

AVALIAÇÃO DE CLASSIFICADORES BASEADOS EM APRENDIZADO DE MÁQUINA PARA A CLASSIFICAÇÃO DO USO E COBERTURA DA TERRA NO BIOMA CAATINGA

Evaluation of Classifiers Based on Machines Learning to Land Use and Cover Classification on Caatinga Biome

Beatriz Fernandes Simplicio Sousa¹
Adunias dos Santos Teixeira¹
Francisco de Assis Tavares Ferreira da Silva³
Eunice Maia de Andrade¹
Arthur Plínio de Souza Braga²

¹ **Universidade Federal do Ceará – UFC**
Departamento de Engenharia Agrícola
Caixa Postal 12.168 - 60021-970 - Fortaleza - CE, Brasil
beatrizsimplicio@gmail.com
adunias@ufc.br
eandrade@ufc.br

² **Universidade Federal do Ceará – UFC**
Departamento de Engenharia Elétrica
Caixa Postal 6001 - 60450 -760 - Fortaleza- CE, Brasil
arthurp@dee.ufc.br

³ **Instituto Nacional de Pesquisas Espaciais – INPE**
Caixa Postal 21- 61760-000 - Eusébio - CE, Brasil
tavares@roen.inpe.br

RESUMO

O manejo adequado dos recursos naturais em ambientes frágeis, como o da Caatinga, requer o conhecimento de suas propriedades e distribuição espacial. Nesse contexto, o trabalho tem por objetivo avaliar o desempenho de dois algoritmos baseados em aprendizado de máquina (*Multi Layer Perceptron* (MLP) e o *Support Vector Machine* (SVM)) e do método da Máxima Verossimilhança na classificação do uso e cobertura da terra no bioma Caatinga. Para o experimento, foi utilizada uma imagem do satélite LANDSAT-5/TM contendo a área de estudo localizada no município de Iguatu-CE e definidas as classes de cobertura da terra, a saber: antropização por agricultura (APA), outros tipos de antropização (OTA), água, caatinga herbácea arbustiva (CHA) e caatinga arbórea densa (CAD). O desempenho dos métodos foi analisado através dos coeficientes de Exatidão Global (EG), Exatidão Específica (EE) e Kappa (K) calculados a partir dos dados da matriz de confusão correspondente à verdade terrestre. Os valores do coeficiente de EG foram de: 86,03%, 82,14% e 81,2% e K de: 0,77, 0,76 e 0,75 nos métodos SVM, MLP e Máxima Verossimilhança, respectivamente. Os valores de EE foram superiores a 70% para todos os classificadores testados. Os resultados obtidos demonstram que os métodos SVM e MLP estão aptos à classificação dos padrões propostos, já que apresentaram resultados semelhantes ao método tradicional da Máxima Verossimilhança. Porém, estes classificadores podem consumir mais tempo na etapa de definição dos parâmetros da rede e de processamento.

Palavras chaves: Inteligência Artificial, Semi-árido, Classificação de Imagens de Satélite.

ABSTRACT

Proper management of natural resources in fragile environments, such as the *Caatinga*, requires knowledge of their properties and spatial distribution. In this context, the study aims at evaluating the performance of two algorithms based on machine learning (*Multi Layer Perceptron* (MLP) and *Support Vector Machine* (SVM)) and the Maximum

Likelihood method to classify land use and land cover in the *Caatinga* biome. For the experiment, it was used a satellite image of LANDSAT-5/TM containing the study area located in the municipality of Iguatu-CE, and classes of land cover, namely: anthropized by agriculture, other types of anthropized, water, herbaceous shrub savanna (CHA) and dense arboreal savanna (CAD) were defined. The performance of the methods was analyzed by the coefficient of Global Accuracy (EG), Accuracy Specific (EE) and Kappa (K) coefficient calculated with data taken from the confusion matrix corresponding to ground truth. The coefficient of EG were: 86.03%, 82.14% and 81.2% and K: 0.77, 0.76 and 0.75 in the methods SVM, MLP and maximum likelihood respectively. EE values were above 70% for all classifiers tested. The results have shown that SVM and MLP methods are suited to the classification of the proposed standards, as it showed similar results to the traditional method of maximum likelihood. However, these methods are more time consuming in the stage of defining the parameters of the network and may require more computation power during stage of processing.

Keywords: Artificial Intelligence, Semi Arid, Satellite Image Classification.

1. INTRODUÇÃO

O domínio do bioma Caatinga abrange uma área de aproximadamente 900 mil Km², o que corresponde a cerca de 10% do território nacional (ANDRADE *et al.*, 2005). Este bioma é um dos menos estudados e protegidos do Brasil. Ainda segundo Andrade *et al.* (2005), o sistema agropastoril apresenta-se como o fator que exerce maior pressão sobre a cobertura vegetal do semi-árido nordestino. Degradações causadas pelo homem à vegetação nativa e aos recursos naturais poderão causar sérios danos a este bioma, como por exemplo, a desertificação. No entanto, o conhecimento mais aprofundado da Caatinga por meio de ferramentas como o sensoriamento remoto pode dar suporte ao planejamento de exploração sustentável.

O mapeamento da cobertura da terra é uma informação essencial em estudos de gestão ambiental, em avaliação de biodiversidade e como suporte à adoção de políticas públicas. Esse tipo de mapeamento ainda pode ser utilizado como dado de entrada para o modelo da equação universal de perda de solo (EUPS), proposto por Wischmeier & Smith (1978), proporcionando identificar áreas suscetíveis ao processo de erosão. A informação gerada é importante especialmente em áreas onde o problema se faz presente, como é o caso do semi-árido.

Na literatura técnica é possível encontrar desde métodos tradicionais para classificação de imagens de satélite como, por exemplo, o da máxima verossimilhança (QUEIROZ *et al.*, 2004) e o da mínima distância (LEDUC *et al.*, 2007) até métodos mais avançados como aqueles que utilizam técnicas baseadas em aprendizado de máquina (CARVALHO *et al.*, 2004). Diversos métodos baseados em aprendizado de máquina têm atraído a atenção da comunidade de sensoriamento remoto (GIGANDET *et al.*, 2005).

O aprendizado de máquina está inserido na área de conhecimento denominada Inteligência Artificial (IA) e têm por objetivo desenvolver métodos, técnicas e ferramentas para construir máquinas inteligentes capazes de realizar as mais diversas tarefas. De acordo com Mitchell (1997), essas máquinas devem treinar sobre determinado conjunto de

amostras e experimentar estratégias mais eficazes para a construção de programas que aprendem a partir da experiência adquirindo conhecimento de forma automática. Como exemplo de algoritmos de aprendizado de máquina pode-se citar que representam analogias aos sistemas biológicos humanos (Redes Neurais Artificiais (RNAs) e Algoritmos Genéticos), sistemas de aprendizado simbólico (Árvores de Decisão) e as teorias estatísticas (Support Vector Machines (SVMs)) (MITCHELL, 1997).

De acordo com Haykin (1999), uma RNA do tipo *Multi Layer Perceptron* (MLP) possui uma arquitetura constituída por, pelo menos, três camadas de neurônios artificiais, sendo uma camada de entrada, uma camada de saída e uma ou mais camadas escondidas com neurônios ocultos. Segundo Egmont *et al.* (2002), uma rede MLP com uma camada escondida é suficiente para aproximar qualquer função contínua. Os pesquisadores Gelete & Volatão (2007) aplicaram uma rede do tipo MLP na classificação do uso da terra e obtiveram resultados satisfatórios com uma única camada oculta de 11 neurônios.

Já o algoritmo SVM constrói um hiperplano como superfície de decisão de tal forma que a margem de separação entre exemplos positivos e negativos seja máxima (HAYKIN, 1999). As SVMs têm sido aplicadas e comparadas a diversos métodos na classificação de dados provenientes do sensoriamento remoto, como pode ser visto nos trabalhos de Pal & Mather (2004); Foody & Mathur (2004); Su *et al.* (2007).

Desta forma, o objetivo do presente trabalho foi avaliar o desempenho de dois algoritmos baseados em aprendizado de máquina na classificação do uso e cobertura da terra no bioma Caatinga em imagens do satélite LANDSAT-5/TM. Foram escolhidos os métodos *Multi Layer Perceptron* (MLP) e *Support Vector Machine* (SVM). O desempenho destes foi ainda comparado ao método tradicional estatístico da Máxima Verossimilhança.

2. CLASSIFICAÇÃO DE IMAGENS MULTIESPECTRAIS

Técnicas de classificação de imagens implicam na implementação de um processo de decisão para que o computador possa atribuir o pixel a uma classe (NOVO, 1989).

Segundo Iwai (2003), são gerados neste processo mapas temáticos correspondentes aos temas de interesse por meio de técnicas de classificação supervisionadas ou não supervisionadas.

Nos métodos de classificação supervisionada, as classes são previamente definidas pelo analista, isto é, caracterizadas através das amostras de treinamento. Assim, os algoritmos realizam o reconhecimento dos padrões espectrais na imagem com base nestas amostras (MOREIRA, 2005). Para que a classificação seja bem sucedida, é necessário que as assinaturas espectrais dos alvos estejam bem definidas e distintas entre si.

2.1 Classificação por Máxima Verossimilhança

A classificação por Máxima Verossimilhança rotula os pixels em uma determinada classe conforme a distância estatística entre cada pixel e a média da resposta espectral da classe (ASSAD & SANO, 1998).

Uma função de probabilidade $P(x)$ define a que classe pertencerá determinado pixel. O cálculo é feito para as várias classes pré-definidas e o pixel é atribuído à classe cuja probabilidade de pertencer seja maior. É considerado um método paramétrico, pois envolve parâmetros (vetor média e matriz de covariância) da distribuição gaussiana multivariada, e é supervisionado, pois estima estes parâmetros através das amostras de treinamento (ERBERT, 2001).

As classes na imagem serão representadas por ω_i , $i = 1, \dots, M$, onde M é o número total de classes. Um pixel X pertencerá à classe ω_i se a probabilidade que associa este pixel à classe ω_i for maior do que a probabilidade que associa o pixel a qualquer outra classe (RICHARDS & JIA, 2006):

$$X_i \in \omega_i, \text{ se } P(\omega_i|X) > P(\omega_j|X), \text{ para todo } \omega_i \neq \omega_j \quad (1)$$

Onde $P(\omega_i | X)$ é a probabilidade de ω_i ocorrer na posição X .

2.2 Rede Neural Artificial do tipo MLP

Segundo Haykin (1999), uma RNA pode ser definida como um processador maciçamente paralelo distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso.

As unidades básicas das redes neurais, os neurônios artificiais, simulam as funções dos neurônios naturais e processam todas as informações das RNA.

A modelagem do neurônio biológico, visando gerar o modelo do neurônio artificial, foi iniciada por McCulloch e Pitts (1943) e seu modelo mais simples está exposto na Figura 1.

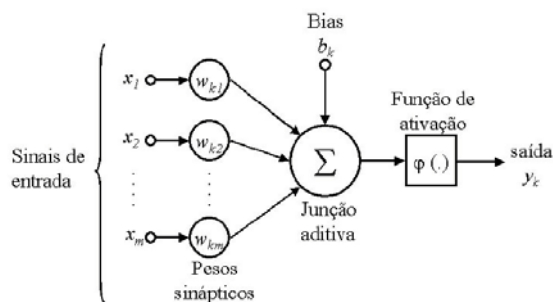


Fig 1 – Modelo do neurônio artificial
Fonte: Haykin, 1999

Este modelo é constituído de um conjunto de sinapses, sendo cada uma dessas caracterizada por um peso (w); um somador (Σ) para somar os sinais de entrada, ponderado pelas respectivas sinapses do neurônio; e uma função de ativação (ϕ) para restringir a amplitude de saída de um neurônio, ou seja, limitar o intervalo permissível de amplitude do sinal de saída a um valor finito (HAYKIN, 1999).

Observando-se a Figura 1, tem-se que: x_1, x_2, \dots, x_m são os elementos do vetor de entrada; w_1, w_2, \dots, w_{kj} são os pesos (ou conexões sinápticas); b_k : limiar de ativação do neurônio; $\phi(\cdot)$ é a função de ativação e y_k é o sinal de saída do neurônio.

Os elementos do vetor de entrada x_1, x_2, \dots, x_m são multiplicados pelos pesos w_1, w_2, \dots, w_{km} . Desta forma, o neurônio k é descrito pelo seguinte par de Equações. Assim:

$$u_k = \sum_{j=1}^m w_{kj} \cdot x_j \quad (2)$$

e

$$y_k = \phi(u_k + b_k) \quad (3)$$

O campo local induzido (v_k) é dado por:

$$v_k = u_k + b_k \quad (4)$$

A função de ativação processa o conjunto de entradas recebidas e o transformar em estado de ativação. Os neurônios podem assumir os seguintes estados de ativação: binários (0 e 1), bipolares (-1 e 1) e reais.

A Equação 5 corresponde à função de ativação do tipo tangente hiperbólica.

$$\phi(u) = \frac{1 - e^{-u}}{1 + e^{-u}} \quad (5)$$

Uma RNA do tipo Multi Layer Perceptron (MLP) possui uma arquitetura constituída por, pelo menos, três camadas de neurônios artificiais (HAYKIN, 1999). A primeira é chamada de camada de entrada e tem a função de armazenar os dados de

entrada na rede. As camadas intermediárias são consideradas extratoras de características. E a camada de saída é onde o resultado final é concluído e apresentado.

O número de neurônios e de camadas é função do problema em questão. Na classificação de imagens de satélite, por exemplo, o número de neurônios na camada de entrada poderá corresponder ao número de bandas utilizadas. Desta forma, para cada neurônio é atribuído o valor da resposta espectral do *pixel* do conjunto de treinamento da banda correspondente. Na camada intermediária o número de neurônios e de camadas ficará a critério do usuário e na camada de saída, a quantidade de neurônios corresponde ao número de classes para as quais os padrões serão designados.

A primeira camada de neurônios usa os padrões de entrada da rede e, por meio do cálculo executado pela aplicação da função de transferência, gera a resposta que alimenta a camada seguinte, e assim sucessivamente até que a última camada seja atingida.

O treinamento das RNAs faz com que esta aprenda por meio de exemplos. Haykin (1999) define aprendizagem como: “Um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente na qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira como a modificação dos parâmetros ocorre”.

Um algoritmo de aprendizagem largamente utilizado nas RNAs do tipo MLP é o *backpropagation*. Neste algoritmo, os vetores de entrada do conjunto de amostras de treinamento são fornecidos à rede. Esta, por sua vez, calcula uma resposta utilizando os valores dos pesos iniciais. A resposta obtida será comparada à saída desejada da rede. A intenção do treinamento é diminuir gradativamente o valor desse erro. Assim, o valor dos pesos será ajustado a cada nova iteração. O processo de aprendizagem continuará acontecendo enquanto a resposta calculada não for igual à desejada ou até que o número de ciclos chegue ao fim (HAYKIN, 1999).

As redes do tipo MLP são utilizadas para a classificação de padrões não linearmente separáveis, ou seja, padrões que caem em determinadas posições de um hiper-espaço nas quais não podem ser separados por um hiperplano. Esta capacidade de classificar dados não lineares a posiciona como uma rede adequada a solucionar problemas mais próximos da realidade, tais como: no processamento de imagens (LIN, 1991), no mapeamento de florestas tropicais (CARVALHO et al., 2004), na classificação de terras para irrigação por meio da análise dos parâmetros de produtividade do solo (BUCENE & RODRIGUES, 2004), dentre outros.

2.3 Support Vector Machine (SVM)

Uma SVM é basicamente uma máquina linear, cuja idéia principal é construir um hiperplano

como superfície de decisão de tal forma que a margem de separação entre exemplos positivos e negativos seja máxima (HAYKIN, 1999). Este método foi desenvolvido pelo pesquisador Vladimir Vapnik e colaboradores (BOSE et al., 1992; CORTES & VAPNIK, 1995; VAPNIK, 1995).

O treinamento da SVM envolve a resolução de um problema quadrático, dependente dos vetores de treinamento e de parâmetros especificados pelo usuário. Neste método, a partir de um espaço de entrada de padrões não-linearmente separáveis é formado um novo espaço de características, em dimensão outra, onde os padrões serão linearmente separáveis. Assim, um hiperplano de separação ótimo entre os exemplos é construído (VAPNIK, 1995).

Como principais vantagens do método pode-se citar: possui teoria bem estabelecida dentro da matemática e estatística; apresenta robustez em dados de grandes dimensões como, por exemplo, imagens de satélite; ótima capacidade de generalização, ou seja, apresenta eficiência para prever dados que não pertençam ao conjunto utilizado no treinamento (SMOLA et al., 2000).

De acordo com Vapnik (1995), para construir o hiperplano ótimo em casos onde os dados não são linearmente separáveis deve-se introduzir as variáveis de folga $\{\xi_i\}_{i=1}^N$ na definição do hiperplano de separação e encontrar os valores ótimos do vetor de peso w e do *bias* b de modo que satisfaçam a restrição:

$$d_i(w^T x_i + b) \geq 1 - \xi_i \quad (6)$$

$$\xi_i \geq 0, i = 1, \dots, N \quad (7)$$

As variáveis de folga ξ_i são responsáveis por medir a quantidade de violações das restrições (OSUNA et al., 1997).

O vetor peso w e as variáveis de folga ξ_i devem minimizar a seguinte função:

$$\Phi(w, \xi) = \left[\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \right] \quad (8)$$

Onde C é o parâmetro de penalização.

Sujeito a:

$$d_i(w^T x_i + b) - 1 + \xi_i \geq 0 \quad (9)$$

$$\xi_i \geq 0 \quad i=1, \dots, N. \quad (10)$$

O primeiro termo na Equação 8 tem a função de controlar a capacidade de aprendizagem, enquanto que o segundo termo visa penalizar as amostras localizadas no lado incorreto do hiperplano. Este termo é visto como um parâmetro de regularização e deve ser definido pelo usuário. Quanto menor o valor de C mais rígida será a penalização dos pontos; caso contrário, a

penalização será menos rígida e a classificação terá maior quantidade de erros (OSUNA et al., 1997).

O problema de otimização gerado é quadrático, com as restrições lineares apresentadas em 9 e 10. Esse problema de otimização é resolvido usando o método dos multiplicadores de Lagrange que maximizem a seguinte função:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j (x_i^T x_j) \quad (11)$$

Sujeito às restrições:

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (12)$$

$$0 \leq \alpha_i \leq C \quad \text{para } i = 1, 2, \dots, N \quad (13)$$

A função de separação do hiperplano é escrita como (VAPNIK, 1995):

$$f(x) = \text{sign} \left(\sum_{i=1}^N d_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}) + b_o \right) \quad (14)$$

A SVM é construída projetando o conjunto de treinamento (\mathbf{x}) de seu espaço original, referenciado como de entradas, através de uma função de mapeamento (Φ), para o espaço de característica de alta dimensionalidade (H). Após essa projeção é construído o hiperplano ótimo que separa as classes (VAPNIK, 1995).

O teorema de Cover (1965) afirma que um espaço multidimensional pode ser transformado com alta probabilidade em um novo espaço de características onde os padrões são linearmente separáveis. Porém, duas condições devem ser satisfeitas: A transformação deve ser não linear e a dimensionalidade do espaço de características deve ser suficientemente alta.

Para realizar o mapeamento, aplica-se a função de mapeamento (Φ) aos exemplos presentes no problema de otimização apresentado na Equação 11, conforme ilustrado a seguir:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) \quad (15)$$

Sujeito às restrições apresentadas em (12) e (13). Assim, têm-se a função de separação do hiperplano é reescrita como:

$$f(x) = \text{sign} \left(\sum_{i=1}^N d_i \alpha_i \Phi(\mathbf{x}_i) \cdot (\mathbf{x}_j) + b_o \right) \quad (16)$$

O algoritmo de aprendizagem por vetor de suporte pode ser usado para construir diversos tipos de máquinas como, por exemplo: as polinomiais, redes de função de base radial e *perceptrons* de duas camadas. A construção destas dependerá de como os dados são projetados do espaço de entrada para o espaço característico (HAYKIN, 1999).

A função usada para realizar essa projeção é chamada de funções kernel. Em outras palavras, um kernel K é uma função que recebe dois pontos \mathbf{x}_i e \mathbf{x}_j do espaço de entradas computa o produto escalar desses dados no espaço de características (HAYKIN, 1999). Assim, têm-se a seguinte Equação:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (17)$$

Neste estudo escolheu-se a função kernel do tipo Gaussiana (OSUNA et al., 1997), que tem a representação matemática apresentada na Equação 18.

$$K(x_i, x_j) = \exp \left(- \frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \quad (18)$$

Sendo que σ é interpretado como a largura da gaussiana e deve ser especificado a priori pelo usuário.

Esta função é uma das mais comumente aplicadas em dados provenientes do sensoriamento remoto, como pode ser observado nos trabalhos de Pal & Mather (2004) e Foody & Mathur (2004). De acordo com Chapelle et al. (1999), estas funções devem ser escolhidas cuidadosamente para evitar desempenhos ruins pelo classificador. O percentual de acertos de uma classificação pelo método SVM é dependente da magnitude do parâmetro C e, quando se faz uso da função kernel do tipo gaussiana, do parâmetro σ .

As SVMs são originalmente classificadores binários, porém, a maioria dos dados reais envolve grande quantidade de informação e pode necessitar ser rotulada em mais de duas classes. Basicamente, duas estratégias são usadas para solucionar este tipo de problema: *One-Against- One* (OAO) e *One-Against-All* (OAA) (WESTON et al., 1999).

A estratégia OAO constrói $N(N-1)/2$ funções discriminantes, em que N é o número de classes. Um processo de escolha será aplicado para definir em que classe o dado x deverá ser classificado e a decisão será tomada pelo número de vezes que este foi associado à determinada classe pela função discriminante apresentada na Equação 27 (WESTON et al., 1999).

Na estratégia OAA cada SVM existente resolve um problema de separação entre duas classes. Por exemplo: em um determinado problema de classificação de padrões com N classes, serão criados N classificadores binários e cada um desses será treinado para distinguir uma classe das restantes, ou seja, das $N-1$ classes (WESTON et al. 1999). A decisão final sobre qual classe o dado pertencerá é dada pela regra “o vencedor leva tudo”, ou seja, a classe

vencedora é aquela que possui maior valor de função discriminante, $f(x)$ (MELGANI, 2004). Maiores detalhes sobre a teoria das estratégias apresentadas podem ser encontrados nos trabalhos de Weston et al. (1999) e Hsu et al. (2002).

Gidudu et al. (2004), avaliaram o uso de estratégias multiclases na classificação de imagens da região de Jinja, Uganda. Os dados de entrada corresponderam às 7 bandas do satélite LANDSAT e as classes de saída foram água, vegetação e antropizada. Os testes realizados envolveram as estratégias one-against-one (OAO) e one-against-all (OAA) em quatro tipos de classificadores SVM, a saber: linear, quadrática, polinomial e gaussiana. A estratégia indicada foi a OAO já que a OAA teve maior propensão de gerar pixels não classificados ou erroneamente classificados. Para a SVM com classificador gaussiana, por exemplo, foram 103 pixels classificados erroneamente para a estratégia OAO e 4645 para a OAA.

2.4 Distância de Jeffries-Matusita

Uma forma de avaliar a classificação se dá por meio de medidas estatísticas. De acordo com Richards & Jia (2006), a distância de Jeffries-Matusita (JM) entre duas classes i e j que apresentem dados com distribuição normal é definida por:

$$JM_{ij} = 2 * (1 - e^{-B_{ij}}) \quad (16)$$

Em que JM_{ij} representa a distância entre as classes i e j ; e B_{ij} é a distância de Bhattacharyya que é calculada por:

$$B_{ij} = \frac{1}{8} (M_i - M_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (M_i - M_j) + \frac{1}{2} \left(\frac{\frac{\Sigma_i + \Sigma_j}{2}}{\sqrt{|\Sigma_i| |\Sigma_j|}} \right) \quad (17)$$

Sendo M_i e M_j os vetores de média das classes i e j , respectivamente e Σ_i e Σ_j as matrizes de covariância das classes i e j , respectivamente.

Quanto maior o valor estimado por B_{ij} , maior a separabilidade entre classes i e j . O valor de JM_{ij} entre duas classes poderá variar entre 0 e 2. Valores acima de 1,8 indicam que o par de classes obteve boa separabilidade. Já valores abaixo de 1,8 indicam a possibilidade de confusão na classificação entre o par de classes (RICHARDS & JIA, 2006).

2.5 Validação da exatidão do mapeamento

Um mapeamento do uso do terra, apoiado em sensoriamento remoto, necessita ser realizado com exatidão para que as informações geradas sejam confiáveis. Para a avaliação da exatidão do mapeamento são usados os coeficientes de

concordância obtidos através de uma matriz de confusão.

A matriz de erro ou matriz de confusão é uma matriz quadrada de números que expressam a quantidade de unidades amostrais, associada a uma dada categoria durante o processo de classificação efetuado, e à categoria real a que pertencem essas unidades. (MANGABEIRA, 2003 apud CONGALTON 1991).

A exatidão global (EG) é obtida pela divisão das somas das entradas que formam a diagonal principal da matriz, ou seja, o número de classificações corretas, pelo número total de amostras utilizadas como mostra a Equação 18 (WATZLAWICK, 2003).

$$EG = \frac{\sum_{i=1}^r x_{ii}}{N} * 100 \quad (18)$$

sendo: r = número de linhas da matriz de erro; x_{ii} = valor da linha i e coluna i ; N = quantidade total de pontos amostrais.

A Exatidão Específica (EE) é obtida pela divisão do número de amostras classificadas corretamente na classe pelo número total de amostras desta, como apresentada a seguir:

$$EE = \frac{x_{ii}}{n} * 100 \quad (19)$$

sendo: x_{ii} = valor na linha i e coluna i e n = quantidade total de pontos amostrais da classe r .

A análise de Kappa, diferentemente da EG, utiliza todos os elementos da matriz de confusão para seu cálculo. O coeficiente Kappa de concordância (K) é uma medida da concordância real (indicado pelos elementos diagonais da matriz de confusão) menos a concordância por chance (indicado pelo produto total da linha e coluna) (MOREIRA, 2005), conforme segue:

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} x_{+i})} \quad (20)$$

Sendo: K = coeficiente Kappa de concordância; N = quantidade total de pontos amostrais; r = número de linhas da matriz de erro; x_{ii} = valor na linha i e coluna i ; x_{i+} = soma da linha i ; x_{+i} = soma da coluna i .

A Tabela 1, desenvolvida por Landis & Koch (1977), é utilizada como referência para classificar mapas resultantes da utilização de imagens de sensoriamento remoto. Nesta, são apresentados níveis de desempenho da classificação para o valor de Kappa obtido.

TABELA 1 - QUALIDADE DA CLASSIFICAÇÃO ASSOCIADA AOS VALORES DA ESTATÍSTICA KAPPA.

Valor de Kappa	Qualidade do mapa temático
<0,00	Péssima
0,00-0,20	Ruim
0,20-0,40	Razoável
0,40-0,60	Boa
0,60-0,80	Muito boa
0,80-1,00	Excelente

FONTE: adaptada de Landis & Koch (1977).

3. MATERIAL E MÉTODOS

3.1 Caracterização da área de estudo

A área de estudo está situada no município de Iguatu, no Estado do Ceará. Este município localiza-se entre os paralelos 6° 17' 42,33''S e 6° 33' 41,04''S e os meridianos 39° 3' 55,64''W e 39° 29' 28,52''W. Possui extensão territorial de 1029 km² e, segundo dados do IBGE (2009), a estimativa da população é de 97203 habitantes. Está inserido em uma região semi-árida com vegetação típica da Caatinga. O relevo é classificado como plano, suave ondulado e ondulado (JACOMINE et al., 1973). Pela classificação de Köppen (1948), o clima é do tipo BSw'h': quente com chuvas de outono e temperaturas médias superiores a 18°C.

Para a realização do experimento foi feito um recorte de uma área situada entre os paralelos 6° 19' 29,91''S e 6° 32' 39,94''S e os meridianos 39° 11' 50,18''W e 39° 24' 52,01''W. A área total é de 576 km². A Figura 2 exibe a localização da área de estudo.

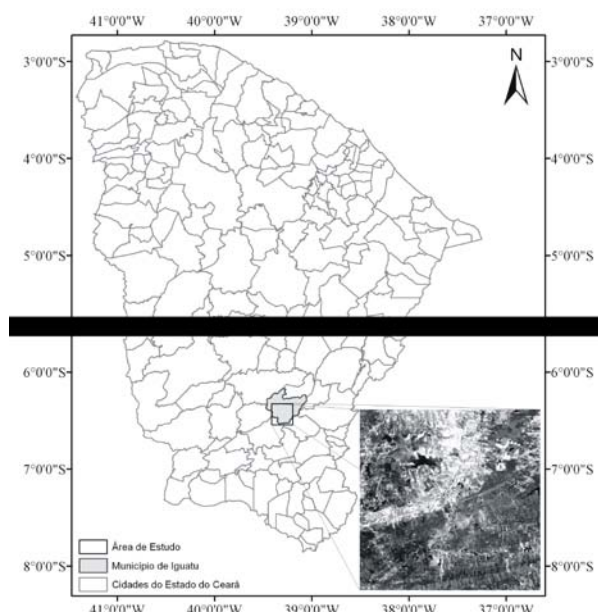


Fig 2 – Localização da área de estudo no Estado do Ceará.

3.2 Abordagem metodológica

A área foi extraída de uma cena do satélite LANDSAT-5/TM (217/65) referente ao dia 20 de agosto de 2008, adquirida por este satélite às 9 horas e cedida pelo Instituto Nacional de Pesquisas Espaciais (INPE). Foram consideradas as bandas 2 (0,50 - 0,60 μm), 3 (0,63 - 0,69 μm) e 4 (0,76 - 0,90 μm) que correspondem ao verde, vermelho e infravermelho próximo, respectivamente, e gerada, no software ENVI 4.3®, a composição colorida 432 do espaço RGB. Optou-se por utilizar somente três bandas do satélite por ser esta a quantidade de entrada padrão no software ENVI 4.3® que, por sua vez, foi utilizado na classificação supervisionada. Desta forma, todos os métodos foram testados com a mesma quantidade de bandas.

Os pontos referenciados em campo por meio do aparelho Leica GPS1200 foram utilizados no software ENVI 4.3® para reajustar o georreferenciamento da imagem do satélite LANDSAT-5/TM. Para este processo foram coletados pontos conhecidos e facilmente visualizados na imagem, como: cruzamentos de rodovias e paredes de açudes.

O fluxograma exibido na Figura 3 mostra, esquematicamente, as etapas realizadas no desenvolvimento do experimento.

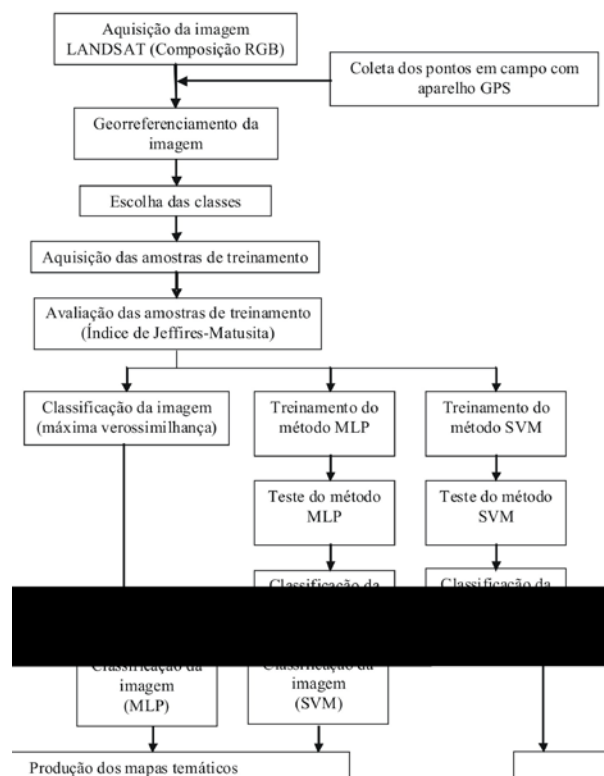


Fig 3 – Etapas e ordem de execução do experimento.

3.2.1 Escolha das classes e processo de amostragem

Na composição colorida foram selecionadas cinco classes genéricas: Classe (1) – Antropização por Agricultura (APA): engloba áreas modificadas para plantio por diferentes tipos de culturas, sendo caracterizada na imagem por polígonos de lados bem definidos; Classe (2) – Água: rios, açudes e lagoas; Classe (3) – Outros Tipos de Antropização (OTA): aglomerados urbanos ou áreas desprovidas de cobertura susceptíveis ou em processo de degradação ou erosão; Classe (4) – Caatinga Herbácea Arbustiva (CHA): vegetação herbácea arbustiva (porte baixo a médio) aberta a densa; Classe (5) – Caatinga Arbórea Densa (CAD): vegetação arbórea densa, de porte mais elevado.

Para o treinamento dos algoritmos, realizou-se o processo de amostragem das classes no software ENVI 4.3®. E, para que esses dados também pudessem ser utilizados nos classificadores baseados em aprendizado de máquina, os valores de RGB das amostras foram exportados para um formato de arquivo legível no MATLAB® 7.0 (*.txt). As amostras coletadas, 2000 pixels por cada classe, foram obtidas na mesma cena, mas fora da área de estudo, de forma a testar a capacidade de generalização dos classificadores.

3.2.2 Análise da separabilidade das amostras de treinamento

Após o processo de amostragem, analisou-se a resposta espectral nas bandas 2, 3 e 4 de cada classe e a separabilidade entre cada par de classe, por meio do índice de Jeffries-Matusita (JM). Para o cálculo, são necessários os dados de entrada M_i e M_j correspondentes aos vetores de média das classes i e j , respectivamente e C_i e C_j correspondentes as matrizes de covariância das classes i e j . O software ENVI 4.3® dispõe de uma ferramenta que realiza o cálculo, restando ao usuário fornecer as amostras de treinamento utilizadas.

3.2.3 Processo de classificação

3.2.3.1 Método da Máxima Verossimilhança

A partir das amostras de treinamento previamente definidas, foi realizada a classificação supervisionada pelo método estatístico da máxima verossimilhança no software ENVI 4.3®.

3.2.3.2 Método Perceptron de Múltiplas Camadas (MLP)

O método MLP treinado com algoritmo *backpropagation* e alimentação *feedforward*, foi implementado no software MATLAB® 7.0, por meio da *Toolbox Neural Network* (The MathWorks Inc., Natick, MA, USA) versão 6.0.

Os dados a serem classificados e o número de classes determinam a quantidade de neurônios na camada de entrada e saída, respectivamente. Desta forma, têm-se três neurônios na camada de entrada e cinco na camada de saída.

Quanto à camada escondida, como sugerido por Egmont *et al.* (2002), optou-se por utilizar uma rede MLP de única camada. Assim, os testes realizados objetivaram encontrar o número de neurônios nesta camada que proporcionassem melhor desempenho na classificação da imagem testando, para isso, 12, 15 e 18 neurônios.

A Figura 4 exemplifica a arquitetura da rede MLP correspondente ao teste com 12 neurônios na camada intermediária.

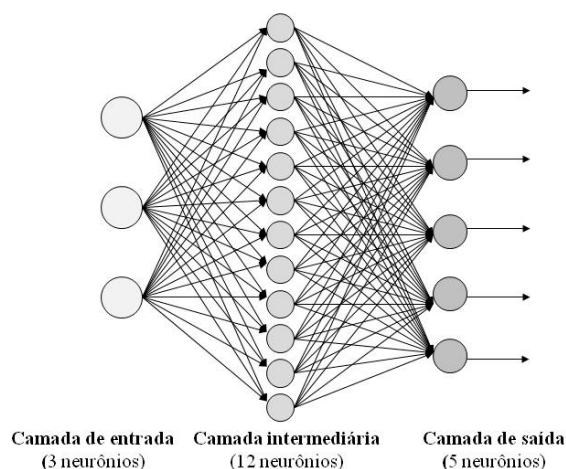


Fig 4 – Arquitetura da MLP com 12 neurônios na camada intermediária.

O critério de convergência da rede foi estabelecido por um número máximo de ciclos igual a 3000 ou erro final de 0,001. A taxa de aprendizagem (η) foi fixada em 0,005 e a função de ativação escolhida foi do tipo tangente hiperbólica (Equação 5) a qual limita os valores de saída de -1 à 1. E

A imagem correspondente à área de estudo foi transformada para o formato de vetor e submetida ao processo de classificação.

O fluxograma apresentado na Figura 5 ilustra as etapas realizadas neste método.

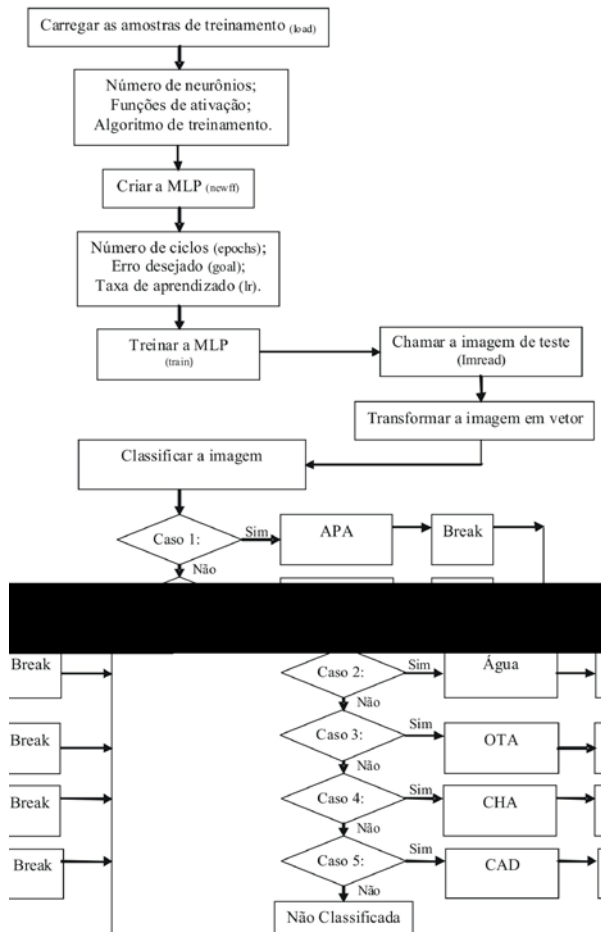


Fig 5 – Fluxograma das atividades relacionadas ao método MLP.

3.2.3.3 Método Máquina de Vetor de Suporte (SVM)

Para a implementação do método SVM no software MATLAB® 7.0 foi utilizada a SVM-KMToolbox (SVM and Kernel Methods MATLAB Toolbox) desenvolvida por Canu *et al.* (2005). Como principais referências bibliográficas utilizadas no desenvolvimento dessa *toolbox* têm-se os trabalhos de Vapnik (1995) e Osuna *et al.* (1997).

Para a classificação das cinco classes propostas adotou-se a estratégia *multiclasse one-against-one* (OAO) tendo como referência os melhores resultados obtidos nos trabalhos de Gidudu *et al.* (2004) e Melgani (2004).

Os 22 testes foram realizados por meio de combinações do parâmetro de penalização C nos valores de 10; 50; 100; 500; 1000; 1300; 1400; 1500 e 3000 com o parâmetro σ da função gaussiana (Equação 18) nos valores de 1,5; 1,8; 2,2; 2,0; 2,5; 5 e 10.

Os tempos de treinamento e teste foram cronometrados para os métodos baseados em aprendizado de máquina. Vale ressaltar que a contagem do tempo é apenas uma estimativa já que acabam sendo incluídos os tempos que o Sistema Operacional usa para que o software MATLAB funcione.

O fluxograma apresentado na Figura 6 ilustra as etapas realizadas neste método.

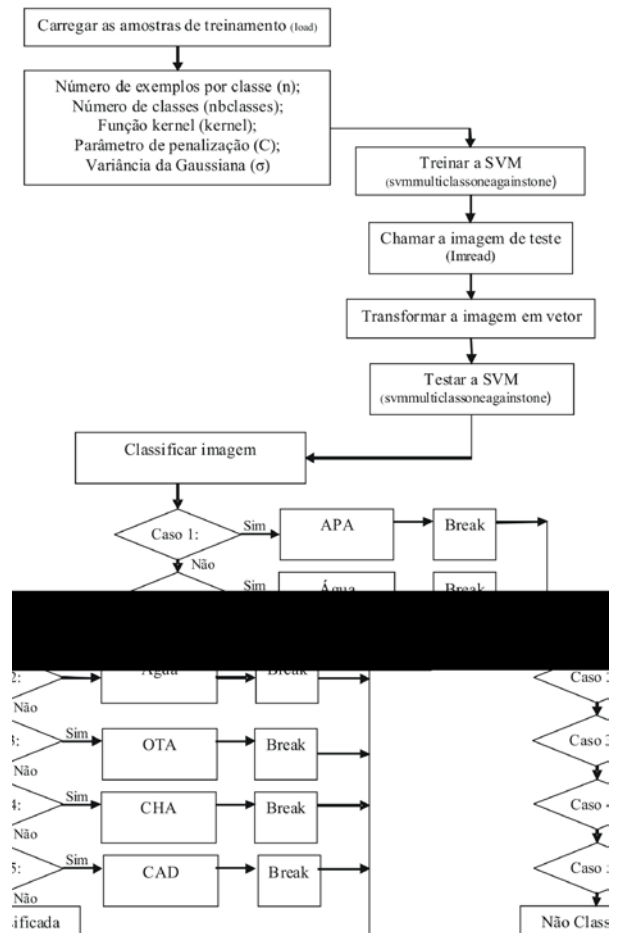


Fig 6 – Fluxograma das atividades relacionadas ao método SVM.

3.2.3.4 Validação da exatidão do mapeamento

O processo de validação foi realizado por meio da construção de uma matriz de confusão. Desta forma, foi-se a campo em busca de referenciar pontos amostrais correspondentes à verdade de campo das classes de saída.

No total, foram referenciados 112 pontos com aparelho GPS Garmim® entre os dias 29 e 31 de julho de 2008, período próximo à data de aquisição da imagem pelo satélite. Foram obtidos 10 pontos para a classe água, 14 para CAD, 27 para CHA, 39 para APA e 22 para OTA. Todos estes pontos estão localizados dentro dos limites da área de estudo e não coincidem com a localização das amostras de treinamento, pois, apesar destas estarem contidas na mesma cena (217/65), foram selecionadas fora desta área visando testar a capacidade de generalização dos classificadores. O desenho esquemático exposto na Figura 7 exemplifica a localização das amostras de treinamento e de validação.

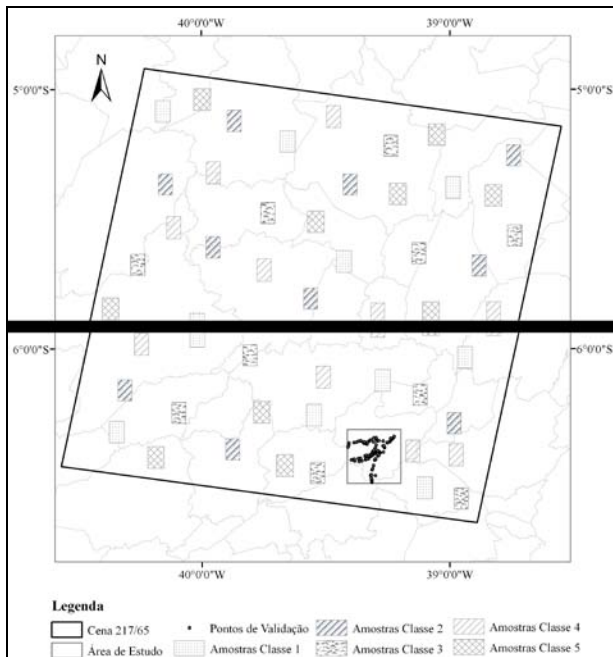


Fig. 7 – Desenho esquemático da localização das amostras de treinamento e de validação.

Para a validação pelo método da Máxima Verossimilhança, os pontos do GPS foram exportados para formato *.shp e sobrepostos à imagem classificada no software ArcGis® 9.3. Desta forma, os dados da verdade de campo foram comparados manualmente à classificação gerada e cada acerto ou erro contabilizado na matriz.

Já para os métodos baseados em aprendizado de máquina foram obtidos na composição colorida os valores dos pixels correspondentes a cada ponto referenciado em campo para as três bandas utilizadas. Esses valores foram digitados em um arquivo no formato *.txt e submetidos aos algoritmos no software MATLAB® 7.0 de forma a servirem de dados de teste para estes métodos. A classificação resultante do processamento foi comparada à verdade de campo e os acertos e erros contabilizados na matriz.

As matrizes geradas forneceram os dados necessários para o cálculo dos coeficientes de Exatidão Global, Exatidão Específica e de Kappa através das Equações 14, 15 e 16, respectivamente.

4. RESULTADOS E DISCUSSÃO

4.1 Análise da separabilidade das classes

As respostas espectrais das amostras de treinamento foram analisadas com o objetivo de avaliar a separabilidade entre as classes de interesse e, por consequência, obter um indicativo sobre a qualidade da classificação. A Figura 8 apresenta a resposta espectral média das amostras de treinamento das classes água, outros tipos de antropização (OTA), antropização por agricultura (APA), CHA e CAD nas bandas 2, 3 e 4 do satélite LANSAT-5.

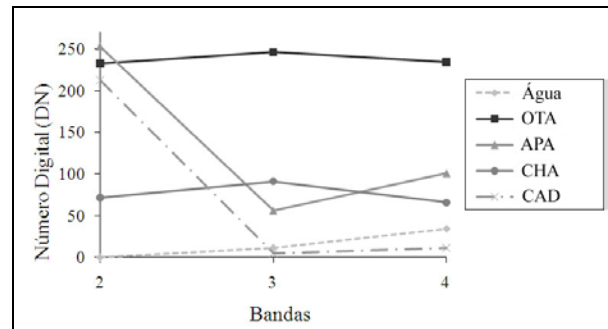


Fig. 8 - Resposta espectral média das diferentes classes de cobertura da terra das amostras de treinamento.

Por meio do gráfico apresentado na Figura 8 é possível notar que, em geral, as classes possuem respostas espectrais distintas. Nota-se ainda que, das classes em estudo, as de APA e CAD compõem o par com os valores mais próximos.

A Tabela 2 expõe os valores dos índices de Jeffries-Matushita entre os pares de classes.

TABELA 2 - ÍNDICE DE SEPARABILIDADE ENTRE AS CLASSES ESTUDADAS

Classes	Água	OTA	APA	CHA	CAD
Água	-	-	-	-	-
OTA	2,0	-	-	-	-
APA	2,0	1,99	-	-	-
CHA	1,97	1,98	1,99	-	-
CAD	1,99	2,0	1,89	1,99	-

Na Tabela 2, nota-se que os pares de classes água e APA, água e OTA e OTA e CAD apresentaram o valor máximo deste índice sendo, desta forma, os que possuem maior separabilidade. Nota-se ainda que o menor valor de JM obtido, de 1,89, ocorreu para o par de classes APA e CAD, confirmando o que foi concluído na Figura 8.

Todos os pares de classes apresentaram valores do índice de Jeffries-Matusita acima do limiar, o que aponta para uma boa separabilidade das classes pelo ponto de vista da análise espectral.

4.2 Classificação por meio do método Perceptron de Múltiplas Camadas (MLP)

Na fase de treinamento, a convergência da rede MLP ocorreu por conta do número máximo de iterações estabelecido. A Tabela 3 mostra os parâmetros fornecidos à rede neural MLP em cada teste realizado, com os respectivos valores de erro final, EG, K e o tempo computacional necessário.

TABELA 3 - RESULTADOS DOS TESTES PELO MÉTODO MLP

Teste	Número de neurônios	Erro final	Exatidão Global (%)	Kappa (K)	Tempo (s)
1	12	0,008	82,14	0,76	657,5
2	15	0,007	81,25	0,75	875,9
3	18	0,008	70,53	0,61	1103,6

Por meio da Tabela 3 é possível observar que a maior capacidade de mapeamento dos padrões pela rede MLP ocorreu no teste 1, ou seja, para a arquitetura 3-12-5, já que apresentou maiores valores de EG e K, 82,14% e 0,76, respectivamente. O desempenho caiu com o aumento do número de neurônios, podendo ter ocorrido um *overfitting*. Nota-se ainda que, além dos maiores destes índices, o tempo de processamento foi o mais rápido quando comparado aos dos testes 2 e 3.

De acordo com a classificação proposta por Landis & Koch (1977), o mapa temático gerado é classificado como muito bom. Gelete & Volatão (2007), assim como no presente estudo, utilizaram a *toolbox Neural Network* contida no software MATLAB e obtiveram resultados satisfatórios com uma rede MLP com 11 neurônios na camada oculta.

4.3 Classificação por meio do método Máquina de Vetor de Suporte (SVM)

A classificação por meio do método SVM foi testada por meio da variação do fator de penalização (C) e da largura da função gaussiana (σ).

Objetivando analisar o comportamento de σ , foram escolhidos para C os valores de 100, 1000 e 1500 e, para cada um destes, foram realizados testes variando σ entre 1,5 e 10 (1,5; 1,8; 2,0; 2,2; 2,5; 5 e 10). A Tabela 4 expõe os valores de EG, K e o custo computacional de cada teste.

TABELA 4 - DESEMPENHO DO CLASSIFICADOR SVM COM A VARIAÇÃO DE Σ

Teste	Fator C	σ	EG (%)	K	Tempo (s)
1	1000	1,8	82,14	0,766	350.9
2	1000	2	86,03	0,778	210.6
3	1000	2,2	81,03	0,752	266.9
4	1500	1,5	80,35	0,741	260.8
5	1500	2	82,14	0,766	207.1
6	1500	2,5	81,25	0,755	268.9
7	1500	5	80,35	0,742	276.6
8	1500	10	79,46	0,731	301.2
9	100	1,5	80,35	0,741	243.2
10	100	2	86,03	0,778	276.6
11	100	2,5	81,25	0,756	269.2
12	100	5	78,57	0,719	283.7
13	100	10	78,57	0,719	325.2

Pela análise desta Tabela nota-se que, para as amostras de treinamento utilizadas, os melhores valores de EG e K ocorrem com $\sigma=2$. É possível constatar que, para os três valores de C testados, e para valores de σ superiores a 2 os coeficientes de EG e K tendem a decrescer, pois com o aumento da largura da função gaussiana pode ocorrer maior confusão na classificação dos dados.

A partir deste resultado, os demais testes foram realizados fixando σ no valor 2 e alterando o fator de penalização C entre 10 e 5000. A Tabela 5 expõe o resultado dos testes com seus respectivos valores de EG, K e o custo computacional de cada teste com a variação de C.

TABELA 5 - DESEMPENHO DO CLASSIFICADOR SVM COM A VARIAÇÃO DO FATOR C

Teste	Fator C	Sigma	EG (%)	K	Tempo (s)
14	10	2	81,25	0,755	215.2
15	50	2	82,14	0,767	202.6
16	100	2	86,03	0,778	276.6
17	500	2	86,03	0,778	206.9
2	1000	2	86,03	0,778	210.6
18	1300	2	83,03	0,778	337.3
19	1400	2	82,14	0,766	294.4
20	1500	2	82,14	0,766	207.1
21	2000	2	82,14	0,766	274.5
22	3000	2	80,35	0,741	303.2
23	4000	2	80,35	0,741	283.7
24	5000	2	80,35	0,741	327.6

Pela análise da Tabela 5, nota-se que o valor de K apresentou tendência de crescimento com C entre 10 e 1300. Porém, após estes valores, o coeficiente K passou a decrescer. Com valores de C maiores do que 1000 os coeficientes de EG e K tendem a decrescer, já que a penalização é menos rígida e a classificação apresenta maior quantidade de erros.

A classificação realizada pelo método SVM, correspondente ao teste 2, obteve valor de Exatidão Global de 86,03% e coeficiente de Kappa igual a 0,77. Segundo Landis & Koch (1977), o mapa temático resultante é rotulado como muito bom.

Outro ponto importante a ser analisado é o custo computacional envolvido no processo de classificação. Para o método SVM, como pode ser observado na Tabela 5, o teste 2 realizou o processo de classificação em 210,51 segundos. Já para no método MLP, o teste 1 classificou a imagem em 657,5 segundos. Desta forma, pode-se afirmar que o SVM apresentou menor custo computacional do que o método MLP ratificando os resultados encontrados por Foody & Mathur (2004).

4.4 Validação das classificações

As matrizes de confusão, correspondentes ao teste 1 do método MLP e ao teste 2 do método SVM, estão expostas nas Tabelas 6 e 7, respectivamente. A Tabela 8 apresenta a matriz de confusão correspondente ao método da Máxima Verossimilhança. Os números 1, 2, 3, 4 e 5 representam as classes APA, água, OTA, CHA e CAD, respectivamente. O total de linhas corresponde à soma dos pontos nas linhas e o total de colunas à soma dos pontos nas colunas.

TABELA 6 - MATRIZ DE CONFUSÃO PARA O MÉTODO MLP

Classes	Verdade de Campo					Total Linhas	EE (%)	
	1	2	3	4	5			
Classificação	1	33	1	4		1	39	84,6
	2	1	7			2	10	70,0
	3	1		19	2		22	86,4
	4			6	21		27	77,8
	5				2	12	14	85,7
Total colunas	35	8	29	25	15	112	-	

TABELA 7 - MATRIZ DE CONFUSÃO PARA O MÉTODO SVM

Classes	Verdade de Campo					Total Linhas	EE (%)	
	1	2	3	4	5			
Classificação	1	30		7		2	39	76,9
	2		10				10	100,0
	3			21	1		22	95,5
	4	3		4	20		27	74,1
	5	1			1	12	14	85,7
Total colunas	34	0	32	22	14	112	-	

TABELA 8 - MATRIZ DE CONFUSÃO PARA O MÉTODO DA MÁXIMA VEROSSIMILHANÇA

Classes	Verdade de Campo					Total Linhas	EE (%)	
	1	2	3	4	5			
Classificação	1	34		4		1	39	87,2
	2		8		2		10	80,0
	3	3		16	3		22	72,7
	4	1		3	23		27	85,2
	5	2			2	10	14	71,4
Total colunas	40	8	23	30	11	112	-	

Por meio dos dados apresentados na Tabela 8 foram calculados os valores de EG e K para o método da Máxima Verossimilhança. O coeficiente de EG permitiu constatar que neste método 81,2% dos pixels foram classificados corretamente, e que o coeficiente de K foi 0,75.

Assim como nos estudos de Foody & Mathur (2004) e Pal & Mather (2004) é possível observar, por meio dos valores de K, que a SVM obteve desempenho satisfatório quando comparada aos métodos MLP e da Máxima Verossimilhança.

Para melhor avaliar o desempenho dos classificadores em cada classe calculou-se, por meio da Equação 15, o coeficiente de Exatidão Específica (EE) para cada método. Os resultados estão sumarizados na Tabela 9.

TABELA 9 - EXATIDÃO ESPECÍFICA DAS CLASSES

Classe	Exatidão Específica (%)		
	Máxima Verossimilhança	MLP	SVM
APA	87,2	84,6	76,9
Água	80,0	70,0	100,0
OTA	72,7	86,4	95,5
CHA	85,2	77,8	74,1
CAD	71,4	85,7	85,7

Para a classe APA, o maior nível de acurácia foi obtido no método da Máxima Verossimilhança, com valor de EE de 87,2%. Para as classes água e

antropizada obteve-se valores de 100 e 95,5%, respectivamente, quando classificadas pelo método SVM sendo este, portanto, o método mais indicado para a classificação destas classes. Já para a classe CHA o maior valor de EE obtido foi de 85,2% pelo método da Máxima Verossimilhança. E para a classe CAD, obteve-se 85,7% de EE para os métodos SVM e MLP. Os resultados expostos apontam para o bom desempenho dos métodos baseados em aprendizado de máquina segundo os critérios de Landis & Kock (1977). Assim como no trabalho de Su *et al.* (2007), os resultados obtidos permitem afirmar que estes métodos estão aptos à classificação do uso e cobertura da terra no bioma Caatinga, já que apresentaram resultados equivalentes e até superiores ao método da Máxima Verossimilhança.

5. CONCLUSÕES

As amostras de treinamento apresentaram boa separabilidade, o que contribuiu para o desempenho satisfatório dos classificadores utilizados.

Nos testes realizados, os classificadores baseados em aprendizado de máquina apresentaram maior acurácia quando comparados ao classificador tradicional da Máxima Verossimilhança.

Os valores dos coeficientes de Exatidão Global (EG), de Exatidão Específica (EE) e de Kappa (K), obtidos a partir da verdade terrestre, revelam o alto potencial dos métodos baseados em aprendizado de máquina na classificação do uso e cobertura da terra no bioma Caatinga.

Em comparação ao método da Máxima Verossimilhança, classificadores baseados em aprendizado de máquina podem consumir maior tempo computacional além da necessidade do usuário depender boa parte do tempo na escolha dos parâmetros do método a ser utilizado, o que pode ser extremamente desgastante. Assim, caso o tempo seja um fator limitante, pode-se utilizar o método da Máxima Verossimilhança.

AGRADECIMENTOS

Às instituições que apoiaram esta pesquisa: Universidade Federal do Ceará (UFC), Instituto Nacional de Pesquisas Espaciais (INPE) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro desta. Ao Dr. Rakotomamonjy por disponibilizar o pacote de software SVM-KMToolbox para este estudo. Aos Engenheiros Agrônomos Clênio Jario, Alexandre Gomes Costa e Amaury pela colaboração na exaustiva coleta dos pontos no campo.

REFERENCIAS BIBLIOGRÁFICAS

ANDRADE, L. A.; PEREIRA, I. M. P.; LEITE, U. T.; BARBOSA, M. R. V. Análise da cobertura de duas fitofisionomias de Caatinga, com diferentes históricos

de uso, no município de São João do Cariri, Estado da Paraíba. **Cerne**, Lavras, v. 11, n. 3, p. 253-262, 2005.

ASSAD, E. D.; SANO, E. E. **Sistema de informações geográficas: aplicações na agricultura**. 2. ed. Brasília: EMBRAPA-SPI/EMBRAPA-CPAC, 1998.

BOSER, B.; GUYON, I; VAPNIK, V. N. A training algorithm for optimal margin classifiers. Annual Workshop on Computational Learning Theory, 5., 1992, Pittsburgh. **Proceedings...** SanMateo, CA: Morgan Kayfmann, 1992. p. 144-152. Disponível em: <<http://portal.acm.org/citation.cfm?id=130401>>. Acesso em: 15 dez. 2008.

BUCENE, L. C.; RODRIGUES, L. H. A. Utilização de redes neurais artificiais para avaliação de produtividade do solo, visando classificação de terras para irrigação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.8, n.2/3, p. 326-329, maio/dez. 2004.

CANU, S.; GRANDVALET, Y.; GUIGUE, V.; RAKOTOMAMONJY, A. **SVM and Kernel Methods MATLAB Toolbox**. Perception Systèmes et Information, INSA de Rouen, Rouen, France, 2005.

CARVALHO, L. M. T; CLEVERES, J. G. P. W. ; SKIDMORE, A. K.; JONG, S. M. J. Selection of imagery data and classifiers for mapping Brazilian semideciduous Atlantic forests. **International Journal of Applied Earth Observation and Geoinformation**, v. 5, n. 5, p. 173-186, 2004.

CHAPELLE, O.; HAFFNER, P.; VAPNIK, V. N. Support Vector Machines for Histogram-Based Image Classification. **IEEE Transactions On Neural Networks**, v. 10, n. 5, p. 1055-1064, 1999.

CORTES, C.; VAPNIK, V. Support vector networks. **Machine Learning**. v. 20, n. 3, p. 273-297, set. 1995

COVER, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. **IEEE Transactions on Electronic Computers**, v. 25, n. 3, p. 326-334, jun. 1965.

EGMONT, M. P.; RIDDER, D. de; HANDELS, H. Image processing with neural networks—a review. **Pattern Recognition**, v. 35, n. 10, p. 2279-2301, 2002.

ERBERT, M. **O Uso da Análise Discriminante Regularizada (RDA) no Reconhecimento de Padrões em Imagens Digitais Hiperespectral de Sensoriamento Remoto**. 2001. 118 f. Dissertação (Mestrado em Sensoriamento Remoto) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001.

FOODY, G. M.; MATHUR, A. A Relative Evaluation of Multiclass Image Classification by Support Vector

Machines. **IEEE Transactions on Geoscience And Remote Sensing**, v. 42, n.6, p. 1335-1343, 2004.

GELETE, C.; VOLATÃO, C. F. de S. Análise do potencial de classificação do uso e cobertura do solo por meio de rede neural. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007. Florianópolis. **Anais...** Florianópolis: INPE, 2007. p. 5779-5785. CD-ROM, On-line. ISBN 978-85-17-00031-7. Disponível em: <<http://urlib.net/dpi.inpe.br/sbsr@80/2006/10.20.13.05>>. Acesso em: 13 out. 2008.

GIGANDET, X.; CUADRA, M. B.; POINTET, A.; CAMMOUN, L.; CALOZ, R.; THIRAN, J. Region-based satellite image classification: method and validation. **Image Processing**, v. 3, n. 4, p. 832-835, Set.2005.

HAYKIN, S. **Neural Networks: A Comprehensive Foudation**. 2. ed. New Jersey, Prentice Hall, 1999. 842 p.

HSU, C. W.; LIN, C. J. A Comparison of Methods for Multiclass Support Vector Machines. **IEEE Transactions On Neural Networks**, v. 13, n. 2, p. 415-425, mar. 2002.

IBGE. 2009. **Estimativas populacionais para os municípios brasileiros em 01.07.2009**. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística (IBGE). Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/estimativa2009/POP2009_DOU.pdf>. Acesso em 01 de abril. 2010.

IWAI, O. K. **Mapeamento do uso do solo urbano no município de São Bernardo do Campo, através de imagens de satélites**. 2003. 127 f. Dissertação (Mestrado em Engenharia de Transportes) – Escola Politécnica da Universidade de São Paulo. São Paulo, 2003.

JACOMINE, P. K. T.; ALMEIDA, J. C.; MEDEIROS, L. A. R. **Levantamento exploratório-reconhecimento de solos do estado do Ceará**. Recife, Ministério da Agricultura/ Superintendência do Desenvolvimento do Nordeste, 1973, v.1. 301p. (Boletim técnico, 28; Série Pedologia, 16).

LANDIS, J. R.; KOCH, G. C. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.

LEDUC, R.; GRAÇA, P.M.L. de A.; PEIXOTO, J.M.A. Classificação da cobertura do solo da Reserva Extrativista do Baixo – Juruá/AM. . In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2007, 13., Florianópolis. **Anais...** Florianópolis: INPE, 2007. p. 5329-5336. CD-ROM, On-line. ISBN 978-85-

17-00031-7. Disponível em: <<http://mar.te.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.16.02.53.01/doc/6787-6789.pdf>>. Acesso em: 13 mar. 2009.

LIN, W. C. A hierarchical multiple view approach to three-dimensional object recognition. **IEEE Transactions On Neural Networks**, v. 2, n. 1, p. 84-92, jan. 1991.

MANGABEIRA, J. A. de C., AZEVEDO, E. C. de, LAMPARELLI, R. A. C. **Avaliação do levantamento do uso das terras por imagens de satélite de alta e média resolução espacial**. Campinas: Embrapa Monitoramento por Satélite, 2003. 16 p, (Comunicado Técnico, 11). Disponível em: <<http://www.cnpem.br/publica/serie.html>>. Acesso em: 15 jan. 2008.

MELGANI, F. Classification of Hyperspectral Remote Sensing Images With Support Vector Machines. **IEEE Transactions on geoscience and remote sensing**, v. 42, n. 8, p. 1778-1790, ago.2004.

MITCHELL, T. **Machine Learning**. New York, McGraw Hill, 1997. 414p.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. 3. ed. Viçosa, UFV, 2005. 320p.

NOVO, E. M. L. M. **Sensoriamento Remoto, Princípios e Aplicações**; 2. ed. São Paulo: Edgrad Blucher Ltda., 1989. 269 p.

OSUNA, E.; FREUND, R.; GIROS, F. **Support vector machines: Training and applications**. Massachusetts, M.I.T Artificial Intelligence Laboratory, 1997. 42 p.

PAL, M.; MATHER, P. M. Assessment of the effectiveness of support vector machines for hyperspectral data. **Future Generation Computer Systems**, v. 20, n. 7, p. 1215–1225, 2004.

QUEIROZ, R. B.; SEVERINO, P. A. R.; RODRIGUES, A. G.; GÓMEZ, A. T. U. Redes Neurais: Um comparativo com Máxima Verossimilhança Gaussiana na Classificação de Imagens CBERS. In: WORKSHOP DE TECNOLOGIA DA INFORMAÇÃO APLICADA AO MEIO AMBIENTE, 2., 2004, Itajaí. **Anais...** Universidade Vale do Itajaí: UVI, 2004. p. 746-749. CD-ROM, On-line. ISBN 1677-2822. Disponível em: <www.niee.ufrgs.br/cbcomp/cbcomp2004/html/pdf/Workshop_Ambiente/Intelig%eancia_Artificial/t170100151_3.pdf>. Acesso em: 05 nov. 2008.

RICHARDS, J.A. JIA, X. **Remote sensing digital image analysis: An introduction.** 4 ed. New York, Springer, 2006. 476 p.

SMOLA, A. J.; BARLETT, P.; SCHÖLKOPF, B.; SCHUURMANS, D. **Advances in Large Margin Classifiers.** Cambridge: MIT Press, 2000. 423 p.

SU, L.; MARK, J. C.; RANGO, A.; MARTONCHIK, J. V.; PETERS, D. P. C. Support vector machines for recognition of semi-arid vegetation types using MISR multi-angle imagery. **Remote Sensing of Environment**, v. 107, n.1 p. 299-311, 2007.

VAPNIK, V.N. **The Nature of Statistical Learning Theory.** New York, Springer-Verlag, 1995. 332 p.

WATZLAWICK, L. F.; KIRCHNER, F. F.; MADRUGA, P. R de A. Avaliação de classificação digital em imagens de vídeo multiespectral utilizando matriz confusão. **Revista de Ciências Exatas e Naturais**, Guarapuava-PR, v. 5, n. 1, p. 47-57, jul./dez. 2003.

WESTON, J.; WATKINS, C. **Multi-Class Support Vector Machines.** Maio, 1998. Comunicado técnico.

WISCHMEIER, W. H. & D. D. SMITH. **Predicting rainfall erosion losses: a guide to conservation planning.** Washington, D.C, United States Department of Agriculture, 1978. 58p. (Agriculture handbook, 537)