

Comprehensive, non-invasive, and quantitative monitoring of the health and nutrition state of crop plants by means of hyperspectral imaging and computational intelligence based analysis

Andreas Backhaus and Udo Seiffert

Fraunhofer Institute for Factory Operation and Automation IFF
Sandtorstr. 22, 39106 Magdeburg, Germany

Abstract Against the background of hyperspectral imaging this paper evaluates a number of different machine learning based classification methods in terms of their performance. All considered methods offer relevance profiles that additionally provide valuable information about the relevance of all acquired wavelengths to get the obtained classification. This relevance profile can be used to select appropriate wavelengths or wavelength bands to customize data acquisition and analysis tailored to the specific application at hand.

1 Introduction

Quantitative assessment of phenotypic properties of crop plants in relation to different genotypes, nutrition, stress tolerance, and fruit quality has become increasingly important in crop plant research, modern plant breeding, and particularly in precision agriculture / smart farming. The required assessment can generally be based either on morphological features, such as plant height, leaf shape, root structure etc., or on biomolecular/biochemical analyses. While the latter one is typically invasive and destroys the intact biological structure, morphological features are often not sufficient to unravel all relevant information at the required level of detail. Moreover, wet lab analyses typically assess only a more or less small number of samples and are not suitable to monitor crop plants in productive operation on the field or for large-scale (high-throughput) phenotyping of many genotypes in plant breeding. The results of wet lab analyses typically cannot be incorporated into on-line monitoring systems.

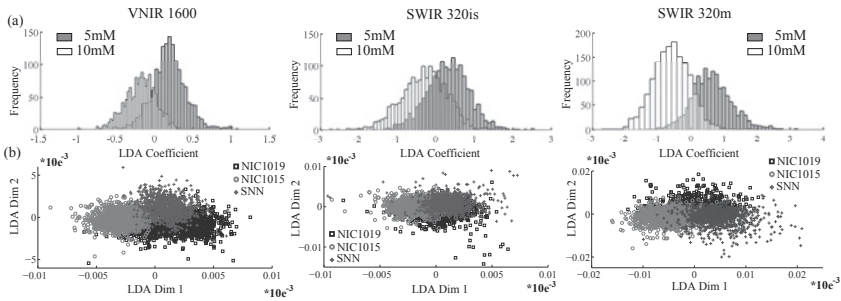


Figure 10.1: (a) Histogram of LDA coefficient for the two class problem of nutrition state; (b) Scatter plot of the LDA coefficients for the three class problem of genotype classification.

There are several non-invasive systems currently available on the market. Typical applications are ground-based or airborne data acquisition for precision agriculture / smart farming as well as automated greenhouses in crop plant research and plant breeding. Technically these systems are based on broadband or selective color imaging (ca. 400-800 nm) that is sometimes accompanied by a few selected spectral bands in IR (>800 nm) or UV (<400 nm) range. From the application's perspective these systems can only monitor for example the level of green color (chlorophyll) of the plants (leaves) as indicator of nitrogen nutrition (e.g. Yara N-Sensor), the water content, or some particular biochemical compounds. Monitoring of a comprehensive health and nutrition state of crop plants is currently not commercially available.

Hyperspectral imaging linked to subsequent computational intelligence based analysis has proven its suitability to unravel complex information in a number of different application areas, such as geology, defense, etc. The extension of this approach to crop plant research, plant breeding, agriculture, and food processing has started just quite recently. Here, the image acquisition ranges from airborne sensing mainly for agricultural applications down to single leaf analysis in the context of precision and high-throughput plant phenotyping. All these applications have in common, that particular relevant compounds of the plant need to be determined by means of hyperspectral signatures as complement or substitute to extensive biochemical analyses.

Often the direct relationship between spectral information and bio-

chemical target value or material category is not known in a closed mathematical form. In this case a machine learning approach is used to acquire an analysis model from reference data, a paradigm often referred to as 'soft-sensor'. Sensor data analysis becomes a pattern recognition task. Regarding pattern recognition and data mining in the acquired spectral data, computational intelligence based methods are still providing powerful tools to cope with this kind of high-dimensional and complex data.

In this paper we assess the ability of machine learning methods to robustly classify nutrition states and genotypic identity from input of three different hyperspectral cameras covering the VNIR/SWIR range.

2 Data acquisition

The data set originated from three genetically different tobacco varieties, namely *Nicotiana tabacum* L., cv. *SamsunNN* (SNN), *Nicotiana tabacum* L., cv. *undulata* (NIC1015) and *Nicotiana tabacum* L., cv. *undulata* (NIC1019). Plants were cultivated on quartz sand and maintained under controlled environmental conditions in a greenhouse. The plants were daily irrigated with a complete nutrient solution containing either 5 mM or 10 mM NH_4NO_3 (ammonium nitrate). Twelve weeks old plants were used in the experimental setup. Hyperspectral images were acquired from whole leaf blade (lamina). Leaves of different age were taken into account. Four leaves per plant were recorded. Leaves are numbered starting from the top downwards along the stem, therefore leaf age increases with the leaf number.

Images were acquired covering the complete VNIR/SWIR-band with three sensors simultaneously (Norsk Elektro Optikk A/S, VNIR 1600, SWIR 320i, SWIR 320m, 0.4-1.0, 0.9-1.7, 1.3-2.5 μm , respectively). The acquired images were read from 16 bit raw data using the vendors software. Blank images of the image background were also taken into account for inhomogeneous pixel sensitivities, which were found negligible. Reflectance calibration values were obtained from a standard optical PTFE (polytetrafluoroethylene) pad.

In a first k-means clustering [1], background pixels and pixels of non-leaf objects were removed. For display purposes and as input to the Support Vector Machine classifier in the classification stage, a Linear Discriminant Analysis (LDA) was performed. Figure 10.1a shows the data distribution in a one-dimensional LDA space per camera for

the two-class problem of nutrition classification (5 mM and 10 mM NH_4NO_3). In Figure 10.2a, the mean spectrum (middle line) is depicted per camera. The seam around the mean spectra depicts the extend of the standard deviation per spectral band. Spectra were normalized to unit length. In both figures data originates from mutant NIC1019 and all leaf ages. Figure 10.1b shows the scatter plot in a two-dimensional LDA space per camera for the three-class problem of genotype classification while Figure 10.3a shows the mean spectra with standard deviation. Data originates from all leaf ages.

3 Theory

For machine learning, four different classification models are considered, a Radial Basis Function (rRBF) Network with Relevance Learning [2, 3], Generalized Relevance Learning Vector Quantization (GRLVQ) [4], Supervised Relevance Neural Gas (SRNG) [5] as well as a Support Vector Machine [6]. In general, rRBF, SRNG, and GRLVQ Networks are similar in terms that they process the input data in a layer of prototypical data points. While the rRBF generates activation due to the similarity with prototypes which is accumulated in a second layer for the network output, the GRLVQ and SRNG directly assign classes to prototypical data points. Prototypes usually represent central positions in a data cloud. In contrast, the Support Vector Machine stores support vectors, e.g. representative data points at the margin between data clouds. The used Support Vector Machine implementation of the ν -SVM variant [7] from the freely available libSVM package¹ takes up a variable amount of support vectors.

In order to compute the distance of spectral data point \mathbf{v} and a prototype \mathbf{w} in the rRBF, SRNG and GRLVQ, we used the weighted Euclidean distance metric

$$d(\mathbf{v}, \mathbf{w}_r, \lambda) = \sum_i \lambda_i (v_i - w_{ir})^2, \quad (10.1)$$

where λ_i is the relevance factor per spectral band which is adapted during the learning process to form the relevance profile. The rRBF, SRNG, and GRLVQ learning approach is essentially an energy minimization problem. In the standard learning scheme, stochastic gradient descent

¹ www.csie.ntu.edu.tw/~cjlin/libsvm/

with step-sizes manually set for different parameters are used. In order to avoid a manually chosen step-size, we used the non-linear conjugate gradient approach with automatic step size from the optimization toolbox 'minFunc'² available for Matlab. For this purpose we had to provide the objective/energy function along with the first derivatives according to the optimization parameters. The derivatives are accumulated for all data points (batch learning).

3.1 Radial basis function network with relevance

For the rRBF [2, 3] the objective function is the accumulated quadratic error of the network output y and target value t across network outputs and data samples \mathbf{v}^j .

$$E(\mathbf{V}, \mathbf{W}, \lambda) = \frac{1}{2} \sum_j \sum_k \left\{ y_k(\mathbf{v}^j) - \mathbf{t}_k^j \right\}^2,$$

with $y_k(\mathbf{v}) = \sum_r u_{rk} \phi(d(\mathbf{v}, \mathbf{w}_r, \lambda))$ and $\phi(x) = \exp\left(-\frac{x}{2\sigma^2}\right)$. The partial derivatives are as follows

$$\begin{aligned} \frac{\partial E}{\partial w_{ir}} &= \sum_j \sum_k \left\{ y_k(\mathbf{v}^j) - \mathbf{t}_k^j \right\} \\ &\quad u_{rk} \phi\left(d(\mathbf{v}^j, \mathbf{w}_r, \lambda)\right) \frac{(x_i^j - w_{ir})}{\sigma_r^2} \\ \frac{\partial E}{\partial \sigma_r} &= \sum_j \sum_k \left\{ y_k(\mathbf{v}^j) - \mathbf{t}_k^j \right\} \\ &\quad u_{rk} \phi\left(d(\mathbf{v}^j, \mathbf{w}_r, \lambda)\right) \frac{\sum_i \lambda_i (v_i - w_{ir})^2}{\sigma_r} \\ \frac{\partial E}{\partial \lambda_i} &= - \sum_j \sum_k \left\{ y_k(\mathbf{v}^j) - \mathbf{t}_k^j \right\} \\ &\quad \sum_r u_{rk} \phi\left(d(\mathbf{v}^j, \mathbf{w}_r, \lambda)\right) \frac{(v_i^j - w_{ir})^2}{2\sigma_r^2}. \end{aligned}$$

² <http://www.di.ens.fr/~mschmidt/Software/minFunc.html>

The output weights u_{rk} are yielded by direct update $\mathbf{U}^T = \Phi^\dagger \mathbf{T}$ where \dagger denotes the pseudo inverse [8]. For the classification task a 1-out-of-N coding scheme for the target vector was used.

3.2 Generalized relevance learning vector quantization

For the GRLVQ the objective function is the accumulated difference in shortest distance of a data point to a prototype representing its class d_r^+ and a prototype representing any other class d_r^- [4]:

$$E(\mathbf{V}, \mathbf{W}, \lambda) = \sum_{v \in V} \frac{d_r^+ - d_r^-}{d_r^+ + d_r^-}.$$

The partial derivatives are as follows

$$\begin{aligned} \frac{\partial E}{\partial w_{ir}^+} &= -\frac{2 \cdot d_r^-}{(d_r^+ + d_r^-)^2} 2(v_i - w_{ir}^+) & \frac{\partial E}{\partial w_{ir}^-} &= \frac{2 \cdot d_r^+}{(d_r^+ + d_r^-)^2} 2(v_i - w_{ir}^-) \\ \frac{\partial E}{\partial \lambda_i} &= \frac{2 \cdot d_r^-}{(d_r^+ + d_r^-)^2} (v_i - w_{ir}^+)^2 - \frac{2 \cdot d_r^+}{(d_r^+ + d_r^-)^2} (v_i - w_{ir}^-)^2. \end{aligned}$$

All partial derivatives not belonging to the winning prototype of same class w_r^+ and any other class w_r^- are set to zero.

3.3 Supervised neural gas

The SRNG [5] is a supervised version of the well known neural gas clustering algorithm [9]. Like in the GRLVQ a number of prototype vectors with pre-assigned class labels are distributed in the input space while minimizing the energy function

$$E(\mathbf{V}, \mathbf{W}, \lambda) = \sum_{\mathbf{v} \in V} \sum_{\mathbf{w}_r \in \mathbf{W}_c} h_\gamma(r, \mathbf{v}, \mathbf{W}_c) \frac{d_r^+ - d_r^-}{d_r^+ + d_r^-},$$

where $h_\gamma(r, \mathbf{v}, \mathbf{W}_c)$ denotes the degree of neighborhood cooperation among all prototypes representing the respective spectral vector class.

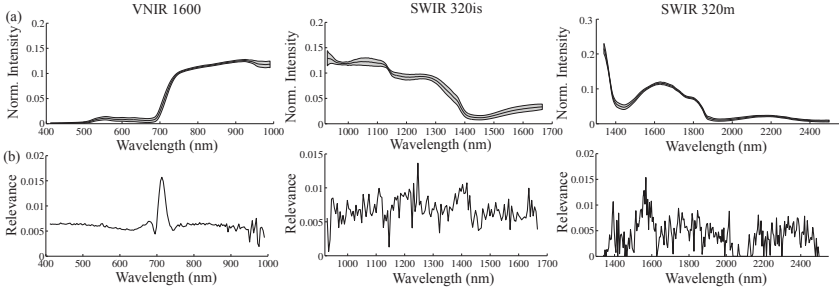


Figure 10.2: Nutrition state classification: (a) Mean spectral profile; (b) rRBF relevance profile for NIC1019; In VNIR 1600, a clear dominant frequency range around the intensity flank can be observed. This is known as “red edge” that corresponds to photosynthesis activity in plants as result of varying nitrogen supply.

The partial derivatives are as follows

$$\begin{aligned}
 \frac{\partial E}{\partial w_{ir}^+} &= -\frac{2 \cdot d_r^-}{(d_r^+ + d_r^-)^2} h_\gamma(r, \mathbf{v}, \mathbf{W}_c) 2 (v_i - w_{ir}^+) \\
 \frac{\partial E}{\partial w_{ir}^-} &= \sum_{\mathbf{w}_r \in \mathbf{W}_c} \frac{2 \cdot d_r^+}{(d_r^+ + d_r^-)^2} h_\gamma(r, \mathbf{v}, \mathbf{W}_c) 2 (v_i - w_{ir}^-) \\
 \frac{\partial E}{\partial \lambda_i} &= \sum_{\mathbf{w}_r \in \mathbf{W}_c} h_\gamma(r, \mathbf{v}, \mathbf{W}_c) \left(\frac{2 \cdot d_r^-}{(d_r^+ + d_r^-)^2} (v_i - w_{ir}^+)^2 \right. \\
 &\quad \left. - \frac{2 \cdot d_r^+}{(d_r^+ + d_r^-)^2} (v_i - w_{ir}^-)^2 \right).
 \end{aligned}$$

4 Spectral band reduction

All three described machine learning modeling approaches optimize their respective energy function by adaptation of per-spectral band weighting. This weight vector is used to order the spectral bands according to their relevance. In order to check model performance on reduced spectral information, an rRBF network model was trained on the largest weighted bands as input and continuously added bands with

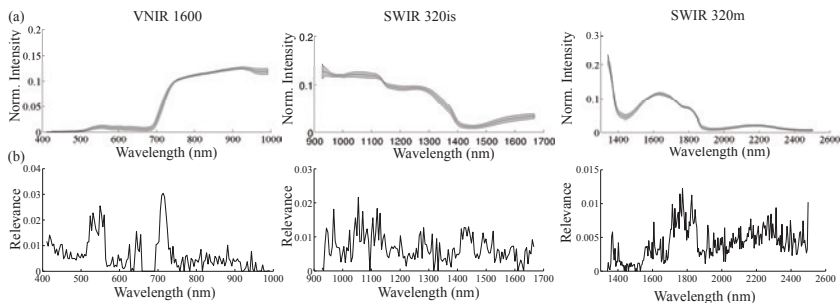


Figure 10.3: Relevance profiles for genotype classification: (a) Exemplary mean spectral profile; (b) rRBF relevance profile.

the highest input dimensionality of 70. The rRBF networks were trained with the same setup and parameters as above. The data is taken from the VNIR range for leaf 1. Test accuracy on unseen data for nutrition and genotype classification is evaluated for each n -dimensional input space. As control, a matching number of spectral bands are chosen at random (uniform distribution).

5 Results

For the prediction of nutrition states and genotypic identity, data sets were obtained from all three cameras for all leaves, for leaf one (youngest), and leaf three (second oldest) separately. Furthermore, for nutrition prediction, data was separated for each mutant condition to remove variation from genotype identity, which is a reasonable assumption since crop plant species should be known in advance in a precision farming environment. For the genotype classification, all nutrition conditions per genotype were included, since these could vary on the field. For each class 3,000 randomly selected spectral samples were chosen. Data was partitioned into 50% training and 50% test data. A 5-fold cross validation with randomized assignment of data samples to test and training data under the given partition and randomized initialization in GRLVQ, SRNG, and rRBF was performed. Classifier accuracy was averaged and standard deviation was calculated.

Table 10.1 on the left shows test data accuracy for classification of nu-

Class. Method	Nutrition			Genotype		
	VNIR 1600	SWIR 320i	SWIR 320m	VNIR 1600	SWIR 320i	SWIR 320m
	Leaf 1-4					
SVM+LDA	79.2 (0.1)	70.8 (0.9)	80.3 (0.2)	79.2 (0.1)	70.8 (0.9)	80.3 (0.2)
SVM	59.0 (0.5)	53.6 (1.7)	55.2 (1.6)	59.0 (0.5)	53.6 (1.7)	55.2 (1.6)
SRNG	56.0 (5.1)	55.5 (3.8)	54.5 (2.0)	56.0 (5.1)	55.5 (3.8)	54.5 (2.0)
GRLVQ	58.4 (0.6)	55.4 (4.0)	53.4 (2.9)	58.4 (0.6)	55.4 (4.0)	53.4 (2.9)
rRBF	83.7 (0.6)	75.2 (0.9)	81.5 (0.7)	83.7 (0.6)	75.2 (0.9)	81.5 (0.7)
	Leaf 1 - youngest					
SVM+LDA	99.3 (0.1)	87.9 (0.7)	85.8 (0.3)	99.3 (0.1)	87.9 (0.7)	85.8 (0.3)
SVM	97.3 (0.1)	77.9 (0.3)	69.2 (1.5)	97.3 (0.1)	77.9 (0.3)	69.2 (1.5)
SRNG	92.6 (1.9)	74.3 (2.7)	57.6 (4.5)	92.6 (1.9)	74.3 (2.7)	57.6 (4.5)
GRLVQ	92.3 (0.3)	61.5 (8.9)	60.3 (0.7)	92.3 (0.3)	61.5 (8.9)	60.3 (0.7)
rRBF	99.5 (0.1)	93.9 (0.6)	91.1 (0.8)	99.5 (0.1)	93.9 (0.6)	91.1 (0.8)
	Leaf 3 - second oldest					
SVM+LDA	78.9 (0.4)	74.6 (0.5)	83.8 (0.5)	78.9 (0.4)	74.6 (0.5)	83.8 (0.5)
SVM	70.6 (0.3)	53.5 (0.4)	59.1 (0.9)	70.6 (0.3)	53.5 (0.4)	59.1 (0.9)
SRNG	70.6 (0.3)	60.3 (8.8)	52.9 (6.2)	70.6 (0.3)	60.3 (8.8)	52.9 (6.2)
GRLVQ	60.7 (5.4)	59.0 (1.3)	54.0 (1.8)	60.7 (5.4)	59.0 (1.3)	54.0 (1.8)
rRBF	83.5 (0.4)	79.8 (1.4)	84.3 (0.5)	83.5 (0.4)	79.8 (1.4)	84.3 (0.5)

Table 10.1: The table contains the classification test accuracy averaged over a five-fold cross validation and according standard deviation in brackets. Best performance per classification is highlighted.

trition from the NIC1019 mutant, other mutants showed comparable levels of performance. It is apparent that both GRLVQ and SRNG show poor performance in data sets containing all leaf ages while improving significantly if just the youngest leaf is considered, especially in the VNIR 1600 camera data. Furthermore, VNIR 1600 and SWIR 320m data performed better in classification than SWIR 320is. As to be expected, the SVM classifier gains much from data transformation by LDA and shows poor performance on plain spectra data for the data set with all leaves, gaining performance if the youngest leaf is considered only. Leaf age seems to be the most prominent confounding factor for prediction of the nutrition state. Finally we have to note that the RBF network performs robustly with levels of performance matching those of an SVM on LDA subspace. Classification performance is affected by the leaf age to a much smaller extend than seen in GRLVQ and SRNG staying around 80% accuracy in all cameras for all leaf ages.

Figure 10.2b shows relevance profiles obtained from the rRBF trained for nutrition classification for all three cameras and the NIC1019 mutant in the youngest leaf (which showed best classification performances). Some interesting properties can be learnt from these relevance profiles.

The relevance profile is altered in a way to minimize the mean square error between network output and target vector, e.g. to maximize network classification performance. In the VNIR 1600 data, a peak of relevance can be observed right at the position of the intensity flank between $0.7\mu\text{m}$ and $7.5\mu\text{m}$. In the SWIR 320m, increased relevance can be observed at the flank between $1.5\mu\text{m}$ and $1.6\mu\text{m}$ as well as the end of the intensity flank between $1.8\mu\text{m}$ and $1.9\mu\text{m}$. A similar behavior can be found for the other mutants.

Figure 10.4a compares the test accuracy of nutrition classification for the rRBF model trained on input spaces of different dimensionality for the three selection strategies (relevance, discriminance, and random selection). The data, taken from the VNIR range, contains mutant NIC1019 and leaf one. In general, accuracy saturated extremely fast for both strategies. From around 20 spectral bands onwards, all strategies converge. However, it is apparent that the selection based on the relevance vector yields test accuracy above 0.9 with even just one spectral band. This shows that the relevance weighting can identify bands that will contribute highly to the classification task at hand.

Table 10.1 on the right side shows test data accuracy for classification results of genotypes. Generally, the classification of genotypes proves to be a harder task than the classification of nutrition. Like in the nutrition classification both GRLVQ and SRNG show poor performance in data sets containing all leaf ages while in contrast to nutrition performance did not improve significantly if just the youngest leaf was considered. There is also no clear difference of performance between VNIR 1600, SWIR 320m and SWIR 320is ranges across the board. The best results for genotypic classification were gained with an rRBF network in the VNIR range if just the youngest leaf is considered, yielding near 90% accuracy. From theory of plant nutrition it is known that young leaves are most active in terms of photosynthesis while older leaves show weaker photosynthesis related signals up to beginning senescence. From the application point of view this fact offers excellent perspectives for smart farming set-ups because typically only the youngest leaves are visible from above.

Figure 10.3b shows relevance profiles obtained from the rRBF network trained for genotype classification for all three cameras in the youngest leaf (which showed best classification performances). Like for the nutrition classification, some interesting properties can be derived from these relevance profiles. In the VNIR 1600 data, a peak of relevance can be observed right at the position of the intensity flank

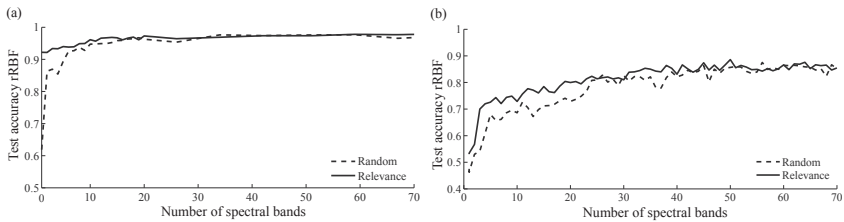


Figure 10.4: Model accuracy on reduced spectral bands: Spectral bands are chosen due to rRBF relevance profile or at random; rRBF Models are trained on reduced data set and test accuracy is depicted for (a) nutrition state prediction and (b) genotype classification.

between $0.7\mu\text{m}$ and $7.5\mu\text{m}$ but additionally, the range between $0.5\mu\text{m}$ and $0.6\mu\text{m}$ gains high relevance factors. In the SWIR 320m, increased relevance can be observed at the flank between $1.7\mu\text{m}$ and $1.9\mu\text{m}$.

Figure 10.4b compares the test accuracy of genotype classification for the rRBF model trained on input spaces of different dimensionality for the three selection strategies (relevance, discriminance, and random selection). The data, taken from the VNIR range, contained all nutrition conditions and leaf one. In general, accuracy saturated slower than for the nutrition classification showing that more spectral information is needed for this classification task. The relevance selection strategy yields a slight advantage in the classification performance. In general, this graph shows that the number of dimensions can be reduced massively and still retaining the classification accuracy.

6 Summary

Leaf age showed a strong impact on reducing classification performance for nutrition states by GRLVQ, SRNG, and SVM (plain spectra) down to near guessing level. Both SVM on LDA subspace and rRBF showed a robust nutrition prediction under leaf age variation. Such performance is needed for the utilization of machine learning methods in precision farming where other than under laboratory conditions recorded data cannot be controlled for individual leaf age for example. The genotype classification task proved to be much harder compared to nutrition classification. Here an rRBF trained on young leaf data in the VNIR range

showed best results while performance dropped significantly when all leaf ages are considered in the same model.

From the application point of view the obtained results clearly demonstrate the usefulness and suitability of this framework for precision phenotyping and smart farming. While detecting unknown genotypes is typically not in the focus of this kind of applications, recognition and quantitative modeling of the abundance of several metabolites, caused by various abiotic and biotic stress factors, is much more relevant. For example, quantitative changes in the plant's metabolism based on different supply of nutrients is the key to a wide range of smart farming applications. This also paves the way for modeling further metabolic effects that are relevant in plant breeding and pathogen response.

References

1. J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, Eds., vol. 1. University of California Press, 1967, pp. 281–297.
2. J. Moody and C. J. Darken, "Fast learning in networks of locally tuned processing units," *Neural Computation*, vol. 1, pp. 281–294, 1989.
3. A. Backhaus, F. Bollenbeck, and U. Seiffert, "Robust classification of the nutrition state in crop plants by hyperspectral imaging and artificial neural networks," in *In Proc. 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Lissabon, Portugal, 2011.
4. B. Hammer and T. Villmann, "Generalized relevance learning vector quantization," *Neural Networks*, vol. 15, pp. 1059–1068, 2002.
5. B. Hammer, M. Strickert, and T. Villmann, "Supervised Neural Gas with general similarity measure," *Neural Processing Letters*, vol. 21, pp. 21–44, 2005.
6. V. Vapnik and A. Chervonenkis, *Theory of Pattern Recognition*. Nauka, 1974.
7. B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms," *Neural computation*, vol. 12, no. 5, pp. 1207–1245, 2000.
8. C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press., 1995.
9. T. M. Martinetz and K. J. Schulten, "A Neural-Gas network learns topologies," in *Artificial Neural Networks*, T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, Eds. North-Holland, Amsterdam, 1991, pp. 397–402.