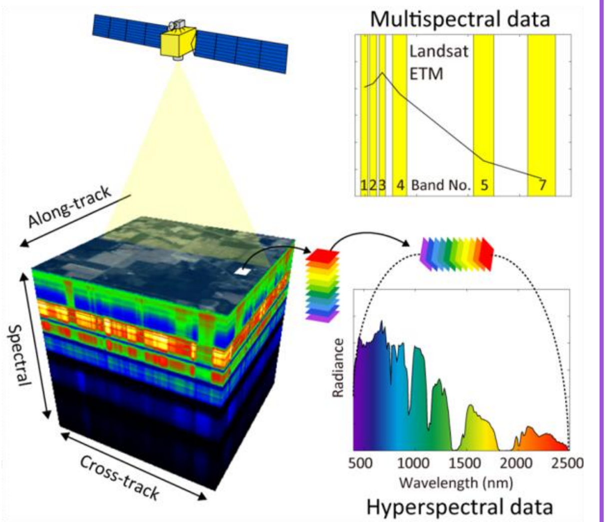
Mutual Information and Entropy based Band Selection for Spectral-Spatial Classification of Hyperspectral Images

Subir Paul (*subirpaul@iisc.ac.in*) and Nagesh Kumar Dasika
Department of Civil Engineering, Indian Institute of Science, Bengaluru, India

EGU2018-7297

Motivation

- Hyperspectral data has great advantage in different types of land surface features identification or classification since this data contains large number of bands with very fine spectral resolution
- But hyperspectral data processing is very challenging task because of the presence of high dimensionality and redundant information in the data
- Plenty of techniques are being developed to deal with the issues of the hyperspectral data
- Here we are proposing the use of a simple and unsupervised band selection approach along with spatial features in order to achieve apt classification performance



Subir Paul (*subirpaul@iisc.ac.in*) and Nagesh Kumar Dasika
Department of Civil Engineering, Indian Institute of Science, Bengaluru, India

EGU2018-7297

Methodology

Hyperspectral data

↓

Spectral segmentation based on MI

↓

Selection of representative spectral bands from each segment based on Entropy

↓

Creation of MPs corresponding to each representative spectral band

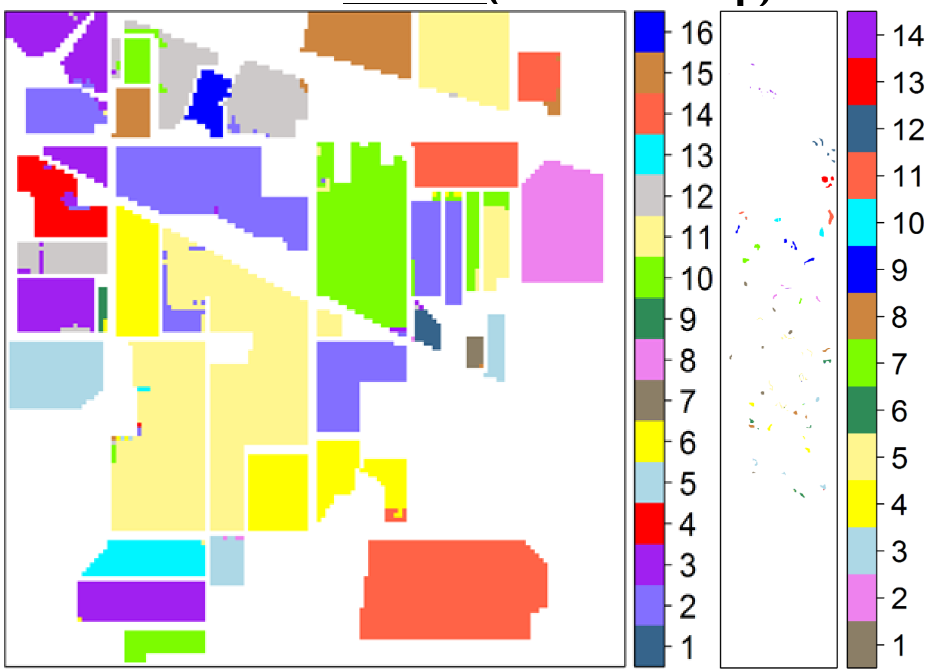
↓

Classification of the data using spectral-spatial features in SVM/RF classifier

Subir Paul (*subirpaul@iisc.ac.in*) and Nagesh Kumar Dasika
Department of Civil Engineering, Indian Institute of Science, Bengaluru, India

EGU2018-7297

Results (classified map)



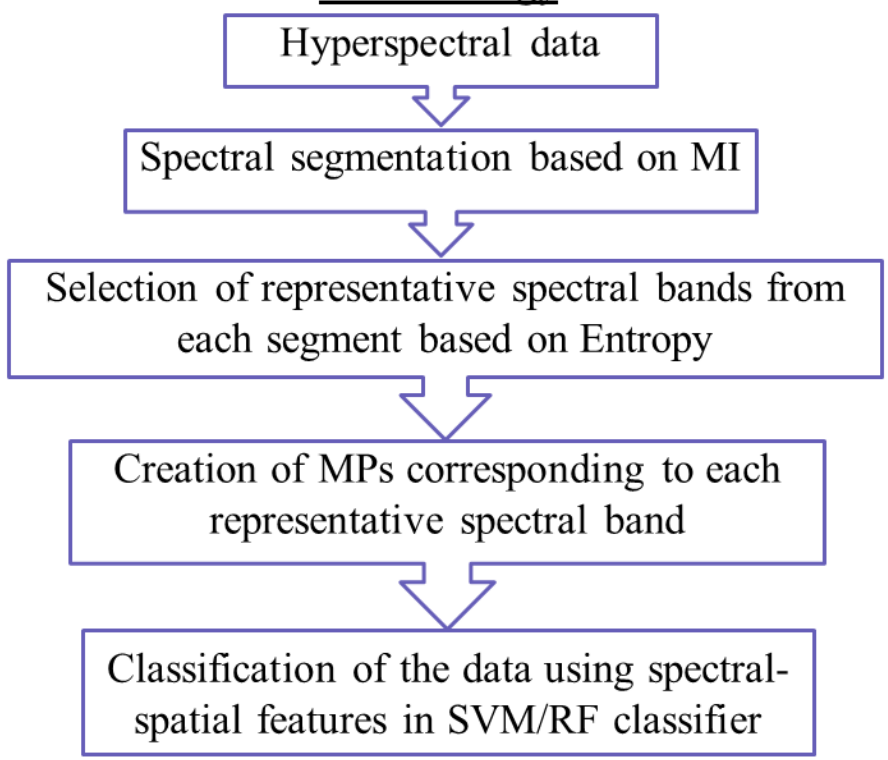
Indian Pines, OA=97.54%   Botswana, OA=98%

MOTIVATION

- Hyperspectral data processing is very challenging because of the presence of high dimensionality and redundant information in the data
- Reduction of high dimensionality and preservation of salient information from the data can be achieved either by feature selection (FS) or by feature extraction (FE) approach
- FS techniques, having the advantages of retaining the original physical information of the spectral bands, often found to be preferable over the FE techniques (Feng et al. 2014; MartÍnez-UsÓMartinez-Uso et al. 2007)
- Use of a simple and unsupervised feature selection approach in order to achieve optimal classification performance in less computational time
- Use of spatial features along with the spectral bands in the classifier model for the improvement of classification performance

METHODOLOGY

Methodology



METHODOLOGY

- All the spectral bands of the HS data are divided into local spectral segments based on their inter-band dependencies (MI).

$$MI(X;Y)=\sum_{y\in Y}\sum_{x\in X}p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

- Representative bands are selected from each spectral segment, having the maximum entropy measure.

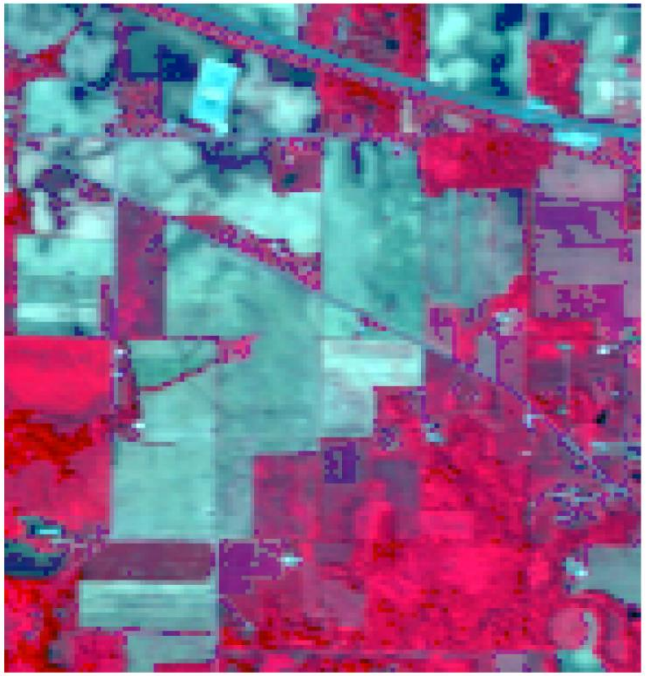
$$H(X)=-\sum_{x\in X}p(x)\log(p(x))$$

- EMPs are created by performing the morphological operations (opening and closing) on the selected representative spectral bands to take into account the spatial information in the classification process.

DATASETS

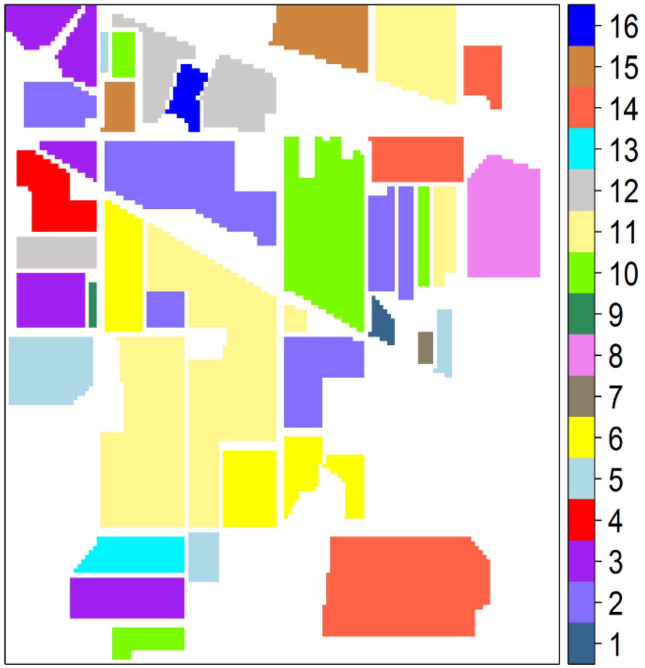
Indian Pines Botswana

Details of the labelled samples



Colour composite image [R: 880, G: 647, B: 548 nm]

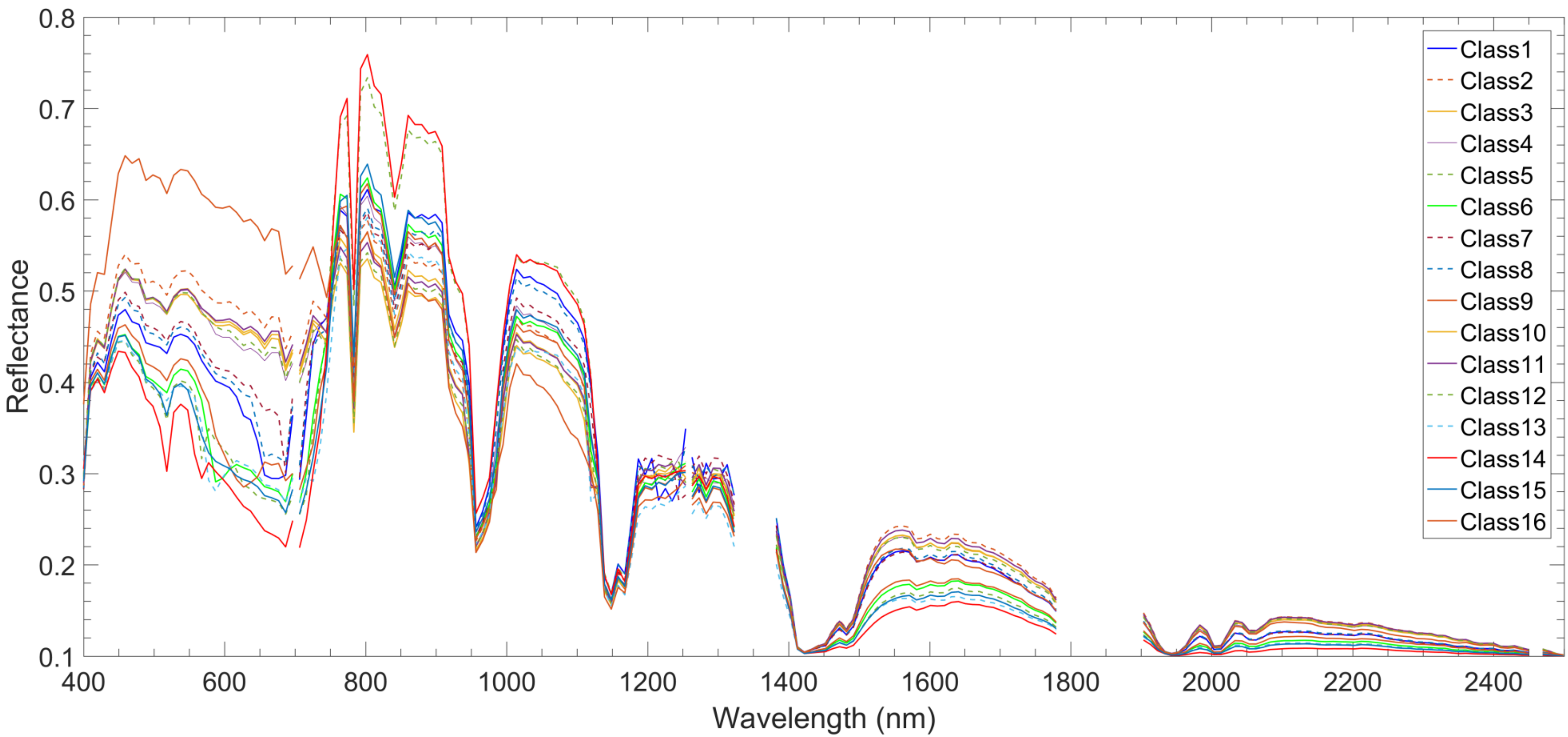
Class Sl. No.	Class Name	Training Samples	Testing Samples	Total Samples
1	Alfalfa	4	42	46
2	Corn-notill	142	1286	1428
3	Corn-mintill	82	748	830
4	Corn	23	214	237
5	Grass-pasture	49	434	483
6	Grass-trees	72	658	730
7	Grass-pasture-mowed	2	26	28
8	Hay-windrowed	48	430	478
9	Oats	2	18	20
10	Soybean-notill	98	874	972
11	Soybean-mintill	245	2210	2455
12	Soybean-clean	59	534	593
13	Wheat	21	184	205
14	Woods	127	1138	1265
15	Buildings-Grass-Trees-Drives	38	348	386
16	Stone-Steel-Towers	9	84	93
Total		1021	9228	10249



Ground-truth or class label map

DATASETS

Indian Pines Botswana



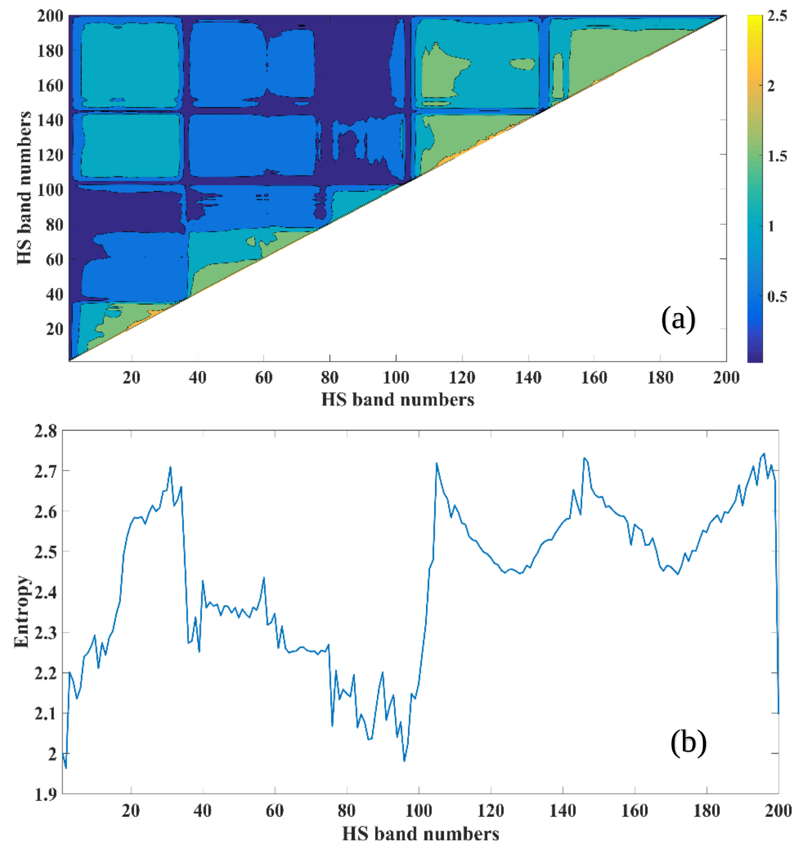
Average Spectral Reflectance Curve of all the classes

RESULTS

Indian Pines Botswana

- All the bands are divided into six spectral segments (Band 1-35, 36-60, 61-79, 80-104, 105-145, 146-200), where all the bands in a segment are highly dependent on each other (Paul and Kumar 2018)
- Representative spectral bands from each segment:

Band number	Wavelength (nm)
6	450
31	696
36	745
57	947
62	995
79	1158
105	1432
144	1933
151	2003
196	2450



(a) MI between all the combinations of two HS bands, and (b) entropy of each HS band.

RESULTS

Indian Pines Botswana

Classification performances using the proposed approach

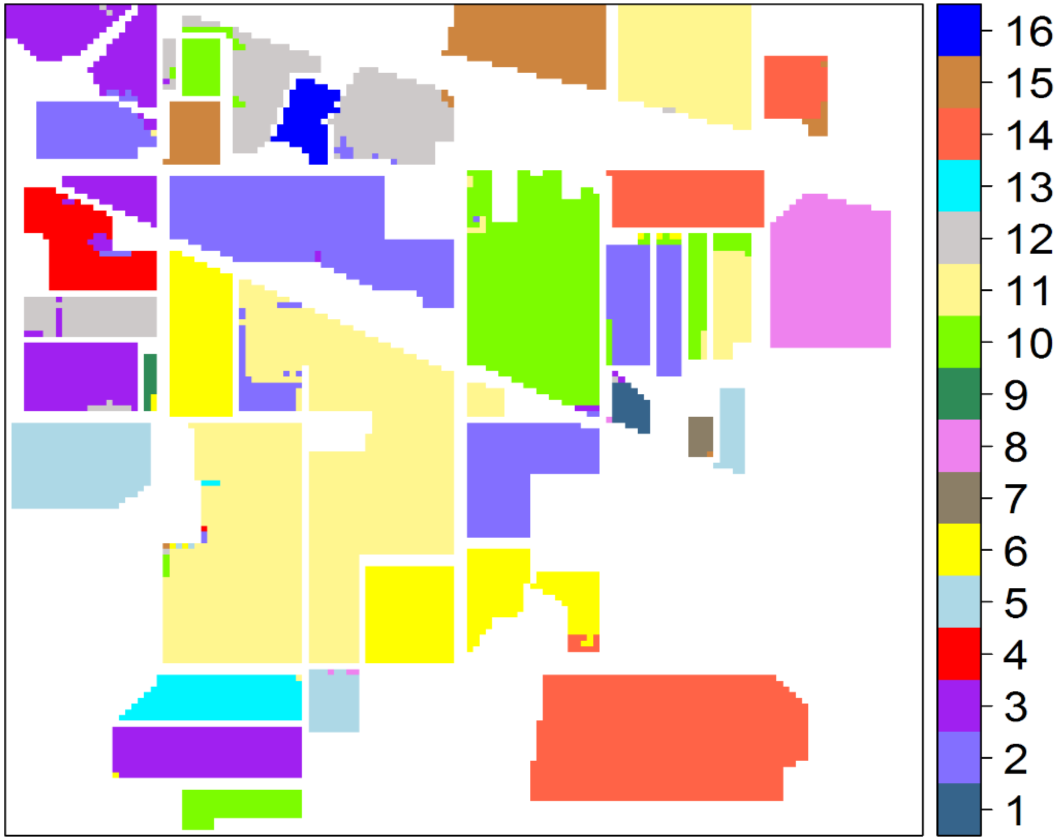
Classifier	OA (%)	k	AA (%)
RBF-SVM	94.05±0.66	0.9321±0.0076	95.13±0.91
RF	**96.51±0.42**	**0.9602±0.0048**	**97.27±0.48**

Comparison of different feature selection based approaches

Method	No. of features	Classification performances		
		OA (%)	k	AA (%)
MBR_MVPCA	15	69.05±0.80	0.6447±0.0099	68.78±2.24
MBR_MI	15	67.44±1.57	0.6250±0.0183	68.05±4.66
MBR_ANR_AP	15	75.88±1.19	0.7239±0.0134	73.47±2.93
Proposed approach	15	90.18±1.89	0.8879±0.0221	91.67±1.38
Proposed approach	30	**96.51±0.42**	**0.9602±0.0048**	**97.27±0.48**
k-means clustering and entropy	30	68.35±0.99	0.6356±0.0123	66.12±3.38

RESULTS

Indian Pines Botswana



OA=97.54%

Classified map (and corresponding OA) prepared from the results of the proposed approach

RESULTS

Indian Pines Botswana

Confusion Matrix

Output Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	UA
1	42	0	2	0	0	0	0	1	0	0	0	1	0	0	0	0	91.3%
2	0	1372	7	0	0	4	0	0	0	16	29	0	0	0	0	0	96.1%
3	0	6	813	0	0	1	0	0	0	0	0	10	0	0	0	0	98.0%
4	0	5	11	221	0	0	0	0	0	0	0	0	0	0	0	0	93.2%
5	0	0	1	0	462	0	0	3	0	2	0	15	0	0	0	0	95.7%
6	0	0	0	0	0	718	0	0	0	0	0	0	0	12	0	0	98.4%
7	0	0	0	0	0	0	27	0	0	0	0	0	0	0	1	0	96.4%
8	0	0	0	0	0	0	0	478	0	0	0	0	0	0	0	0	100%
9	0	0	0	0	0	3	0	0	17	0	0	0	0	0	0	0	85.0%
10	0	3	4	0	0	0	0	0	0	953	12	0	0	0	0	0	98.0%
11	0	15	0	1	2	2	0	0	0	23	2405	3	3	0	1	0	98.0%
12	0	9	9	0	0	0	0	0	0	15	0	555	0	0	4	1	93.6%
13	0	0	0	0	0	0	0	0	0	0	1	0	204	0	0	0	99.5%
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1251	14	0	98.9%
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	386	0	100%
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	93	100%
PA	100%	97.3%	96.0%	99.5%	99.6%	98.6%	100%	99.2%	100%	94.4%	98.3%	95.0%	98.6%	99.0%	95.1%	98.9%	97.5%

Target Class

RESULTS

Indian Pines Botswana

Classification performances of the proposed approach with the change of training sample size

Training sample size	10%	20%	30%	40%	50%
Training time (min)	1.84±0.28	2.98±0.42	5.16±1.07	8.03±1.59	10.76±2.14
OA (%)	96.51±0.42	98.19±0.24	98.94±0.16	99.18±0.25	99.33±0.20
k	0.9602±0.0048	0.9794±0.0027	0.9880±0.0018	0.9907±0.0029	0.9924±0.0023
AA (%)	97.27±0.48	98.51±0.44	99.25±0.13	99.50±0.20	99.52±0.28

RESULTS

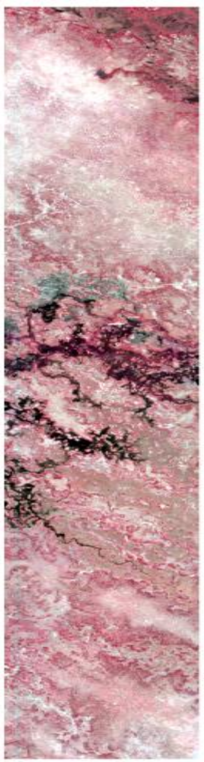
Indian Pines Botswana

Performances of feature extraction based different spectral and spectral-spatial classification approaches (Paul and Kumar 2018)

Approaches	Spectral classification		Spectral-spatial classification			
FE methods	**AE**	**SAE**	**PCA**	**AE**	**S-AE**	**S-SAE**
Classifier	RBF-SVM	RBF-SVM	RF	RF	RF	RF
OA (%)	80.46±0.73	80.16±0.63	94.77±0.70	95.63±0.62	96.07±0.60	**96.66±0.66**
k	0.7765±0.0082	0.7730±0.0070	0.9404±0.0080	0.9501±0.0071	0.9552±0.0069	**0.9619±0.0076**
AA (%)	79.63±2.71	80.53±2.77	95.92±0.73	95.93±1.53	97.03±0.79	**97.42±0.91**

DATASETS

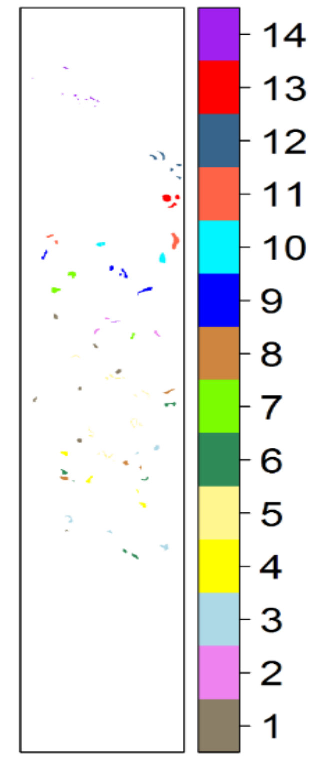
Indian Pines   Botswana



Colour composite image [R: 875, G: 650, B: 550 nm]

Details of the labelled samples

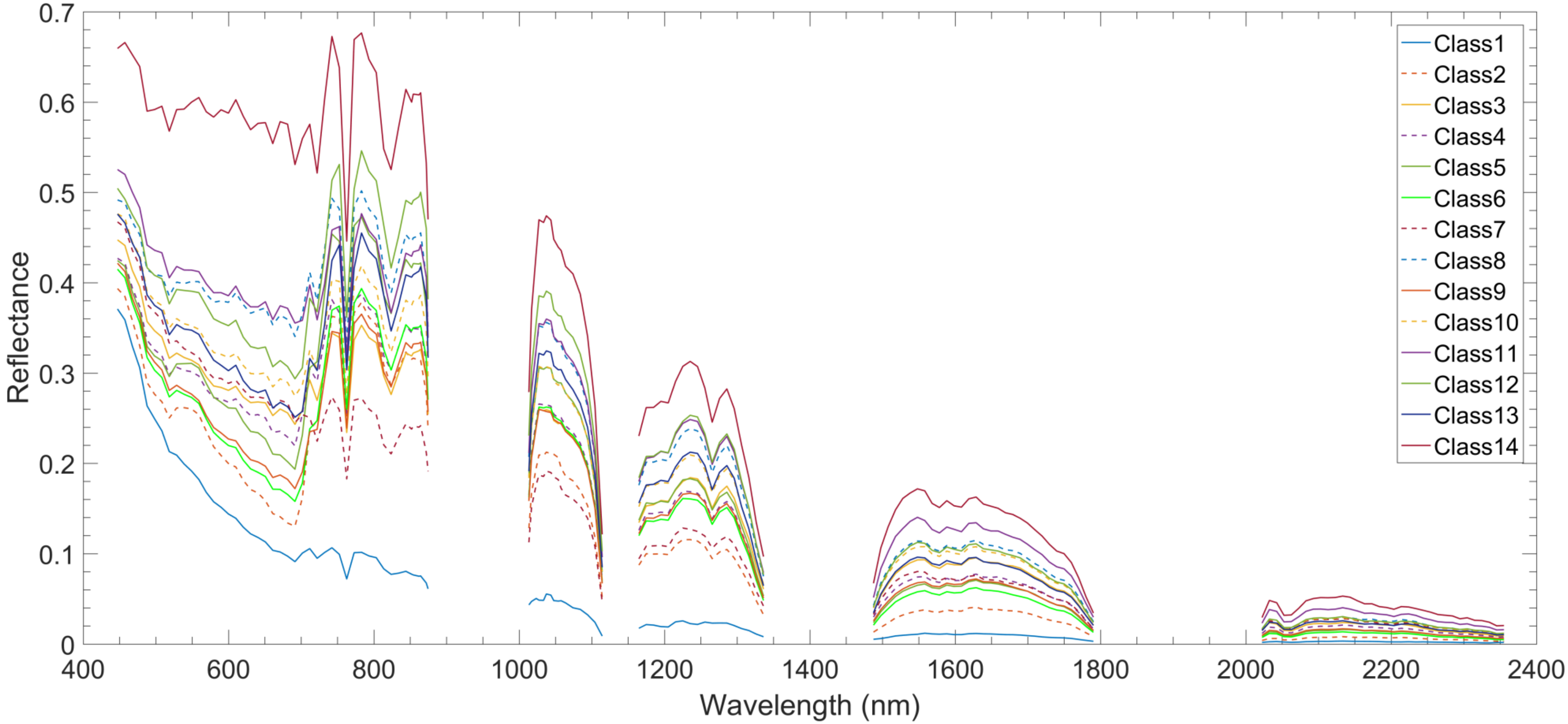
Class Sl. No.	Class Name	Training Samples	Testing Samples	Total Samples
1	Water	26	244	270
2	Hippo grass	11	90	101
3	Floodplain grasses 1	25	226	251
4	Floodplain grasses 2	21	194	215
5	Reeds	27	242	269
6	Riparian	27	242	269
7	Firecar	25	234	259
8	Island interior	21	182	203
9	Acacia woodlands	32	282	314
10	Acacia shrublands	24	224	248
11	Acacia grasslands	31	274	305
12	Short mopane	19	162	181
13	mixed mopane	26	242	268
14	Exposed soils	9	86	95
Total		324	2924	3248



Ground-truth or class label maps

DATASETS

Indian Pines Botswana



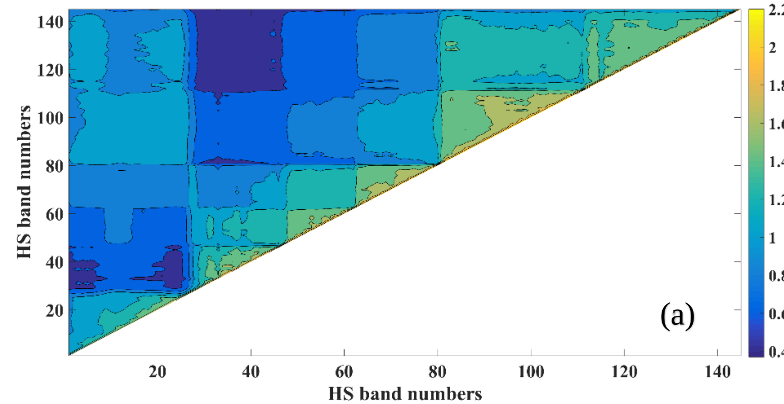
Average Spectral Reflectance Curve of all the classes

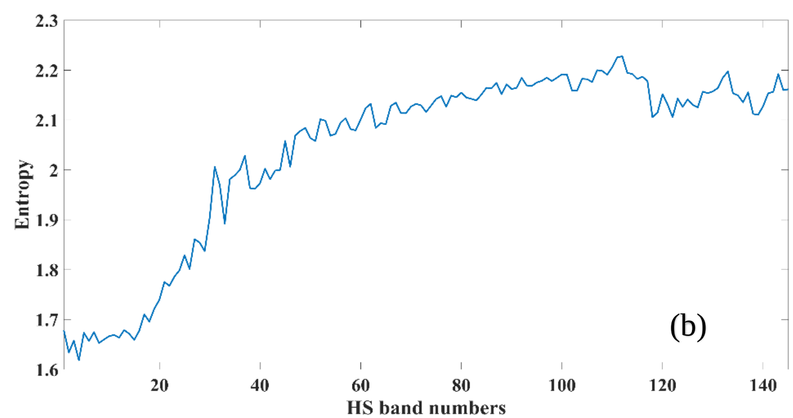
RESULTS

Indian Pines Botswana

- All the bands are divided into six spectral segments (Band 1-27, 28-47, 48-63, 64-81, 82-111, 112-145), where all the bands in a segment are highly dependent on each other (Paul and Kumar 2018)
- Representative spectral bands from each segment:

Band number	Wavelength (nm)
27	712
47	1013
62	1114
80	1336
111	1790
112	2022





(a) MI between all the combinations of two HS bands, and (b) entropy of each HS band.

RESULTS

Indian Pines Botswana

Classification performances using the proposed approach

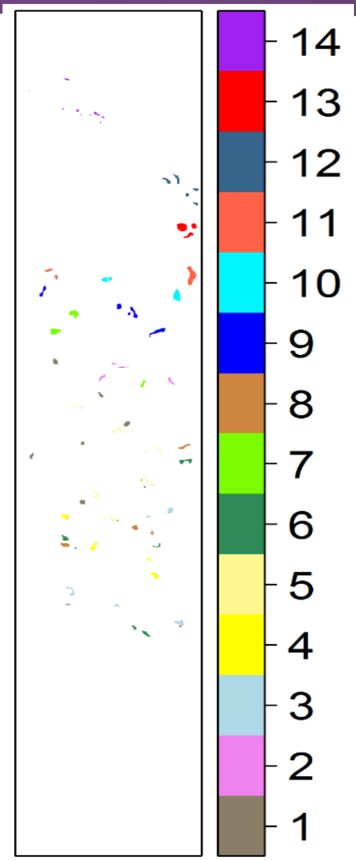
Classifier	OA (%)	k	AA (%)
RBF-SVM	**97.01±0.77**	**0.9676±0.0084**	**97.13±0.84**
RF	94.43±1.32	0.9397±0.0144	94.64±1.26

Comparison of different feature selection based approaches

Method	No. of features	Classification performances		
		OA (%)	k	AA (%)
MBR_MVPCA	15	84.54±0.94	0.8324±0.0101	85.84±1.45
MBR_MI	15	81.29±1.44	0.7972±0.0157	82.60±1.37
MBR_ANR_AP	15	88.23±1.28	0.8724±0.0139	89.34±1.11
Proposed approach	18	**97.01±0.77**	**0.9676±0.0084**	**97.13±0.84**
k-means clustering and entropy	18	88.69±1.54	0.8775±0.0167	89.61±1.51

RESULTS

Indian Pines Botswana



OA=98%

Classified maps (and corresponding OA) prepared from the results of the proposed approach

RESULTS

Indian Pines Botswana

Confusion Matrix

Output Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	**UA**
1	270	0	0	0	0	0	0	0	0	0	0	0	0	0	91.3%
2	0	101	0	0	0	0	0	0	0	0	0	0	0	0	96.0%
3	0	0	239	0	0	0	6	0	1	0	0	0	5	0	97.8%
4	0	0	4	211	0	0	0	0	0	0	0	0	0	0	92.9%
5	0	1	0	1	257	8	0	0	2	0	0	0	0	0	95.5%
6	0	0	0	1	14	252	0	0	2	0	0	0	0	0	98.2%
7	0	0	0	0	0	0	259	0	0	0	0	0	0	0	96.5%
8	0	0	0	0	0	0	0	202	0	1	0	0	0	0	100%
9	0	0	0	0	0	3	0	0	311	0	0	0	0	0	81.0%
10	0	0	1	0	0	0	0	0	0	244	3	0	0	0	98.0%
11	0	0	0	0	0	0	0	0	0	4	301	0	0	0	98.0%
12	0	0	0	0	0	0	0	0	0	0	0	181	0	0	94.1%
13	0	0	1	0	1	0	0	0	0	0	0	0	266	0	99.1%
14	0	0	0	0	6	0	0	0	0	0	0	0	0	89	99.9%
PA	100%	99.0%	97.6%	99.1%	92.4%	95.8%	97.7%	100%	98.4%	98.0%	99.0%	100%	98.2%	100%	**98.0%**

Target Class

RESULTS

Indian Pines Botswana

Classification performances of the proposed approach with the change of training sample size

Training sample size	10%	20%	30%	40%	50%
Training time (min)	1.14±0.16	1.25±0.19	1.61±0.13	1.94±0.23	2.06±0.16
OA (%)	97.01±0.77	98.55±0.47	99.13±0.32	99.30±0.24	99.51±0.15
k	0.9676±0.0084	0.9843±0.0051	0.9906±0.0035	0.9924±0.0026	0.9947±0.0016
AA (%)	97.13±0.84	98.64±0.43	99.17±0.29	99.37±0.22	99.54±0.13

RESULTS

Indian Pines Botswana

Performances of feature extraction based different spectral and spectral-spatial classification approaches (Paul and Kumar 2018)

Approaches	Spectral classification		Spectral-spatial classification			
FE methods	**AE**	**SAE**	**PCA**	**AE**	**S-AE**	**S-SAE**
Classifier	RBF-SVM	RBF-SVM	RBF-SVM	RBF-SVM	RBF-SVM	RBF-SVM
OA (%)	91.77±0.71	91.72±0.95	96.42±0.81	97.15±0.72	97.28±0.74	**97.61±1.04**
k	0.9109±0.0077	0.9103±0.0103	0.9612±0.0088	0.9691±0.0078	0.9705±0.0080	**0.9741±0.0112**
AA (%)	92.43±0.74	92.41±0.88	96.74±0.60	97.42±0.58	97.17±0.80	**97.74±0.85**

CONCLUSIONS

- Information theory criteria, MI (non-parametric dependency measure) and entropy measure are used for selecting the representative HS bands and their corresponding EMPs are created to consider the spatial information in the spectral-spatial classification approach.

- The selected features are used in the RBF-SVM and RF classifiers, where parameters of these models are optimized using Bayesian optimization technique.

- RF classifier is performing better for the Indian Pines dataset, whereas RBF-SVM is performing better for the Botswana dataset.

- Comparing the results of different approaches and applying the statistical test, it is confirmed that this approach is providing statistically better classification performances for both the datasets.

REFERENCES

- Chang, Chein-I, Qian Du, Tzu-Lung Sun, and Mark LG Althouse. 1999. "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification." *IEEE Transactions on geoscience and remote sensing* 37 (6):2631-41.
- Feng, Jie, LC Jiao, Xiangrong Zhang, and Tao Sun. 2014. "Hyperspectral band selection based on trivariate mutual information and clonal selection." *IEEE Transactions on geoscience and remote sensing* 52 (7):4092-105.
- Guo, Baofeng, Steve R Gunn, Robert I Damper, and James DB Nelson. 2006. "Band selection for hyperspectral image classification using mutual information." *IEEE Geoscience and Remote Sensing Letters* 3 (4):522-6.
- Jia, Sen, Zhen Ji, Yuntao Qian, and Linlin Shen. 2012. "Unsupervised band selection for hyperspectral imagery classification without manual band removal." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (2):531-43.
- MartÍnez-UsÓMartinez-Uso, Adolfo, Filiberto Pla, José Martínez Sotoca, and Pedro García-Sevilla. 2007. "Clustering-based hyperspectral band selection using information measures." *IEEE Transactions on geoscience and remote sensing* 45 (12):4158-71.
- Paul, Subir, and D Nagesh Kumar. 2018. "Spectral-spatial classification of hyperspectral data with mutual information based segmented stacked autoencoder approach." *ISPRS Journal of Photogrammetry and Remote Sensing* 138:265-80.
- Santos, ACS, and H Pedrini. 2016. "A combination of k-means clustering and entropy filtering for band selection and classification in hyperspectral images." *International Journal of Remote Sensing* 37 (13):3005-20.