

Received May 23, 2021, accepted May 28, 2021, date of publication June 7, 2021, date of current version June 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3087231

Matrix Relevance Learning From Spectral Data for Diagnosing Cassava Diseases

GODLIVER OWOMUGISHA^{1,2,3}, FRIEDRICH MELCHERT^{1,4},
ERNEST MWEBAZE², JOHN. A. QUINN², AND MICHAEL BIEHL¹

¹Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, 9700 AK Groningen, The Netherlands

²College of Computing and Information Sciences, Makerere University, Kampala, Uganda

³Faculty of Engineering, Busitema University, Tororo, Uganda

⁴Fraunhofer Institute for Factory Operation and Automation IFF, 39106 Magdeburg, Germany

Corresponding author: Godliver Owomugisha (g.owomugisha@rug.nl)

This work was supported by the Bill and Melinda Gates Foundation under Grant OPP1112548.

ABSTRACT We discuss the use of matrix relevance learning, a popular extension to prototype learning algorithms, applied to a three-class classification task of diagnosing cassava diseases from spectral data. Previously this diagnosis has been done using plant image data taken with a smartphone. However for this method disease symptoms need to be visible. Unfortunately for some cassava diseases, once symptoms have manifested on the aerial part of the plant, the root which is the edible part of the plant has been totally destroyed. This research is premised on the hypothesis that diseased crops without visible symptoms can be detected using spectral information, allowing for early interventions. In this paper, we analyze visible and near-infrared spectra captured from leaves infected with two common cassava diseases (cassava brown streak disease and cassava mosaic virus disease) found in Sub-Saharan Africa. We also take spectra from leaves of healthy plants. The spectral data come with thousands of dimensions, therefore different wavelengths are analyzed in order to identify the most relevant spectral bands for diagnosing these disease. To cope with the nominally high number of input dimensions of data, functional decomposition of the spectra is applied. The classification task is addressed using Generalized Matrix Relevance Learning Vector Quantization and compared with the standard classification techniques performed in the space of expansion coefficients.

INDEX TERMS Cassava disease diagnosis, feature selection, matrix relevance learning, spectral data.

I. INTRODUCTION

The ability to quickly diagnose disease in the field is of critical importance in most agro-reliant economies the world over. For places where the crop grown is not only of economic but also food security importance, this is particularly crucial. In this study we investigate improved ways of accurately diagnosing plants in the field by leveraging a unique dataset: spectral data from plant leaves, and using improved algorithms that not only provide higher accuracy but also a profile of wavelengths that are most important for the disease classification task.

We particularly focus on cassava (*Manihot esculenta*), a staple crop in Sub-Saharan Africa that feeds over 500 million people daily. Cassava suffers from two serious diseases: cassava brown streak disease (CBSD) and cassava mosaic virus disease (CMD). According to [1], CBSD and CMD

together account for over 90% of yield losses in cassava production systems in Sub-Saharan Africa. This in turn greatly affects smallholder farmers.

We present an approach to detection and diagnosis of CBSD and CMD based on image spectroscopy to extract representative features from example leaves manifesting these diseases, and machine learning for building the predictive models based on such data. This work is an early step in our endeavor to run experiments on diseased but non-symptomatic cassava plants using spectral data collected from spots on a diseased leaf that are symptomatic and spots that are non symptomatic. The novelty in our approach is not only applying spectroscopy in field-level diagnostics of cassava but also optimizing the models for fewer features of the data while maintaining accuracy. This is very important for deployment. Our eventual goal is to deploy these models on low-cost sensor devices to capture spectral data from the leaves of a plant and provide a reading of its state of health.

The associate editor coordinating the review of this manuscript and approving it for publication was Tomasz Trzcinski¹.

The field of feature engineering and feature selection provides techniques for reducing the number of required features of input data, usually making the model simpler and less prone to bias from noisy features. Here we employ Generalized Matrix Relevance Learning Vector Quantization (GMLVQ), a highly intuitive algorithm that can be optimized in the training to give high accuracy and also detect the most important features relevant for the classification task.

We specifically apply GMLVQ here for a number of reasons: It has been used successfully in previous related studies and displayed favorable performance e.g [2]–[4]. It is, however, not our aim to show that GMLVQ outperforms other classifiers in the problem at hand. Prototype-based systems in general are natural tools for the analysis of multi-class datasets. GMLVQ is particularly suitable when combined with efficient dimensional reduction methods as highlighted in Section III. Importantly, GMLVQ offers great interpretability and insight into the importance of different input features for the classification task at hand and in our case, can serve as a tool for the identification of the most relevant spectral wavelengths. Here, we exploit these important aspects to a large extent as outlined in greater detail in Section III-A.

The sections that follow describe the experimental procedure we followed to provide evidence of the efficacy of using spectral data for this task and optimizing the models for a reduced featureset of the data. Specifically, Section II gives a small synopsis of the literature related to the use of spectroscopy for classification and some examples, Section III describes the GMLVQ algorithm, the feature selection process and dimensionality reduction techniques and Section IV describes the experimental set-up we employed. Results and discussion are presented in Section V and VI respectively.

II. RELATED WORK

The de facto way to diagnose crops in the field has been through leaf images, for example taken with a smartphone. Several recent studies have demonstrated the efficacy of these methods on visual diagnosis of different crops. Our work builds on previous studies in [5], [6] that focused on the use of conventional smartphone camera plant images to diagnose disease in the field. Most of the earlier work considers the use of leaf images as the key data input into the model. For these techniques to be effective, diseases symptoms need to be visible on the aerial part of the plant. However, once symptoms have manifested, for some particular diseases, a lot of damage has already been inflicted particularly to the root of the plant and it can no longer be used as food. Our hypothesis is that spectral data collected from parts of the plant can offer a better signal of the inherent disease in the plant. This presents better opportunities for early detection of disease in the plant than image data. Examples of successful use of spectral data in early disease detection include work in Belasque *et al.* [7] where fluorescence spectroscopy was used to detect mechanical and disease stresses in citrus plants; a similar methodology was also employed in [8] for detection of diseases in citrus plants in the USA and Brazil.

Yang *et al.* [9] also presents work in the early detection of rice blast disease using near-infrared hyper-spectral imaging. Some previous work has focused on a combination of pests and diseases, for example in [10] where a multi-spectral imaging methodology for the diagnosis of plant diseases and insect pests was employed. The use of near-infrared spectroscopy to analyze cold rice blast is also discussed in [11]. Similarly, hyper-spectral data were used for the pre-symptomatic detection of infections in sugar beet in [12].

The uniqueness of our work is the focus on the early detection of diseases in plants based on spectral data. Using images only works once the diseases have manifested physically on leaves of the diseased plant. Using hyper-spectral imaging techniques is infeasible in our context because of the costs associated with acquiring the hyper-spectral cameras. Using a hand-held Spectrometer gives us a cheaper option to detect diseases in cassava plants before they are symptomatic giving the smallholder farmer in our case a window of time to apply an intervention to control the disease.

This work constitutes a significant extension of previous work in [13] that investigated diagnosing cassava diseases using spectral data from visibly infected leaves compared with use of image-based features extracted from crop leaves taken by a mobile camera.

A key aim of this earlier work was to gain a first understanding of whether the location of where spectra were taken from an infected plant or leaf matters. In particular, we studied the potential difference between taking spectral data from visibly infected parts of the leaf or from parts of the leaf that are not visibly infected. Results of that study showed a significant increase in performance for the experiment using the spectral data from the non-visibly infected part. Here, we extend the scope of the investigation significantly by comparing different methods for dimension reduction of spectral data and by exploiting the interpretability of the GMLVQ systems explicitly.

Spectral data by its nature is very high-dimensional, nominally comprising more than 3600 features or dimensions. A feature pre-processing step is thus essential in this case to ensure our models do not suffer from the large p small n problem; a common problem in machine learning when the data consists of more features than the number of examples.

To do this, we apply and compare different pre-processing strategies. As the baseline approach, we consider the use of the original high-dimensional data. To reduce data dimensionality, we employ functional representation of spectra using polynomial approximations and formulate the machine learning in the space of the corresponding coefficients. The basic approach is introduced and investigated using spectral data from different contexts and more generally functional data [3], [4].

We also employ standard principal component analysis (PCA) as a dimension reduction technique and compare it to the other schemes.

The ultimate aim is to identify a specific feature representation or particular features (i.e. wavelengths or ranges of

wavelengths) which contain most information for the classification and that will facilitate technical solutions using simple sensors.

The selection of features is mainly addressed within the example framework of GMLVQ, e.g. [14], [15]. This prototype and distance-based classifier was previously studied for detecting cassava diseases on the basis of relatively few features directly derived from camera images [16], [17] and [6]. We apply a similar methodology here to show the comparative advantage of the different featureset.

III. THE GMLVQ MACHINE LEARNING FRAMEWORK

Here we briefly introduce the machine learning framework employed, GMLVQ.

Generally, we will consider datasets of the form:

$$\{\mathbf{x}^\mu, y^\mu\}_{\mu=1}^P \quad (1)$$

where $\mathbf{x}^\mu \in \mathbb{R}^N$ are feature vectors and the labels $y^\mu \in \{1, 2, \dots, C\}$ specify their class membership.¹

These data are generally standardized by performing a z-score operation as shown in Eq. (2). This is computed by subtracting the sample mean ϑ_i from data point components x_i^μ and dividing by the corresponding standard deviation δ_i :

$$z_i = \frac{x_i^\mu - \vartheta_i}{\delta_i} \quad (2)$$

where $i \in \{1, 2, \dots, N\}$ and N is the dimension of the feature vectors.

Learning Vector Quantization (LVQ) is a family of prototype-based supervised classification algorithms first introduced in 1986 [18]. It has been applied in a variety of practical contexts; its key advantage being ease of interpretation of the trained model. Several modifications of the original LVQ algorithm have been proposed in the literature, aiming at faster convergence or better generalization behavior.

The LVQ system is defined by a set of M prototypes $W = \{\mathbf{w}^j, c(\mathbf{w}^j)\}_{j=1}^M$ with vectors $\mathbf{w}^j \in \mathbb{R}^N$ which carry labels $c(\mathbf{w}^j) \in \{1, 2, \dots, C\}$. The system can be set up with one or more prototype vectors per class. Prototype vectors are identified in the feature space and ideally serve as typical representatives of their classes.

A nearest prototype classifier (NPC) assigns a given feature vector $\mathbf{x} \in \mathbb{R}^N$ to the *closest* prototype with respect to some meaningful distance measure.

Most frequently, standard Euclidean distance $d(\mathbf{w}, \mathbf{x})$ is employed. The corresponding NPC assigns \mathbf{x} to the class $c(\mathbf{w}^L)$ of the closest prototype with $d^\Lambda(\mathbf{x}, \mathbf{w}^L) \leq d^\Lambda(\mathbf{x}, \mathbf{w}^j)$ for all j .

One important extension of the basic LVQ concept is relevance learning in which an adaptive distance d^Λ is used, where Λ denotes a set of adjustable parameters which are

adapted, together with the prototypes, in a data-driven training process. The output is a trained model and a vector Λ that denotes how relevant each feature is for the classification.

The GMLVQ algorithm proposed by Schneider *et al.* [14] is a further extension that employs a full matrix $\Lambda \in \mathbb{R}^{N \times N}$ of relevances that describes the importance of the individual features in the classification task. Here, the distance measure $d^\Lambda(\mathbf{x}, \mathbf{w})$ is defined as:

$$d^\Lambda(\mathbf{x}, \mathbf{w}) = (\mathbf{x} - \mathbf{w})^\top \Lambda (\mathbf{x} - \mathbf{w}) \quad (3)$$

where the parameterization $\Lambda = \Omega^\top \Omega$ guarantees that $d^\Lambda(\mathbf{x}, \mathbf{w}) \geq 0$ for unrestricted matrices $\Omega \in \mathbb{R}^{N \times N}$. In order to avoid numerical degeneracies, a normalization constraint of the following form is imposed:

$$\sum_{i=1}^N \Lambda_{ii} = \sum_{i,j=1}^N \Omega_{ij}^2 = 1.$$

In GMLVQ, the training process is guided by the optimization of a cost function of the form suggested in [19]:

$$E(W) = \sum_{\mu=1}^P \Phi \left(\frac{d_J^\Lambda(\mathbf{x}^\mu) - d_K^\Lambda(\mathbf{x}^\mu)}{d_J^\Lambda(\mathbf{x}^\mu) + d_K^\Lambda(\mathbf{x}^\mu)} \right). \quad (4)$$

where the sum is over all examples in the data set, d_J^Λ denotes the distance of \mathbf{x}^μ from the closest correct prototype with $c(\mathbf{w}^J) = y^\mu$ and d_K^Λ is the distance from the closest incorrect prototype ($c(\mathbf{w}^J) \neq y^\mu$). The modulation function Φ is frequently chosen to be a sigmoidal function. Here, we consider a simple function $\Phi(x) = x$. Training constitutes the minimization of $E(W)$ with respect to the model parameters, i.e. the prototype W and the relevance matrix Λ .

The learning algorithm defined in [19] uses stochastic gradient descent. For the experiments presented in this work, we use the publicly available LVQ toolbox [20], which implements the batch gradient minimization of the cost function, Eq. (4), with adaptive step size control [20], [21]. If not specified otherwise, we use default parameters as suggested in [20].

Training yields the GMLVQ classifier in terms of the prototype vectors and the relevance matrix Λ . Its diagonal elements Λ_{ii} can be interpreted as the relevance of the corresponding feature dimensions for the classification [14], [22].

A. DIMENSIONALITY REDUCTION

Spectral data of the type considered here are nominally high-dimensional. As a consequence, the naive application of machine learning techniques will result in classifiers with a very large number of adjustable parameters, which causes problems ranging from computationally expensive training to a potentially increased risk of over-fitting.

The former is disadvantageous for efficient deployment of the model, for instance in mobile systems. The latter point could result in inferior generalization performance.

It is important to realize that spectra, like other functional data, comprise highly correlated features, as such the intensities of neighboring wavelengths can be expected to be very similar in a more or less smooth spectrum.

¹Throughout the following we denote high-dim. vectors by boldface letters, e.g. \mathbf{x} , while low-dim. projections are denoted as, for instance, \bar{c} or \bar{y}

We consider three different approaches for the dimensionality reduction of the data in order to circumvent the above mentioned problems.

1) FUNCTIONAL APPROXIMATION BY CHEBYSHEV POLYNOMIALS

The functional nature of the data can be exploited systematically by using appropriate representations. For instance, polynomial approximations have been used on spectral data [3], [4].

In particular, Chebyshev polynomials of the first kind [23] have been employed for a set of basis functions and showed good classification performance in several applications.

We interpret original features x_i as discretized observations of an underlying continuous spectrum, represented by a function $f(v)$ with $v \in \mathbb{R}$, i.e.

$$x_i = f(v_i), \quad i = 1, 2, \dots, N. \quad (5)$$

Given a suitable set of basis function g_k it is possible to expand f as

$$f(v) = \sum_{k=0}^{\infty} c_k g_k(v) \quad \text{with coefficients } c_k \in \mathbb{R}. \quad (6)$$

Restricting the maximum number of basis functions to a finite number n , Eq. (6) yields an approximation $\hat{f}(v)$ of the original spectrum.

The computation of the coefficients $c_k^\mu, k = 1, 2, \dots, n$ for a given observation \mathbf{x}^μ can be formulated as an optimization problem, achieving best approximation quality (e.g. mean square error) for a given number of basis functions. For Chebyshev polynomials of the first kind [23], the computation of coefficients can be done in an effective manner employing a linear transformation:

$$\tilde{\mathbf{c}}^\mu = \mathbf{C}\mathbf{x}^\mu \quad \text{where } \mathbf{x}^\mu \in \mathbb{R}^N, \quad \mathbf{C} \in \mathbb{R}^{n \times N} \quad \text{and } \tilde{\mathbf{c}}^\mu \in \mathbb{R}^n. \quad (7)$$

In practice, setting $n \ll N$ yields an efficient dimensionality reduction which does not require prior knowledge of the data. Furthermore, this includes an implicit denoising of the spectral information by discarding higher order polynomials [4].

2) PRINCIPAL COMPONENT ANALYSIS

PCA is a widely used standard technique for correlation analysis and dimensional reduction, e.g. [24]. PCA yields a linear projection of the data onto the eigenvectors of its covariance matrix, ordered according to the observed variations in the data set. Consider a matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$, where M is the number of samples and N is the data dimension. PCA transforms \mathbf{X} into $\mathbf{Y} \in \mathbb{R}^{M \times N'}$ with, in general, $N' \leq N$.

Whereas both PCA and Chebyshev polynomials are applicable for our problem, it is important to note the following properties. In PCA, the linear transformation depends on the actual training dataset. The emerging transformation matrix is then applied to novel data. For the Chebyshev case, polynomial coefficients are determined individually for each data point and do not depend on other data.

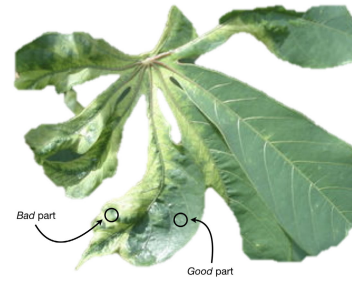


FIGURE 1. Depiction of asymptomatic(good) and symptomatic(bad) part of a leaf.

Both dimension reduction schemes can be interpreted as a linear transformation of the general form:

$$\tilde{\mathbf{y}} = \Psi \mathbf{x} \quad \text{where } \mathbf{x} \in \mathbb{R}^N, \quad \Psi \in \mathbb{R}^{M \times N} \quad \text{and } \tilde{\mathbf{y}} \in \mathbb{R}^M. \quad (8)$$

which projects the original data, potentially centered in the case of PCA, to an M -dimensional space with $M < N$.

A low-dimensional $\tilde{\mathbf{y}}$ corresponds to M expansion coefficients in the polynomial representation. When applying PCA, the components of $\tilde{\mathbf{y}}$ are the projections of \mathbf{x} on the M leading principal components.

We can compare the form of the distance measure in both spaces:

$$\begin{aligned} (\tilde{\mathbf{v}} - \tilde{\mathbf{y}})^\top \hat{\Lambda} (\tilde{\mathbf{v}} - \tilde{\mathbf{y}}) &= (\mathbf{w} - \mathbf{x})^\top \underbrace{\Psi^\top \hat{\Lambda} \Psi}_{\Lambda} (\mathbf{w} - \mathbf{x}) \\ &= (\mathbf{w} - \mathbf{x})^\top \Lambda (\mathbf{w} - \mathbf{x}). \end{aligned} \quad (9)$$

Here we denote a prototype and the relevance matrix in the low-dimensional space by $\tilde{\mathbf{v}} \in \mathbb{R}^M$ and $\hat{\Lambda} \in \mathbb{R}^{M \times M}$, respectively. We observe that formally

$$\tilde{\mathbf{v}} = \Psi \mathbf{w} \quad \text{and} \quad \Lambda = \Psi^\top \hat{\Lambda} \Psi. \quad (10)$$

Hence, we can back-transform the relevance matrix $\hat{\Lambda}$ and although the training is performed in terms of Chebyshev coefficients or principal components, we can identify the relevance of the original features, i.e. in terms of wavelengths or ranges thereof.

3) PEAK SELECTION

Peak selection is a method of feature selection in which we select a set of wavelengths with the highest peaks from the relevance profile obtained from running GMLVQ on the original data spectrum. It has a very simple intuition; the wavelengths with higher peaks represent areas in the spectrum where the sensor had a strong response to the item being measured. Harvesting wavelengths where there is a high response for the different classes provides an intuitive way of selecting features that may be relevant. This technique has commonly been used in many signal-processing applications, e.g. [25]. Like the two methods (PCA and Chebyshev polynomials) mentioned above, a subset of the original dimensions is selected by constructing new dimensions.

Given our feature matrix Λ , we put an intensity threshold on $\Lambda_{;i}$ to eliminate low-ranked features and select out wavelengths with a response above the threshold. We employ the

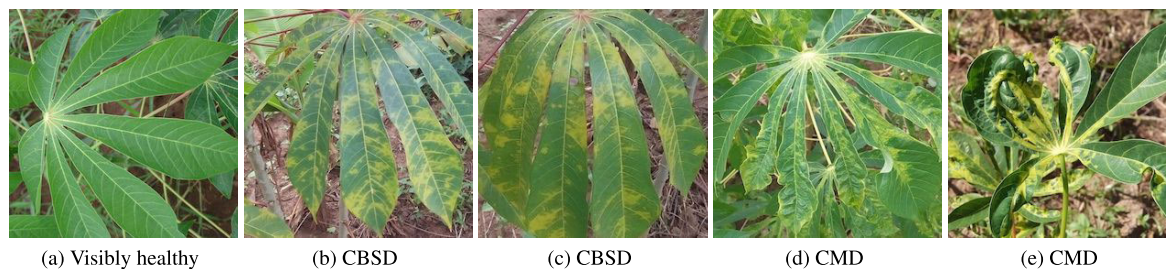


FIGURE 2. Example images of leaves of cassava manifesting the different diseases.

convenient function *findpeaks* defined in MATLAB(R2016a) which works by finding local peaks or valleys (local extrema) in a noisy vector using a user-defined magnitude threshold to determine if each peak is significantly larger (or smaller) than the data around it.

IV. EXPERIMENTS

Here we discuss the experiments carried out with spectra collected from leaves of cassava plants and how the methods described were applied. First we describe collection of the spectral data and the pre-processing applied. Next we present the machine learning techniques and discuss how they are combined with polynomial expansions and PCA.

A. EXPERIMENT DESIGN AND DATA COLLECTION

Our goal was to collect representative spectral data from the leaves of cassava plants under two conditions: when plants are healthy and when they are infected by the two different diseases CBSD and CMD with visibly symptomatic leaves. In Fig. 2, we provide some example images of leaves from the two diseases. The manifestation of disease on the leaf is determined to large extent by the variety of cassava and the severity of disease. In future work one goal is to relate the spectra extracted to the severity of disease. For this current work, however, we only looked at the binary case, disease vs (visibly) healthy for the two diseases.

These data were acquired using a CI-710 miniature leaf spectrometer [26]. The device is USB powered from a device (e.g. tablet or laptop) that makes the setup mobile and able to collect data in the field. To collect data, the device is clamped onto a leaf of a particular plant and the profile of the amount of light absorbed or reflected is captured as a spectrogram on the device for each position of the clamped leaf.

Several ambient factors influence the intensity and shape of the spectra, illumination being particularly important. For this reason, we collected data directly infields under similar lighting conditions.

We used the reflectance mode of operation of the spectrometer based on previous experiments where reflectance and absorption modes of operation gave the same performance for cassava leaves.

We collected data for plants aged 6 to 9 months from several cassava varieties including *Nase 3*, *Nase 4*, *Nase 14*, *Nase 19*, *Alado Alado*, *Magana*, *Orera* and *NAROCass 2* [27].

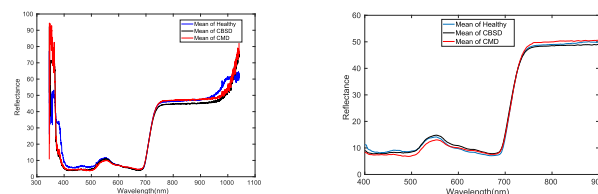


FIGURE 3. Illustration for class-conditional means of Cassava spectral data not individual spectra. The left panel displays raw, full signal, the right panel shows the corresponding pre-processed spectra.

For each variety, three plants were considered; and for each plant, three leaves were considered. For each leaf, two spectral readings were taken on each leaf lobe: one on the *best* part (least affected/non-symptomatic) and the *worst* part (most affected/symptomatic) (Fig. 1). Because the spectrometer takes readings on a small area of the plant about 2cm in diameter, readings for every leaf lobe were recorded in order to achieve a representative and reliable sampling. Note that this was considered during validation, such that we never trained and tested on data from the same plant. In total, 1656 data points were collected for evenly distributed classes: healthy, CMD and CBSD.

B. DATA PRE-PROCESSING

A typical spectrogram for each of the three classes is shown in Fig. 3. The intensities corresponding to the smallest and largest wavelengths are affected by significant noise. By truncating the spectrogram, we selected a wavelength range of 400 - 900 nm for subsequent analysis. This truncation provided a range of 500 nm, corresponding to 2500 equally spaced feature dimensions, which was still quite high. The spectrogram had many perturbations from small noise added to each wavelength. Consequently, the next pre-processing step aimed at smoothing the data over a small window of wavelengths. We compared two filtering techniques: median [28], [29] and average [30]. For both, we used a window size of 15 nm. Our experiments showed that average filtering yielded better classification results for this window size. As a consequence, average filtering was applied on all the data. An example of the final pre-processed spectrogram is shown in Fig. 3.

The dimensionality reduction could also be interpreted as optional pre-processing steps. However, because they are closely linked with training of the model we present them separately.

C. TRAINING AND VALIDATION

The data collection involved picking more than one sample from a particular plant, therefore it was important to choose a validation strategy that matched this condition in order to avoid training and testing on data from the same plant. We kept track of the class label (healthy, CBSD and CMD) as well as the unique plant labels (also called groups). During training, partitioning was based on plant groups and the validation scheme was Shuffle-Group(s)-Out cross-validation.

We employed the standard Scikit-learn implementation of this cross-validation scheme for the algorithms that were implemented using Scikit-learn [31]. In a similar way, this validation strategy was implemented for LVQ in MATLAB(R2016a) for the open source GMLVQ toolbox [20] that we employed for the GMLVQ algorithm. For all the models we train, we carry out a 10-fold crossvalidation and average the performance over the folds. We employ parameter $K = 15$ for the KNN algorithm, $C = 1$ for the linear SVC and 200 estimators for the Extra trees algorithm. For the GMLVQ algorithm we employ standard parameters used in the GMLVQ tool box, which is available online [20].

V. RESULTS

In this section we present results of training the LVQ family of methods and a standard algorithm (SVC) on the full spectral dataset and on the reduced dataset with different kinds of feature reduction: PCA, Chebyshev and peak methods. As a baseline we also applied a Convolutional Neural Network (CNN) using 1-D convolutional filters given the nature of the data. CNNs are models that have been shown to have superior performance on many tasks from computer vision, to natural language processing. One of the complexities of implementing CNNs is the choice of the architecture required for a particular problem.

In our case, for comparison with other base algorithms we employed a convolutional neural net (CNN) with an architecture based on the principles of popular models for image classification, but adapted to be suitable for 1D inputs. We use repeated convolution and ReLU blocks, with a max-pooling operation at the end of each block. A final fully connected softmax layer is used to perform the classification. This architecture is analogous to the VGG-16 architecture for 2D (image) inputs [32], using the same principle that the initial layers are intended to capture 'local' patterns within the input, and the successive convolution and max-pooling blocks successively downsample the input so that more global patterns can also be captured. Note, however, that because in our case the model operates on 1D spectral data, there would be no way to utilise existing CNN models such as VGG-16 for 2D data directly as starting points for training; we therefore trained this model from scratch starting from a random initialisation.

A. FULL SPECTRAL DATA

Our goal is two-fold: (1) to develop an algorithm that can perform well on spectral leaf data and (2) to engineer the algo-

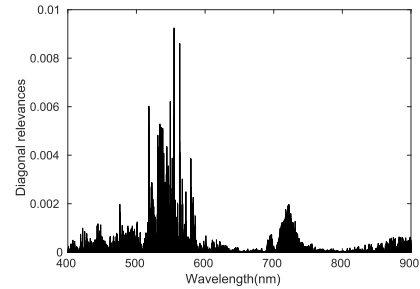


FIGURE 4. Feature relevance as quantified by diagonal elements of Λ , cf. Eq. (3), for original spectra as feature vectors.

rithm with a reduced featureset to a comparable performance. The first goal builds a baseline for building models for disease detection on cassava plants that are non-symptomatic and the second provides the base for design and implementation of low-cost devices that can use the narrow bands discovered as relevant in this study. The spectrometer we used costs in the order of thousands of dollars, and we aim to build one costing tens of dollars.

To investigate model performance with the full spectral data (goal 1) we pre-processed the data by truncating the signal at the extreme ends of the spectrogram and using the band 400 - 900 nm, as earlier described. We trained six algorithms, three from the family of LVQ, two from Scikit-learn and a CNN. The LVQ algorithms included Generalized Learning Vector Quantization (GLVQ) that does not train for relevances of the feature vector, Generalized Relevance Learning Vector Quantization (GRLVQ) that is similar to GLVQ but that trains a vector λ which represents the relevance of the features and GMLVQ which uses a matrix of relevances as described in an earlier section. The other two algorithms were a Linear Support Vector Classifier (SVC) and CNN.

Table 1 shows the overall cross-validation accuracies for a multiclass classification problem for the five different algorithms with the full dataset. We obtained good performance for the GMLVQ, SVC and CNN algorithms, with the SVC algorithm showing the best performance on this dataset. One important consideration to note in Table 1 is that feature reduction techniques produce reduced featuresets that are not immediately amenable to calculating convolutions as is the case for CNNs, and performance thus degrades.

The profile of the wavelengths most relevant for the classification (Fig. 4) was derived from the diagonal matrix of Λ from the GMLVQ algorithm. Results in Table 1 indicate that this provides some level of advantage over the SVC algorithm which performed best with the original dataset.

Projections of the feature vectors onto the leading eigenvectors of Λ allows depicting the spatial location of the training data points in relation to the prototypes per class see (Fig. 5). This 2-D representation shows good placement of the prototypes in the space of the training data.

B. REDUCED FEATURE SPACE

For the reduced features we tried the three methods described in the previous sections: PCA, Chebyshev and a method based

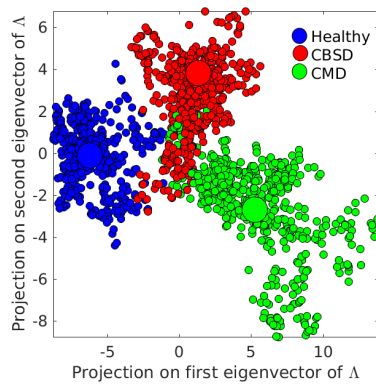


FIGURE 5. Visualization of the dataset depicting the three major classes in the dataset plotted as projections of feature vectors (original spectra) on the two leading eigenvectors of GMLVQ relevance matrix.

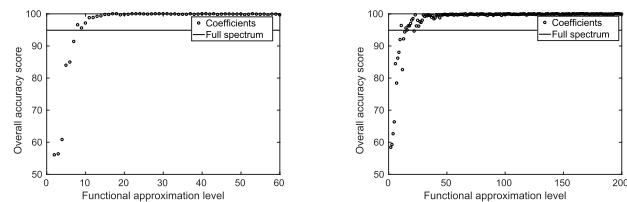


FIGURE 6. Performance of classifiers based on N Principal Component (left) and n coefficients in the polynomial representation (right panel).

TABLE 1. Overall cross-validation accuracy score for a multiclass classification problem (Healthy Vs CBSD Vs CMD).

Classifier	Original space	PCA	Chebyshev	Peaks
GLVQ	0.605	0.998	0.719	0.702
GRLVQ	0.845	0.856	0.738	0.699
GMLVQ	0.949	1.000	0.996	0.952
Linear SVC	0.972	0.976	0.753	0.709
CNN	0.969	0.925	0.840	0.330

TABLE 2. Confusion matrix for GRLVQ with PCA.

	Healthy	CBSD	CMD
Healthy	95.1	2.3	2.6
CBSD	20.4	75.9	3.7
CMD	10.6	3.7	85.7

on truncating the peaks of the relevance profile produced from training the algorithms on the full spectral data.

The challenge with feature reduction is to reduce the features from the 2500 features to a suitable number N that can represent the full spectrogram and still perform relatively well on the dataset. To determine a suitable N , we ran experiments of a set of different N s and plotted the accuracy of the GMLVQ algorithm and compared these with the Chebyshev method and PCA methods for different values of N (Fig. 6).

The results of this experiment were an optimal N for PCA of 30 coefficients and Chebyshev of close to 200 components. We then trained our set of five algorithms on these reduced featuresets of the data. Table 1 shows results of the PCA and Chebyshev methods of feature reduction. Furthermore, Table 2, 3 and 4 are confusion matrices for GRLVQ, Linear SVC and CNN with PCA method respectively.

TABLE 3. Confusion matrix for Linear SVC with PCA.

	Healthy	CBSD	CMD
Healthy	100	0	0
CBSD	0	100	0
CMD	4.8	2.2	93

TABLE 4. Confusion matrix for CNN with PCA.

	Healthy	CBSD	CMD
Healthy	88	4	8
CBSD	3.4	94.6	2
CMD	3.2	1.8	95

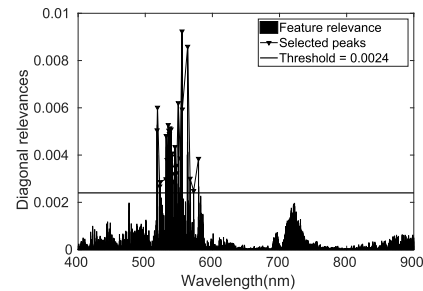


FIGURE 7. Selection of features with diagonal relevances (GMLVQ) above a threshold.

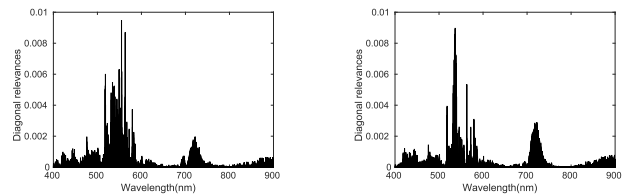


FIGURE 8. Diagonal relevances of GMLVQ in original feature space as reconstructed after performing the training in terms of 30 (left panel) and 5 (right panel) principal components.

The peak method of selecting a set of wavelengths from the relevance profile of the full spectral data is a fairly intuitive way of reducing the set of features. For this method we used a threshold on the relevance profile to select the 30 most relevant features from the profile to train our algorithms. Results of the peak method are shown in Fig. 7 and Table 1. We obtained the best performance with these reduced featuresets for the GMLVQ algorithm with the PCA feature reduction method performing the best overall for all five algorithms.

A key idea of our experimentation was to determine how well we could derive the relevance profile from the original spectral data profile using the reduced featuresets. By Eq. 10, we reconstructed the relevance profile from the reduced featuresets for PCA method using five coefficients (very few) and 30 coefficients (optimal performance) according to Fig. 8.

The results of this back transformation are shown in Fig. 8. The resultant shapes of the relevance profiles for all wavelengths tended to follow the shape of the relevance profile of the original spectral data with the bi-modal distribution of relevance. In a general way this justifies our choice of N for the PCA method.

VI. DISCUSSION

We presented a method of diagnosing disease from plant leaves in the field using spectral data which is different from previous methods that are based solely on image data. Our study provides the first step in our search for a method that can be used to diagnose disease in leaves before they are visibly symptomatic. Although the use of spectrometry for classification is not a new idea, its application for this particular problem in combination with the machine learning based analysis is novel. We obtained a competitive level of classification accuracy in the difficult problem of discriminating between a healthy cassava plant and those affected by CMD and CBSD. The results showed improved classification accuracy when using a reduced featureset, particularly when PCA was used for dimension reduction. The spectral data were very noisy and it is likely that the feature reduction removes most noise from the signal. The work also investigated different techniques for disease classification. As expected, performance of the CNN degrades with reduction in the number of features since deep neural networks particularly excel with a large amount of data and many features. We also observed an interesting change in top performance with SVC performing best for the full spectrum data, but the roles changed and GMLVQ performed best consistently with all the other reduced representations of the featureset. One explanation for this is that with the reduced featureset, GMLVQ is able to obtain a more reliable relevance profile, which in turn enhances performance of the algorithm. However, with SVC, which does not calibrate for relevance profiles of the features, the decision boundaries are thrown off with the relatively fewer data points. In addition, we have performed experiments using Linear Discriminant Analysis (LDA) [33] as a classifier, yielding performances comparable to the combination of PCA and GMLVQ. While GMLVQ and LDA appear similar on a conceptual level, the LVQ approach offers greater flexibility in terms of extensions with respect to the number of prototypes per class and the use of local relevance matrices [22], [34]. Furthermore, LDA is restricted to employing $(C - 1)$ -dim. internal representations of the data in C -class problems, while in GMLVQ the rank of the adaptive relevance matrix emerges from the training process and reflects the complexity of the classification problem.

In future work we intend to look at spectral data collected from a more controlled environment where plants can be inoculated and data collected before they are visibly symptomatic. The experiments of this study started in our work in [35] and were guided by chemical analysis from the biochemist. The findings of the current work contributed to this study by identifying the most spectral band for the disease classification. Results indicate that the presence of the disease can be detected from leaf spectra six weeks earlier before the appearance of visual symptoms. Another aspect of our future work is to build a low-cost smartphone add-on spectrometer. In [36], we present our initial proof of concept towards this area. We built a low-cost spectral artifact (less than 5 USD) instead of using an off-the-shelf and expensive spectrometer

(approximately 1000 USD). One possible way to improve the artifact would be to use specific diodes that are sensitive at those particular relevant feature wavelengths and building a light emitting, absorption and measurement system around that. The success of this technology will provide a cheap and user friendly diagnostic tool to be used by smallholder farmers in developing countries.

ACKNOWLEDGMENT

The authors would like to thank the Directors of the Uganda National Crop Resources Research Institute (NaCRRI), for granting us permission to access cassava fields. They extend their thanks to Dr. Ephraim Nuwamanya of NaCRRI for the support he showed us in the data collection process. They also thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. The data used in this work is available at: https://github.com/godliver/IEEE_ACCESS_Cassava_Spectral_Data.

REFERENCES

- [1] B. Zeyimo, M. Eric, L. Boykin, M. Macharia, M. Nzola, T. Hangy, L. Diankenda, M. Godefroid, J. Harvey, J. Ndunguru, C. Kayuki, J. Pita, L. Munseki, and T. Kanana, "Attempts to identify cassava brown streak virus in western democratic republic of congo," *J. Agricult. Sci.*, vol. 11, no. 2, p. 31, Jan. 2019.
- [2] P. Schneider, F.-M. Schleif, T. Villmann, and M. Biehl, "Generalized matrix learning vector quantizer for the analysis of spectral data," in *Proc. Eur. Symp. Artif. Neural Netw. (ESANN)*, M. Verleysen, Ed. Bruges, Belgium: D-Side Publishing, 2008, pp. 451–456.
- [3] F. Melchert, U. Seiffert, and M. Biehl, "Functional representation of prototypes in LVQ and relevance learning," in *Proc. 11th Int. Workshop Adv. Self-Organizing Maps Learn. Vector Quantization (WSOM)*, Houston, TX, USA, Jan. 2016, pp. 317–327.
- [4] F. Melchert, U. Seiffert, and M. Biehl, "Functional approximation for the classification of smooth time series," in *Proc. Workshop New Challenges Neural Comput.*, 2016, pp. 24–31.
- [5] J. R. Aduwo, E. Mwebaze, and J. A. Quinn, "Automated vision-based diagnosis of cassava mosaic disease," in *Proc. Workshop Data Mining Agricult. (ICDM)*, 2010, pp. 114–122.
- [6] E. Mwebaze and M. Biehl, "Prototype-based classification for image analysis and its application to crop disease diagnosis," in *Advances in Self-Organizing Maps and Learning Vector Quantization*, vol. 428. Cham, Switzerland: Springer, 2016, pp. 329–339.
- [7] J. J. Belasque, M. C. G. Gasparoto, and L. G. Marcassa, "Detection of mechanical and disease stresses in citrus plants by fluorescence spectroscopy," *Appl. Opt.*, vol. 47, no. 11, pp. 1922–1926, 2008.
- [8] C. B. Wetterich, R. Kumar, S. Sankaran, J. Belasque Junior, R. Ehsani, and L. G. Marcassa, "A comparative study on application of computer vision and fluorescence imaging spectroscopy for detection of Huanglongbing citrus disease in the USA and Brazil," *J. Spectrosc.*, vol. 2013, pp. 1–6, Oct. 2013.
- [9] Y. Yang, "Early detection of rice blast (*Pyricularia*) at seedling stage in nipponbare rice variety using near-infrared hyper-spectral image," *Afr. J. Biotechnol.*, vol. 11, no. 26, pp. 6809–6817, Mar. 2012.
- [10] J. Feng, N.-F. Liao, M.-Y. Liang, B. Zhao, and Z.-F. Dai, "Multispectral imaging system for the plant diseases and insect pests diagnosis," *Guang Pu Xue Yu Guang Pu Fen Xi*, vol. 29, no. 4, pp. 1008–1012, 2009.
- [11] F. Tan, X. Ma, C. Wang, and T. Shang, "Data analysis of cold rice blast based on near infrared spectroscopy," in *Proc. 5th Comput. Technol. Agricult. (CCTA)*, Beijing, China, 2012, pp. 64–71.
- [12] N. Arens, A. Backhaus, S. Döll, S. Fischer, U. Seiffert, and H.-P. Mock, "Non-invasive presymptomatic detection of cercospora beticola infection and identification of early metabolic responses in sugar beet," *Frontiers Plant Sci.*, vol. 7, p. 1377, Sep. 2016.

- [13] G. Owomugisha, F. Melchert, E. Mwebaze, J. A. Quinn, and M. Biehl, "Machine learning for diagnosis of disease in plants using spectral data," in *Proc. 20th Int. Conf. Artif. Intell.*, 2018, pp. 9–15.
- [14] P. Schneider, M. Biehl, and B. Hammer, "Relevance matrices in LVQ," in *Proc. 15th Eur. Symp. Artif. Neural Netw.*, Bruges, Belgium, 2007, pp. 37–42.
- [15] P. Schneider, M. Biehl, and B. Hammer, "Adaptive relevance matrices in learning vector quantization," *Neural Comput.*, vol. 21, no. 12, pp. 3532–3561, Dec. 2009.
- [16] E. Mwebaze, P. Schneider, F.-M. Schleif, J. R. Aduwo, J. A. Quinn, S. Haase, T. Villmann, and M. Biehl, "Divergence-based classification in learning vector quantization," *Neurocomputing*, vol. 74, no. 9, pp. 1429–1435, Apr. 2011.
- [17] E. Mwebaze, G. Bearda, M. Biehl, and D. Zühlke, "Combining dissimilarity measures for prototype-based classification," in *Proc. 23rd Eur. Symp. Artif. Neural Netw. (ESANN)*, 2015, pp. 31–36.
- [18] T. Kohonen, "Learning vector quantization for pattern recognition," Helsinki University Technol., Espoo, Finland, Tech. Rep. TKKF-A601, 1986.
- [19] A. Sato and K. Yamada, "Generalized learning vector quantization," in *Proc. 8th Int. Conf. Neural Inf. Process. Syst.*, 1995, pp. 423–429.
- [20] M. Biehl, *A no-Nonsense GMLVQ Toolbox*. Groningen, The Netherlands: Univ. Groningen, 2017. [Online]. Available: <http://www.cs.rug.nl/~biehl/gmlvq>
- [21] G. Papari, K. Bunte, and M. Biehl, "Waypoint averaging and step size control in learning by gradient descent," Leipzig Univ., Leipzig, Germany, Tech. Rep. MLR-2011-06, 2011.
- [22] M. Biehl, B. Hammer, and T. Villmann, "Prototype-based models in machine learning," *Wiley Interdiscipl. Rev. Cognit. Sci.*, vol. 7, pp. 92–111, Jan. 2016.
- [23] T. Driscoll, N. Hale, and L. Trefethen, Eds., *Chebfun Guide*. Oxford, U.K.: Pafnuty Publications, 2014.
- [24] K. K. Vasan and B. Surendiran, "Dimensionality reduction using principal component analysis for network intrusion detection," *Perspect. Sci.*, vol. 8, pp. 510–512, Sep. 2016.
- [25] A. Liutkus, "Scale-space peak picking," Inria Nancy—Grand Est, Villers-lès-Nancy, France, Res. Rep., Jan. 2015. [Online]. Available: <https://hal.inria.fr/hal-01103123>
- [26] CID Bio-Science, Inc. (2010). *Ci-710 Miniature Leaf Spectrometer*. [Online]. Available: <http://www.cid-inc.com>
- [27] G. Nakabonge, C. Samukoya, and Y. Baguma, "Local varieties of cassava: Conservation, cultivation and use in Uganda," *Environ., Develop. Sustainability*, vol. 20, no. 6, pp. 2427–2445, Dec. 2018.
- [28] E. Arias-Castro and D. L. Donoho, "Does median filtering truly preserve edges better than linear filtering?" *Ann. Statist.*, vol. 37, no. 3, pp. 1172–1206, Jun. 2009.
- [29] A. B. Hamza, P. L. Luque-Escamilla, J. Martínez-Aroza, and R. Román-Roldán, "Removing noise and preserving details with relaxed median filters," *J. Math. Imag. Vis.*, vol. 11, no. 2, pp. 161–177, 1999.
- [30] S. W. Smith, "Moving average filters," in *The Scientist and Engineer's Guide to Digital Signal Processing Moving Average Filters*, vol. 15. San Diego, CA, USA: California Technical Publishing, 1999, pp. 277–284.
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [33] A. J. Izenman, "Linear discriminant analysis," in *Modern Multivariate Stat. Techniques: Regression, Classification, and Manifold Learning*. Springer, 2008, pp. 237–280, doi: [10.1007/978-0-387-78189-1_8](https://doi.org/10.1007/978-0-387-78189-1_8).
- [34] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl, "Limited rank matrix learning, discriminative dimension reduction and visualization," *Neural Netw.*, vol. 26, pp. 159–173, Feb. 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0893608011002632>
- [35] G. Owomugisha, E. Nuwamanya, J. A. Quinn, M. Biehl, and E. Mwebaze, "Early detection of plant diseases using spectral data," in *Proc. 3rd Int. Conf. Appl. Intell. Syst. (APPIS)*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–6.
- [36] G. Owomugisha, P. K. B. Mugagga, F. Melchert, E. Mwebaze, J. A. Quinn, and M. Biehl, "A low-cost 3-D printed smartphone add-on spectrometer for diagnosis of crop diseases in field," in *Proc. 3rd ACM SIG-CAS Conf. Comput. Sustain. Societies*, Jun. 2020, pp. 331–332, doi: [10.1145/3378393.3402252](https://doi.org/10.1145/3378393.3402252).



GODLIVER OWOMUGISHA received the bachelor's and master's degrees in computer science from Makerere University, in 2011 and 2015, respectively, and the Ph.D. degree in computer science from the University of Groningen, in 2020. She is currently a Senior Lecturer with the Department of Computer Engineering, Busitema University. Her research interests include machine learning and computational intelligence in relation to solving real world problems.



FRIEDRICH MELCHERT received the master's degree in electrical engineering and the master's in medical systems engineering from the University of Magdeburg, Magdeburg, Germany, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree with the University of Groningen, The Netherlands. He worked with the Biosystem Engineering Department, Fraunhofer Institute for Factory Operation and Automation IFF, Magdeburg. Besides his research in the field of functional data approximation and classification, he founded a machine learning focused start-up company, in late 2018.



ERNEST MWEBAZE received the bachelor's and Graduate Diploma degrees in electrical engineering and computer science from Makerere University, Uganda, in 2003 and 2005, respectively, and the Ph.D. degree in computer science from the University of Groningen, in 2014. His research interests include machine learning and computer vision. He is particularly interested in using computational methods to address developing world problems.



JOHN. A. QUINN received the B.A. degree in computer science from the University of Cambridge, in 2000, and the Ph.D. degree from the University of Edinburgh, in 2007. He coordinates the Machine Learning Group, Makerere University. His research interests include pattern recognition and computer vision particularly applied to developing world problems.



MICHAEL BIEHL received the Ph.D. degree in physics from the University of Gießen, Germany, in 1992, and the Habilitation degree in theoretical physics with the University of Würzburg, Germany, in 1996. He is currently a Professor of computer science with the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence, University of Groningen, The Netherlands. His research interests include theoretical investigation, development, and application of machine learning methods.

...