

Classificação da capacidade produtiva de sítios florestais
utilizando máquina de vetor de suporte e rede neural artificialSite classification with support vector
machine and artificial neural networkDiogo Nepomuceno Cosenza¹, Helio Garcia Leite², Gustavo Eduardo Marcatti³,
Daniel Henrique Breda Binoti³, Aline Edwiges Mazon de Alcântara³ e Rafael Rode⁴

Resumo

Pesquisadores da área de mensuração florestal têm incluído com frequência em seus estudos o uso das técnicas de inteligência computacional (IC) para realização de trabalhos de modelagem por serem capazes de manipular um grande conjunto de dados e criar modelos robustos. Dentre essas técnicas, se destacam a Rede Neural Artificial (RNA) e a recente Máquina de Vetor de Suporte (MVS). Dessa forma, buscou-se nesse trabalho avaliar a aplicação dessas técnicas (RNA e MVS) no processo de classificação da capacidade produtiva de povoamentos florestais com a inclusão de variáveis edáficas, de manejo e do povoamento, comparando os seus resultados com os obtidos pelo método da curva guia. Foi possível concluir que as técnicas de IC avaliadas são capazes de classificar a capacidade produtiva do local de forma satisfatória, desde que utilizadas as variáveis adequadas; o uso conjunto das variáveis “tipo de solo”, “espaçamento do plantio”, “idade” e “altura dominante”, foi suficiente para classificar os sítios; a RNA foi mais precisa para classificar a capacidade produtiva do que a MVS; a inclusão de muitas variáveis pouco significativas pode prejudicar ou ser indiferente no desempenho das técnicas.

Palavras-chave: classificação da capacidade produtiva; redes neurais artificiais; máquina de vetor de suporte; inteligência computacional.

Abstract

Researchers in forest measurement have often included in their studies the use of computational intelligence (CI) techniques for modeling by being able to manipulate a large data set and create robust models. Among these techniques stands out Artificial Neural Network (ANN) and the latest Support Vector Machine (SVM). Therefore this study aimed to evaluate the use of these techniques (ANN and SVM) in site classification including some characteristics of soil, management and forest, comparing their results with those obtained by the guide curve method. It was concluded that CI techniques evaluated are able to classify sites satisfactorily since the appropriate variables are used; the combination of variables “soil type”, “planting spacing”, “age” and “dominant height” was sufficient to classify the sites; the ANN is better than SVM to site indexing; the inclusion of many low significance variables can be either detrimental or indifferent to the techniques performances.

Keywords: site classification; artificial neural networks; support vector machine; computational intelligence.

INTRODUÇÃO

Para o correto gerenciamento das florestas é necessário compreender os processos de crescimento e de produção, uma vez que as decisões a serem tomadas serão baseadas na previsão da produção a partir de informações correntes. Para isso os manejadores buscam modelos estatísticos que possibilitam fazer estas previsões. A classificação de terras se torna neste caso uma das principais etapas no processo de planejamento da floresta a ser instalada, sendo considerada por Campos e Leite (2013) como um dos três elementos do manejo, junto com a prognose da produção e a prescrição de tratamentos silviculturais.

¹Mestrando em Ciências Florestais. UFV – Universidade Federal de Viçosa. Av. P.H. Rolfs, s/n – Departamento de Engenharia Florestal – 36570-000 – Viçosa, MG. E-mail: dncosenza@gmail.com

²Professor Titular do Departamento de Engenharia Florestal. UFV – Universidade Federal de Viçosa – 36570-000 – Viçosa, MG. E-mail: hgleite@gmail.com

³Doutorando(a) em Ciências Florestais – UFV – Universidade Federal de Viçosa – 36570-000 – Viçosa, MG. E-mail: gustavomarcatti@gmail.com; danielhbbinoti@gmail.com; alimazon@gmail.com

⁴Professor Adjunto – UFOPA – Universidade Federal do Oeste do Pará – R. Vera Paz, s/n – Salé – CEP: 68035-110 - Santarém, PA – E-mail: rafaelrode@gmail.com

Na literatura existem diversas metodologias para realizar essa modelagem, sendo que as mais comuns envolvem a relação entre idade e altura dominante. Devido à fragilidade de se utilizar somente essas duas variáveis (CAMPOS; LEITE, 2013), outros trabalhos avaliaram com êxito a inclusão de diferentes dados de entrada, como o tipo de solo, relevo e precipitação, para aumentar a qualidade da classificação (OLIVEIRA, 1998; ORTIZ et al., 2006; BORGES, 2012). Além disso, estas pesquisas se justificam pela necessidade de compreender os efeitos dos diversos fatores de produção, tais como o clima, o manejo das áreas ou o genótipo adotado, dentre outros (CAMPOS, 1970; BRAGA et al., 1999).

Entretanto, equacionar o comportamento dessas variáveis bem como o seus efeitos sobre a produtividade do povoamento pode ser uma tarefa árdua, e por vezes complexa, por se tratar de dados de natureza distinta. Dessa forma cria-se a oportunidade para verificar a eficiência das ferramentas de Inteligência Computacional (IC) no processo de classificação de sítio, por serem capazes de lidar com grande número de variáveis quantitativas e qualitativas simultaneamente e gerar modelos robustos.

A Inteligência Computacional tem sido utilizada em diversas aplicações florestais, com destaque para a modelagem do crescimento e a classificação de padrões. Dentre essas técnicas, a Rede Neural Artificial (RNA) já se tornou uma ferramenta consolidada no setor, sendo aplicadas em diversos ramos da ciência florestal, para estimar parâmetros dendrométricos, classificar e manipular dados georreferenciados e em estudos de dinâmica de pragas e doenças (PENG; WEN, 1999; DIAMANTOPOULOU, 2005; GORGENS et al., 2009; SILVA et al., 2009; BINOTI et al., 2014).

Conforme exposto por Braga et al., (2007) o mecanismo de funcionamento da rede é inspirada no funcionamento dos neurônios biológicos. Segundo o autor, cada neurônio constitui uma unidade de processamento simples onde se opera uma equação matemática (função de ativação), em geral não linear, que processa a informação recebida e a transmite para outro neurônio.

Didaticamente, cada uma dessas unidades é disposta em camadas com conexões entre si, que simbolizam os sinais a serem transmitidos de um neurônio para outro (Figura 1). A primeira camada é referente aos dados de entrada, que seria uma sequência de neurônios não computacionais, que servirão como fonte de dados para a rede, sendo uma unidade para cada variável quantitativa e uma para cada categoria das variáveis qualitativas, as quais são tratadas como variáveis binárias, de forma que cada categoria de uma variável recebe um código específico. A segunda camada (camada oculta) é onde o processamento se realiza para ser transmitido à última camada. Esta por sua vez fará o processamento para apresentação dos valores da rede, sendo a quantidade de neurônios igual à quantidade de variáveis a serem estimadas ou à quantidade de categorias da classificação (HAYKIN, 2001).

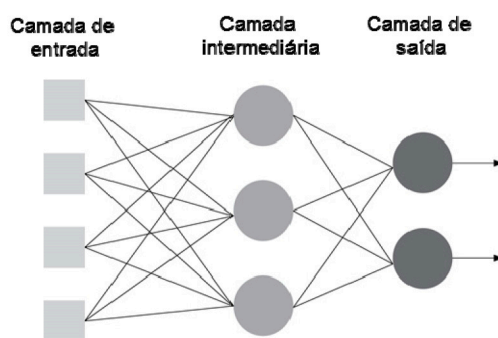


Figura 1. Exemplo simplificado de uma rede neural multicamadas (HAYKIN, 2001).
Figure 1. Exemplification of a simple multilayer neural network (HAYKIN, 2001).

Durante o treinamento, o erro associado aos valores estimados é então retornado aos neurônios da rede, os quais reajustam os seus parâmetros em um processo iterativo para minimizar os erros segundo uma função de custo. Dessa forma o tipo de rede será determinado pela sua arquitetura, dada pela disposição dos neurônios, as suas conexões, o número de camadas e as funções de ativação utilizadas. Dentre as arquiteturas mais comuns estão as redes do tipo Perceptron Multicamadas (MLP) e as de Função de Base Radial (RBF) (BROOMHEAD; LOWE, 1988; BRAGA et al., 2007).

Outra ferramenta de IC bastante promissora é a Máquina de Vetor de Suporte (MVS) desenvolvida por Vapnik (1995) a partir dos estudos de Vapnik e Chervonenkis (1971); e Boser et al., (1992). O seu mecanismo de classificação consiste na separação ótima de um grupo de dados, independentemente da sua dimensionalidade, através de um problema de programação quadrática que permite boa generalização (VAPNIK, 1998; HAYKIN, 2001; LORENA, 2006). Esse procedimento garante que a MVS encontre um mínimo global na superfície de custo, o que pode ser considerado como vantagem do método (HAYKIN, 2001).

Essa técnica vem sendo utilizada com sucesso em trabalhos que envolvem classificação de dados e por isso é amplamente aplicada na bioinformática para classificação de genes e biomoléculas, modelagem de distribuição geográfica de espécies, identificação de imagens e vários outros objetivos (HEARST et al., 1998; GUYON et al., 2002; NOBLE, 2006; BARRETO, 2008). Apesar disso, ainda não foram encontrados estudos envolvendo a aplicação da MVS na área de modelagem de florestas para fins de manejo, sendo mais comuns os trabalhos que envolvem sensoriamento de imagens e análises químicas (SOUSA et al., 2010; NUNES et al., 2011).

Sendo assim, devido ao sucesso destes métodos em lidar com dados não lineares, corriqueiros na modelagem de florestas, este trabalho teve por objetivo comparar os resultados obtidos pelo processamento de dados florestais com a Máquina de Vetor de Suporte e a Rede Neural Artificial, visando à classificação da capacidade produtiva de povoamentos de eucalipto.

MATERIAL E MÉTODOS

Para realizar esse estudo utilizou-se o cadastro de 926 propriedades de uma empresa florestal contendo as informações sobre o tipo de solo e o tipo de preparo antes do plantio, o espaçamento utilizado, a idade do povoamento, altura dominante, área basal, volume com casca, diâmetro à 1,3 m de altura (dap) mínimo, médio e máximo do povoamento, o número de indivíduos por hectare e a classe de local, previamente determinada por curvas de índice de local geradas através do método da curva guia (Tabela 1). Porém, essa última variável não foi utilizada para realização dos testes, mas somente para fins comparativos. Assim, cada propriedade foi classificada entre as classes I para o melhor sítio, classe II para o intermediário e classe III para o pior sítio.

Tabela 1. Descrição do banco de dados usado para realização dos testes.

Table 1. Description of the database used in the tests.

Características	Classe I		Classe II		Classe III	
	Média	Desv. Pad.	Média	Desv. Pad.	Média	Desv. Pad.
Nº de casos	215		598		113	
Índice de Local (m)	14,3	0,6	12,1	0,8	9,3	1,0
Altura dominante (m)	9,9	1,3	9,2	1,5	6,5	1,2
Idade (anos)	3,3	0,8	3,8	1,0	3,3	0,8
Dap mínimo (cm)	6,0	0,7	5,9	0,8	5,5	0,6
Dap médio (cm)	9,2	1,5	9,3	1,7	7,4	1,2
Dap máximo (cm)	13,4	2,6	13,5	2,8	10,4	2,4
Área basal (m ²)	8,1042	2,9778	7,7974	3,3078	3,3104	2,1251
Volume com casca (m ³)	35,3126	17,4358	32,7159	17,5286	10,4514	7,6632
Árvores/ha	1193,9063	190,3723	1095,8816	270,2025	710,3259	380,1747

Dentro das variáveis categóricas, estavam descritos dez tipos de solos diferentes, distinguidos até o quinto nível categórico (EMBRAPA, 2006), dois tipos de preparo do solo, sendo diferenciados pelo implemento utilizado na subsolagem, e cinco tipos de espaçamento diferentes. Cabe destacar que maiores detalhes sobre as variáveis categóricas são irrelevantes neste trabalho, uma vez se objetiva avaliar o seu uso na classificação do sítio e não a análise individual da produtividade em cada situação em que elas ocorrem.

Foram realizados cinco testes para cada uma das duas ferramentas usando grupos de variáveis diferentes, somando ao todo 10 testes (Tabela 2). Inicialmente, foi feito o teste com a RNA e a MVS utilizando todas as variáveis de entrada existentes no banco de dados, com exceção do índice de local, como mencionado anteriormente. No segundo, realizou-se o mesmo processo utilizando

somente as variáveis “altura dominante”, “idade”, “espaçamento” e “tipo de solo” na tentativa de reduzir o número de entradas e avaliar a contribuição delas nas estimativas. O terceiro teste foi feito usando somente as variáveis “espaçamento”, “tipo de solo” e “Idade”; e no quarto teste usaram-se as variáveis “espaçamento”, “tipo de solo” e “altura dominante”, para verificar se existe diferença na utilização da variável “Idade” em detrimento da “altura dominante”. Por último, foram testadas somente as entradas “espaçamento” e “tipo de solo” para fazer a classificação por meio de elementos ambientais e de manejo, sem a necessidade de dados do povoamento.

Tabela 2. Descrição dos testes executados.
Table 2. Description of the tests performed.

Teste	Técnica	Variáveis utilizadas
1	RNA	Todas*
	MVS	Todas*
2	RNA	Tipo de solo, espaçamento, Idade, altura dominante
	MVS	Tipo de solo, espaçamento, Idade, altura dominante
3	RNA	Tipo de solo, espaçamento, idade
	MVS	Tipo de solo, espaçamento, idade
4	RNA	Tipo de solo, espaçamento, altura dominante
	MVS	Tipo de solo, espaçamento, altura dominante
5	RNA	Tipo de solo, espaçamento
	MVS	Tipo de solo, espaçamento

* Com exceção da variável “índice de local”.

O processamento dos dados foi feito no programa Statistica 10 (STATSOFT INC, 2015) através do localizador automático de rede, onde foram treinadas 400 redes com diferentes configurações e selecionada a que apresentava o melhor resultado de classificação.

Com deste método, permitiu-se que o número de neurônios da camada oculta variasse de 2 a 15 unidades, bem como a arquitetura da rede, podendo ser do tipo Perceptron Multicamada ou Função de Base Radial. Ainda, no caso da rede MLP, variou-se a função de ativação, podendo ser a função identidade, logística, tangencial hiperbólica ou exponencial, para saber qual se adaptaria melhor aos dados.

Para o caso da rede RBF esta variação não é permitida, uma vez que seu funcionamento é definido pelo uso da função gaussiana na camada intermediária. Já a função da camada de saída será determinada conforme a função de custo. Dentre elas foram testadas a do tipo soma de quadrados e entropia cruzada, estando respectivamente associadas às funções de ativação identidade e Softmax para a camada de saída (BISHOP, 1995).

Uma vez que não foram encontradas na literatura recomendações sobre o uso deste tipo de dado para o treinamento de uma RNA, optou-se por fracioná-lo em treino, com 70%, teste e validação, ambos com 15%.

Os dados de treinamento são destinados ao ajuste dos pesos da rede, sendo desejável, a princípio, maior percentual para abranger a diversidade dos dados em todas as suas categorias. O segundo grupo são dados não apresentados durante o treino. Os erros de classificação do teste são obtidos para cada ciclo do treinamento, com tendências a reduzir no início do processo e a aumentar à medida que a rede se torna mais específica para os dados de treino (BRAGA et al., 2007). Este fato é utilizado como artifício do programa para que a rede não realize sobreajuste, permitindo que o programa finalize o treinamento quando o erro de teste do ciclo atual for menor do que no ciclo anterior. Os dados da validação também não são apresentados durante o treinamento, sendo utilizados ao fim do processo para calcular o erro de generalização da rede, servindo-se então de mais um parâmetro para avaliar a sua qualidade. Assim, mesmo com pouco percentual, estas duas repartições (teste e validação) estarão aplicadas a um banco de dados extenso, abrangendo um volume de ocorrências considerável.

Após o treinamento foram retidas as 15 melhores redes para a avaliação dos resultados e foi escolhida aquela que apresentava a melhor acurácia dos resultados na validação, que seria o percentual das classificações corretas de cada teste, dando preferência para a rede que apresentava menor quantidade de neurônios na camada oculta, a fim de evitar sobreajuste.

Para o caso da Máquina de Vetor de Suporte escolheu-se o kernel mais comum nos trabalhos de classificação, a Função de Base Radial (BROOMHEAD; LOWE, 1988), utilizando o padrão do próprio sistema para definir o parâmetro Gama e o otimizador *v-fold-cross-validation* para encontrar o fator Capacidade, como recomendado em Statsoft Inc (2015).

Este otimizador promove a repartição dos dados em *v*-pastas, neste caso 10, e promove o treinamento com *v*-1 pastas, no caso 9. A pasta restante é utilizada para calcular o erro de classificação como forma de validação. Em seguida a pasta utilizada como validação é englobada ao conjunto de pastas de treinamento e outra é selecionada para se obter novamente o erro de validação. Esse procedimento é então repetido para todas as pastas (*v* vezes) para se estimar o erro médio de validação do grupo de dados. Dessa forma permitiu-se que o programa variasse unitariamente o fator Capacidade entre 1 e 40, para que fosse escolhido aquele que causasse a menor estimativa do erro médio de validação.

Nessa etapa os dados também foram divididos em treinamento (70%) e teste (30%), sendo o teste utilizado para calcular o erro de generalização do treinamento da MVS.

RESULTADOS E DISCUSSÃO

As configurações obtidas após os testes para cada uma dos casos descritos estão apresentados na Tabela 3 e 4 para a RNA, e na Tabela 5 para a MVS.

Tabela 3. Configuração das Redes Neurais Artificiais após o treinamento.

Table 3. Summary of network parameters after training.

RNA	Variáveis de entrada ¹	Arquitetura ²	Função de custo	Função de ativação	
				Camada intermediária	Camada de saída
rede-1	Todas	MLP 24-9-3	Entropia Cruzada	Identidade	Softmax
rede-2	Alt_Dom.; Idade; Espaç.; Solo	MLP 17-2-3	Entropia Cruzada	Identidade	Softmax
rede-3	Idade; Espaç.; Solo	MLP 16-8-3	Soma de Quadrados	Logística	Tangencial hiperbólica
rede-4	Alt_Dom.; Espaç.; Solo	MLP 16-3-3	Soma de Quadrados	Exponencial	Logística
rede-5	Solo e Espaç.	RBF 15-5-3	Entropia Cruzada	Gaussiana	Softmax

¹Espaç. = Espaçamento, Alt_dom = Altura dominante. ²Tipo da rede e distribuição dos neurônios na camada de entrada, oculta e de saída.

Tabela 4. Descrição dos resultados obtidos para as redes.

Table 4. Description of the results obtained with the neural networks.

RNA	Variáveis de entrada ¹	Acurácia (%)			
		Treino	Teste	Validação	Total
rede-1	Todas	99,54	97,10	98,55	99,03
rede-2	Solo; Espaç.; Idade; Alt_Dom	98,92	96,38	98,55	98,49
rede-3	Solo; Espaç.; Idade;	65,69	67,39	68,84	66,41
rede-4	Solo; Espaç.; Alt_Dom	71,85	67,39	76,81	71,92
rede-5	Solo; Espaç.	64,31	68,84	61,59	64,58

¹Espaç. = Espaçamento, Alt_dom = Altura dominante.

Tabela 5. Descrição dos parâmetros das MVSs obtidas.

Table 5. Summary of SVMs parameters after training.

MVS	Variáveis de entrada ¹	Capacidade	Gama	Acurácia (%)			
				Treino	Teste	Validação	Total
mvs-1	Todas	38	0,1	95,21	89,93	89,51	93,63
mvs-2	Solo; Espaç.; Idade; Alt_Dom	38	0,25	95,83	93,88	92,43	95,24
mvs-3	Solo; Espaç.; Idade;	10	0,33	69,29	64,39	65,43	67,82
mvs-4	Solo; Espaç.; Alt_Dom	15	0,33	71,30	72,66	67,90	71,71
mvs-5	Solo; Espaç.	1	0,5	64,35	65,11	61,88	64,58

¹Espaç. = Espaçamento, Alt_dom = Altura dominante.

Após o processamento dos dados e a seleção das melhores redes, foi observada que a função de custo Entropia Cruzada e, conseqüentemente, a função da camada de saída Softmax (BISHOP, 1995), foi a melhor para a rede-1, rede-2 e rede-5, indicando que esta função é adequada para lidar com esse tipo de tarefa (Tabela 3).

Estas funções utilizadas pelo programa são dadas por:

$$E_{ec} = - \sum_{i=1}^N t_i * \ln \left(\frac{y_i}{t_i} \right)$$

$$f(a_j) = \frac{e^{(a_j)}}{\sum_{j=1}^K e^{(a_j)}}$$

Onde E_{ec} é o erro da equação entropia cruzada; N é o número de dados de treinamento; y_i é a estimativa da rede para o valor observado t_i ; $f(a_j)$ é o sinal produzido pela função de ativação Soft-max; a_j é a soma dos sinais recebidos pelo neurônio j ; e K é o total de neurônios na camada de saída.

Com essa arquitetura a classificação dos talhões pela rede terá caráter probabilístico, assumindo distribuição multinomial das classes (STATSOFT INC, 2015). Assim, o valor emitido por cada um dos três neurônios da camada de saída representará a probabilidade de um dado talhão pertencer à sua respectiva classe, sendo então dada a classe do neurônio que apresentar maior probabilidade de acerto.

Outro ponto importante é a qualidade dos ajustes para as diferentes variáveis de entrada. No caso da RNA, a eficiência para o treinamento, teste e validação, foram semelhantes na rede-1 e rede-2 (Tabela 6), mostrando que a presença das variáveis: preparo do solo, área basal, número de árvores por hectare, volume com casca, Dap mínimo, médio e máximo, não interferem significativamente na classificação dos dados. Além disso, o percentual de acerto para ambas as redes foram semelhantes.

Tabela 6. Resultado da classificação.

Table 6. Classifications results.

Variáveis de entrada ¹	Classificação pela RNA				Classificação pela MVS				
	I	II	III	Total	I	II	III	Total	
Todas	Observado:	215	598	113	926	215	598	113	926
	Correto:	212	593	112	917	194	583	90	867
	Incorreto:	3	5	1	9	21	15	23	59
	Correto (%):	98,60	99,16	99,12	99,03	90,23	97,49	79,65	93,63
	Incorreto (%):	1,40	0,84	0,88	0,97	9,77	2,51	20,35	6,37
Alt_Dom.; Idade; Espaç.; Solo	Correto:	208	593	111	912	198	582	102	882
	Incorreto:	7	5	2	14	17	16	11	44
	Correto (%):	96,74	99,16	98,23	98,49	92,09	97,32	90,27	95,25
	Incorreto (%):	3,26	0,84	1,77	1,51	7,91	2,68	9,73	4,75
Idade; Espaç.; Solo	Correto:	45	531	39	615	55	532	41	628
	Incorreto:	170	67	74	311	160	66	72	298
	Correto (%):	20,93	88,80	34,51	66,41	25,58	88,96	36,28	67,82
	Incorreto (%):	79,07	11,20	65,49	33,59	74,42	11,04	63,72	32,18
Alt_Dom.; Espaç.; Solo	Correto:	5	579	82	666	38	553	73	664
	Incorreto:	210	19	31	260	177	45	40	262
	Correto (%):	2,33	96,82	72,57	71,92	17,67	92,47	64,60	71,71
	Incorreto (%):	97,67	3,18	27,43	28,08	82,33	7,53	35,40	28,29
Solo e Espaç.	Correto:	0	598	0	598	0	598	0	598
	Incorreto:	215	0	113	328	215	0	113	328
	Correto (%):	0,00	100,00	0,00	64,58	0,00	100,00	0,00	64,58
	Incorreto (%):	100,00	0,00	100,00	35,42	100,00	0,00	100,00	35,42

¹Espaç. = Espaçamento; Alt_dom = Altura dominante.

Resultado similar foi encontrado com a Máquina de Vetor de Suporte (Tabela 5), em que houve inclusive pequena redução da precisão quando se utilizou todas as variáveis de entrada (mvs-1). Isso pode ter ocorrido pela dificuldade do sistema em fazer a correspondência entre a variável de saída e um grande número de variáveis de entrada, as quais não influenciavam significativamente no processo de classificação.

Para os testes com três variáveis os resultados foram semelhantes para ambas as técnicas (rede-3, rede-4, mvs-3 e mvs-4). O valor da precisão reduziu consideravelmente, em especial nos casos onde se usou a idade ao invés da altura dominante, evidenciando que esta última variável possui maior peso no processo de classificação.

O uso somente do tipo de solo com o espaçamento não permitiu, entretanto, que a RNA (rede-5) e nem a MVS (mvs-5) alcançassem bons resultados, mostrando que a presença das variáveis "idade" e "altura dominante" influenciam fortemente na precisão. Além disso, as estimativas ficaram concentradas apenas para a classe intermediária (Tabela 6). Este tipo de erro seria grave, pois iria superestimar todos os sítios ruins (classe III) e subestimaria todos os sítios considerados melhores (classe I).

Contudo os testes mostraram que o uso dessas quatro variáveis juntas foi suficiente para fazer boas estimativas da classe de local, apresentando erros de 1,51% na rede-2 (Tabela 4) e 4,75% na mvs-2 (Tabela 5).

Tal situação pode ser explicada pela eficiência de se classificar a capacidade produtiva considerando o tipo de solo do povoamento, uma vez que o desenvolvimento das árvores está intrinsecamente ligado aos fatores edáficos do local (BARROS, 1974; BRAGA et al., 1999; ORTIZ et al., 2006). A inclusão da idade neste caso se torna útil por dar sentido ao valor da altura dominante, de forma que um talhão com idade avançada e baixa altura dominante indicaria baixo potencial produtivo, e *vice e versa*. Já a variável espaçamento, contribui para melhoria do treinamento por ser intimamente ligada à densidade do povoamento, sendo possível atribuir pesos adequados para um local onde a produtividade ao longo do tempo ($m^3 ha^{-1} ano^{-1}$) não seja condizente com adensamento do plantio.

A metodologia de IC para esse tipo de banco de dados foi bem sucedida em avaliar a qualidade do local baseado não só nas informações tradicionais de idade e altura dominante, mas também incluindo uma variável ambiental, o tipo de solo, e outra referente ao manejo, que seria o espaçamento. Entretanto Campos (1970) e Braga et al., (1999) mostram que existem vários outros fatores capazes de interferir no desenvolvimento da planta, porém estes não foram contemplados no estudo.

Já com relação ao uso dos dois sistemas, a Máquina de Vetor de Suporte teve vantagem por dispensar avaliação *a posteriori* que é feita na RNA para seleção da melhor rede. Isto se deve à otimização quadrática ocorrida durante o treinamento da MVS (BOSER et al., 1992) que permite a cada configuração do sistema obter o mesmo resultado, sempre que aplicada a um mesmo banco de dados. Já as redes neurais possuem mais elementos a serem manipulados, além da inicialização dos parâmetros dos neurônios ocorrerem de maneira aleatória (HAYKIN, 2001). Assim cada rede treinada apresentará pequenas diferenças nas estimativas, mesmo se mantida a mesma arquitetura. Esta diferença entre as duas técnicas evidencia a praticidade das MVS por excluir a subjetividade do operador em ter que escolher a melhor rede a ser aplicada no banco de dados.

CONCLUSÕES

Ao fim deste trabalho foi possível concluir que:

- As técnicas de Inteligência Computacional avaliadas são capazes de classificar a capacidade produtiva do local de forma satisfatória, desde que utilizadas as variáveis de entrada adequadas.
- O uso conjunto das variáveis tipo de solo, espaçamento, idade e altura dominante, foi suficiente para classificar os locais.
- As Redes Neurais Artificiais é mais precisa que a Máquina de Vetor de Suporte na classificação de sítios florestais para produção.
- A inclusão de muitas variáveis pouco significativas pode prejudicar ou ser indiferente no desempenho da RNA e da MVS.

AGRADECIMENTOS

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), ao Ministério de Ciência, Tecnologia e Inovação do governo brasileiro e à Universidade Federal de Viçosa pelo apoio financeiro e pela oportunidade concedida para se produzir ciência.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARRETO, F. C. C. **Modelagem de distribuição potencial de espécies como ferramenta para conservação: seleção e avaliação de algoritmos e aplicação com *Heliconius nattereri* Felder, 1865 (Nymphalidae: Heliconiinae)**. 2008. 57 p. Tese (Doutorado em Entomologia) - Universidade Federal de Viçosa, Viçosa, 2008.
- BARROS, N. F. **Contribuição ao relacionamento de características pedológicas e topográficas com altura de *Eucalyptus alba*, na região de Santa Barbara, MG**. 1974. 89 p. Dissertação (Mestrado em Fitotecnia) - Universidade Federal de Viçosa, Viçosa, 1974.
- BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação de volume de árvores. *Ciência da Madeira*. Pelotas, v. 5, n. 01, p. 58-67, 2014.
- BISHOP, C. M. **Neural networks for pattern recognition**. Oxford: Cladenron Press, 1995. 482 p.
- BORGES, J. S. **Modulador edáfico para uso em modelo ecofisiológico e produtividade potencial de povoamentos de eucaliptos**. 2012. 70 p. Tese (Doutorado em Solos e Nutrição de Plantas) - Universidade Federal de Viçosa, Viçosa, 2012.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: WORKSHOP ON COMPUTATIONAL LEARNING THEORY, 5., 1992, Pittsburgh. **Proceedings...** New York: ACM Press, 1992. p. 144-152.
- BRAGA, A. D. P.; CARVALHO, A. P. D. L. F. D.; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2.ed. Rio de Janeiro: LTC, 2007. 260 p.
- BRAGA, F. A.; BARROS, N. F.; SOUZA, A. L.; COSTA, L. M. Características ambientais determinantes da capacidade produtiva de sítios cultivados com eucalipto. *Revista Brasileira de Ciência do Solo*, Viçosa, v. 23, n. 2, p. 291-298, 1999.
- BROOMHEAD, D. S.; LOWE, D. Multivariable functional interpolation and adaptive networks. *Complex Systems*, v. 2, p. 321-355, 1988.
- CAMPOS, J. C. C. Principais fatores do meio que afetam o crescimento das árvores. *Floresta*, Curitiba, v. 2, n. 3, p. 45-52, 1970.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração Florestal: perguntas e respostas**. 4.ed. Viçosa: Editora UFV, 2013. 605 p.
- DIAMANTOPOULOU, M. J. Artificial neural networks as an alternative tool in pine bark volume estimation. *Computers and Electronics in Agriculture*, v. 48, n. 3, p. 235-244, 2005.
- EMBRAPA. EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Sistema Brasileiro de Classificação de Solos**. 2.ed. Rio de Janeiro: CNPS. 2006. 306 p.
- GORGENS, E. B.; LEITE, H. G.; SANTOS, H. N.; GLERIANI, J. M. Estimação do volume de árvores utilizando Redes Neurais Artificiais. *Revista Árvore*, Viçosa, v. 33, n. 6, p. 1141-1147, 2009.
- GUYON, I.; WESTON, J.; BARNHILL, S. Gene selection for cancer classification using Support Vector Machines. *Machine Learning*, v. 46, n. 1, p. 389-422, 2002.
- HAYKIN, S. **Redes Neurais: princípios e práticas**. 2.ed. Porto Alegre: Bookman, 2001. 900 p.

HEARST, M. A.; SCHOLKOPF, B.; DUMAIS, S.; OSUNA, E.; PLATT, J. Trends and controversies - support vector machines. *IEEE Intelligent Systems*, v. 13, n. 4, p. 18-28, 1998.

LORENA, A. C. **Investigação de estratégias para a geração de máquinas de vetores de suporte multiclases**. 2006. 203 p. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Universidade de São Paulo, São Carlos, 2006.

NOBLE, W. S. What is a support vector machine? *Nature Biotechnology*, v. 24, n. 12, p. 1565-1567, 2006.

NUNES, C. A.; LIMA, C. F.; BARBOSA, L. C. A.; COLODETTE, J. L.; FIDÊNCIO, P. H. Determinação de constituintes químicos em madeira de eucalipto por Pi-CG/EM e calibração multivariada: comparação entre redes neurais artificiais e máquinas de vetor suporte. *Química Nova*, São Paulo, v. 34, n. 2, p. 279-283, 2011.

OLIVEIRA, R. A. **Classificação de sítios em plantações de eucalipto pelo método de índices de local e por classes de solos e precipitação**. 1998. 84 p. Dissertação (Mestrado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa, 1998.

ORTIZ, J. L.; VETTORAZZI, C. A.; COUTO, H. T. Z.; GONÇALVES, J. L. M. Relações espaciais entre o potencial produtivo de um povoamento de eucalipto e atributos do solo e do relevo. *Scientia Florestalis*, Piraciaba, n. 72, p. 67-79, 2006.

PENG, C.; WEN, X. Recent Applications of artificial neural networks in forest resource management: an overview. In: AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE WORKSHOP, Orlando, FL. 1999. Proceedings... AAAI Technical Reports WS-99-07. Menlo Park, CA: AAAI Press, 1999. p. 15-22. Disponível em: < <http://www.aaai.org/Library/Workshops/1999/ws99-07-003.php> >. Acesso em: 18 dez. 2015.

SILVA, M. L. M.; BINOTI, D. H. B.; GLERIANI, J. M.; LEITE, H. L. Ajuste do modelo de Schumacher e Hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto. *Revista Árvore*, Viçosa, v. 33, n. 6, p. 1133-1139, 2009.

SOUSA, B. F. S.; TEIXEIRA, A. S.; SILVA, F. A. T. F.; ANDRADE, E. M.; BRAGA, A. P. S. Avaliação de classificadores baseados em aprendizado de máquina para a classificação do uso e cobertura da terra no bioma caatinga. *Revista Brasileira de Cartografia*, Rio de Janeiro, v. 2, n. 62, p. 385-399, 2010.

STATSOFT INC. **Statistica (data analysis software system), version 10**. 2015. Disponível em: <www.statsoft.com>. Acesso em: 18 abr. 2015.

VAPNIK, V. N. **The nature of statistical learning theory**. New York: Springer-Verlag, 1995. 188 p.

VAPNIK, V. N. **Statistical Learning Theory**. New York: John Wiley and Sons, 1998. 768 p.

VAPNIK, V. N.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*. v. 16, n. 2, p. 264-280, 1971.

Recebido em 28/07/2014

Aceito para publicação em 11/06/2015

