

Classificação de Imagens Tomográficas de Ciência dos Solos Utilizando Redes Neurais e Combinação de Classificadores

Fabricio A. Breve, Nelson D. A. Mascarenhas

*Departamento de Computação – Universidade Federal de São Carlos, São Paulo, SP, Brasil
{fabricio, nelson}@dc.ufscar.br*

Abstract

Pattern Recognition is a subject being used in a multidisciplinary scope, with different approaches. One of them is its application in computerized tomography images, commonly acquired in order to do medical diagnosis, but they have been used in several other applications as well, including Soil Science. The objective of this work is to study and to discuss the performance of neural network-based classifiers (Multilayer Perceptron and Radial Basis Functions) and classifier combiners (Bagging, Decision Templates and Dempster-Shafer) applied to identify materials in Soil Science multispectral images, acquired using Computerized Tomography. The results were evaluated by error estimation by Hold-Out and the Kappa coefficient.

1. Introdução

Reconhecimento de Padrões é um processo de descobrimento, de identificação e entendimento de padrões que sejam relevantes à realização de uma tarefa que pode ser baseada em imagens [1]. As técnicas de Reconhecimento de Padrões podem ser aplicadas às mais diversas áreas e aos mais diversos tipos de imagens e sinais, sendo que a Tomografia Computadorizada é uma das áreas onde elas vêm sendo amplamente utilizadas.

A EMBRAPA Instrumentação Agropecuária da cidade de São Carlos projetou e construiu um minitomógrafo para obtenção de imagens de Ciências de Solos. Com este equipamento, imagens podem ser adquiridas com o uso de diferentes fontes de energia: raios gama provenientes de fontes radioativas como Césio (137Cs) e Amerício (241Am) ou raios-X com múltiplas energias. Alguns trabalhos foram desenvolvidos aplicando-se técnicas de classificação multiespectral com o intuito de identificar materiais em imagens obtidas através de testes não destrutivos com tecnologia do minitomógrafo, utilizando raios- Γ e raios-X com múltiplas energias [2][3][4][5]. Com as crescentes pesquisas em técnicas de

reconhecimento de padrões baseadas em redes neurais, redes de base radial e sistemas de combinação de classificadores, abre-se espaço para propor e apresentar novos algoritmos para classificação das imagens e para combinação de classificadores.

Dentro deste contexto as Redes Neurais surgem como uma extensão das muitas técnicas convencionais que vêm sendo desenvolvidas por décadas. Em determinados tipos de problema as Redes Neurais apresentam muitas vantagens com relação aos classificadores tradicionais. Neste trabalho são estudados e aplicados os classificadores Perceptron de Múltiplas Camadas e Redes de Base Radial, ambos desenvolvidos a partir do estudo das redes neurais.

Combinação de Classificadores é uma área muito dinâmica e que vem crescendo de forma extremamente rápida. Portanto se torna bastante interessante aplicar alguns de seus métodos para avaliar seus possíveis benefícios. Neste trabalho são utilizadas as técnicas de Bagging e os combinadores Decision Templates e Dempster-Shafer.

O objetivo deste trabalho é apresentar métodos para identificação de materiais em imagens tomográficas de solo utilizando técnicas de reconhecimento de padrões baseadas em redes neurais, dando continuidade aos trabalhos anteriores [2][3][4][5]. Ao classificar amostras de solo e verificar o desempenho de diversos algoritmos de classificação e combinação, esta dissertação deverá contribuir para as pesquisas e trabalhos nas áreas de Ciência dos Solos, Reconhecimento de Padrões e Redes Neurais.

2. Classificação

Classificação de Imagens é o tema principal deste trabalho. Neste capítulo são apresentados os classificadores baseados em redes neurais que foram utilizados.

2.1 Classificadores e Redes Neurais

Uma rede neural artificial é uma máquina projetada para modelar a maneira como o cérebro realiza uma determinada tarefa; é implementada por componentes eletrônicos ou simulada por programação em um computador digital. Para alcançar um bom resultado as redes neurais utilizam diversas células computacionais simples (denominadas *neurônios*) e altamente interligadas. Seu objetivo é armazenar conhecimento experimental e torná-lo disponível para uso. O conhecimento é adquirido pela rede através de um processo de aprendizagem. O conhecimento é armazenado nos neurônios em unidades denominadas *pesos sinápticos*. O processo de aprendizagem consiste em modificar os pesos sinápticos da rede de uma forma ordenada para alcançar o objetivo desejado.

A combinação de diversos neurônios artificiais forma a chamada *rede neural*. A saída de cada neurônio pode ser conectada à(s) entrada(s) de outro(s) neurônio(s) formando assim uma rede. Cada neurônio pode então ser definido como um “nó” dessa rede. A formação da rede pode assumir muitas estruturas diferentes, das quais se destacam duas de particular interesse na tarefa de reconhecimento de padrões que serão descritas a seguir.

2.1.1 Perceptron de Múltiplas Camadas

O *Perceptron de Múltiplas Camadas* é um tipo de rede que consiste de um conjunto de unidades sensoriais, que constituem a *camada de entrada*, uma ou mais *camadas ocultas* e uma *camada de saída* de nós computacionais. O sinal que é colocado nas entradas se propaga para frente camada por camada até chegar à camada de saída.

Este tipo de rede vem sendo utilizado com sucesso para resolver problemas difíceis através de seu treinamento utilizando o *algoritmo de retropropagação de erro*, que consiste basicamente em dois passos: um passo para frente em que o sinal aplicado à entrada vai se propagando pelos nós computacionais da rede até chegar aos nós de saída; e um passo para trás, no qual todos os pesos sinápticos são ajustados de acordo com uma regra de correção de erro [6][7][8].

2.1.2 Redes de Função de Base Radial

Diferentemente das redes neurais do tipo Perceptron (que utiliza uma técnica recursiva conhecida na estatística como aproximação estocástica), as Redes de Função de Base Radial

utilizam uma abordagem que vê a rede neural como um problema de ajuste de curva em um espaço de alta dimensionalidade. Neste contexto, aprender equivale a encontrar uma superfície num espaço multidimensional que forneça o melhor ajuste para os dados de treinamento do ponto de vista estatístico.

A construção de uma rede de função de base radial envolve três camadas com papéis totalmente diferentes. A primeira camada é constituída de nós sensoriais que recebem os dados do ambiente. A segunda camada, que é a única camada oculta da rede, faz uma transformação não-linear do espaço de entrada para um espaço oculto, normalmente de alta dimensionalidade. Finalmente a camada de saída é linear e fornece a resposta da rede ao sinal de entrada [6].

2.2 Avaliação de Desempenho dos Classificadores

Através de métodos de avaliação podemos obter números que indicam qual foi o desempenho dos classificadores que treinamos. Neste capítulo são descritos os métodos de avaliação utilizados.

2.2.1 Método Hold-Out

Neste método, tradicionalmente se separa o conjunto de dados disponíveis (amostras pré-rotuladas) em duas metades. Usa-se uma para treinar o classificador e a outra para testá-lo, obtendo-se a taxa de erro. Separações em outras proporções também são utilizadas. Também é possível inverter os conjuntos, obter outra taxa de erro e então tirar uma média dos dois [7][9].

2.2.2 Coeficiente Kappa

O *Coeficiente Kappa* pode ser utilizado para medir o grau de concordância entre os rótulos das amostras de teste e as saídas do classificador.

Para calcular o Coeficiente Kappa podemos utilizar a seguinte fórmula [10]:

$$K = \frac{P(A) - P(E)}{1 - P(E)}, \quad (2.1)$$

onde $P(A)$ é a proporção de vezes em que o classificador acerta e $P(E)$ é a proporção de vezes que se espera que um classificador acertasse se as classificações fossem feitas ao acaso. Por exemplo, se temos quatro classes equiprováveis o valor de $P(E)$ será 0,25. Os valores de K podem variar entre -1 e 1. Quando não há concordância alguma além da

esperada com a classificação sendo feita ao acaso K será igual a zero. E quando há uma concordância total entre os rótulos pré-estabelecidos e a saída do classificador a saída será um. Valores menores que zero indicam concordância menor do que a esperada com a classificação sendo feita ao acaso.

3. Combinação de Classificadores

Ao combinar classificadores, as deficiências de um determinado classificador são compensadas pelas qualidades de outro [11]. A seguir são descritos os métodos utilizados neste trabalho.

3.1 Combinação de Classificadores com Valores de Saída Contínuos

O grau de suporte para uma dada entrada x pode ser interpretado de diferentes maneiras, sendo que as mais comuns são conhecidas como *confiança* nos rótulos sugeridos e *estimação de probabilidades a posteriori* para as classes [7].

Seja $x \in \mathfrak{R}^n$ o vetor de características e $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ o conjunto de rótulos das classes. Cada classificador D_i do conjunto $D = \{D_1, \dots, D_L\}$ fornece c graus de suporte. Sem perda de generalidade pode-se assumir que todos os c graus estão no intervalo $[0, 1]$, isto é, $D_i : \mathfrak{R}^n \rightarrow [0, 1]^c$. A notação $d_{i,j}(x)$ representa o suporte que o classificador D_i dá para a hipótese de que x vem da classe ω_j . As L saídas dos classificadores para uma determinada entrada x pode ser organizado em um *perfil de decisão* ($DP(x)$) como a matriz:

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & \dots & d_{1,j}(x) & \dots & d_{1,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{i,1}(x) & \dots & d_{i,j}(x) & \dots & d_{i,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{L,1}(x) & \dots & d_{L,j}(x) & \dots & d_{L,c}(x) \end{bmatrix}. \quad (3.1)$$

Vários métodos de combinação utilizam $DP(x)$ para encontrar o suporte geral para cada classe e rotular a entrada x de acordo com a classe de maior suporte.

3.2 Combinador Decision Templates

Neste método os combinadores derivam $\mu_j(x)$ usando todos os $L \times c$ graus de suporte em $DP(x)$. Cada vetor no espaço de atributos intermediário é uma versão expandida das $DP(x)$ obtidas concatenando-se suas L linhas.

São criados *padrões de decisão* (DT) que armazenam o perfil de decisão mais típico para cada classe ω_j . Os perfis de decisão $DP(x)$ são então comparados com os DT. O rótulo de x é dado então a partir do DT mais similar.

O treinamento de um padrão de decisão é feito calculando-se a média dos perfis de decisão $DP(z_k)$ de todos os membros de ω_j do conjunto de dados Z [7]:

$$DT_j = \frac{1}{N_j} \sum_{\substack{z_k \in \omega_j \\ z_k \in Z}} DP(z_k), \quad (3.2)$$

onde N_j é o número de elementos de Z vindos de ω_j .

Após o treinamento, dada uma entrada, é possível construir seu perfil de decisão $DP(x)$ e calcular a similaridade S entre $DP(x)$ e cada DT_j :

$$u_j(x) = S(DP(x), DT_j) \quad j = 1, \dots, c. \quad (3.3)$$

A similaridade S pode ser a distância Euclidiana ou qualquer outra medida de distância. A classe cujo padrão de decisão (DT) for mais próximo ao perfil de decisão $DP(x)$ irá dar rótulo à amostra.

3.3 Combinador de Dempster-Shafer

Este método é baseado na teoria da evidência, introduzida por Glenn Shafer como uma maneira de representar conhecimento cognitivo. Neste formalismo a melhor representação de uma possibilidade é uma função de crença em vez de uma distribuição Bayesiana. [12]

O treinamento do combinador de Dempster-Shafer é igual ao do Decision Templates. Porém, em vez de calcular a similaridade entre o padrão de decisão e o perfil de decisão, são calculadas a proximidade entre o modelo de decisão e a saída de cada classificador. Estas são usadas para calcular o grau de crença para

cada classe. Finalmente os graus de suporte para cada classe são calculados a partir dos graus de crença. Estes passos estão descritos abaixo: [7]:

1. Seja DT_j^i a i ésima linha do padrão de decisão DT_j e $D_i(x)$ a saída de D_i , ou seja, $D_i(x) = [d_{i,1}(x), \dots, d_{i,c}(x)]^T$: a i ésima linha do perfil de decisão $DP(x)$. Calculamos a “proximidade” ϕ entre DT_j^i e a saída do classificador D_i para a entrada x :

$$\phi_{j,i}(x) = \frac{\left(1 + \|DT_j^i - D_i(x)\|^2\right)^{-1}}{\sum_{k=1}^c \left(1 + \|DT_k^i - D_i(x)\|^2\right)^{-1}}, \quad (3.4)$$

onde $\| \cdot \|$ é qualquer norma de matriz. Por exemplo, pode-se usar a distância euclidiana entre os dois vetores. Portanto para cada padrão de decisão existem L proximidades.

2. Usando a equação anterior calculamos todas as classes, $j = 1, \dots, c$; e para cada classificador, $i = 1, \dots, L$, os seguintes graus de crença

$$b_j(D_i(x)) = \frac{\phi_{j,i}(x) \prod_{k \neq j} (1 - \phi_{k,i}(x))}{1 - \phi_{j,i}(x) \left[1 - \prod_{k \neq j} (1 - \phi_{k,i}(x))\right]} \quad (3.5)$$

3. Os graus de suporte finais são

$$\mu_j(x) = K \prod_{i=1}^L b_j(D_i(x)) \quad j = 1, \dots, c. \quad (3.6)$$

onde K é uma constante de normalização.

3.4 Bagging

Este método consiste em construir réplicas do tipo bootstrap do conjunto de treinamento e treiná-las individualmente. A saída dos diversos classificadores é então combinada usando votação, ou seja, a saída apontada pela maioria dos classificadores é a escolhida [13].

A partir de um conjunto de treinamento pré-rotulado são gerados aleatoriamente diversos outros conjuntos utilizando substituição. Para tirar proveito deste método é necessário que o classificador base seja instável, ou seja, um classificador em que pequenas mudanças no conjunto de treinamento levam a grandes mudanças na saída do classificador. Se esta regra não for obedecida o que se obtém é somente um conjunto de classificadores quase

idênticos, que dificilmente irão melhorar a performance do classificador.

Neste trabalho os classificadores utilizados são baseados em redes neurais, de natureza instável, portanto espera-se que o uso do método Bagging possa melhorar significativamente o desempenho dos classificadores obtidos.

4. Metodologia

Este trabalho propõe métodos para identificação de materiais em imagens multiespectrais ruidosas de tomografia de transmissão utilizando múltiplas energias. A proposta consiste em explorar técnicas de classificação de imagem e combinação de classificadores nas imagens tomográficas adquiridas de um corpo de estudo através de múltiplas energias. Através dos experimentos são feitas comparações e corroborações teóricas sobre as técnicas utilizadas.

4.1 Aquisição de Imagens

O minitomógrafo utilizado para a aquisição das imagens deste trabalho foi desenvolvido no EMBRAPA [14] e é um equipamento de primeira geração em que a fonte de raios-X ou raios- Γ (que pode ser proveniente de materiais com decaimento radioativo como Césio ou Amerício) e o detector (de cristal) são fixos, enquanto que a amostra em estudo sofre movimentos de rotação e translação.

As imagens utilizadas neste trabalho foram geradas através do imageamento de um corpo de prova nesse minitomógrafo. O corpo de prova (ou *phantom*) é um objeto confeccionado especialmente para ser utilizado em experimentos. O *phantom* utilizado neste trabalho foi construído em *plexiglass* – que é um tipo de polímero – e tem um formato cilíndrico contendo quatro espaços (também de forma cilíndrica) onde são colocados os materiais a serem analisados. Nele foram inseridos 4 tipos de materiais: cálcio, fósforo, alumínio e água, conforme ilustrado na Figura 1.

Para obtenção das projeções foram empregadas duas fontes de raios- Γ (Césio e Amerício) e duas fontes de raios-X disponíveis. As energias provenientes do césio e do amerício são 662keV e 60keV, respectivamente e as duas energias provenientes de raios-X são 40keV e 85keV, sendo que as duas últimas foram escolhidas a partir da análise do comportamento dos coeficientes de atenuação (Figura 2) de cada material a ser identificado, de forma a obter melhor capacidade discriminatória. A Tabela I mostra os valores teóricos

dos coeficientes de atenuação linear de cada material inclusive o plexiglass.

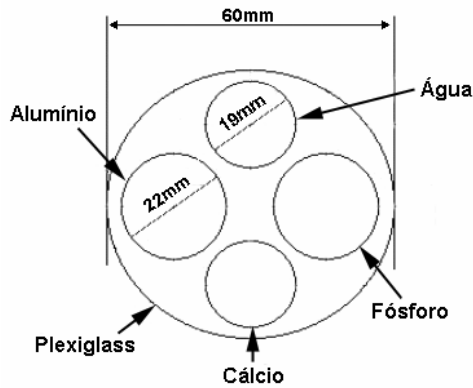


Figura 1 – Diagrama de construção do corpo de prova (phantom)

Tabela 1 – Coeficientes de atenuação linear em função das energias 40, 60, 85 e 662 keV.

Energia (keV)	Coeficientes de Atenuação Linear dos Energias Materiais				
	Água	Alumínio	Cálcio	Fósforo	Plexiglass
40	0,2683	1,5349	2,8365	1,4734	0,2880
60	0,2052	0,7479	1,0013	0,6315	0,2270
85	0,1750	0,6250	0,7750	0,5250	0,1000
662	0,0837	0,2019	0,1205	0,1365	0,0570

O procedimento de obtenção das imagens foi realizado de forma que fossem geradas imagens com alto nível de ruído, utilizando translação total de 7 cm, passo linear de 0,109 cm, rotação total de 180°, passo angular de 2,81° e tempo de contagem de fótons em 3 segundos. Com o alto nível de ruído pretende-se verificar a eficiência dos classificadores propostos mesmo com imagens de baixa qualidade.

Uma vez obtidas as projeções, as imagens foram reconstruídas pelo método de retroprojeção filtrada. Assim, foram geradas quatro imagens no tamanho 65x65 pixels, cada uma correspondendo a uma das energias utilizadas. Estas imagens foram convertidas para o formato BMP (Bitmap), para serem utilizadas no Matlab de forma que pudessem ser classificadas. Nesta conversão os valores do coeficiente de atenuação contidos em cada imagem foram normalizados para um intervalo de 256 valores (intensidades). As imagens utilizadas nesse trabalho são as mesmas utilizadas por Ponti Jr [4][5] e são mostradas na Figura 3. As imagens mostradas são imagens negativas das originais e foram trabalhadas dessa forma para obter uma melhor visualização, uma vez que esta inversão na escala de cinzas não altera o

resultado da classificação. Os valores observados para as bandas de 40, 60, 85 e 662 keV foram, respectivamente, 4129, 1740, 1148 e 336. Estes valores estão abaixo dos valores teóricos e esta diferença é bastante expressiva em alguns casos. Ponti Jr [4] atribui esta diferença às impurezas presentes no material em estudo e ao ruído, incluindo o ruído inerente ao próprio processo de aquisição das projeções tomográficas e o ruído decorrente do dispositivo eletrônico de aquisição (normalmente ruído gaussiano).

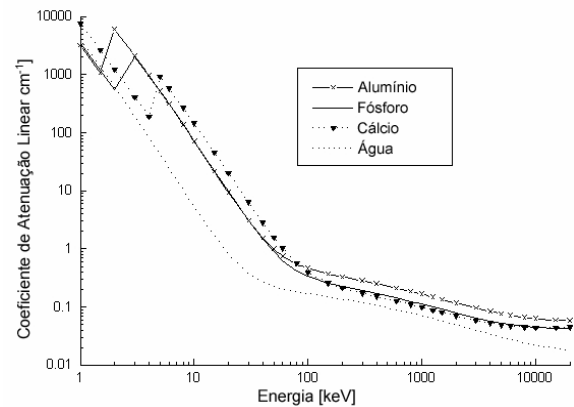


Figura 2 – Curvas do coeficiente de atenuação linear em função da energia para a alumínio, fósforo, cálcio e água [2]

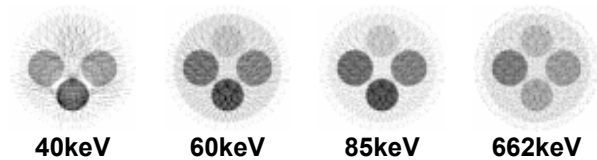


Figura 3 - Imagens adquiridas com energias de 40, 60, 85 e 662 keV.

4.2 Classificação das Imagens

Uma das ferramentas utilizadas no processo de classificação das imagens e avaliação dos classificadores foi o MATLAB, um software de alto desempenho para computação técnica aplicado em diversas áreas, incluindo o Reconhecimento de Padrões. É um sistema interativo com diversas facilidades para trabalhar com vetores e matrizes, de forma que a implementação de um algoritmo que utiliza esses elementos se torna bem mais rápida no MATLAB do que em linguagens tradicionais como C, Fortran e Pascal.

Ele também conta com diversas *toolboxes* específicas para vários tipos de aplicações, incluindo aí uma *toolbox* específica para redes neurais, que é

utilizada neste trabalho e o PRTools, que é uma *toolbox* para o MATLAB desenvolvida pelo grupo de Reconhecimento de Padrões da Universidade de Delft na Holanda [15]. A versão atual (4.0) do PRTools implementa aproximadamente 200 rotinas de reconhecimento de padrões, cobrindo uma grande parte das técnicas dessa área.

Foram realizados experimentos de classificação com as imagens descritas na seção anterior. No procedimento de classificação as quatro imagens BMP são combinadas como bandas de uma única imagem multiespectral. Foram consideradas 6 classes: água, alumínio, fósforo, cálcio, plexiglass e o fundo da imagem. A probabilidade a priori foi considerada igual para todas as classes por não haver informações que pudessem ser utilizadas.

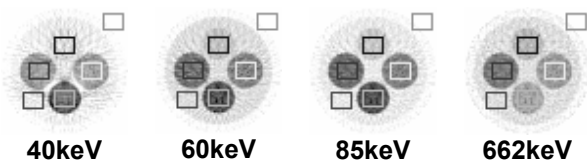


Figura 4 - Janelas de pixels escolhidos como amostras em cada imagem

Para o treinamento e teste foram escolhidas 80 amostras de cada classe, em janelas de 10x8 pixels que podem ser vistas na Figura 4, totalizando 480 amostras. Este número de amostras é reduzido devido ao tamanho das imagens geradas pelo tomógrafo, que têm apenas 65x65 pixels cada. O conjunto de amostras foi subdividido em dois subconjuntos de 240 amostras cada (40 de cada classe), sendo que um subconjunto foi destinado ao treinamento dos classificadores e o outro subconjunto foi destinado ao teste dos classificadores treinados, seguindo assim a técnica de *hold-out*.

Os classificadores escolhidos para este trabalho são o Perceptron de Múltiplas Camadas e as Redes de Função de Base Radial, que foram treinados com diferentes quantidades de neurônios na camada oculta, pois a quantidade ideal de neurônios depende dos dados a serem classificados e não há nenhuma maneira precisa de determinar esta quantidade a priori [16]. Desse modo, todos os classificadores treinados têm uma camada de entrada com 4 neurônios (nós), que correspondem a cada uma das bandas da imagem, uma camada intermediária (oculta) com número de neurônios variável e uma camada de saída com 6 neurônios, que correspondem a cada uma das classes consideradas.

Para o Perceptron de Múltiplas Camadas foram treinados classificadores contendo de 2 a 25

neurônios em uma única camada oculta. Os classificadores baseados no Perceptron de Múltiplas Camadas tendem a apresentar resultados bastante diferentes dependendo dos parâmetros de iniciação da rede, que são aleatórios, portanto para cada tipo de estrutura testada foram treinados 100 classificadores diferentes (com diferentes parâmetros de iniciação), de modo que fosse possível obter estatísticas confiáveis de seu desempenho. Assim foram obtidas as médias e desvio padrão do erro estimado e do coeficiente Kappa para cada tipo de classificador.

Para as Redes de Função de Base Radial também foram treinados classificadores contendo de 2 a 25 neurônios na camada oculta. Este tipo de rede neural não apresenta a mesma instabilidade do Perceptron de Múltiplas Camadas, pois os classificadores gerados com a mesma estrutura e as mesmas amostras tendem a ser sempre muito semelhantes. Apesar dessa observação os experimentos também foram repetidos 100 vezes com diferentes iniciações para manter um padrão.

Cada classificador treinado teve seu desempenho medido através de um teste com as amostras previamente designadas. Dessa forma foram obtidos os erros estimados pelo método *hold-out*, bem como os Coeficientes Kappa de cada classificador.

O erro estimado representa a taxa de erro que se espera do classificador e é obtido pela razão entre a quantidade de erros do classificador e o número de amostras testadas. O Coeficiente Kappa é obtido utilizando-se a equação 2.1, onde o $P(E)$ terá sempre o valor 0,166... pois temos 6 classes equiprováveis, e $P(A)$ é facilmente obtido subtraindo-se o erro estimado de 1.

Em seguida foram executados os experimentos utilizando combinação de classificadores. Foi utilizada primeiramente a técnica de Bagging, e em seguida foram utilizados os combinadores de Decision Templates e Dempster-Shafer. Todos foram utilizados sobre os classificadores base Perceptron de Múltiplas Camadas e Redes de Função de Base Radial, com as mesmas configurações e quantidade de neurônios de seus respectivos testes individuais. Para todos os experimentos com combinadores foram treinados 10 classificadores-base diferentes, variando sempre a iniciação da rede (e também o conjunto de amostras nos experimentos com Bagging). A combinação foi feita a partir das saídas desses 10 classificadores. Todos os experimentos também foram repetidos 100 vezes para que as estatísticas fossem confiáveis. Assim nos experimentos utilizando Decision Templates e Dempster-Shafer foram variados os parâmetros de iniciação em cada uma das 100 execuções. E nos experimentos com

Bagging além da variação dos parâmetros de iniciação tivemos também a variação no conjunto de treinamento, através dos diferentes conjuntos tipo *bootstrap* gerados para cada classificador. Para cada experimento com combinação também foram obtidos as estimativas de erro por Hold-Out e os Coeficientes Kappa, para que fosse possível comparar os resultados com o dos classificadores individuais.

A partir dos resultados obtidos foi observado que seria interessante estudar o que aconteceria ao combinar a técnica de Bagging com os combinadores Decision Templates e Dempster-Shafer, isto é, utilizar Decision Templates e Dempster-Shafer para fazer a combinação dos classificadores gerados para cada amostra do tipo bootstrap, e assim comparar estes resultados com os obtidos com cada técnica de combinação aplicada individualmente. Dessa forma foram conduzidos experimentos utilizando Bagging com Dempster-Shafer e Bagging com Decision Templates, ambos sobre os classificadores base Perceptron de Múltiplas Camadas e Redes de Função de Base Radial, com as mesmas configurações e quantidade de neurônios dos experimentos anteriores, e também fazendo a combinação a partir de 10 classificadores-base. Estes experimentos também foram repetidos 100 vezes, variando as iniciações e os conjuntos do tipo bootstrap utilizados, e assim foram obtidos as estimativas de erro por Hold-Out e os coeficientes Kappa.

5. Resultados

Neste capítulo são apresentados os resultados obtidos com os classificadores Perceptron de Múltiplas Camadas e Redes de Função de Base Radial, tanto para os classificadores aplicados individualmente (IND) como para a aplicação utilizando os combinadores: Bagging (BAG), Decision Templates (DT), Dempster-Shafer (DS), Bagging c/ Decision Templates (BAGDT) e Bagging c/ Dempster-Shafer (BAGDS). As tabelas apresentadas resumem o desempenho apresentado pelos classificadores com cada tipo de estrutura interna (nós na camada oculta), tanto individualmente quanto com o uso de combinadores. Os melhores resultados em cada coluna estão marcados em negrito, e os piores estão marcados em itálico. Os gráficos das tabelas e algumas imagens temáticas podem ser encontradas na dissertação completa, disponível na URL <http://www.dc.ufscar.br/~fabricio/artigos/DissFAB.pdf>

Observando as tabelas III e IV é possível observar que todas as técnicas de combinação melhoraram significativamente o desempenho dos classificadores Perceptron de Múltiplas Camadas, pois os erros estimados diminuíram e os coeficientes Kappa aumentaram. O combinador Decision Templates apresentou os melhores resultados, seguido por Dempster-Shafer, Bagging c/ Decision Templates e Bagging com Dempster-Shafer, nesta ordem e com pouca diferença entre os quatro. A técnica de Bagging apresentou o pior desempenho entre os combinadores, principalmente quando foram utilizados poucos neurônios na camada oculta.

A tabela V indica que o desvio padrão dentro das 100 repetições de cada experimento diminuiu significativamente com a aplicação de técnicas de combinação de classificadores, mostrando que todas as técnicas são eficientes para estabilizar classificadores Perceptron de Múltiplas Camadas. Também neste caso o combinador Decision Templates apresentou o melhor desempenho, seguido por Bagging c/ Decision Templates, Dempster-Shafer e Bagging c/ Dempster-Shafer.

Nos resultados para as Redes de Função de Base Radial, nas tabelas VI e VII, é possível observar que a aplicação de técnicas de combinação melhora a classificação na maioria dos casos, porém essa melhoria não é tão grande quanto no Perceptron de Múltiplas Camadas. A técnica de Bagging apresentou os melhores resultados para as Redes de Função de Base Radial, com o combinador Bagging c/ Decision Templates sendo o segundo colocado apresentando resultados bem próximos ao primeiro. Em seguida vem o combinador Bagging c/ Dempster-Shafer c/ desempenho inferior aos dois primeiros, mas ainda bastante superior a Decision Templates e Dempster-Shafer, que foram os menos eficientes.

A tabela VIII mostra que as Redes de Função de Base Radial são bastante estáveis nas 100 repetições de cada experimento quando comparadas ao Perceptron de Múltiplas Camadas, sendo que em vários casos o desvio padrão era zero, o que não foi alterado com o uso dos combinadores Bagging, Decision Templates e Dempster-Shafer. O uso dos combinadores Bagging c/ Decision Templates e Bagging c/ Dempster-Shafer causaram uma leve instabilidade nas Redes de Função de Base Radial.

TABELA III – Erro Estimado médio para o Perceptron de Múltiplas Camadas

Nós	IND	BAG	DT	DS	BAG DT	BAG DS
2	0,5720	0,3675	0,0349	0,0613	0,0358	0,0552
3	0,2689	0,0493	0,0163	0,0275	0,0166	0,0269
4	0,1318	0,0200	0,0141	0,0177	0,0145	0,0203
5	0,0976	0,0170	0,0123	0,0151	0,0150	0,0172
6	0,0741	0,0168	0,0127	0,0139	0,0155	0,0171
7	0,0681	0,0175	0,0129	0,0138	0,0148	0,0164
8	0,0636	0,0179	0,0130	0,0137	0,0163	0,0177
9	0,0511	0,0190	0,0134	0,0138	0,0167	0,0174
10	0,0570	0,0190	0,0135	0,0139	0,0172	0,0183
11	0,0497	0,0191	0,0136	0,0143	0,0179	0,0188
12	0,0603	0,0203	0,0136	0,0143	0,0181	0,0191
13	0,0525	0,0196	0,0137	0,0145	0,0185	0,0196
14	0,0477	0,0204	0,0140	0,0151	0,0187	0,0199
15	0,0470	0,0210	0,0143	0,0150	0,0190	0,0198
16	0,0403	0,0211	0,0145	0,0147	0,0196	0,0205
17	0,0426	0,0205	0,0148	0,0153	0,0185	0,0194
18	0,0402	0,0211	0,0146	0,0149	0,0192	0,0202
19	0,0340	0,0207	0,0145	0,0154	0,0198	0,0209
20	0,0451	0,0213	0,0151	0,0153	0,0201	0,0209
21	0,0275	0,0203	0,0155	0,0156	0,0201	0,0209
22	0,0366	0,0215	0,0150	0,0158	0,0200	0,0212
23	0,0430	0,0213	0,0150	0,0156	0,0200	0,0207
24	0,0393	0,0217	0,0151	0,0156	0,0202	0,0211
25	0,0348	0,0214	0,0148	0,0162	0,0204	0,0209

TABELA IV – Coeficiente Kappa médio para o Perceptron de Múltiplas Camadas

Nós	IND	BAG	DT	DS	BAG DT	BAG DS
2	0,3137	0,5591	0,9581	0,9265	0,9570	0,9338
3	0,6773	0,9409	0,9805	0,9671	0,9801	0,9678
4	0,8419	0,9760	0,9831	0,9788	0,9826	0,9757
5	0,8829	0,9796	0,9853	0,9819	0,9821	0,9794
6	0,9111	0,9799	0,9848	0,9833	0,9814	0,9795
7	0,9183	0,9790	0,9845	0,9835	0,9822	0,9803
8	0,9237	0,9786	0,9844	0,9836	0,9804	0,9788
9	0,9387	0,9773	0,9840	0,9835	0,9800	0,9792
10	0,9316	0,9773	0,9838	0,9833	0,9794	0,9780
11	0,9404	0,9771	0,9837	0,9829	0,9785	0,9775
12	0,9277	0,9757	0,9837	0,9829	0,9783	0,9771
13	0,9371	0,9765	0,9836	0,9827	0,9778	0,9765
14	0,9428	0,9756	0,9832	0,9819	0,9776	0,9762
15	0,9436	0,9748	0,9828	0,9821	0,9773	0,9762
16	0,9516	0,9747	0,9827	0,9824	0,9765	0,9755
17	0,9489	0,9755	0,9823	0,9817	0,9778	0,9768
18	0,9518	0,9747	0,9825	0,9822	0,9770	0,9758
19	0,9592	0,9752	0,9826	0,9816	0,9762	0,9749
20	0,9459	0,9745	0,9819	0,9817	0,9759	0,9749
21	0,9671	0,9756	0,9815	0,9813	0,9759	0,9750
22	0,9561	0,9743	0,9820	0,9811	0,9760	0,9746
23	0,9484	0,9744	0,9821	0,9813	0,9761	0,9752
24	0,9529	0,9740	0,9819	0,9813	0,9758	0,9747
25	0,9583	0,9744	0,9822	0,9806	0,9756	0,9750

TABELA V – Desvio Padrão do Erro Estimado e Coeficiente Kappa para o Perceptron de Múltiplas Camadas

Nós	IND	BAG	DT	DS	BAG DT	BAG DS
2	0,2554	0,1729	0,0286	0,0348	0,0228	0,0281
3	0,2892	0,0559	0,0061	0,0133	0,0064	0,0117
4	0,2073	0,0079	0,0044	0,0065	0,0046	0,0069
5	0,1504	0,0062	0,0029	0,0046	0,0038	0,0048
6	0,1199	0,0039	0,0020	0,0028	0,0030	0,0043
7	0,1065	0,0039	0,0013	0,0024	0,0026	0,0038
8	0,0842	0,0037	0,0014	0,0020	0,0029	0,0039
9	0,0777	0,0039	0,0018	0,0021	0,0037	0,0039
10	0,0807	0,0042	0,0018	0,0023	0,0032	0,0037
11	0,0770	0,0038	0,0020	0,0026	0,0036	0,0039
12	0,0887	0,0050	0,0020	0,0024	0,0038	0,0042
13	0,0698	0,0044	0,0021	0,0025	0,0044	0,0047
14	0,0623	0,0047	0,0023	0,0026	0,0038	0,0044
15	0,0704	0,0042	0,0025	0,0027	0,0042	0,0042
16	0,0680	0,0041	0,0023	0,0027	0,0042	0,0043
17	0,0618	0,0041	0,0027	0,0030	0,0034	0,0038
18	0,0638	0,0042	0,0026	0,0025	0,0035	0,0036
19	0,0480	0,0046	0,0026	0,0035	0,0041	0,0042
20	0,0756	0,0044	0,0027	0,0034	0,0040	0,0039
21	0,0360	0,0036	0,0029	0,0031	0,0041	0,0041
22	0,0648	0,0039	0,0025	0,0031	0,0045	0,0047
23	0,0650	0,0043	0,0027	0,0027	0,0037	0,0042
24	0,0667	0,0039	0,0030	0,0033	0,0036	0,0038
25	0,0538	0,0045	0,0027	0,0033	0,0041	0,0040

TABELA VI – Erro Estimado médio para Redes de Função de Base Radial

Nós	IND	BAG	DT	DS	BAG DT	BAG DS
2	0,5125	0,3720	0,4458	0,4333	0,3391	0,4218
3	0,3958	0,2118	0,3583	0,3583	0,2334	0,3605
4	0,2958	0,1675	0,2583	0,2583	0,2094	0,1776
5	0,2292	0,1953	0,2333	0,2333	0,2100	0,2083
6	0,1333	0,1269	0,1292	0,1292	0,1332	0,1341
7	0,0750	0,0900	0,0708	0,0708	0,0853	0,0909
8	0,0833	0,0733	0,0833	0,0833	0,0717	0,0719
9	0,0750	0,0589	0,0750	0,0750	0,0596	0,0595
10	0,0833	0,0500	0,0875	0,0875	0,0497	0,0487
11	0,0708	0,0444	0,0667	0,0667	0,0432	0,0422
12	0,0500	0,0382	0,0500	0,0500	0,0365	0,0364
13	0,0390	0,0329	0,0302	0,0292	0,0321	0,0319
14	0,0334	0,0265	0,0333	0,0333	0,0252	0,0251
15	0,0252	0,0201	0,0250	0,0250	0,0197	0,0196
16	0,0232	0,0186	0,0213	0,0208	0,0181	0,0183
17	0,0219	0,0175	0,0197	0,0208	0,0175	0,0170
18	0,0177	0,0160	0,0170	0,0167	0,0161	0,0160
19	0,0173	0,0150	0,0166	0,0167	0,0145	0,0145
20	0,0145	0,0135	0,0125	0,0125	0,0137	0,0135
21	0,0130	0,0132	0,0125	0,0125	0,0130	0,0131
22	0,0126	0,0129	0,0125	0,0125	0,0127	0,0127
23	0,0126	0,0126	0,0125	0,0125	0,0127	0,0125
24	0,0125	0,0126	0,0125	0,0125	0,0126	0,0126
25	0,0125	0,0127	0,0125	0,0125	0,0126	0,0126

TABELA VII – Coeficiente Kappa médio para Redes de Função de Base Radial

Nós	IND	BAG	DT	DS	BAG DT	BAG DS
2	0,3850	0,5536	0,4650	0,4800	0,5931	0,4939
3	0,5250	0,7459	0,5700	0,5700	0,7199	0,5675
4	0,6450	0,7990	0,6900	0,6900	0,7488	0,7869
5	0,7250	0,7656	0,7200	0,7200	0,7480	0,7500
6	0,8400	0,8478	0,8450	0,8450	0,8402	0,8391
7	0,9100	0,8921	0,9150	0,9150	0,8976	0,8909
8	0,9000	0,9120	0,9000	0,9000	0,9140	0,9137
9	0,9100	0,9294	0,9100	0,9100	0,9285	0,9286
10	0,9000	0,9400	0,8950	0,8950	0,9404	0,9416
11	0,9150	0,9468	0,9200	0,9200	0,9482	0,9494
12	0,9400	0,9542	0,9400	0,9400	0,9562	0,9563
13	0,9532	0,9605	0,9638	0,9650	0,9615	0,9617
14	0,9600	0,9682	0,9600	0,9600	0,9698	0,9699
15	0,9698	0,9759	0,9700	0,9700	0,9764	0,9765
16	0,9722	0,9777	0,9744	0,9750	0,9783	0,9781
17	0,9738	0,9790	0,9764	0,9750	0,9790	0,9796
18	0,9788	0,9808	0,9797	0,9800	0,9807	0,9808
19	0,9793	0,9820	0,9801	0,9800	0,9826	0,9826
20	0,9826	0,9838	0,9850	0,9850	0,9836	0,9838
21	0,9844	0,9842	0,9850	0,9850	0,9844	0,9843
22	0,9849	0,9845	0,9850	0,9850	0,9848	0,9848
23	0,9849	0,9849	0,9850	0,9850	0,9848	0,9850
24	0,9850	0,9849	0,9850	0,9850	0,9849	0,9849
25	0,9850	0,9848	0,9850	0,9850	0,9849	0,9849

TABELA VIII – Desvio Padrão do Erro Estimado e Coeficiente Kappa para Redes de Função de Base Radial

Nós	IND	BAG	DT	DS	BAG DT	BAG DS
2	0,0000	0,0000	0,0000	0,0000	0,0432	0,0171
3	0,0000	0,0000	0,0000	0,0000	0,0357	0,0123
4	0,0000	0,0000	0,0000	0,0000	0,0185	0,0287
5	0,0000	0,0000	0,0000	0,0000	0,0139	0,0119
6	0,0000	0,0000	0,0000	0,0000	0,0068	0,0067
7	0,0000	0,0000	0,0000	0,0000	0,0097	0,0097
8	0,0000	0,0000	0,0000	0,0000	0,0097	0,0100
9	0,0000	0,0000	0,0000	0,0000	0,0079	0,0097
10	0,0000	0,0000	0,0000	0,0000	0,0066	0,0073
11	0,0000	0,0000	0,0000	0,0000	0,0063	0,0057
12	0,0000	0,0000	0,0000	0,0000	0,0051	0,0056
13	0,0066	0,0021	0,0019	0,0021	0,0050	0,0041
14	0,0004	0,0000	0,0000	0,0000	0,0053	0,0050
15	0,0012	0,0000	0,0000	0,0000	0,0029	0,0033
16	0,0039	0,0015	0,0015	0,0015	0,0023	0,0022
17	0,0040	0,0019	0,0020	0,0019	0,0021	0,0020
18	0,0025	0,0014	0,0011	0,0014	0,0026	0,0021
19	0,0015	0,0008	0,0006	0,0008	0,0024	0,0022
20	0,0029	0,0000	0,0004	0,0000	0,0020	0,0020
21	0,0015	0,0000	0,0000	0,0000	0,0014	0,0015
22	0,0006	0,0000	0,0000	0,0000	0,0008	0,0009
23	0,0006	0,0000	0,0000	0,0000	0,0009	0,0004
24	0,0004	0,0000	0,0000	0,0000	0,0007	0,0006
25	0,0004	0,0000	0,0000	0,0000	0,0006	0,0006

6. Conclusões

Analisando os resultados é possível concluir que o uso de combinação de classificadores produziu classificadores mais estáveis, além de melhorar significativamente os resultados medidos com o Erro Estimado pelo método Hold-Out e o Coeficiente Kappa, principalmente nas redes com poucos neurônios na camada oculta.

Enquanto os combinadores de Decision Templates e Dempster-Shafer mostraram melhor desempenho que a técnica de Bagging no Perceptron de Múltiplas Camadas, o contrário ocorreu nos testes com Redes de Função de Base Radial. Nelas o Bagging teve um melhor desempenho, enquanto Decision Templates e Dempster-Shafer não se saíram tão bem. Isto pode ser explicado pela maior estabilidade das Redes de Função de Base Radial quando comparadas ao Perceptron de Múltiplas Camadas, o que faz com que os classificadores fiquem muito parecidos e consequentemente a combinação seja menos vantajosa.

O combinador Decision Templates mostrou ser o mais indicado para combinar o Perceptron de Múltiplas Camadas, apresentando o melhor desempenho e mostrando que as diferenças nas redes causadas pela inicialização aleatória dos parâmetros livres já são suficientes para produzir classificadores bastante diferentes e bons para combinação. Além disso, Decision Templates também foi o combinador que deixou o Perceptron mais estável.

Para as Redes de Função de Base Radial a técnica de Bagging mostrou ser o método mais indicado, pois foi o que apresentou o melhor desempenho. Bagging c/ Decision Templates também apresentou um bom desempenho, entretanto foi um dos métodos que trouxe uma leve instabilidade para as Redes de Função de Base Radial, de modo que seu uso deve ser feito com cautela. Decision Templates e Dempster-Shafer foram os combinadores menos eficientes, e isto provavelmente se deve ao fato das Redes de Função de Base Radial serem mais estáveis, produzindo classificadores muito parecidos, o que os torna menos indicados para combinação. A técnica de Bagging levou vantagem por utilizar conjuntos de amostras do tipo bootstrap, produzindo classificadores diferentes para a combinação, e consequentemente melhores resultados.

O uso de combinação de classificadores para combinar redes neurais se mostrou bastante útil para identificar materiais em imagens de ciências dos solos. Outras áreas também podem ser exploradas utilizando esta mesma abordagem, pois as regras de

combinação normalmente melhoraram a performance e aumentam a capacidade de generalização dos classificadores.

Diversos trabalhos podem ser desenvolvidos como extensão deste, na área de combinação de classificadores, como fazer uma pré-seleção dos classificadores mais relevantes para a combinação; combinar redes neurais de tipos diferentes (por exemplo: Perceptron de Múltiplas Camadas com Redes de Função de Base Radial); utilizar outros combinadores com a técnica de Bagging, como o Produto, Máximo e Mínimo, por exemplo; combinar a saída dos combinadores, utilizando votação por maioria ou qualquer outra técnica; utilizar redes neurais para combinar resultados de classificação, dentre outros.

7. Agradecimentos

Gostaríamos de agradecer ao Dr. Paulo E. Cruvinel por nos fornecer as imagens multiespectrais utilizadas nesses experimentos, e a CAPES e a FAPESP (04/05316-7) pelo apoio financeiro. Este trabalho também é parcialmente apoiado pelo Projeto Temático da Fapesp 2002/07153-2.

8. Referências

- [1] GONZALEZ, R. F.; WOODS, R. E., “Processamento de Imagens Digitais”, *Edgar Blücher LTDA*, São Paulo, 2000.
- [2] HOMEM, M. R. P., “Técnicas de Reconhecimento de Padrões Aplicadas a Imagens Tomográficas Adquiridas com Múltiplas Energias”, Dissertação de mestrado, DC-UFSCar: São Carlos, 1998, 200p.
- [3] HOMEM, M. R. P.; MASCARENHAS, N. D. A.; CRUVINEL, P. E., “The Linear Attenuation Coefficients as Features of Multiple Energy CT Image Classification”, *Nuclear Instruments and Methods in Physics Research*. V. 452, p. 351-360, 2000.
- [4] PONTI JR, M. P., “Combinação de Múltiplos Classificadores para Identificação de Materiais em Imagens Ruidosas”, Dissertação de mestrado, DC-UFSCar: São Carlos, 2004, 76p.
- [5] PONTI JR., M.P.; MASCARENHAS, N.D.A, “Material Analysis on Noisy Multispectral Images Using Classifier Combination”, *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation, IEEE Computer Society Press*, Lake Tahoe, Nevada, p. 28-30, 2004.
- [6] HAYKIN S., “Redes Neurais – Princípios e Prática”, *Bookman*, 2 ed. Porto Alegre: 2001.

- [7] KUNCHEVA, L. I., “Combining Pattern Classifiers : Methods and Algorithms”, *Hoboken, NJ :Wiley-Interscience*, 2004.
- [8] BISHOP, C. M., “Neural Networks for Pattern Recognition”, *Oxford*, New York: 1995.
- [9] KOHAVI, R., “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”, *Proc. of the 14th Int. Joint Conf. on A.I.*, Vol. 2, Canada, p. 1137 –1143, 1995.
- [10] CARLETTA, J., “Assessing Agreement on Classification Tasks: the Kappa Statistic”, *Computational Linguistics*. V. 22(2) p. 249- 254, 1996
- [11] PRAMPERO, P.; CARVALHO, A., “Recognition of Vehicles Using Combination of Classifiers”, *Proceedings of the IEEE World Congress on Computational Intelligence*, WCCI’98, Anchorage, USA, 1998.
- [12] AHMADZADEH, M. R.; PETRON M.; SASIKALA, K. R., “The Dempster-Shafer Combination Rule as a Tool to Classifier Combination”, *Geoscience and Remote Sensing Symposium. Proc. IGARSS 2000 IEEE International*, p. 2429-2431, 2000.
- [13] BREIMAN, L., “Bagging Predictors”, *Machine Learning*. V. 26(2) p.123-140, 1996
- [14] CRUVINEL, P. E.; CESAREO, R.; MASCARENHAS, S., “X and γ - Rays Computerized Minitomograph Scanner for Soil Science.” *IEEE Transactions on Instrumentation and Measurements*. v. 39, n. 5, p. 745-750, 1990.
- [15] DUIN, R. P. W; JUSZCZAK, P.; PACLIK, P.; PERALSKA, D.; RIDDER, D.; TAX, D. M. J., “PRTools4: A Matlab Toolbox For Pattern Recognition”, *Delft University of Technology*, Delft, 51p., 2004.
- [16] DUDA, R. O.; HART, P. E., STORK, D. G., “Pattern Classification”, *Wiley*, 2.ed. New York, 2000.