

Sistemas multi-redes para classificação de imagens multitemporais

Rafael de Oliveira Valle dos Santos¹
Raul Queiroz Feitosa^{2,3}
Marley Maria Bernardes Rebuszi Vellasco^{2,3}
Ricardo Tanscheit²

¹ TI / Pesquisa Operacional, PETROBRAS S.A., Brasil
rvsantos@petrobras.com.br

² Depto. de Engenharia Elétrica, PUC-Rio, Brasil

³ Depto. Engenharia de Sistemas e Computação, UERJ, Brasil
{raul,marley,ricardo}@ele.puc-rio.br

Abstract. This work investigates single and multiple multilayer perceptron (MLP) neural networks for the land-use-land-cover classification of single- and multidate remotely sensed data. The multiple neural network (MNN) schemes considered in this work were built upon individual neural networks trained with different data sets, as a way to assure diversity among the MNN members; their outputs were combined by a voting scheme that produces a single classification result. The analysis on multidate classification considers two types of earlier information with different discrimination ability: the previous spectral response and the previous classification of the image objects. Experiments conducted on a sequence of 5 LANDSAT-5 images of the county of Rio de Janeiro, separated by approximately 4 years, have shown that the multidate approaches performed considerably better than the corresponding single date schemes. Moreover, the MNN schemes presented throughout the experiments an improved performance in terms of classification accuracy as well as of stability in relation to the single neural network schemes. The correct classification rates achieved by the multidate MNNs were notably higher than for single neural networks working on single-date data.

Palavras-chave: imagens multitemporais, redes neurais, comitês de redes neurais

1. Introdução

A maioria dos métodos propostos até o momento para a classificação de imagens multitemporais de sensores remotos são abordagens do tipo “pós-classificação” (*post-classification*). Esses métodos consistem em realizar classificações separadas de imagens de uma mesma área geográfica obtidas em duas datas distintas e em seguida comparar os resultados obtidos (e.g. Weismiller et al. (1977)). As técnicas chamadas “classificação em cascata” (Swain (1978)) apresentam potencialmente maior desempenho, pois utilizam toda a informação contida na seqüência de imagens, e procuram explorar a correlação temporal entre elas.

Várias técnicas em cascata propostas até o momento tratam a classificação multitemporal como um problema de classificação de múltiplas fontes, onde a questão central está em como combinar conjuntos de dados multitemporais. De fato, muitos métodos de classificação multitemporal propostos até aqui são evoluções de pesquisas anteriores sobre fusão de imagens de múltiplos sensores (Jeon. e Landgrebe (1992), Solberg et al. (1996), Bruzzone e Prieto (2000), Bruzzone e Prieto (2002)).

A maioria destes métodos assume independência entre os conjuntos de dados multitemporais. Como consequência, estes métodos não capturam a correlação entre a informação temporal e espacial/espectral. Neste contexto, redes neurais aparecem como uma alternativa atraente para a fusão de informação multitemporal (Benediktsson et al. (1990)), uma vez que não requerem nenhum modelo explícito que represente a dependência temporal entre os conjuntos de dados, mas aprendem relações complexas a partir de exemplos.

Paralelamente à pesquisa em análise de imagens multitemporais, tem havido um crescente interesse por sistemas de múltiplos classificadores (SMC) desde a década passada (MCS2005 (2005), Kittler et al. (1998), Xu et al. (1992)). Evidências teóricas e experimentais (Kittler et al. (1997)) indicam que SMCs são potencialmente mais exatos do que classificadores individuais. Isto se verifica especialmente quando os classificadores que compõem o sistema são exatos (melhor do que sorteio aleatório) e complementares, no sentido de que seus erros ocorrem em regiões distintas do espaço de atributos (Opitz e Maclin (1999)).

Muitas publicações descrevem sistemas de classificação que usam múltiplas redes neurais (MRN) (Ueda (2000), Opitz e Maclin (1999), Liu e Yao (2000), Drucker (1999), Cho e Kim (1995), Breiman (1999)). Sistemas MRN procuram aumentar o desempenho da classificação, ou pelo menos reduzir a variância inerente ao processo de treinamento, combinando redes neurais que são potencialmente complementares (Sharkey (1999), Witten e Frank (2005)).

Os benefícios de combinar classificadores especificamente para a interpretação de imagens de sensores remotos têm sido constatados em vários estudos (p. ex. Giacinto et al. (2000)), mas a pesquisa sobre SMCs para a classificação de imagens multitemporais é ainda incipiente. Entre estes merecem menção os estudos descritos em (Melgani e Serpico (2003)), (Bruzzone et al. (2002), Bruzzone e Cossu (2002)) e (Bruzzone et al. (2004)). Os resultados experimentais relatados nestes trabalhos indicam a superioridade dos SMC relativamente aos métodos baseados em classificadores individuais. Contudo, nenhum deles explora o potencial das redes neurais para a fusão de dados multitemporais. Da mesma forma, a pesquisa sobre SMCs baseados em múltiplas redes neurais (MRN) tem sido escassa.

O presente trabalho aborda estas duas questões. Primeiramente se investiga o uso de redes neurais MLP (*multilayer perceptrons*) para a fusão de dados multitemporais. Em seguida, o desempenho de múltiplas redes neurais (MRN) na classificação em cascata de imagens multitemporais é avaliado experimentalmente. Neste estudo cada rede neural que compõe o MRN é treinada com conjuntos de treinamento distintos visando obter a complementaridade entre os membros do sistema; as diversas saídas são combinadas através de um esquema de votação (Witten e Frank (2005)), que produz o resultado final da classificação.

O texto a seguir está organizado da seguinte maneira: a seção 2 apresenta os métodos de classificação investigados neste trabalho; a seção 3 descreve os experimentos e discute seus resultados; e a seção 4 conclui o texto resumindo as principais conclusões deste trabalho.

2. Esquemas de classificação

2.1 Fusão de dados multitemporais com redes neurais

A maneira mais fácil de se obter fusão de dados multitemporais é a abordagem conhecida como vetores empilhados (*stacked vectors*) (Leckei (1990)). Nesta técnica, os objetos da imagem são representados por vetores “estendidos”, obtidos pelo empilhamento de vetores de atributos do mesmo objeto em datas diferentes (Bruzzone et al. (1999), Bruzzone et al. (2004)).

Usualmente os vetores empilhados não podem ser modelados de maneira apropriada através de uma distribuição de probabilidade analiticamente simples, o que prejudica o tratamento estatístico convencional. Redes neurais se apresentam como opções atraentes nesses casos, na medida em que não dependem de modelagem paramétrica do espaço de entrada, mas “aprendem” tais modelos por meio de exemplos. Uma dificuldade nesta abordagem é que o aumento da dimensão do espaço de atributos redundante também num aumento da demanda por padrões de treinamento.

2.2 Comitês de redes neurais

A Figura 1a mostra um classificador formado por uma rede neural individual (RNI). Ela mapeia vetores de entrada $\mathbf{x}=[x_1, \dots, x_p]$, compostos por valores de p atributos que descrevem o objeto a ser classificado, num vetor de valores reais $\mathbf{v}=[v_1, \dots, v_n]$ de dimensão igual ao número (n) de classes do problema. A saída $\mathbf{w}=[w_1, \dots, w_n]$ do classificador é dada por uma função \mathbf{H} definida em \mathbb{R}^n ($\mathbf{w}=\mathbf{H}(\mathbf{v})$) tal que $w_i=1$ se $v_i=\max\{v_1, \dots, v_n\}$, com todos os demais componentes w_j ($j \neq i$) iguais a zero. Se o vetor \mathbf{v} tem mais de um componente igual a $v_i=\max\{v_1, \dots, v_n\}$, então o classificador não pode dizer nada a respeito do objeto e retorna um rótulo de “rejeição”.

Um comitê de redes neurais é um arranjo de redes independentes, trabalhando em paralelo, mas no sentido de uma classificação única e consensual. De maneira geral, a resposta final é o resultado de algum tipo de procedimento de votação envolvendo os classificadores membros do sistema, como ilustrado na Figura 1b.

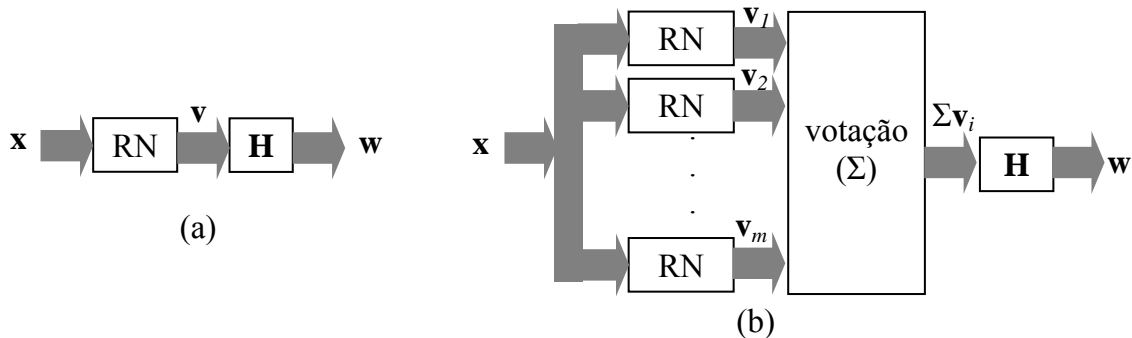


Figura 1: (a) rede neural individual (b) comitê de redes neurais.

Neste trabalho, os comitês são construídos em três passos:

- Seleciona-se uma arquitetura básica, completa e independente para as RN que compõem o comitê;
- m RNs são treinadas com conjuntos distintos gerados via *bootstrap* - um conjunto gerado via *bootstrap* possui o mesmo número de padrões que o conjunto original, gerado a partir deste último por meio de sorteio com reposição (Efron e Tibshirani (1993), Witten e Frank (2005)). As m arquiteturas assim treinadas compõem os *membros* do comitê.
- Todos os membros são usados para classificação final consensual do objeto de entrada, através de um esquema de votação *por pluralidade* (a classe mais “votada” “ganha” (Hansen e Salamon (1990)). Esse terceiro passo é comumente referido na literatura como *bagging* (Witten e Frank (2005)).

Em termos formais, se \mathbf{v}_k denota a resposta do k -ésimo membro do comitê, então, a resposta consensual do comitê é dada por

$$\mathbf{w} = [w_1, \dots, w_n] = \mathbf{H} \left(\sum_{k=1}^m \mathbf{v}_k \right),$$

3 Experimentos

3.1 Região de Teste

Os experimentos basearam-se em um conjunto de cinco imagens multiespectrais, adquiridas pelo satélite Landsat 5, sensor TM (*thematic mapper*). A área de teste corresponde ao

município do Rio de Janeiro, região sudeste do Brasil, coberto pela cena 217_76 (Figura 2). As imagens foram adquiridas nos dias 30 de maio de 1984, 18 de janeiro de 1988, 18 de abril de 1992, 16 de junho de 1996 e 12 de agosto de 1999.



Figura 2: A imagem LANDSAT e a visão aérea da área de teste

Todas as cinco imagens foram co-registradas e classificadas visualmente por um foto-intérprete experiente. Uma imagem SPOT-5 e dados de SIG foram utilizados como subsídio. Essa classificação visual serviu como referência ou “padrão ouro” na análise de desempenho que se seguiu. Consiste de um mapa temático para cada data definindo a classe de cada pixel.

Nos experimentos realizados, os objetos da imagem a serem classificados foram segmentos multi-pixel e não pixels individuais. A resposta espectral média (REM) do segmento nas bandas 3, 4 e 5 foi usada para formar o vetor de atributos de cada objeto. Os segmentos foram extraídos a partir das classificações de referência, como descrito a seguir.

As classificações de cada ano foram empilhadas de modo que cada pixel foi representado por um vetor composto de 5 rótulos, cada um relativo à classe a que pertence o pixel em cada data. Os segmentos foram definidos por componentes conectados contendo pixels com o mesmo vetor de rótulos. Para reduzir o efeito de possíveis imperfeições no registro das imagens, segmentos muito pequenos e muito estreitos e alongados foram descartados.

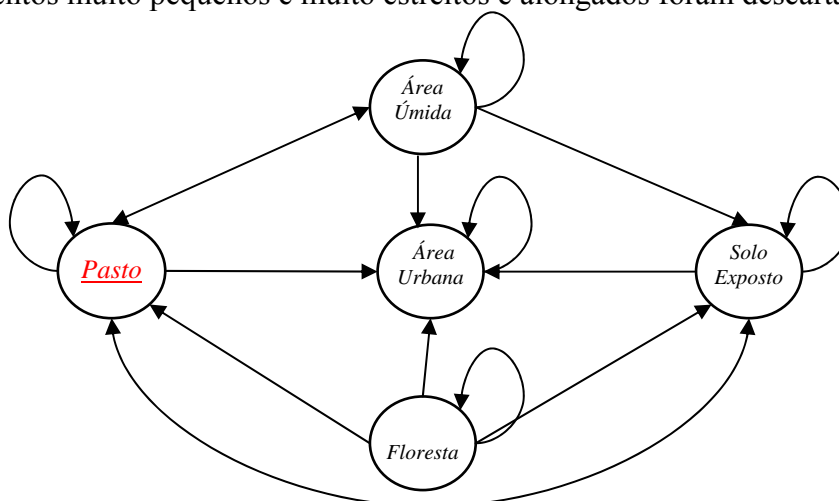


Figura 3: Diagrama de transição de classes na região de teste.

área de teste foi caracterizada inicialmente por sete classes de cobertura do solo (*rocha, corpos d’água, área úmida, urbana, pasto, floresta e solo exposto*). As classes *rocha* e *corpos d’água* são estáveis no banco de dados utilizado e, para evitar que a classificação multitemporal fosse superestimada em nossos experimentos, estas classes foram excluídas da análise. O conhecimento prévio relativo à dinâmica destas classes está representado no

diagrama de transição de classes da Figura 3, em que os nós representam classes e os arcos as transições possíveis de classes num período de aproximadamente 4 anos.

3.2 Arquiteturas Investigadas

Três arquiteturas básicas de redes MLP foram usadas, conforme mostra a Tabela 1, onde os dígitos representam, pela ordem, o número de processadores nas camadas de entrada, oculta e de saída da rede neural, respectivamente.

Tabela 1: Arquitetura das redes neurais

#Arquitetura	Processadores por Camada
1	3-3-5
2	6-6-5
3	8-8-5

Tabela 2: Número de segmentos por ano, por classe e por partição.

CLASSE	ANO							
	1988		1992		1996		1999	
	treino	teste	treino	teste	treino	teste	treino	teste
<i>área úmida</i>	102	52	84	42	366	186	656	330
<i>urbana</i>	104	52	85	45	365	185	655	330
<i>pasto</i>	104	52	84	44	368	188	657	330
<i>floresta</i>	102	51	84	42	365	185	655	330
<i>solo exposto</i>	104	52	84	44	366	186	657	330

O número de processadores nas camadas de entrada foi determinado pela dimensão do vetor de entrada, formado respectivamente pela resposta espectral média (REM) do objeto nas 3 bandas em uma data (arquitetura #1), pela REM do objeto em 3 bandas em duas datas consecutivas (arquitetura #2) e pela REM em 3 bandas numa data, concatenado com o vetor de rótulos relativo à classificação numa data anterior (arquitetura #3). A camada oculta contém em todas as arquiteturas o mesmo número de processadores que a correspondente camada de entrada. Todas as três arquiteturas consideram um vetor de saída binário, com uma componente para cada uma das cinco classes, conforme descrito na seção 2.2.

Foram construídos comitês de redes neurais para todas as três arquiteturas, considerando-se até 101 redes no sistema. A Tabela 2 apresenta o número de padrões considerados nos experimentos, por ano, por classe e por partição (treinamento e teste). Alguns padrões foram duplicados de modo a manter para cada classe aproximadamente a mesma proporção de padrões, tanto na partição de teste como na de treinamento, e o mesmo número de exemplos de transições.

Em todos os casos, o algoritmo básico de treinamento foi o *gradiente conjugado escalável* (*scaled conjugated gradient*), uma versão do *back-propagation* descrita em (Moller (1993)). Cabe mencionar que, no caso da arquitetura #3, na qual se utiliza como dado de entrada a classificação de um objeto na data imediatamente anterior, só foram considerados os resultados de classificação produzidos consistentes com o diagrama de transição de classes da figura 3.

3.3 Resultados

A Tabela 3 mostra, para cada arquitetura e para os dados de cada ano, as taxas médias de classificação (em negrito), e os respectivos desvios padrões (em itálico, entre parênteses), sempre medidos sobre 10 rodadas de experimentos.

As linhas correspondentes aos esquemas RNI mostram que ambos os esquemas multitemporais analisados (arquiteturas #2 e #3) têm desempenho superior à arquitetura monotemporal (arquitetura #1). O melhor resultado ocorre para a imagem de 1988, em que o

uso da classificação anterior (arquitetura #3) trouxe um aumento de aproximadamente 18% em relação ao esquema monotemporal (arquitetura #1).

Os resultados apresentados na Tabela 3 confirmam que a informação relativa à classificação anterior (arquitetura #2) foi, em todos os casos, mais valiosa que a REM anterior (arquitetura #3). Isto se justifica pelo fato de que a classificação anterior foi produzida levando em conta outras fontes de informação além da REM. Isto incluiu a experiência do fotointérprete e seu conhecimento específico sobre a área de teste. Obviamente, este conhecimento não está inteiramente codificados na REM. Vale notar que uma análise dos resultados ao longo das linhas referentes aos esquemas MRN levam às mesmas conclusões.

Tabela 3: Taxas médias de sucesso na classificação

	ANO	Monotemporal	Multitemporal	
		Arquitetura # 1	Arquitetura # 2 (usando REM de 4 anos antes)	Arquitetura # 3 (usando Classificação de 4 anos antes)
Rede Neural Individual (RNI)	1988	45.64 (5.78)	52.01 (9.68)	63.51 (8.78)
	1992	50.69 (7.31)	55.58 (6.52)	66.08 (6.35)
	1996	50.08 (4.56)	50.32 (10.61)	53.60 (5.90)
	1999	49.38 (6.49)	57.22 (8.49)	60.23 (5.63)
Múltiplas Redes Neurais (MRN) 101 redes	1988	49.27 (5.14)	57.41 (0.41)	72.39 (2.19)
	1992	64.75 (2.65)	62.26 (2.16)	73.00 (1.13)
	1996	55.12 (1.04)	56.63 (2.05)	69.47 (1.40)
	1999	53.71 (1.01)	62.04 (2.71)	67.21 (3.84)

Novos experimentos foram realizados para verificar a influência do lapso de tempo entre as imagens multitemporais no desempenho de classificação. Intuitivamente, quanto mais recente é a informação passada, mais relevante será sua contribuição para a classificação presente. Para avaliar essa questão, foram realizados experimentos com MRN para classificar a imagem de 1999, desta vez utilizando como informação passada, separadamente, a classificação anterior de 1992, 1988 e 1984, além da de 1996. Os resultados obtidos estão expressos na **Erro! A origem da referência não foi encontrada.**

Tabela 4: Taxas médias de classificação para 1999

ano anterior	Arquitetura # 3 (MRN, 101 redes)
1984	53.63 (1.93)
1988	52.02 (0.22)
1992	59.01 (4.19)
1996	67.21 (3.84)

Como já era esperado, notou-se que o aumento de desempenho alcançado ao se incluir informação de uma data anterior à da imagem a ser classificada não depende apenas dos atributos escolhidos para representar o estado do objeto no passado, mas também do lapso de tempo que separa as duas datas. Confirmando as expectativas, o desempenho de classificação tende a cair à medida que se recorre à informação de um passado mais remoto. É razoável, portanto, supor que resultados ainda melhores dos que os apresentados nas duas colunas mais à direita da Tabela 3 seriam alcançados, se a informação utilizada nestes esquemas fosse de um passado mais recente.

Um outro aspecto importante revelado pela Tabela 3 é a superioridade dos esquemas MRN (múltiplas redes neurais) sobre os esquemas RNI (rede neural individual) em todos os anos. O aumento na taxa de classificação fica entre 4% e 16%, sendo em média 7%. Os

resultados confirmam também a menor variância nas taxas de classificação dos sistemas MRN medidas ao longo de 10 rodadas de cada experimento. A variância menor indica maior estabilidade em relação aos fatores aleatórios do processo de aprendizado das redes.

Ressalte-se como comentário final que os sistemas MRN multitemporais tiveram desempenho amplamente superior aos sistemas monotemporais RNI. A Tabela 3 apresenta um ganho de desempenho entre 18% e 27% para cada ano.

4 Conclusões

Este trabalho investigou o uso de redes neurais para fusão de dados multitemporais, bem como de comitês de redes neurais para a classificação de imagens multitemporais.

Experimentos realizados a partir de cinco imagens LANDSAT-5 do município do Rio de Janeiro, de datas separadas por aproximadamente 4 anos, demonstraram que redes neurais foram eficazes na fusão de dados multitemporais. Comitês de redes neurais apresentaram, na maioria dos casos analisados, desempenho significativamente superior ao de redes neurais individuais. Além disso, os comitês demonstraram ser muito mais robustos contra os aspectos aleatórios intrínsecos ao processo de treinamento das redes neurais.

Em relação a redes individuais monotemporais, os comitês multitemporais apresentaram um ganho de desempenho de 18% a 27% usando informação de 3 a 4 anos no passado. Os resultados experimentais ressaltaram ainda que o desempenho dos esquemas multitemporais depende decisivamente do poder de discriminação da informação passada. Neste sentido, quanto mais recente for a informação passada, maior será o desempenho da classificação multitemporal.

Agradecimento

Os autores agradecem o apoio financeiro da FAPERJ para a realização deste trabalho.

Referências

6th International Workshop on Multiple Classifier Systems (MCS2005). Disponível em: <<http://web.dice.unica.it/mcs/>>. Acesso em: nov. 2005.

Benediktsson, J. A., Swain, P.H. and Ersoy, O. K.. **Neural network approaches versus statistical methods in classification of multisource remote sensing data.** IEEE Trans. Geosci. Remote Sensing, vol. 28, No. 4, July 1990, pp. 540-552.

Breiman, L. **Combining Predictors**, in **Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems.** Perspectives in Neural Computing, Springer Verlag, 1999.

Bruzzone, L., Cossu, R., Vernazza, G. **Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images.** Information Fusion 3 (2002) 289–297.

Bruzzone, L., Prieto, D. F. and Serpico, S. B. **A Neural-Statistical Approach to Multitemporal and Multisource Remote-Sensing Image Classification.** IEEE Trans. Geosci. Remote Sensing, Vol. 37, No. 3, pp. 1350-1359, May 1999.

Bruzzone, L. and Prieto, D. F. **Automatic analysis of the difference image for unsupervised change detection.** IEEE Trans. Geosci. Remote Sensing, vol. 38, pp. 1171–1182, May 2000.

Bruzzone, L. and Prieto, D. F. **An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images,** IEEE Trans. Image Processing, vol. 11, pp. 452–466, Apr. 2002.

- Bruzzone, L., Cossu, R. **A Multiple-Cascade-Classifer System for a Robust and Partially Unsupervised Updating of Land-Cover Maps**, IEEE Trans. Geosci. Remote Sensing, V. 40, No. 9, pp. 1984-1996, Sept. 2002.
- Bruzzone, L., Cossu, R., Vernazza, G. **Detection of land-cover transitions by combining multirate classifiers**. Pattern Recognition Letters 25 (2004) 1491–1500, pp. 1491-1500.
- Cho, S., Kim, J.H. **Combining Multiple Neural Networks by Fuzzy Integral for Robust Classification**. IEEE Transactions on Systems, Man and Cybernetics, Vol.25, No.2, 1995.
- Drucker, H. **Boosting Using Neural Networks**, in **Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems**. Perspectives in Neural Computing, Springer Verlag, 1999.
- Efron, B., Tibshirani, R. **An Introduction to the bootstrap**. Chapman & Hall, 1993.
- Giacinto, G., Roli, F., Bruzzone, L. **Combination of neural and statistical algorithms for supervised classification of remote-sensing images**. Pattern Recognit. Lett., Vol. 21, Issue 5, May 2000, Pages 385-397.
- Jeon, B. and Landgrebe, D. A. **Classification with spatio-temporal interpixel class dependency contexts**. IEEE Trans. Geosci. Remote Sensing, vol. 30, pp. 663–672, July 1992.
- Kittler, J., Hatef, M., Duin, R.P.W., Matas, J. **On Combining Classifiers**. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.20, No.3, 1998.
- Kittler, J., Hojjatoleslami, A. and Windeatt, T., **Strategies for combining classifiers employing shared and distinct pattern representations**, Pattern Recognit. Letters., vol. 18, pp. 1373–1377, 1997.
- Leckei, D.G. **Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination**. Photogramm. Eng. Remote Sens., vol. 56, pp. 1237–1246, 1990.
- Liu, Y., Yao, X. **Evolutionary Ensembles with Negative Correlation Learning**. IEEE Transactions on Evolutionary Computation, 4(4):380-387, November 2000.
- Melgani, F., Serpico, S.B. **A Markov Random Field Approach to Spatio-Temporal Contextual Image Classification**. IEEE Trans. Geosci. Remote Sensing, Vol. 41, No. 11, pp. 2478-2487, Nov 2003.
- Moller, M. F. **A scaled conjugate gradient algorithm for fast supervised learning**. Neural Networks, vol. 6, pp. 525-533, 1993.
- Sharkey, A.J.C (Ed.). **Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems**. Perspectives in Neural Computing, Springer Verlag, 1999.
- Solberg, A. H. S., Taxt, T. and Jain, A. K. **A Markov random field model for classification of multisource satellite imagery**, IEEE Trans. Geosci. Remote Sensing, vol. 43, pp. 100–113, Jan. 1996.
- Swain, P.H. **Bayesian classification in time-varying environment**. IEEE Trans. Syst. Man Cybern., vol. 8, pp. 880–883, 1978.
- Ueda, N. **Optimal Linear Combination of Neural Networks for Improving Classification Performance**. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.2, 2000.
- Weismiller, R. A., Kristoof, S. J., Scholz, D. K., Anuta, P. E., Momen, S. A., 1977. Change Detection in Coastal Zone Environments. Photogrammetric Engineering and Remote Sensing, 43, pp. 1533-1539.
- Witten, I. H., Frank, E. Data Mining – Practical machine learning tools and techniques. 2nd Edition, Elsevier / Morgan Kaufman, 2005.
- Xu, L., Krzyzak, A., Suen, C.Y. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. IEEE Transactions on Systems, Man and Cybernetics, Vol.22, No.3, 1992.