

14. Markham, B. L. and Barker, J. L., LANDSAT MSS and TM post-calibration dynamic rangers, exoatmospheric reflectance and at-satellite temperatures. *EOSAT LANDSAT Tech. Notes*, 1986, 3–8.
15. Kulkarni, A. V., Srinivasulu, J., Manjul, S. S. and Mathur, P., Field based spectral reflectance studies to develop NDSI method for snow cover monitoring. *J. Indian Soc. Remote Sensing*, 2002, **30**, 73–80.
16. Hall, D. K. and Riggs, G. A., Normalized-difference snow index (NDSI). *Encyclopedia of Snow, Ice and Glaciers*, Springer-Verlag, The Netherlands, 2011, pp. 779–780.
17. Gupta, R. P., Haritashya, U. K. and Singh, P., Mapping dry/wet snow cover in the Indian Himalayas using IRS multispectral imagery. *Remote Sensing Environ.*, 2005, **97**, 458–469.
18. Boris, S., Andrea, M., Jain, S. K., Toby, W. and Ian, H., A method for monthly mapping of wet and dry snow using Sentinel-1 and MODIS: Application to a Himalayan river basin. *Int. J. Appl. Earth Obs. Geoinf.*, 2019, **74**, 222–230.
19. Tsai, Y.-L., Dietz, A., Natascha, O. and Kuenzer, C., Wet and dry snow detection using Sentinel-1 SAR data for mountainous areas with a machine learning technique. *Remote Sensing*, 2019, **11**(8), 895.
20. Antarctica melts under its hottest days on record; <https://earthobservatory.nasa.gov/images/146322/antarctica-melts-under-its-hottest-days-on-record>
21. Flocco, D., Schroeder, D., Feltham, D. L. and Elizabeth, C. H., Impact of melt ponds on Arctic sea ice simulations from 1990 to 2007. *J. Geophys. Res.*, 2012, **117**(C9), 1–17.

Received 30 March 2020; revised accepted 15 January 2021

doi: 10.18520/cs/v120/i5/932-936

Wavelength selection and classification of hyperspectral non-imagery data to discriminate healthy and unhealthy vegetable leaves

Anjana N. Ghule^{1,*} and Ratnadeep R. Deshmukh²

¹Department of Information Technology, Government College of Engineering, Aurangabad 431 005, India

²Department of Computer Science and Information Technology, Dr B.A.M. University, Aurangabad 431 005, India

Being the largest vegetarian population across the globe, vegetables are an integral part of Indian meals. The proposed research finds significant wavelengths to discriminate healthy and unhealthy vegetable plants. Spectral-reflectance (SR) and first-derivative (FD) in the visible, red edge and near infrared region (350–1000 nm) of three vegetables brinjal, cluster beans and long beans were used. The significant wavelengths were selected using ReliefF and Support-

Vector-Machine (SVM). Random forest algorithm was used for classification. The binary classification was used for each vegetable separately, and multiclass classification was applied for all the samples. The most significant spectral wavelengths, for the prediction of diseased brinjal, correspond primarily to the red edge in SR. Long beans samples were classified accurately in the red-edge. In the case of cluster beans, SR is more effective than FD in the red-edge. The results substantiate the utility of HS data for discrimination of healthy and unhealthy vegetable plants and even vegetable types.

Keywords: Classification accuracy, healthy and unhealthy vegetable plants, hyperspectral measurements, spectral reflectance, wavelength selection.

HYPERSPECTRAL (HS) data are the spectral reflectance of target objects spread over a large number of narrow and continuous wavelengths over different portions of the electromagnetic spectrum (EMS). Crops exhibit different biophysical and biochemical characteristics like chlorophyll *a* and *b*, total chlorophyll, nitrogen content, carotenoid pigment, anthocyanin, plant stress, plant moisture and cell structure. HS remote sensing has the potential to detect subtle variations in these characteristics, which provide significant information about plant health, plant stress, crop yield and availability of nutrients. Analytical Spectral Devices, Inc. (ASD) spectroradiometer records reflectance spectra ranging from 350 to –2500 nm, covering the visible (Vis), near infrared (NIR) and shortwave infrared (SWIR) regions of the EMS help in the study and analysis of crops. A large number of bands ultimately result in multicollinearity and high correlation along many adjacent wavelengths^{1–3}. The selection of significant spectral region and optimal wavelengths can alleviate the dimensionality, and reduce classification complexity and improve classification accuracy^{4–6}.

For the classification of peatland vegetation, random forest, support vector machine (SVM), regularized logistic regression and partial least square-discriminant analysis (PLS-DA) have been employed, where non-imagery HS data were used⁷. PLSR and linear discriminant analysis (LDA) methods are employed to find the most significant wavebands for discrimination of similar weeds and different species of crops. Diago *et al.*⁸ captured HS images of three types of grapevines leaves for species discrimination. Images were captured using a camera and leaf reflectance was measured over 1040 wavelengths. Almost 92% classification accuracy was obtained by PLS classifier. Zapolska *et al.*⁹ applied LDA, PLS regression and principal component regression (PCR) to find optimal wavelengths for discrimination between healthy and diseased *Olea europaea* L.

Researchers have used analysis of variance (ANOVA) for selecting optimal spectral bands from both spectral reflectance (SR) and first derivative (FD) for differentiating

*For correspondence. (e-mail: anjanaghule@gmail.com)

healthy and unhealthy oil palms, and for the segregation of papyrus plants from other species^{10,11}. *Lycium barbarum* is an economically important plant in China with various pharmacological effects. Wang *et al.*¹² selected 37 significant wavelengths from the full spectra of 288 wavelengths using ANOVA, to find the origin of the plant.

Kumar *et al.*¹³ used PCA and stepwise discriminant analysis (SDA) for selecting optimum wavelengths from ground-based HS data to analyse spectral behaviour of tea plantations. The yield of tea plants depends upon factors like age, type of plant, bush health and growth stage. It was found that bands for the NIR and green region were best for discrimination between types of tea plants. The wavelengths from the red and NIR regions were most appropriate for discrimination between pruned and unpruned tea plants. Whereas the blue region was found most significant for determining plant health and plant growth.

Manjunath *et al.*¹⁴ carried out an experiment for selecting optimum bands to discriminate ornamental plants, cole crops and pulses. Field HS data analysis using the SDA technique showed that the most decisive wavelengths for cole crop discrimination were in the green (550 nm), red (690 nm) and NIR (740, 770 and 980 nm) regions. The four best wavelengths for pulse crops were found in the NIR region, and for flowers, ten wavelengths from the Vis–NIR regions were the most significant for discrimination. ReliefF was used for wavelength subset selection in the development of new spectral indices for diseases in three types of wheat¹⁵. Researchers have further proved that an ensemble of different dimensionality reduction techniques shows better results than a single technique. Maximum likelihood classifier, SVM classifier, random forest classifier, etc. are widely used for HS data classification. Random forest is a tree-based ensemble of the classifier which has shown good classification accuracy in different applications^{16–18}.

The present study was conducted for finding an optimum number of decisive wavelengths for discrimination of healthy and infected vegetable plants grown in Maharashtra, India. Considering the previous work done by researchers, it was decided to use the 350–1000 nm Visible and Near Infrared (VNIR) region of the EMS and divide this spectrum into three sub-regions. SVM and ReliefF were employed to find the optimum wavelengths from each region. Random forest ensemble classifier was used for accuracy assessment and comparing the performance of SVM and ReliefF attribute selection methods.

A farm of 100 acres in close proximity to Aurangabad city, Maharashtra was selected for the study. Different vegetables are grown here and supplied to the city. Field sample collection was carried out during rainy season on a sunny day of 27 August 2016. Brinjal, long beans and cluster beans vegetable plants were used for the study. Two hundred and four samples of brinjal (38 healthy + 22

unhealthy), cluster beans (40 healthy + 20 unhealthy) and long beans (40 healthy + 44 unhealthy) were selected from different locations of the farm. Leaves were clipped from the vegetable plants and kept in polythene bags separately. The samples were transferred to the laboratory for spectral measurements. In comparison with canopy measurements, laboratory measurements reduce the effect of atmospheric disturbances, unstable light and wind conditions, especially during rainy season.

The leaf spectral reflectance of the three vegetable plants was measured between 350 and 2500 nm with non-imaging ASD FieldSpec4 spectroradiometer. The instrument has a sampling interval of 1.4 nm for the region between 350 and 1000 nm, and 2 nm for the spectral region between 1000 and 2500 nm. The spectral resolution (Full-Width Half-Maximum; FWHM) of the spectroradiometer is 3 nm @ 700 nm, 8 nm @ 1400 and 8 nm @ 2100 nm for High Resolution. SR measurements were done in the darkroom with 8° field-of-view. The leaf was placed 15 cm beneath the sensor taking care that the complete leaf blade gets covered. The sample was illuminated using a 50 W halogen lamp. The white reference panel was used for calibration between the scans. Spectral measurements of the target leaf were recorded using RS3 software¹⁹.

Raw .asd data files using the RS3 software were further processed using ViewSpec Pro version 6.2. The spectral correction was performed for noise removal and calibration in the sensors of the spectroradiometer by splice correction and filtering. Mean spectral reflectance for multiple scans was recorded and first-order derivative with a gap of three was also calculated, for each sample. SR and FD measurements were converted to .csv files. To study the impact of diseases on different regions of the EMS, the SR and FD datasets were further divided into three regions, viz. vis (350–670 nm), red edge (671–780 nm) and NIR region (781–1000 nm) and datasets were prepared for every vegetable type separately.

The SR and FD data comprising 652 wavelengths were used for actual experimentation, but the sample size was comparatively small. The aim of this study was to select optimum significant wavelengths from different spectral regions, to discriminate between healthy and unhealthy vegetable plants.

Non-imagery HS signature describes the input object using hundreds to thousands of attributes in the form of wavelengths, of which most of the attributes provide little information. The task of selecting a small subset of the most significant wavelengths for discrimination is a key issue. It is proved that dimensionality reduction, prior to classification, reduces the complexity and improves classification accuracy. ReliefF and SVM supervised attribute selection methods along with ranker were applied for selecting discriminative wavelengths from SR and FD, to achieve high classification accuracy with reduced complexity. The binary classification was carried out

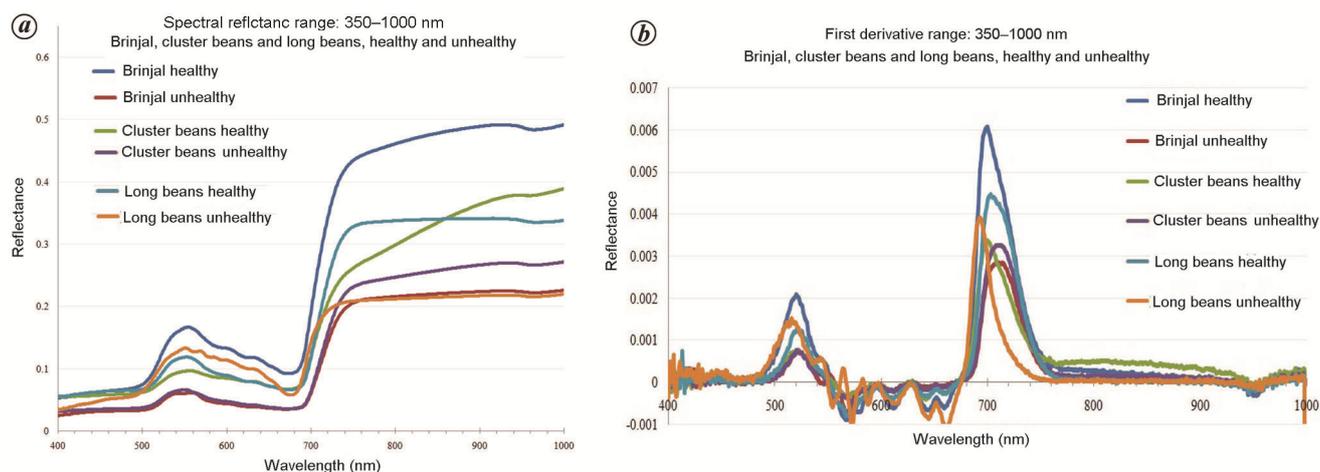


Figure 1. *a*, Spectral reflectance curve of healthy and unhealthy vegetable leaves. *b*, First derivative curve of healthy and unhealthy vegetable leaves.

Input: set of attributes (*A*) for each instances(*R*) and the class label

Output: set of feature weight *W* of estimations of the qualities of attributes

1. set all weights $W[A] := 0.0$
2. for $i := 1$ to m do begin
3. randomly selects an instance R_i ;
4. find k nearest hits H_j ;
5. for each class $C \neq class(R_i)$ do
6. from class C find k nearest misses $M_j(C)$;
7. for $A := 1$ to a do
8. $W[A] := W[A] - \sum_{j=1}^k \frac{diff(A,R_i,H_j)}{m.k} + \sum_{C \neq class(R_i)}^n \left[\frac{p(C)}{1-p(class(R_i))} \right] \sum_{j=1}^k \frac{diff(A,R_i,M_j(C))}{m.k}$;
9. end;

Figure 2. Pseudo code for ReliefF algorithm.

using the random forest algorithm by choosing 5, 10, 20 and all wavelengths in the three regions for healthy and unhealthy pairs of three vegetable plants²⁰.

Figure 1 *a* and *b* depicts SR and FD curves in the range of 350–1000 nm for both healthy and unhealthy leaves of brinjal, long beans and cluster beans respectively.

ReliefF is used to estimate a subset of features from a set of features (*A*), according to how well their values make a distinction among instances (*R*) that are close to each other. For this, sample R_i was selected randomly from the sample space. The ReliefF algorithm searches the two nearest neighbours for the selected instance by calculating the distance. Every neighbourhood contains k samples. The set of k nearest samples of the same class is regarded as a ‘hit’ (*H*) and k neighbours from a diverse class are considered a ‘miss’ (*M*). Figure 2 gives the pseudo-code for the ReliefF algorithm. The value of quality estimator weight vector $W[A]$ is updated for all the attributes *A*, which depend on instance R_i , hit(*H*) and

miss(*M*). The prior probability $P(C)$ of each class of misses is estimated from the training set, which contributes to the weight vector. In every step, we need the equal contributions from both hits and misses and they must be normalized to [0, 1]. It is necessary to have the probability weights of ‘miss’ sum up to 1. To ensure this, each of the probability weights is divided by $(1 - P(class(R_i)))^{21}$.

Supervised attribute selection and classification were performed using Weka 3.6, which is free and open source machine learning software. Considering the meagre sample size and high dimensionality of the dataset, attribute selection methods were used. As shown in Table 1 *a*, ReliefF and SVM algorithms with ranker search were used for selecting significant wavelengths and then random forest classification was applied to brinjal leaves. Classification accuracy was checked for the first 5, 10, 20 and all ranked wavelengths separately in the three regions on all the datasets.

The highest accuracy of 96.67% was obtained in the red edge region. It was also observed that both ReliefF and SVM feature selection demonstrated similar results: SR red edge region, spectral separability of SVM was better in the SR visible region to ReliefF. The contribution of NIR region was comparatively less in both SR and FD. Figure 3 *a* provides a summary of binary classification accuracy for brinjal.

Table 1 *b* shows the classification results of healthy and unhealthy cluster beans leaves. Using ReliefF attribute selection, the highest classification accuracy of 94.91 was observed in the red edge with 20 and all wavelengths. In the FD results, red edge accuracy was better than NIR accuracy. In both SR and FD classification, accuracy of the Vis region was less. Figure 3 *b* depicts variation in the classification accuracies.

For long beans, the highest classification accuracy of 96.43 was observed in the red edge and NIR regions of

SR. A decrease in SR and blue shift in the FD curve were also noticed. Table 1 *c* displays the results. In the FD results, red edge accuracy was better than NIR accuracy. In both SR and FD classification, accuracy of the Vis region was less. Figure 3 *c* shows a comparison of the results.

The classification accuracy of the ReliefF attribute selection method was more effective in the red edge region, whereas accuracy in the Vis and NIR red region was almost similar. The accuracy assessment of SVM indicates that the red edge is an important region, but the Vis region also provides significant wavelengths.

Table 2 shows confusion matrices of 20 ranked wavelengths selected using ReliefF – random forest multiclass classification for all six sample types for Vis, red edge and NIR regions in SR and FD data. Each sample is categorized into one of the six classes.

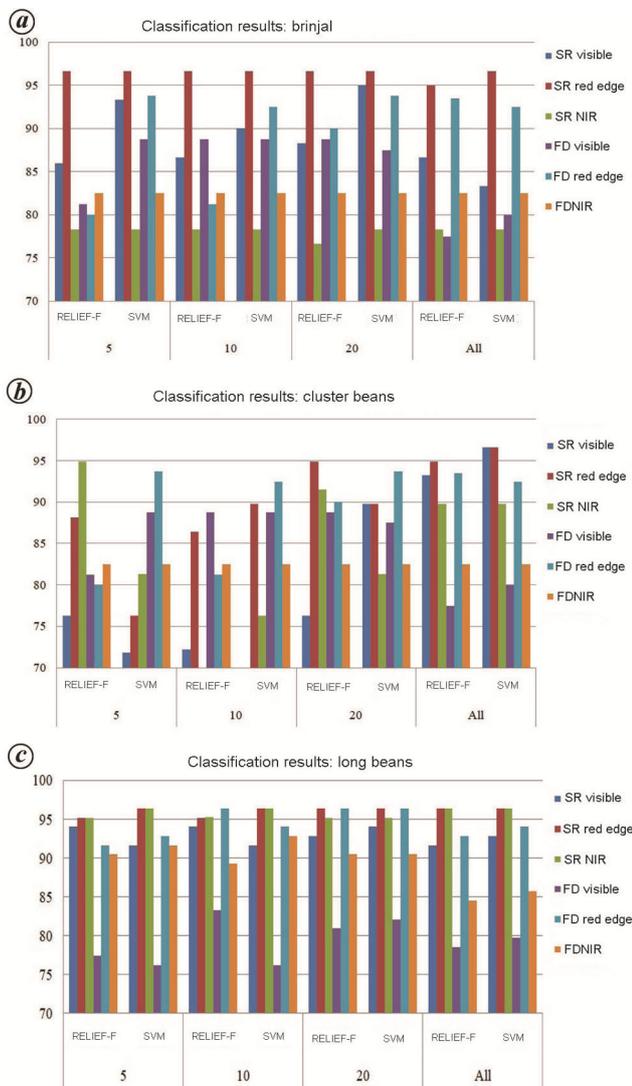


Figure 3. Comparison between classification accuracy and the number of wavelengths selected by support vector machine (SVM) and ReliefF for (a) brinjal; (b) cluster beans and (c) long beans.

The values extracted from the weighted average of precision, recall, F1-score and accuracy calculations of confusion matrices are listed in Table 3 along with kappa statistics. Precision shows how many positively predicted samples are truly positive. Recall shows how many truly positive samples are predicted positive by the classifier. F1-score is a harmonic mean of precision and recall, while accuracy is a measure of correctly classified samples for the total number of samples. A weighted average of these values was calculated by considering the sample size of each class. Table 3 shows the highest classification accuracy of 95.5% in the red edge region of SR.

Chlorophyll is a fundamental indicator of leaf health. Loss of chlorophyll content severely affects the red edge region and also affects the Vis region. From previous studies, it can be concluded that the NIR region is sensitive to the leaf area, leaf density and deformation of leaf structure²²⁻²⁴.

In all the three vegetable plants, the red edge region was dominantly affected (Figure 4). It can be concluded that all of them suffer from chlorophyll stress. In the case of long beans, disease resulted in leaf deformation, which affected the NIR region as well. The same framework can also be used for the classification of different types of vegetables considering their leaf health.

Non-imagery HS remote sensing provides spectral measurements in the form of a large number of contiguous wavelengths. This study was carried out to find the most important region and minimum subset of wavelengths from SR as well as FD, which can be used for discrimination between healthy and unhealthy vegetable leaves. Dividing the spectral signature into different regions helps study the biophysical and biochemical impact of a disease on different regions of the EMS individually. Selecting a subset of bands reduces the complexity of the classification task.

From the results of this study, it can be observed that we can attain good results by focusing on the red edge region. However, the effect on the Vis region was also

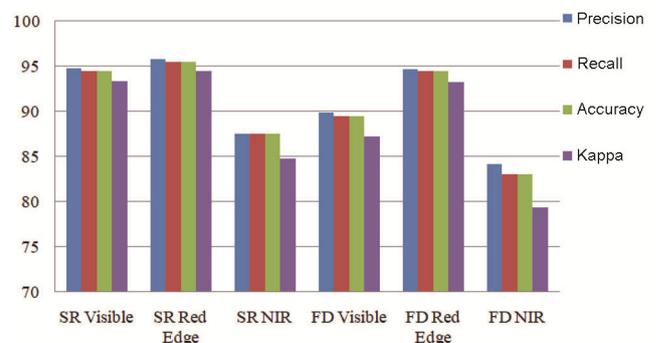


Figure 4. Comparison of classification accuracy with 20 wavelengths selected using ReliefF for visible, red edge and NIR regions of spectral reflectance and first derivative.

Table 1. Classification results of healthy and unhealthy leaves for 5, 10, 20 and all wavelengths

Number of wavelengths	Attribute selection method	Spectral reflectance (SR; accuracy)			First derivative (FD; accuracy)		
		Visible	Red edge	Near infrared	Visible	Red edge	Near infrared
<i>a, Brinjal</i>							
5	ReliefF	86.00	96.67	78.33	81.25	80.00	82.50
	SVM	93.33	96.67	78.33	88.75	93.75	82.50
10	ReliefF	86.67	96.67	78.33	88.75	81.25	82.50
	SVM	90.00	96.67	78.33	88.75	92.50	82.50
20	ReliefF	88.33	96.67	76.67	88.75	90.00	82.50
	SVM	95.00	96.67	78.33	87.50	93.75	82.50
All	ReliefF	86.67	95.00	78.33	77.50	93.50	82.50
	SVM	83.33	96.67	78.33	80.00	92.50	82.50
<i>b, Cluster beans</i>							
5	ReliefF	76.27	88.13	94.91	81.25	80.00	82.5
	SVM	71.81	76.27	81.35	88.75	93.75	82.5
10	ReliefF	72.22	86.44	91.52	88.75	81.25	82.5
	SVM	66.10	89.83	76.27	88.75	92.5	82.5
20	ReliefF	76.27	94.91	91.52	88.75	90.00	82.5
	SVM	89.83	89.83	81.35	87.50	93.75	82.5
All	ReliefF	93.22	94.91	89.83	77.50	93.5	82.5
	SVM	96.61	96.61	89.83	80.00	92.5	82.5
<i>c, Long beans</i>							
5	ReliefF	94.05	95.24	95.24	77.38	91.67	90.48
	SVM	91.67	96.43	96.43	76.19	92.86	91.67
10	ReliefF	94.05	95.24	95.29	83.33	96.43	89.29
	SVM	91.67	96.43	96.43	76.19	94.05	92.86
20	ReliefF	92.86	96.43	95.24	80.95	96.43	90.48
	SVM	94.05	96.43	95.24	82.14	96.43	90.48
All	ReliefF	91.67	96.43	96.43	78.57	92.86	84.52
	SVM	92.86	96.43	96.43	79.76	94.05	85.71

Table 2. Confusion matrices of multi classification carried out on all six types of samples with 20 wavelengths selected using ReliefF in visible, red edge and Infrared regions of SR and FD

SR-visible (a)	SR-red edge (b)	SR-near infrared (c)
a b c d e f ← classified as	a b c d e f ← classified as	a b c d e f ← classified as
38 0 0 0 0 0 a = BH	37 0 0 1 0 0 a = BH	33 4 0 0 1 0 a = BH
1 20 0 0 0 1 b = BUH	1 21 0 0 0 0 b = BUH	6 15 1 0 0 0 b = BUH
1 0 37 0 2 0 c = CBH	1 0 38 0 1 0 c = CBH	0 0 37 1 2 0 c = CBH
0 0 1 39 0 0 d = CBUH	1 0 0 39 0 0 d = CBUH	1 0 0 38 1 0 d = CBUH
3 0 0 1 35 1 e = LBH	2 0 0 2 36 0 e = LBH	0 0 2 2 35 1 e = LBH
0 0 0 0 0 20 f = LBUH	0 0 0 0 0 20 f = LBUH	0 0 0 0 3 17 f = LBUH
FD-visible (d)	FD-red edge (e)	FD-near infrared (f)
a b c d e f ← classified as	a b c d e f ← classified as	a b c d e f ← classified as
35 2 0 1 0 0 a = BH	38 0 0 0 0 0 a = BH	34 3 0 1 0 0 a = BH
1 21 0 0 0 0 b = BUH	1 21 0 0 0 0 b = BUH	2 20 0 0 0 0 b = BUH
0 0 37 2 1 0 c = CBH	1 0 37 0 2 0 c = CBH	1 0 37 0 2 0 c = CBH
0 0 9 31 0 0 d = CBUH	0 0 4 36 0 0 d = CBUH	0 0 12 28 0 0 d = CBUH
1 1 0 1 35 2 e = LBH	1 0 0 0 38 1 e = LBH	0 0 0 0 33 7 e = LBH
0 0 0 0 0 20 f = LBUH	0 0 0 0 1 19 f = LBUH	0 0 0 0 6 14 f = LBUH

noticeable for chlorophyll and other pigments. It was also observed that the NIR region was more sensitive to leaf structure. In majority of cases, the classification performance with 20 wavelengths was comparable to all the wavelengths in that region. Performance of classification for ReliefF and SVM attribute selection was similar

in some of the cases. However, taken as a whole, random forest classification accuracy was better for the first 20 ranked wavelengths selected using the ReliefF algorithm. The performance of ReliefF–Random forest was good in both binary as well as multiclass classification.

Table 3. Summary of weighted average of classification results derived from the use of the full 20 wavelengths selected using ReliefF supervised attribute selection method

		Precision	Recall	F-score	Accuracy	Kappa
SR	Visible – 320 (351–670 nm)	94.8	94.5	94.65	94.5	93.31
	Red edge – 110 (671–780 nm)	95.8	95.5	95.65	95.5	94.52
	Near infrared – 220 (781–1000 nm)	87.5	87.5	87.5	87.5	84.75
FD	Visible – 320 (351–670 nm)	89.9	89.5	89.7	89.5	87.25
	Red edge – 110 (671–780 nm)	94.7	94.5	94.6	94.5	93.3
	Near infrared – 220 (781–1000 nm)	84.2	83	83.6	83	79.34

Further efforts are needed in the SWIR region to study the impact of diseases on water content of the leaves. In order to monitor the damage caused by diseases and pest control, measurements can be taken at different levels of infestation.

- Thenkabail, P. S. *et al.*, Selection of hyperspectral narrowbands (HNBS) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and hyperion/EO-1 data. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sensing*, 2013, **6**(2), 427–439; doi:10.1109/jstars.2013.2252601.
- Sahoo, S. S. *et al.*, Hyperspectral remote sensing of agriculture. *Curr. Sci.*, 2015, **108**(5), 848–859.
- Nagaraja, S. *et al.*, Spectral discrimination of healthy and malformed mango panicles using spectroradiometer. *Indian J. Hortic.*, 2014, **72**(1), 40–44.
- Lehmann, J. *et al.*, Field spectroscopy in the VNIR–SWIR region to discriminate between Mediterranean native plants and exotic-invasive shrubs based on leaf tannin content. *Remote Sensing*, 2015, **7**(2), 1225–1241; doi:10.3390/rs70201225.
- Pal, M. and Foody, G. M., Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sensing*, 2010, **48**(5), 2297–2307; doi:10.1109/tgrs.2009.2039484.
- Ray, J. P., Use of hyperspectral remote sensing data for crop stress detection: ground-based studies. *Int. Arch. Photogramm., Remote Sensing Spat. Inf. Sci.*, 2010, **38**(8), 562–570.
- Erudel, T. *et al.*, Criteria comparison for classifying Peatland vegetation types using *in situ* hyperspectral measurements. *Remote Sensing*, 2017, **9**(7), 748; doi:10.3390/rs9070748.
- Diago, M. P., Fernandes, A. M., Millan, B., Tardaguila, J. and Melo-Pinto, P., Identification of grapevine varieties using leaf spectroscopy and partial least squares. *Comput. Electron. Agric.*, 2013, **99**, 7–13; doi:10.1016/j.compag.2013.08.021.
- Zapolska, A. *et al.*, Linear discriminant analysis of spectral measurements for discrimination between healthy and diseased trees of *Olea europaea* L. artificially infected by *Fomitiporia mediterranea*. *Int. J. Remote Sensing*, 2020, **41**(14), 5388–5398; doi:10.1080/01431161.2020.1731931.
- Shafri, H. Z. M. *et al.*, Spectral discrimination of healthy and ganoderma-infected oil palms from hyperspectral data. *Int. J. Remote Sensing*, 2011, **32**(22), 7111–7129; doi:10.1080/01431161.2010.519003.
- Adam, E. *et al.*, Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands Ecol. Manage.*, 2009, **18**(3), 281–296; doi:10.1007/s11273-009-9169-z.
- Wang, L., Li, J., Qin, H., Xu, J., Zhang, X. and Huang, L., Selecting near-infrared hyperspectral wavelengths based on one-way ANOVA to identify the origin of *Lycium barbarum*. In International Conference on High Performance Big Data and Intelligent Systems, Shenzhen, China, 2019, pp. 122–125.
- Kumar, A. *et al.*, Field hyperspectral data analysis for discriminating spectral behavior of tea plantations under various management practices. *Int. J. Appl. Earth Obs. Geoinf.*, 2013, **23**, 352–359; doi:10.1016/j.jag.2012.10.006.
- Manjunath, K. R. *et al.*, Discrimination of spectrally-close crops using ground-based hyperspectral data. *J. Indian Soc. Remote Sensing*, 2011, **39**(4), 599–602; doi:10.1007/s12524-011-0099-x.
- Huang, W. *et al.*, New optimized spectral indices for identifying and monitoring winter wheat diseases. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing*, 2014, **7**(6), 2516–2524; doi:10.1109/jstars.2013.2294961.
- Mountrakis, G. *et al.*, Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sensing*, 2011, **66**(3), 247–259; doi:10.1016/j.isprsjrs.2010.11.001.
- Lin, X. *et al.*, Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics. *Molecules*, 2017, **23**(1), 52; doi:10.3390/molecules23010052.
- Huang, L. *et al.*, Identification of Fusarium head blight in winter wheat ears based on Fisher's linear discriminant analysis and a support vector machine. *Appl. Sci.*, 2019, **9**(18), 3894; doi:10.3390/app9183894.
- <https://www.malvernanalytical.com/en/products/product-range/asd-range/fieldspec-range/fieldspec4-hi-res-high-resolution-spectroradiometer>
- Hall, M. and Smith, L. A., Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In Florida Artificial Intelligence Research Society Conference, 1999.
- Robnik-Šikonja, M. and Kononenko, I., Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.*, 2003, **53**(1), 23–69.
- Curran, P. J. *et al.*, Exploring the relationship between reflectance red edge and chlorophyll concentration in slash pine leaves. *Tree Physiol.*, 1995, **15**(3), 203–206; doi:10.1093/treephys/15.3.203.
- Kollenkark, J. C. *et al.*, Effects of cultural practices on agronomic and reflectance characteristics of soybean canopies I. *Agron. J.*, 1982, **74**(4), 751–758; doi:10.2134/agronj1982.00021962007400-040035x.
- Slaton, M. R. *et al.*, Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am. J. Bot.*, 2001, **88**(2), 278–284; doi:10.2307/2657019.

ACKNOWLEDGEMENTS. The Department of Science and Technology-Funds for Infrastructure under Science and Technology (DST-FIST) has supported the Department of Computer Science and Information Technology, Dr Babasaheb Ambedkar Marathwada University, Aurangabad, under section No. SR/FST/ET1340 to carry out this work. We thank the Department and University authorities for providing the required resources.

Received 30 November 2020; revised accepted 14 January 2021

doi: 10.18520/cs/v120/i5/936-941