, Member, IEEE, , Senior Member, IEEE

***Abstract*—This paper addresses the problem of the classification of hyperspectral remote sensing images by support vector machines (SVMs). First, we propose a theoretical discussion and experimental analysis aimed at understanding and assessing the potentialities of SVM classifiers in hyperdimensional feature spaces. Then, we assess the effectiveness of SVMs with respect to conventional feature-reduction-based approaches and their performances in hypersubspaces of various dimensionalities. To sustain such an analysis, the performances of SVMs are compared with those of two other nonparametric classifiers (i.e., radial basis function neural networks and the K-nearest neighbor classifier). Finally, we study the potentially critical issue of applying binary SVMs to multiclass problems in hyperspectral data. In particular, four different multiclass strategies are analyzed and compared: the one-against-all, the one-against-one, and two hierarchical tree-based strategies. Different performance indicators have been used to support our experimental studies in a detailed and accurate way, i.e., the classification accuracy, the computational time, the stability to parameter setting, and the complexity of the multiclass architecture. The results obtained on a real Airborne Visible/Infrared Imaging Spectroradiometer hyperspectral dataset allow to conclude that, whatever the multiclass strategy adopted, SVMs are a valid and effective alternative to conventional pattern recognition approaches (feature-reduction procedures combined with a classification method) for the classification of hyperspectral remote sensing data.**

***Index Terms*—Classification, feature reduction, Hughes phenomenon, hyperspectral images, multiclass problems, remote sensing, support vector machines (SVMs).**

R

first-

second

A. SVM Mathematical Formulation

1) Linear SVM: Linearly Separable Case:

N d

$\mathbf{x_i} \in \Re^d \ (i = 1, 2, \ldots, N)$ $y_i \in \{-1, +1\}$

$\mathbf{x_i}$

$\mathbf{w} \in \Re^d$

$b \in \Re$

$[\mathrm{f}(\mathbf{x})]$ $\mathrm{f}(\mathbf{x})$

$$\mathrm{f}(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b.$$

$\mathbf{w}$ b

$$y_i(\mathbf{w} \cdot \mathbf{x_i} + b) > 0, \qquad i = 1, 2, \ldots, N.$$

$1/\|\mathbf{w}\|$

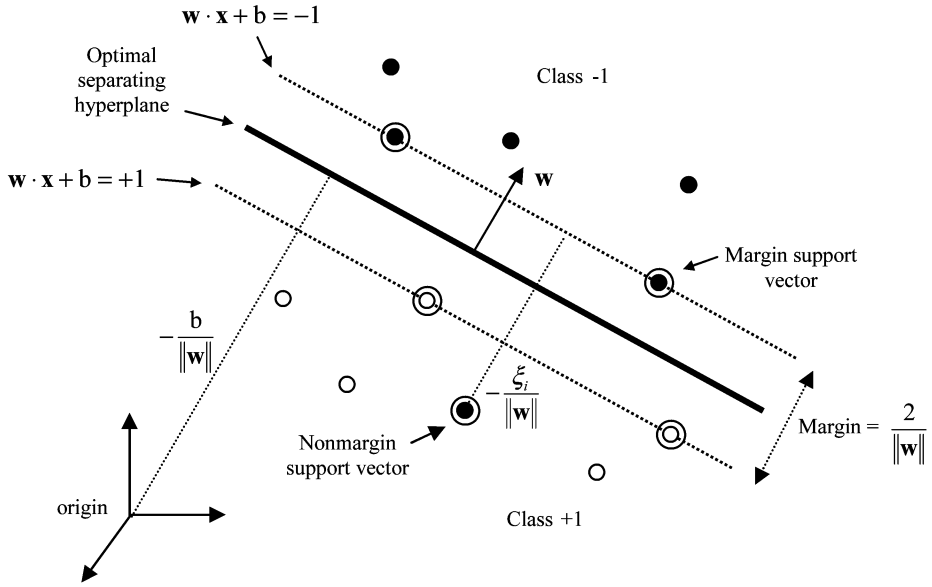
$\mathbf{w}$ b

$$\min_{i=1,2,\ldots,N} y_i(\mathbf{w} \cdot \mathbf{x_i} + b) \geq 1.$$

$2/\|\mathbf{w}\|$

$$\begin{cases} \frac{1}{2}\|\mathbf{w}\|^2 \\ y_i(\mathbf{w} \cdot \mathbf{x_i} + b) \geq 1, \qquad i = 1, 2, \ldots, N \end{cases}$$

$$\begin{cases} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (\mathbf{x_i} \cdot \mathbf{x}_j) \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \quad \alpha_i \geq 0, \qquad i = 1, 2, \ldots, N. \end{cases}$$



$+1$ -1

$\alpha_i \quad (i = 1,2,\ldots,N)$

$$y_i(\mathbf{w}\cdot\mathbf{x_i}+b)\geq 1-\xi_i, \qquad i=1,2,\ldots,N$$
$$\xi_i \geq 0, \qquad i=1,2,\ldots,N.$$

$$\mathrm{f}(\mathbf{x})=\sum_{i\in S}\alpha_i y_i(\mathbf{x_i}\cdot\mathbf{x})+b$$

S α_i

support vectors

$1/\|\mathbf{w}\|$

2) Linear SVM: Linearly Nonseparable Case:

$$\Psi(\mathbf{w},\xi)=\frac{1}{2}\|\mathbf{w}\|^2+\sum_{i=1}^{N}\xi_i$$

ξ_i *slack variables*

3) Nonlinear SVM: Kernel Method:

$\Phi(\cdot)$ $\Phi(\mathbf{x})\in\Re^{d'}$ $(d'>d)$

$\mathbf{w}\in\Re^{d'}$ $b\in\Re$

$(\mathbf{x}_i\cdot\mathbf{x}_j)$

$[\Phi(\mathbf{x_i})\cdot\Phi(\mathbf{x}_j)]$

$\Phi(\mathbf{x})$

$$K(\mathbf{x_i},\mathbf{x})=\Phi(\mathbf{x_i})\cdot\Phi(\mathbf{x}).$$

$[\Phi(\mathbf{x_i}) \cdot \Phi(\mathbf{x}_j)]$

$$\begin{cases} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x}_j) \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \qquad 0 \leq \alpha_i \leq \quad , \\ \qquad\qquad\qquad\quad i = 1, 2, \ldots, N. \end{cases}$$

$\mathrm{f}(\mathbf{x})$

$$\mathrm{f}(\mathbf{x}) = \sum_{i \in S} \alpha_\mathbf{i} y_i K(\mathbf{x_i}, \mathbf{x}) + b.$$

$$K(\mathbf{x_i}, \mathbf{x}) = \exp(-\gamma \|\mathbf{x_i} - \mathbf{x}\|^2)$$

γ

p

$$K(\mathbf{x_i}, \mathbf{x}) = [\mathbf{x_i} \cdot \mathbf{x} + \mathbf{1}]^p.$$

γ p

B. SVMs in Hyperspectral Feature Spaces

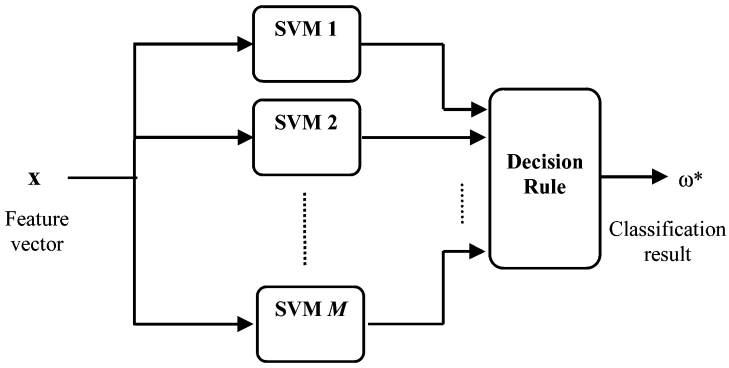
R_V R

$[-R, R]$ d

$$R_V = \frac{\pi^{d/2}}{d 2^{d-1} \Gamma(d/2)}$$

$\Gamma(\cdot)$

If you possess a limited amount of information to solve a problem, try solving it directly and never solve a more general problem as an intermediate step. The available information may be sufficient for a direct solution, though insufficient to solve a more general intermediate problem.



$\Omega = \{\omega_1, \omega_2, \ldots, \omega_T\}$ T d $\mathbf{X}$ d $\mathbf{x}$ Ω

two Ω_A Ω_B $\Omega_A \subset \Omega$ $\Omega_B \subset \Omega$ $+1$ -1 Ω_A Ω_B

A. Parallel Approach

1) One-Against-All Strategy: T two $\omega_i \in \Omega$

$$\begin{cases} \Omega_A = \omega_i \\ \Omega_B = \Omega - \omega_i. \end{cases}$$

2) One-Against-One Strategy:

M T M $T(T-1)/2$

$T(T-1)/2$ ω_i ω_j $(\omega_i \in \Omega, \omega_j \in \Omega, i \neq j)$ $\mathrm{f}_{ij}(\mathbf{x})$

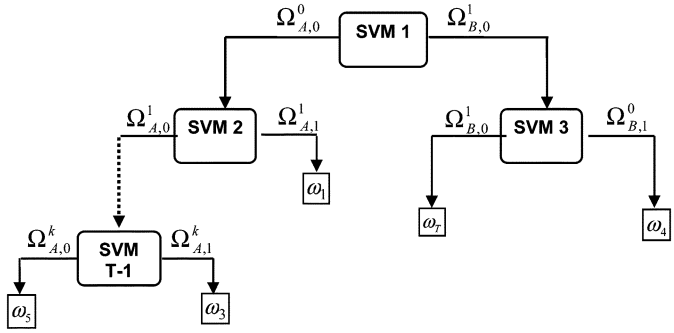
$$\begin{cases} \Omega_A = \omega_i \\ \Omega_B = \omega_j. \end{cases}$$

$\omega_i \in \Omega$ $S_i(\mathbf{x})$

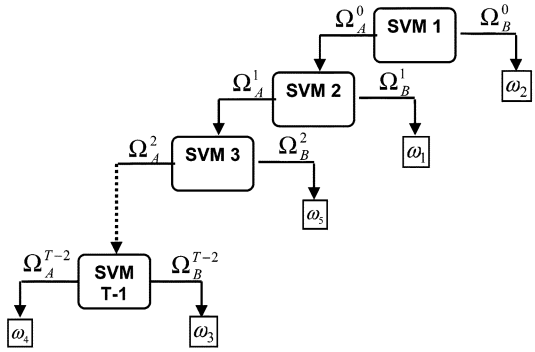
$$S_i(\mathbf{x}) = \sum_{\substack{j=1 \\ j \neq i}}^{T} \{\mathrm{f}_{ij}(\mathbf{x})\}.$$

$$\omega^* = \arg\max_{i=1,\ldots,T} \{S_i(\mathbf{x})\}.$$

B. Hierarchical Tree-Based Approach



(a)



(b)

1) BHT-Balanced Branches Strategy:

Ω_A Ω_B

T

Step 0: **Root Node**

$k = 0$

Ω $\Omega^k_{A,0}$ $\Omega^k_{B,0}$ $\sum_{\omega_i \in \Omega^k_{A,0}} P(\omega_i) \approx \sum_{\omega_j \in \Omega^k_{B,0}} P(\omega_j)$

*Step 1:***k-Level Branching**

$q = 0, \ldots, 2k-1 \; (q \geq 0)$

$\{\Omega^k_{A,q}\} \geq 2$ $\Omega^k_{A,q}$ $\Omega^{k+1}_{A,2q}$ $\Omega^{k+1}_{A,2q+1}$ $\sum_{\omega_i \in \Omega^{k+1}_{A,2q}} P(\omega_i) \approx \sum_{\omega_j \in \Omega^{k+1}_{A,2q+1}} P(\omega_j)$

$\{\Omega^k_{B,q}\} \geq 2$ $\Omega^k_{B,q}$ $\Omega^{k+1}_{B,2q}$ $\Omega^{k+1}_{B,2q+1}$ $\sum_{\omega_i \in \Omega^{k+1}_{B,2q}} P(\omega_i) \approx \sum_{\omega_j \in \Omega^{k+1}_{B,2q+1}} P(\omega_j)$

$k = k + 1$

Step 2: **Stop Condition**

$\exists \Omega^k_{A,q}$ $\Omega^k_{B,2q}$ $\{\Omega^k_{A,q}\} \geq 2$ $\{\Omega^k_{B,q}\} \geq 2$ $(q = 0, \ldots, 2^k - 1)$

Step 1.

2) BHT-One Against All Strategy:

Ω_A Ω_B Ω_B

$\Omega_A \cup \Omega_B$

T

CLASS	TRAINING	TEST
ω_1 – Corn-no till	742	692
ω_2 – Corn-min till	442	392
ω_3 - Grass/Pasture	260	237
ω_4 - Grass/Trees	389	358
ω_5 - Hay-windrowed	236	253
ω_6 - Soybean-no till	487	481
ω_7 - Soybean-min till	1245	1223
ω_8 - Soybean-clean till	305	309
ω_9 – Woods	651	643
Total	**4757**	**4588**

Step 0: **Root Node**

$k = 0$

Ω Ω^k_A Ω^k_B $P(\Omega^k_B)_{\Omega^k_B \in \Omega} = \max_{\omega_j \in \Omega}\{P(\omega_j)\}$ $\Omega^k_A = \Omega - \Omega^k_B$

Step 1: **k-Level Branching**

Ω^k_A Ω^{k+1}_A Ω^{k+1}_B $P(\Omega^{k+1}_B)_{\Omega^{k+1}_B \in \Omega^k_A} = \max_{\omega^k_j \in \Omega_A}\{P(\omega_j)\}$

$\Omega^{k+1}_A = \Omega^k_A - \Omega^{k+1}_B$

$k = k + 1$

Step 2: **Stop Condition**

$\{\Omega^k_A\} \geq 2$ *Step 1.*

T $T(T-1)/2$

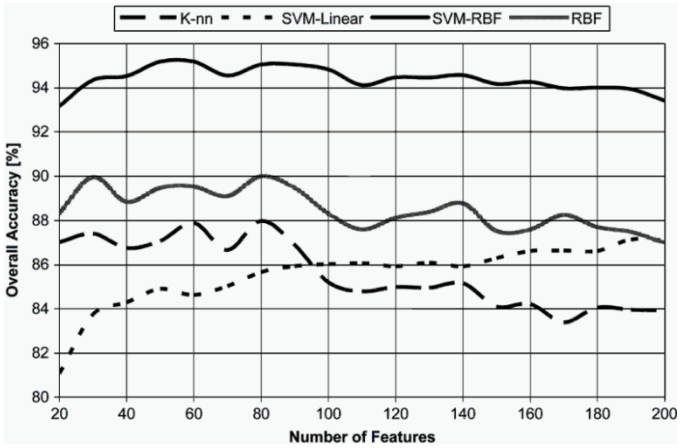
$T - 1$

METHOD	CLASSIFICATION ACCURACY [%]										TIME
	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	OA	[s]
SVM-Linear	89.02	69.13	94.51	98.60	100	75.47	83.48	83.17	99.22	**87.10**	**40342**
SVM-RBF	91.47	87.76	94.94	98.88	100	88.57	91.25	95.79	99.38	**93.42**	**2702**
K-nn classifier	96.73	61.16	86.59	80.46	99.60	98.88	90.72	65.82	74.42	**83.94**	**2618**
RBF classifier	98.44	74.11	88.47	79.83	99.21	98.04	91.98	73.72	80.06	**86.99**	**4743**

A. Dataset Description and Experiment Design

B. Results of Experiment 1: Classification in the Original Hyperdimensional Feature Space

METHOD	PARAMETER RANGE	OVERALL ACCURACY [%]		MEAN TOTAL TIME [s]
		Mean	Variance	
SVM-Linear	$C \in [1, 100]$	85.38	4.94	20785
SVM-RBF	$C \in [1, 100]$; $\gamma = 1$	92.64	0.84	1695
SVM-RBF	$\gamma \in [0.1, 3]$; $C = 40$	92.51	0.50	2412
K-nn classifier	$K \in [1, 25]$	82.42	1.56	2630
RBF classifier	N° clusters $\in [20, 200]$	85.59	1.12	1505



METHOD	OVERALL ACCURACY [%]	
	Mean	Variance
SA + SVM-Linear	85.56	2.04
SA + SVM-RBF	94.38	0.30
SA + K-nn classifier	85.60	2.27
SA + RBF classifier	88.49	0.81

C. Results of Experiment 2: Feature Reduction and Classification

METHOD	CLASSIFICATION ACCURACY [%]										DIFF-OA [%]
	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	OA	
SA + SVM-Linear	98.29	73.79	81.11	72.35	100	99.16	89.45	57.4	86.27	**83.74**	**-3.36**
SA + SVM-RBF	99.69	92.23	93.30	91.48	99.6	99.72	97.89	88.52	91.62	**94.35**	**0.93**
SA + K-nn classifier	80.93	71.43	94.52	99.44	99.61	87.53	88.06	71.20	96.42	**87.40**	**3.46**
SA + RBF classifier	83.24	77.30	93.25	97.77	98.03	86.08	90.03	77.67	98.29	**89.95**	**2.96**

+10.69%, +7.21% +5.82%

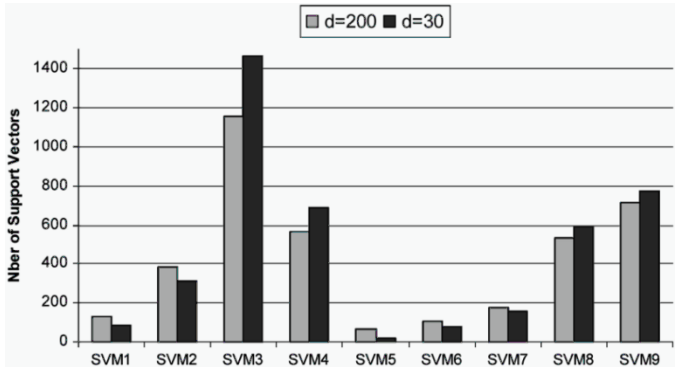
%

%

%

%

ω_3



$(d = 200)$

$(d = 30)$

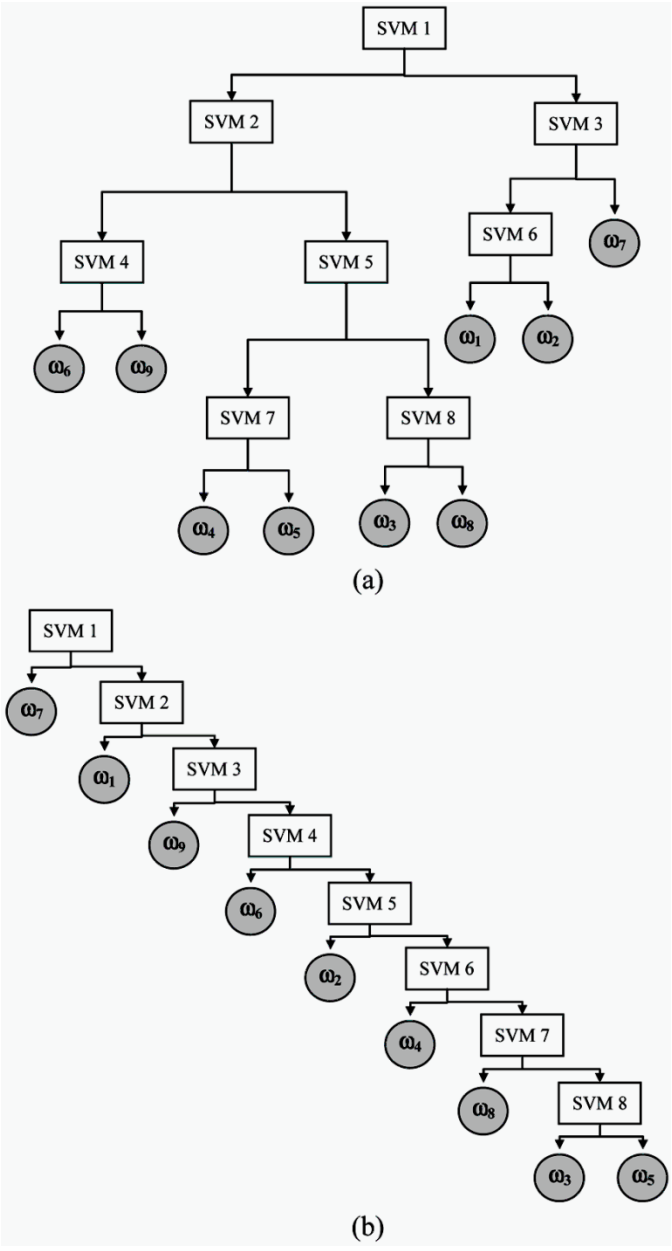
D. Results of Experiment 3: SVM and Multiclass Strategies

+2.72%, +1.72% +0.54%

two

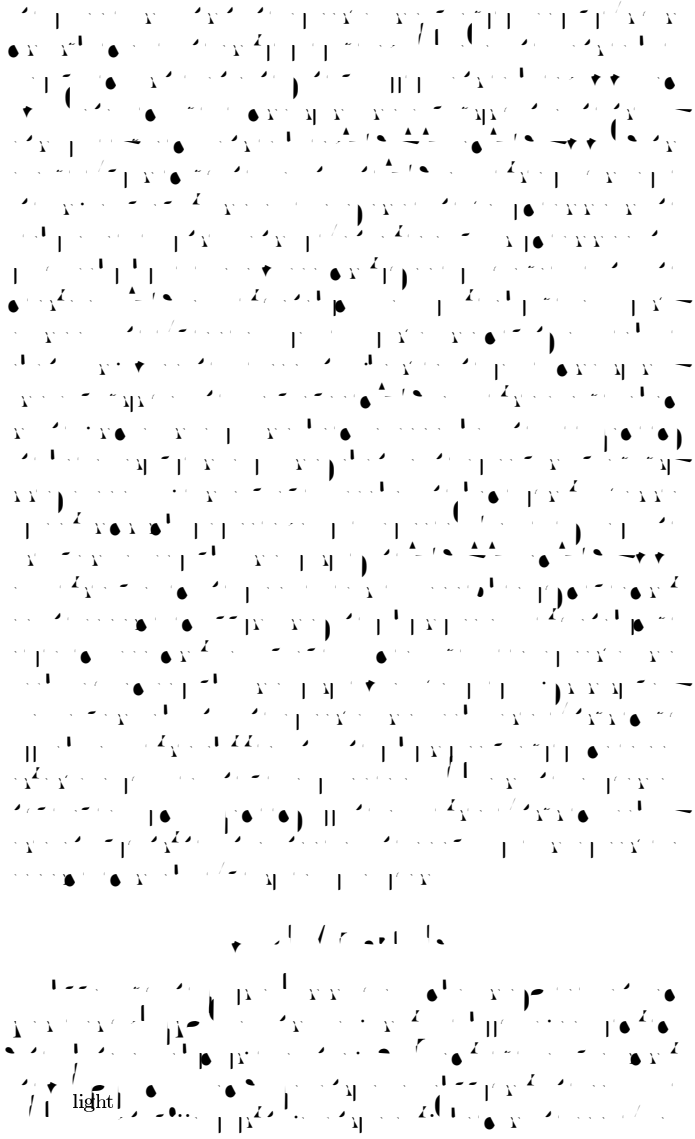
/ / %

Multiclass Strategy	CLASSIFICATION ACCURACY [%]									
	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	OA
OAA	91.47	87.76	94.94	98.88	100	88.57	91.25	95.79	99.38	**93.42**
OAO	90.32	89.54	94.51	99.72	100	88.36	93.54	94.82	99.38	**93.96**
BHT-BB	89.70	87.25	95.78	99.16	100	87.73	91.17	85.11	98.60	**92.24**
BHT-OAA	88.87	85.97	95.36	98.88	100	85.24	87.82	90.62	99.07	**91.24**



Multiclass Strategy	TIME [S]		NUMBER OF SVMS	number of Support Vectors		
	Train	Test		Min	Max	Average
OAA	2361	341	9	62	1159	**424**
OAO	212	554	36	9	569	**130**
BHT-BB	311	125	8	31	977	**334**
BHT-OAA	410	155	8	45	1270	**333**

Multiclass Strategy	OVERALL ACCURACY [%]							
	SVM1	SVM2	SVM3	SVM4	SVM5	SVM6	SVM7	SVM8
BHT-BB	95,66	99,25	94,1	100	99,57	98,62	100	98,53
BHT-OAA	93,77	97,74	99,7	99,16	98,9	99,57	99,12	100



light

IEEE Trans. Inform. Theory

IEEE Trans. Pattern Anal. Machine Intell.

IEEE Trans. Geosci. Remote. Sensing

IEEE Trans. Geosci. Remote. Sensing

IEEE Trans. Geosci. Remote. Sensing

J. R. Statist. Soc.

Signal Process. Mag.

Remote Sensing Digital Image Analysis

IEEE Trans. Geosci. Remote Sensing

Pattern Recognition and Signal Processing

Pattern Recognit. Lett.

IEEE Trans. Geosci. Remote Sensing

IEEE Trans. Pattern Anal. Machine Intell.

IEEE Trans. Geosci. Remote Sensing

IEEE Trans. Geosc. Remote. Sensing

Remote Sens. Environ.

IEEE Trans. Geosc. Remote. Sensing

IEEE Trans. Geosci. Remote Sensing

IEEE Trans. Geosci. Remote Sensing

Proc. IGARSS

Proc. SPIE

Int. J. Remote Sens.

Proc. SPIE

Summaries 8th JPL Airborne Earth Science Workshop

Proc. IGARSS

Proc. IGARSS

Statistical Learning Theory

Proc. 5th Annl. ACM Workshop Computationdl Learning Theory

Data Mining Knowl. Discov.

IEEE Trans. Med. Imag.

IEEE Trans. Neural Networks

IEEE Trans. Pattern Anal. Machine Intell.

IEEE Trans. Signal Processing

IEEE Trans. Neural Networks

Mach. Learn.

Neural. Comput.

IEEE Trans. Syst., Man, Cybern. C

A Course in the Geometry of n-Dimensions

Introduction to Statistical Pattern Recognition

Proc. Int. Conf. Pattern Recognition

Advances in Kernel Methods: Support Vector Learning

IEEE Trans. Geosci. Electron.

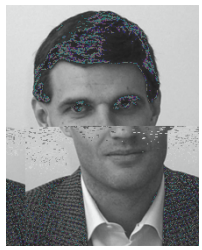
IEEE Trans. Geosci. Remote Sensing

Proc. 3rd Int. Workshop on Multiple Classifier Systems—MCS 2002

IEEE Trans. Geosci. Remote Sensing

IEEE Trans. Geosci Remote. Sensing

Farid Melgani



Lorenzo Bruzzone

Recognition of IEEE Transactions on Geoscience and Remote Sensing Best Reviewers