteria as our approach to selecting the discriminative bands for class separation.

II. Proposed Method

As shown in Figure 1, the proposed hierarchical clustering approach is a two-step procedure. First, the wavelengths are merged based on their similarity by a bottom-up clustering into several bands that respect a width constraint, as presented in Section II-A. Each band is characterized by an interval that combines several weighted wavelengths to emulate the response of an optical filter. The pseudocode outline is shown in Algorithm 1. Following that, we assess the relevancy of the obtained bands for classes differentiation, as elaborated in Section II-B. A selection among these bands is done that is relevant according to the available classes. Finally, we choose the most effective bands for class discrimination, which are the final output of our algorithm, bands that allow the best classification, indicating whether they are relevant to the classes we have.

The inputs are the P NIR spectra, each acquired for a class $c \in C$ and having n wavelengths. A spectrum $S(\lambda_i) \in \mathbb{R}^P$ gives the reflectance of the target for each wavelength λ_i in the range $[\lambda_1, \ldots, \lambda_n]$.

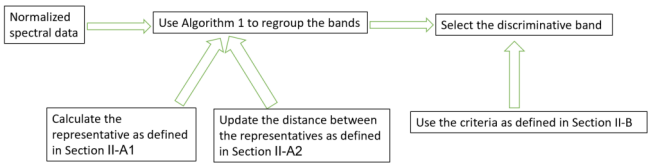

Fig. 1: The pipeline of our band selection approach.

A. Searching for relevant similarity while preserving adjacency

Initially, each band B_i consists of a single wavelength λ_i. The merging process is based on the similarity between adjacent bands, which is calculated on the representatives R_i of each band B_i. It begins by combining the most similar adjacent bands at each step and continues until no feasible band combination exceeds the specified maximum number Max of wavelengths. Here, Max represents the largest size of the desired optical filter response. If the merging of the most similar adjacent bands is constrained by Max, the algorithm will not merge them and consider them as independent bands, it will merge the next most similar adjacent bands that are not constrained by Max, and so on. This program accomplishes the task by calculating the distances between each pair of adjacent bands and generating a sorted list of bands. The algorithm attempts to merge bands, starting with the top selection and working down the list. As soon as two bands are

merged, the sorted list loop is closed, and the next iteration is started. At each iteration, it checks the total number of bands. If the number remains constant for two consecutive iterations, it indicates that the model cannot merge two bands without violating constraints. At this point the main loop is closed and the optimal set of bands is obtained. Additionally, each band should consist of at least Min wavelengths. At the end of the process, a check is performed and any bands that do not meet this criterion are removed.

Since similarity calculations are exclusively performed between a band and its two neighbors, the search for maximum similarity follows a linear pattern. To speed up the algorithm, similarities are not calculated at each iteration. Instead, the list of similarities between adjacent bands is computed at the beginning and is only updated between the new merged band and its neighbors in each iteration. Compared to the bottom-up hierarchical groupings mentioned above, this implementation is more adaptable to our case for selecting the groups of neighboring wavelengths. It also has quadratic complexity compared to cubic complexity.

Data: List of wavelengths $\lambda_1, \ldots, \lambda_n$
Data: Reflectance $S(\lambda_i) \in \mathbb{R}^P$, with P the number total of spectra, each belonging to a class $c \in C$
Result: Set of m bands $\mathcal{B} = \{B_1, \ldots, B_m\}$
`/* Initialization */`
$B_i = \{\lambda_i\}, i = 1, \ldots, n$;
Representatives $R_i = S(B_i)$ cf section II-A1;
Distances $d(R_i, R_{i+1})$ between the representatives of bands adjacent, $i = 1, \ldots, n-1$ cf section II-A2;
`/* Hierarchical grouping */`
while $length(B) > 1$ **do**
  $list\_sort \leftarrow argsort_i d(R_i, R_{i+1})$;
  $bandsize = length(B)$;
  while $a \in list\_sort$ **do**
    if $length(B(a)) + length(B(a+1)) < Max$ **then**
      $B_a \leftarrow B_a \cup B_{a+1}$ and remove B_{a+1} from $\mathcal{B}$;
      Compute the representative $R_a = R(B_a)$ of B_a;
      Update the distances $d(R_{a-1}, R_a)$ and $d(R_a, R_{a+1})$;
      `/* adjacent bands found */`
      break;
  if $length(B) = bandsize$ **then**
    `/* no aggregation found */`
    break;
for i ***from*** *1* ***to*** *length*(B) **do**
  if *length*$(B[i]) < Min$ **then**
    Delete $B[i]$;
return $\mathcal{B}$;

Algorithm 1: Our hierarchical bottom-up classification

1) Representative of a band: Our approach consists of weighting spectral information at wavelengths composing a band B_i. We use a Hanning window that emulates a real optical filter physically, but other windows can also be used. The R_i representative of the B_i band is calculated as the weighted average of the spectral information in this window

$$R_i = \frac{1}{L} S(B_i) \cdot W(L), \tag{1}$$

where $W(L) \in \mathbb{R}^L$ is the Hanning window of size L, corresponding to the number of wavelengths contained in the band B_i, and $S(B_i) \in \mathbb{R}^{P \times L}$ is the spectral information at the wavelengths contained in this band.

2) Distance between the representatives: We estimate the distance between the representatives of these bands as the similarity between these bands. Four distances between band representatives are considered in our work: the commonly used Euclidean distance D_{L2}; the adaptive hyperbolic distance D_{HY} proposed by [2]; the Jensen-Shannon divergence D_{JS}, which takes advantage of the distribution of spectral information rather than spectral values; and the conditional mutual information D_{CMI}, which takes into account the labels $c \in C$ associated to each spectrum. Only the last two are described in detail below.

a) Jensen-Shannon (JS) divergence: can be interpreted as a kind of similarity between two probability distributions that symmetries the Kullback-Leibler distance. We assume that R_i and R_j are the representatives of the bands B_i and B_j. The distance based on the divergence of JS, D_{JS} is defined by

$$D_{JS}(R_i, R_j) = \frac{1}{2} D_{KL}(R_i \mid M) + \frac{1}{2} D_{KL}(R_j \mid M), \tag{2}$$

where D_{KL} is the Kullback-Leibler divergence between the probability densities of the variables R_i, R_j and M, the mean distribution of the distributions of R_i and R_j. The Jensen-Shannon divergence is always positive or zero. It cancels out when R_i and R_j have the same probability distribution. Thus, two bands B_i and B_j with a low value of D_{JS} are combined if the representatives of these bands have close densities.

b) Conditional Mutual Information (CMI): measures the dependence between the spectral information of the B_i bands in each class $c \in C$. Each class c has a probability $p(c)$ corresponding to its proportion among the data set. The mutual information $I(R_i; R_j \mid c)$ between the representatives of B_i and B_j conditioned to class c is given by the weighted average

$$I(R_i; R_j \mid C) = \sum_{c \in C} I(R_i; R_j \mid c) \times p(c). \tag{3}$$

and the CMI distance is defined by

$$D_{CMI}(R_i, R_j) = \frac{1}{1 + I(R_i; R_j \mid C)}. \tag{4}$$

This distance offers a key advantage because it incorporates spectral information from two bands within different classes.

B. Selecting relevant bands for classification

Since the bandwidth of the built bands is constrained, the procedure may yield too many representative bands.

1) Examining the variance (VAR): Examining the variance between each band and the classes allows us to prioritize bands with higher values. The ranking of them involves computing the variance between the final representative of each band and the class, and the selection of discriminating representatives is based on this ranking.

2) Leverage state-of-the-art methods: Using band selection techniques such as MRMR and RFE, we are able to select discriminating representatives with respect to the class.

We chose these selection criteria because they allow us to evaluate the variance, mutual information and accuracy of the classifier, three different aspects that allow us to evaluate the discriminatory power of the bands identified for class separation. Once we have found those bands, given that they are the most relevant bands for class separation that we have, we can then specify the filters of a multispectral camera suitable for the application considered.

III. Results

We have carried out several acquisition campaigns on vine leaves of the Chardonnay grape variety at the Comité Champagne's Plumecoq experimental estate. The acquisitions were carried out after the harvest periods of the years 2021 to 2023. NIR spectra were acquired with a LabSpec4 portable spectrometer on around 700 grapevine leaves per year that were picked and arranged on polystyrene boards. Two spectra were acquired on each leaf, which gives a collection of $P = 4282$ spectra acquired between 2021 and 2023. This spectrometer provides spectral information every 1 nm from 350 to 2500 nm, but only the $400 - 1000$ nm region was retained. The collection $S(\lambda_i) \in \mathbb{R}^P$ thus represents the reflectance of the leaves for each wavelength λ_i in the range $[\lambda_1, \ldots, \lambda_n]$ with $n = 600$. The aim of this study was to identify the spectral bands that allow us to distinguish the $C = 2$ classes: healthy plants and grapevine yellows. To identify the discriminating bands, from the collection of $P = 4282$ spectra, 1709 spectra correspond to grapevine yellows and 2573 spectra to healthy plants.

We set $Min = 10$ wavelengths and $Max = 100$ wavelengths because the bands of the optical filters available on the market or the color filter array that can be specified are generally within these limits. All available $P = 4282$ spectra were used to identify discriminative bands with the proposed approach. For comparison, as mentioned in the introduction, we adapted the spectral clustering to group the adjacent wavelengths into bands, without bandwidth restriction, and then applied the selection techniques to obtain the discriminative bands.

Once the bands have been identified, we divided the data, i.e. the spectral information at the selected bands, into two sets, one training and one test. The training set contains 90% of the data, i.e. 3854 spectra, and the test set contains the remaining 10%, i.e. 428 spectra. In the former set, a Linear Discriminant Analysis (LDA) classifier was trained, and classification performance was evaluated in the latest set considering spectral information at the bands identified by the selection methods. We decided to use the LDA classifier for the evaluation because, compared to other classifiers such as Support Vector Machine (SVM) or k-Nearest Neighbors (KNN), LDA is more often used for feature extraction, where it identifies the most relevant features for class discrimination.

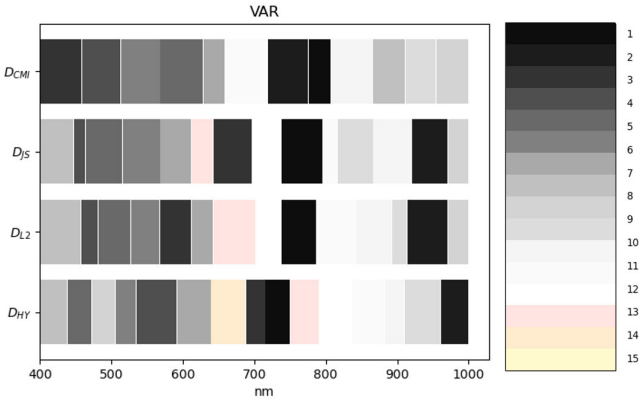

Fig. 2: Band grouping and VAR selection.

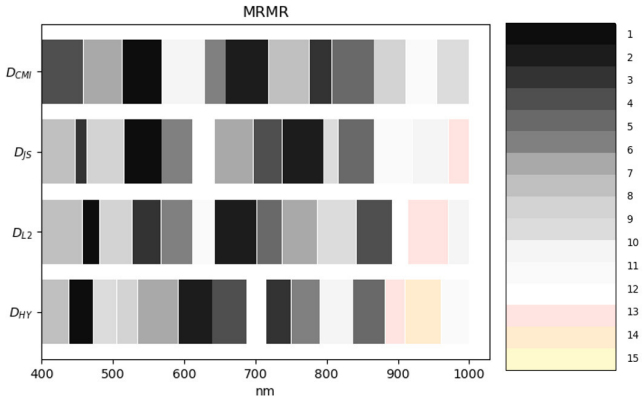

Fig. 3: Band grouping and MRMR selection.

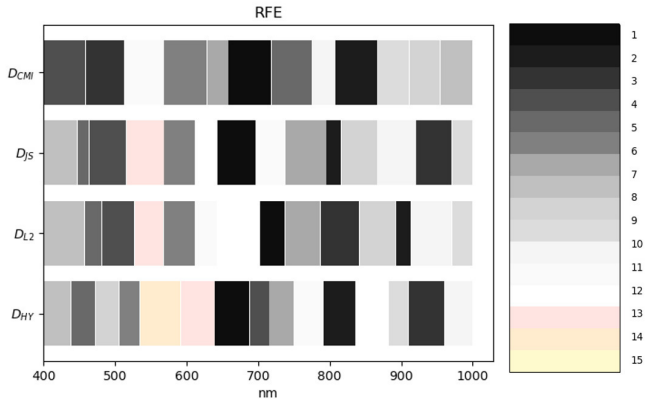

Fig. 4: Band grouping and RFE selection.

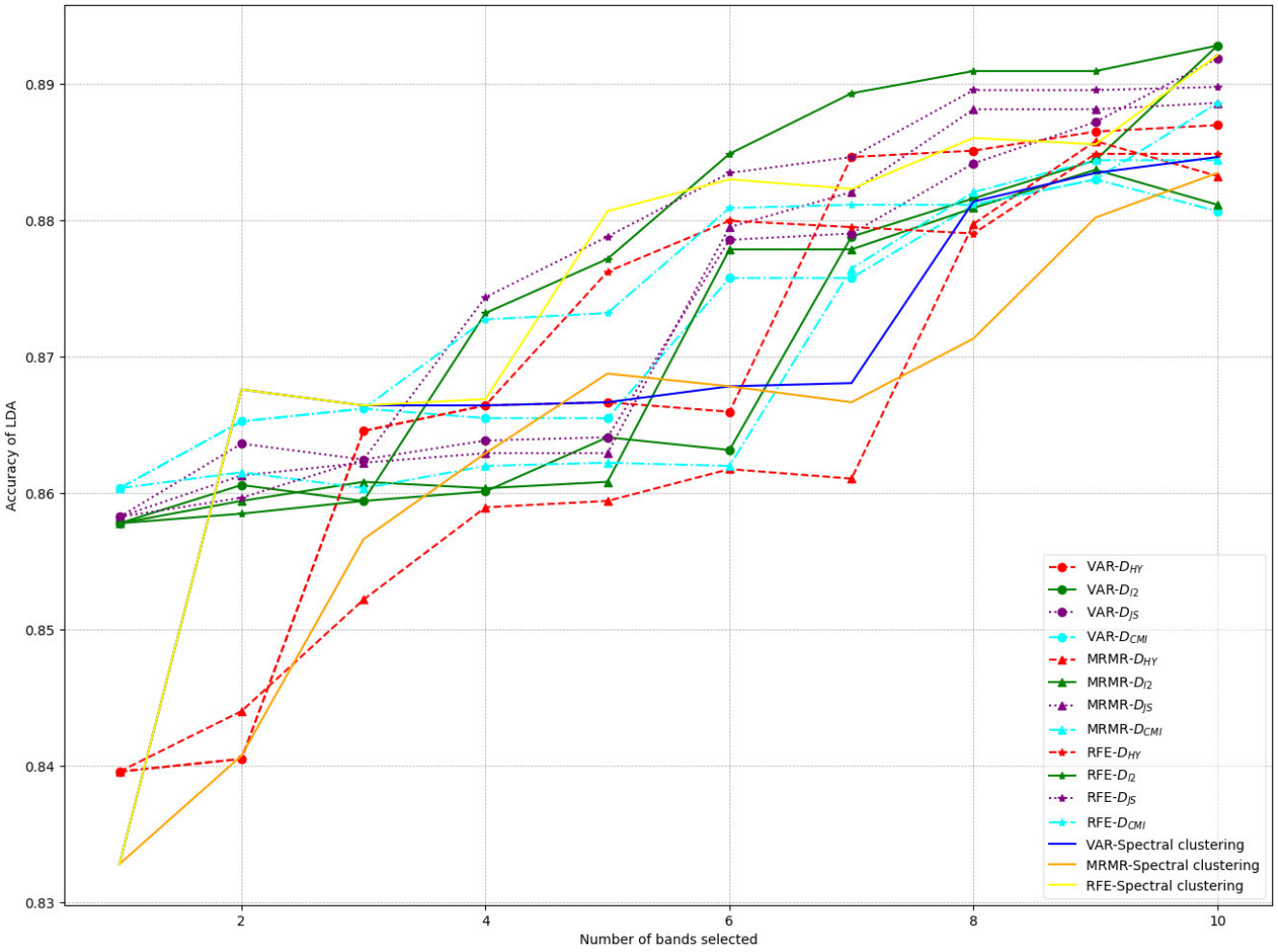

Fig. 5: Accuracy values for LDA with all methods.

	D_{HY}	D_{L2}	D_{JS}	D_{CMI}
VAR	0.8659	0.8631	0.8785	0.8757
MRMR	0.8617	0.8778	0.8794	0.8619
RFE	0.8799	**0.8848**	0.8834	0.8808

(a) 6 discriminating bands

	D_{HY}	D_{L2}	D_{JS}	D_{CMI}
VAR	0.8846	0.8787	0.8790	0.8757
MRMR	0.8610	0.8778	0.8820	0.8764
RFE	0.8794	**0.8893**	0.8846	0.8811

(b) 7 discriminating bands

Fig. 6: Accuracy of LDA for 6 and 7 discriminating bands

Figure 2, 3 and 4 show the bands obtained in relation to the different distances proposed in section II-A2. The colors assigned convey the discriminatory quality of the bands for class discrimination in relation to the different criteria proposed in section II-B. A darker color indicates the discriminatory nature of a band and the accompanying number indicates its position as the n^{th} band selected by the specified method. A consistent trend is observed across all methods, showing a higher frequency of selection for bands around 700 nm and an intermediate frequency of selection for bands between 400 nm and 500 nm across all three selection methods. In contrast, bands between 900 nm and 1000 nm are rarely selected in the MRMR selection. We could also note that, as discussed in the introduction, D_{HY} has more uniform bands than the other methods. In addition, each distance provides a different number of bands: D_{HY} has 15 bands, D_{L2} and D_{JS} each have 14 bands, and D_{CMI} has 12 bands.

In Figure 5, the LDA accuracy is shown for each method, with distinctive markers assigned based on the selection technique: variance (VAR) selection is marked with 'o', MRMR selection is marked with a triangle, and RFE selection is marked with '*'. Different colors are also assigned to each distance metric: red for the hyperbolic distance (D_{HY}), green for the L2 distance (D_{L2}), purple for Jenson-Shannon distance (D_{JS}), and cyan for conditional mutual information distance (D_{CMI}). Straight curves are used for spectral clustering with

selection techniques. The results show that increasing the number of bands is associated with an expected increase in LDA accuracy due to the additional spectral information. Note that the accuracy of LDA using all 600 available wavelengths (from 400 to 1000 nm) is 0.9523. In particular, techniques that use D_{JS}, D_{L2}, and D_{CMI} show superior performance over different numbers of bands. In contrast, methods that use D_{HY} show less effectiveness, especially when we select a small number of bands, possibly due to the narrow discriminating bands at other distances compared to the uniform nature of all bands in D_{HY}.

However, the hyperbolic distance works well in our previous methods [6], which is to be expected because, as we mentioned in the introduction, this distance produces almost uniform bands. However, even though it is designed to avoid this type of problem without imposing a bandwidth limit, this distance leads to quite heterogeneous bands in terms of width. Even if the classification results obtained with LDA are satisfactory, the practical realization of a multispectral camera using a color-array filter could still be difficult, since the amount of light absorbed by the same sensor depends on the bandwidth of the filters.

For spectral clustering, we regrouped the adjacent wavelengths in 15 bands using the spectral clustering, then we applied the three criteria of selection,. We could notice that its performance is less than our methods except for RFE-Spectral clustering in the case of selecting 5 discriminating bands.

Compared to the original hierarchical clustering method, our model imposes limits and stops as soon as we can't merge bands within the size limit. This approach prevents us from producing a band that exceeds the camera's range. However, spectral clustering does not have the constraint of the camera's band range, meaning it is possible to obtain a band that falls outside the camera's range in the results.

The evaluation is limited to 1 to 10 bands, in line with the objective of constructing multispectral cameras with relevant characteristics, including dimension, volume, and spatial resolution. Beyond this threshold, the associated cost increases significantly and widely available cameras typically offer between 6 and 7 band configurations. Consequently, figure 6 summarizes the accuracy for two classical configurations. Using 6 and 7 representative bands, the most effective method is RFE-D_{L2} with an accuracy of 0.8848 and 0.8893.

IV. Conclusion and outlook

We have proposed a new hierarchical bottom-up approach that merges the bands and then selects the most discriminating ones. This approach, which imposes a constraint on the bandwidths, aggregate similar bands by assessing the similarity between band representatives using different distances (L2, hyperbolic distance, Jensen-Shannon divergence, conditional mutual information). Finally, it selects the most relevant bands for the separation of classes using criteria based on maximization of the variance between bands and classes or using state-of-the-art selection techniques, MRMR and RFE.

This approach is useful in specifying a multispectral camera suitable for a specific application. It requires spectra to be recorded on samples belonging to different classes and to indicate the number of desired bands. In our case, we used it to identify bands distinguishing the vine yellows from healthy plants.

Work is underway to make the algorithm robust to multi-year acquisitions and to configure it so that it identifies bands that are insensitive to the acquisition year.

References

[1] A. Martínez-Usó, F. Pla, J. M. Sotoca, P. García-Sevilla, Clustering-Based Hyperspectral Band Selection Using Information Measures, IEEE Trans. Geosci. Remote Sens. 45(12), pp. 4158-4171, Dec. 2007. DOI: 10.1109/TGRS.2007.904951

[2] H. Sun, L. Zhang, J. Ren, H. Huang, Novel hyperbolic clustering-based band hierarchy (HCBH) for effective unsupervised band selection of hyperspectral images, Pattern Recognition, 130 :108788, October 2022. DOI: 10.1016/j.patcog.2022.108788

[3] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27(8), pp. 1226-1238, Aug. 2005. DOI: 10.1109/TPAMI.2005.159

[4] P. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products, Chemometrics and Intelligent Laboratory Systems, 83(2), 2006, DOI: 10.1016/j.chemolab.2006.01.007

[5] D. Aha, R. Bankert, A comparative evaluation of sequential feature selection algorithms, Proc. Fifth International Workshop on Artificial Intelligence and Statistics, PMLR, 1995.

[6] S. Zhang, E. Perrin, A. Goupil, V. Vrabie, M.-L. Panon, Nouveau modèle hiérarchique ascendant pour la sélection des bandes spectrales discriminant les maladies de la vigne, Colloque GRETSI, Grenoble, 2023.

[7] U. V. Luxburg, A tutorial on spectral clustering, Statistics and Computing, 17(4):395–416, December 2007. DOI: 10.48550/arXiv.0711.0189