

PERCEPTRON DE MÚLTIPLAS CAMADAS OTIMIZADO PARA A CLASSIFICAÇÃO GEOGRÁFICA E GENOTÍPICA DE QUATRO GENÓTIPOS DE CAFÉ ARÁBICA

Jade Varaschim Link^{1*}, André Luis Guimarães Lemes¹, Herily Pereira Sato¹, Maria Brígida dos Santos Scholz², Evandro Bona¹.

¹ Universidade Tecnológica Federal do Paraná - Campus Campo Mourão

² Instituto Agronômico do Paraná - Londrina

Resumo: As condições climáticas específicas de cada região de cultivo do café fornecem atributos especiais para a bebida e podem assim aumentar seu valor agregado. No entanto, é essencial provar a origem geográfica e genotípica do cultivar por meio de métodos confiáveis. Vários métodos estatísticos foram desenvolvidos na tentativa de reproduzir a capacidade humana de reconhecimento de padrões. Os perceptrons de múltiplas camadas (MLP) são um tipo de rede neural artificial (RNA) com aprendizagem supervisionada que são muito utilizados para classificação de padrões. Este estudo objetivou o desenvolvimento de uma RNA para a classificação geográfica e genotípica de café arábica. Para esse fim os espectros obtidos no equipamento de espectroscopia no infravermelho com transformada de Fourier (FTIR) foram analisados através do emprego de MLPs otimizados através do simplex sequencial. As redes que utilizaram a faixa de 1900-800 cm^{-1} do espectro FTIR obtiveram menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica tanto por região de cultivo (100%) quanto por genótipo (77,78%). Os MLPs otimizados foram capazes de classificar geograficamente as amostras de café arábica. Entretanto, para a classificação genotípica o desempenho não foi totalmente satisfatório. Além disso, o MLP desenvolvido para classificação genotípica apresenta um número muito elevado de pesos sinápticos, não proporcionando número suficiente de graus de liberdade para que a rede neural possa aprender de maneira confiável. Portanto, para melhorar o desempenho da classificação genotípica, os autores sugerem a utilização de outro tipo de RNA e informações do infravermelho próximo.

Palavras-chave: Café Verde, Espectro Infravermelho, MLP, Redes Neurais Artificiais.

Optimized Multilayer perceptron for the geographical and genotypic classification of four genotypes of arabica coffee: The climatic conditions of the coffee crop give special attributes to the beverage and could increase its value. However, it is essential to prove the geographical and genotype origin of the cultivar using reliable methods. Several statistical methods have been developed in an attempt to reproduce the human capability of pattern recognition. The multilayer perceptron (MLP) is an artificial neural network (ANN) with supervised learning that is widely used for pattern classification. This study aimed to develop a MLP to classify the geographic origin and the genotypic of the arabica coffee. For this purpose, spectra obtained in the Fourier transform infrared (FTIR) were analyzed using MLPs optimized by sequential simplex. The networks that used the range 1900-800 cm^{-1} of the raw spectrum had lower mean squared error (MSE) and a higher percentage of correct classification for geographical (100%) and genotypic (77,74%) segmentation. After the results it was concluded that the optimized multilayer perceptrons were able to classify the samples of arabica coffee geographically. However, for genotypic classification the performance was not satisfactory. Also, the MLP developed for genotypic classification has a high number of synaptic weights, thus a large degrees of freedom database is necessary to produce a network with generalization capability. Therefore, to improve the genotype classification performance the authors suggest the use of other type of ANN and information from the near infrared.

Keywords: Green Coffee; Infrared Spectrum; MLP; Artificial Neural Networks.¹

* E-mail: jadejvl@hotmail.com

1 Introdução

O reconhecimento de padrões é um processo pelo qual um sinal recebido é atribuído a uma classe dentre um número predeterminado de categorias. Os seres humanos possuem excelência nata para a aprendizagem e reconhecimento de padrões. Nesse sentido, os métodos matemáticos têm sido desenvolvidos na tentativa de mimetizar, mesmo de maneira rudimentar, tal aptidão humana (BISHOP, 1995; HAYKIN, 2001). As redes neurais artificiais (RNAs) são um conjunto de técnicas baseadas em princípios estatísticos, que vem crescentemente ganhando espaço para realizar tarefas de regressão e reconhecimento de padrões. As RNAs são extremamente versáteis para realizar o mapeamento de relações complexas e não-lineares entre múltiplas variáveis de entrada e saída (BISHOP, 1995). Originalmente as RNAs foram projetadas para ser um esquema, tão preciso quanto possível, do modelo da atividade do cérebro humano. O córtex cerebral é capaz de armazenar padrões de comportamento, mesmo em presença de dados ruidosos, tornando-o mais poderoso do que qualquer computador existente (MARINI, 2009). Para alcançarem bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples denominadas neurônios. Ela se assemelha ao cérebro em dois aspectos principais; o primeiro é que o conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem. O segundo aspecto que assemelha as RNAs com o cérebro são as forças de conexão entre os neurônios, conhecidas como pesos sinápticos, as quais armazenam o conhecimento e servem para ponderar a entrada recebida de cada neurônio. Tal estrutura conduz a uma capacidade de generalização, ou seja, uma RNA pode produzir saídas adequadas para entradas que não estavam presentes durante o treinamento. Esses atributos tornam possível a aplicação das RNAs em problemas complexos. Os perceptrons de múltiplas camadas (MLP) são um tipo de rede neural artificial (RNA) com aprendizagem supervisionada que são muito utilizados para classificação de padrões (HAYKIN, 2001; LUCIA e MINIM, 2006).

A importância do café na economia mundial é indiscutível, é um dos mais valiosos produtos primários comercializados no mundo. Existem cerca de 72 países produtores do grão no mundo, essa multiplicidade de produtores, deve-se à extensa faixa apta à produção do cafeeiro, graças à sua versatilidade. Os principais países produtores concentram-se na América do Sul (Brasil e Colômbia), América Central e Ásia (Vietnã) (SINDICAFÉ, 2011). O consumo brasileiro de café aumenta a cada ano, no período entre maio de 2011 e abril de 2012 foi registrado o consumo de 19,975 milhões de sacas, representando um acréscimo de 3,05% em relação ao período anterior correspondente (ABIC, 2012). Além do mercado interno, as exportações de café verde do Brasil somaram 116,63

mil toneladas em março de 2012, baixa de 19,2% em relação ao mesmo mês de 2011 (CECAFÉ, 2012).

O café tem procedência de uma árvore do gênero *Coffea* e pertence à família das rubiáceas englobando mais de 500 gêneros e 6.000 espécies. Dentro do gênero *Coffea* existem duas espécies principais, o *Coffea arabica* variedade Arábica e o *Coffea canephora* variedade Robusta (RUBAYIZA; MEURENS, 2005). As espécies *Coffea arabica* e *Coffea canephora* (robusta) apresentam uma composição química muito distinta e o café arábica fornece um bebida com qualidade e aroma superior ao café robusta (FARAH, 2009). A mistura de grãos de alto valor com outros de menor valor comercial e adição de substitutos são formas de adulteração em cafés (BRIANDET *et al.*, 1996). Além disso, o Brasil apresenta diversidade de clima e solo, assim os mais variados tipos de cafés podem ser produzidos. Essas condições climáticas conferem atributos especiais para a bebida do café (corpo, acidez e sabor). Quando a qualidade do produto pode ser atribuída ao seu local de produção, ocorre um aumento do valor agregado do café. No entanto, neste processo, devem haver mecanismos confiáveis que comprovem a origem geográfica e genotípica do café (BORSATO *et al.*, 2011). Assim, torna-se imperativo a utilização de técnicas analíticas rápidas e de baixo custo que permitam a diferenciação e certificação da composição e qualidade dos cafés comercializados. Uma metodologia que reúne tais atributos é a espectroscopia com transformada de Fourier na região do infravermelho médio (FTIR). O FTIR permite a detecção dos componentes de maior importância na caracterização de alimentos, sendo uma metodologia que origina rapidamente uma informação sumária sobre o produto. Outra vantagem reside no fato do preparo da amostra, geralmente, não requerer a extração com solventes diminuindo assim o tempo de análise e o impacto ambiental (KAROUI *et al.*, 2010). Ressalta-se ainda que o FTIR já foi utilizado com sucesso para autenticar misturas de cafés arábica e robusta ou adição de adulterantes (KEMSLEY *et al.*, 1995; BRIANDET *et al.*, 1996; DOWNEY *et al.*, 1997; WANG *et al.*, 2009; WANG *et al.*, 2011).

Este estudo objetivou a utilização de uma rede neural artificial para a classificação geográfica e genotípica de café arábica. Para esse fim os espectros obtidos no FTIR foram analisados através do emprego de perceptrons de múltiplas camadas otimizados através do simplex sequencial.

2 Materiais e Métodos

Todos os pré-processamentos dos espectros foram realizados no software IRsolution 1.5 (Shimadzu Corporation, Kyoto, Japão) que acompanha o equipamento. As análises matemáticas e/ou estatísticas, assim como a ACP, as redes neurais artificiais e o simplex sequencial foram realizadas no software

MATLAB R2007b (The MathWorks Inc., Natick, USA).

2.1 Amostras de café arábica

Foram utilizados 4 genótipos de café arábica catalogados pelo Instituto Agrônomo do Paraná - Londrina (IAPAR) da safra de 2010 (Tabela 1).

Tabela 1 - Genótipos de café arábica utilizados.

Genótipo	Local	Total Amostras
IPR 99	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IPR 105	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IPR 106	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IA 59	Cornélio Procópio	2
	Paranavaí	2
	Mandaguari	1
	Londrina	1

No IAPAR, as amostras foram imediatamente colocadas em caixas de madeira com uma malha de fundo e movidas oito vezes por dia até que a umidade dos grãos chegasse a 11-12% e logo após as amostras foram beneficiadas (remoção da casca e pergaminho) (BRASIL, 2003). Os grãos verdes, moídos, peneirados e embalados dos genótipos de café arábica fornecidos pelo IAPAR, foram recebidos e armazenados em um freezer a -18°C e utilizados posteriormente para análise. Foram utilizados cafés cultivados em 4 diferentes regiões (Cornélio Procópio, Paranavaí, Mandaguari e Londrina) totalizando 18 amostras. A Figura 1 apresenta a localização das cidades aonde os cafés foram cultivados e a distância máxima entre as mesmas.



Figura 1 - Mapa da distância entre as cidades dos cafés estudados.

2.2 Espectroscopia de infravermelho com transformada de Fourier (FTIR)

Para preparar as pastilhas foram adicionados em torno de 100 mg de KBr seco (padrão cromatográfico) e aproximadamente 1 mg de amostra finamente moída. A mistura foi, então, prensada em uma prensa hidráulica

(Bovenau, P15 ST) usando um molde (ICL, ICL's Macro/Micro KBr die) empregando aproximadamente 360 kgf/cm^2 . Produziu-se, assim, uma pastilha transparente. Antes da análise de cada amostra, o FTIR (Shimadzu, FTIR - 8300) foi programado para realizar um espectro de *background* do ar, sendo o mesmo utilizado para descontar a influência dos componentes do ar no espectro. Na sequência, a pastilha foi posicionada no feixe do instrumento e os espectros foram obtidos na faixa de 3750 a 750 cm^{-1} . Foram usadas 32 varreduras acumuladas para formar o espectro final e realizadas 5 repetições (pastilhas) para cada amostra, totalizando assim 90 espectros. Foram utilizados 72 desses espectros (80%) como amostras de treinamento das redes neurais artificiais empregadas e 18 espectros (20%) como amostras de teste (uma repetição, escolhida aleatoriamente, de cada um dos cafés estudados).

2.3 Transformação dos espectros

Após obtenção dos espectros foram aplicadas algumas transformações nos mesmos. Primeiramente, foi realizada a normalização do espectro (a maior banda obteve absorbância 1 e a menor 0), correção da linha de base, a suavização do espectro ($17 \text{ smoothing points}$) e uma remoção da banda referente a concentração de CO_2 , desprezando assim as variações de CO_2 entre as amostras. Para a ACP e a RNA foi considerado o uso da região entre 3750 e 750 cm^{-1} do espectro desconsiderando ruídos que se apresentavam além desta região. Também foi avaliada a região entre 1900 e 800 cm^{-1} do espectro pois a mesma contém as bandas de absorção devidas à deformação axial simétrica da carbonila (ésteres, aldeídos e cetonas), deformação angular simétrica do metileno, deformações axiais e angulares simétricas do C-O (ésteres e alcoóis). Portanto, essa região contém a informação de *fingerprint* para discriminação de diferentes amostras de cafés (BRIANDET *et al.*, 1996; LYMAN *et al.*, 2003; WANG *et al.*, 2009; WANG *et al.*, 2011).

Após estes pré-processamentos foi utilizada a análise de componentes principais (ACP) para a redução da dimensionalidade dos dados (SÁ, 2007). A ACP foi realizada nos dados normalizados assim como na 1ª e 2ª derivada dos espectros. A aplicação da 1ª e da 2ª derivada sobre os dados espectrais brutos é um procedimento que pode destacar ombros espectrais, bem como minimizar o efeito de inclinações provocadas na linha de base dos espectros. Ao aplicarmos as operações de derivação as informações contidas ao longo dos diferentes números de onda são geralmente acentuadas (os ruídos também tornam-se acentuados), portanto deve-se ter cuidado com a qualidade dos espectros com os quais se deseja aplicar o cálculo das derivadas (COSTA FILHO; POPPI, 2002). A quantidade de componentes principais empregadas foi um dos parâmetros otimizados através do simplex sequencial.

2.4 Normalização

Os vetores de entrada (as componentes principais escolhidas) foram pré-processados antes de serem alimentados na rede neural para que as funções de ativação dos neurônios artificiais não fossem facilmente saturadas ou ocorresse um erro de *overflow* (quando um número é grande demais para a estrutura de dados reservada para ele) (HAYKIN, 2001). Os métodos de pré-processamento utilizados foram: máximo e mínimo (minimax), transformação para uma escala entre -1 e 1 e autoescalonamento, vetor de entrada com média zero e variância unitária (PÉREZ-MAGARIÑO *et al.*, 2004).

2.5 Perceptron de múltiplas camadas (MLP)

Foi empregada uma rede neural artificial do tipo Perceptron de Múltiplas Camadas (MLP - *Multi Layer Perceptron*) que é muito utilizada para classificação de padrões (BONA *et al.*, 2011; BORSATO *et al.*, 2011; GALÃO *et al.*, 2011; MARINI, 2009). A arquitetura desta rede é constituída por uma camada de entrada com n neurônios (um para cada componente principal utilizada), uma camada oculta (responsável pela separação dos padrões através da formação de fronteiras de decisão) contendo uma quantidade de neurônios a ser definida e uma camada de saída com 4 neurônios, um para cada região ou genótipo de café, que constrói combinações das fronteiras de decisão formadas pelos neurônios ocultos. Assim, o vetor de resposta desejada apresenta dimensão igual a 4, para uma amostra pertencente a classe k o k -ésimo valor é igual a 1 e todos os outros são zerados. A Figura 2 apresenta um esquema representativo de um perceptron de múltiplas camadas (HAYKIN, 2001).

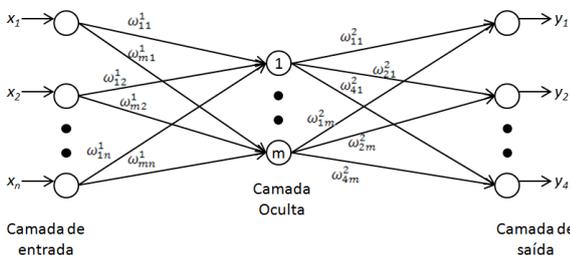


Figura 2 - Representação de um perceptron de múltiplas camadas (MLP).

Para cada conexão entre os neurônios (Figura 2) existe um peso w_{ij}^l associado, sendo i o índice do neurônio de entrada do sinal, j o neurônio de saída do sinal e l a camada onde está localizado o neurônio de entrada. Em

cada um dos N^l neurônios das L camadas da rede MLP é realizado um somatório ponderado pelos pesos sinápticos dos sinais provenientes dos neurônios da camada anterior. Esta soma, chamada de campo local induzido (1), é aplicada a uma função de ativação não linear (2) que irá produzir a saída do neurônio (BISHOP, 1995; HAYKIN, 2001).

$$v_i^l = \sum_{j=0}^{m_i} w_{ij}^l y_j^{l-1} \quad i = 1, 2, \dots, N^l; l = 1, 2, \dots, L. \quad (1)$$

$$y_i^l = \varphi(v_i^l) \quad i = 1, 2, \dots, N^l; l = 1, 2, \dots, L. \quad (2)$$

Nas equações (1) e (2) v_i^l é o campo local induzido do neurônio i na camada l , w_{ij}^l é o peso sináptico associado ao sinal de saída y_j^{l-1} do neurônio j da camada anterior ($l-1$), m_i é o número de entradas associadas ao neurônio i . Para $j=0$, temos $y_0^{l-1} = 1$ e w_{i0}^l é chamado de termo de polarização, ou *bias*, que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo. Para a rede MLP desenvolvida foram testadas para a camada oculta de neurônios artificiais as funções do tipo logística e tangente hiperbólica sigmoide, equações (3) e (4), respectivamente,

$$\varphi = \frac{1}{1 + \exp(-\alpha v)} \quad (3)$$

$$\varphi = \frac{2}{1 + \exp(-2v)} - 1 \quad (4)$$

Para a função logística (3) o símbolo α representa a inclinação e pode ser considerado como um parâmetro a ser otimizado pelo projetista da rede, embora em muitos casos é utilizado um valor unitário para este parâmetro (HAYKIN, 2001; HUANG *et al.*, 2007; MARINI, 2009). Para a camada de saída além destas também a função de ativação linear foi testada.

Uma rede neural extrai seu poder computacional através de sua habilidade de aprender e, portanto de generalizar. A generalização se refere ao fato de a rede neural produzir saídas adequadas para entradas que não estavam presentes durante o treinamento (aprendizagem). O processo utilizado para realizar o processo de aprendizagem é chamado de algoritmo de aprendizagem, cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo de projeto desejado (GALÃO *et al.*, 2011; HAYKIN, 2001; LUCIA; MINIM, 2006). Para a rede neural artificial construída foi utilizado o algoritmo de Levenberg-Marquardt. De maneira geral, o algoritmo propõe uma solução de compromisso entre o algoritmo do gradiente descendente (retropropagação) e o método de Gauss-Newton, que utiliza uma expansão da série de Taylor para aproximar

o modelo de regressão não linear com termos lineares e, então, aplica mínimos quadrados para estimar os parâmetros. No algoritmo de Levenberg-Marquardt existe um parâmetro que regula o tamanho do passo das correções de peso. Para valores muito pequenos desse parâmetro a regra da atualização de pesos leva em consideração o método Gauss-Newton, mas para valores grandes a regra da atualização de pesos leva em consideração o gradiente descendente (BISHOP, 1995).

O processo de aprendizagem é mantido até os pesos sinápticos e os níveis de *bias* se estabilizarem e o erro quadrado médio (equação 5) convergir para um valor mínimo (BONA *et al.*, 2011; HAYKIN, 2001).

$$EQM = \frac{1}{2N} \sum_{n=1}^N \sum_{i=1}^C e_i^2(n) \quad (5)$$

Na equação (5), N é igual ao tamanho do conjunto de treinamento e C igual ao número de neurônios na camada de saída. Tendo como objetivo uma boa generalização, foi utilizado o método de treinamento com parada antecipada para evitar que a rede acabe sendo excessivamente ajustada aos dados de treinamento. É possível identificar o início do excesso de treinamento através do uso de amostras de validação (11 amostras ou 15% das amostras de treinamento), onde além dos dados de treinamento é usado um segundo conjunto de dados que após cada época (uma apresentação completa do conjunto de treinamento) é submetido à rede para verificar a capacidade de generalização desta com dados não utilizados para o ajuste de pesos. Normalmente, o modelo não apresenta resultados tão bons sobre o conjunto de validação quanto sobre o de treinamento, em relação ao qual foi baseado. A curva de aprendizagem, ou curva do erro quadrado médio, decresce monotonamente para um número crescente de épocas. Diferentemente, para as amostras de validação, a curva do erro quadrado médio decresce para um mínimo e então começa a crescer conforme o treinamento continua, ou seja, o que a rede aprende após este ponto é essencialmente o ruído contido nos dados de treinamento. Esta heurística sugere que o ponto mínimo na curva de aprendizagem de validação seja usado como critério de parada da sessão de treinamento da rede (BISHOP, 1995; HAYKIN, 2001).

2.6 Média de ensemble

Um modo de melhorar a capacidade de generalização de uma rede neural é a combinação das respostas de várias RNAs, também chamada de máquinas de comitê, as quais foram treinadas de maneira semelhante de modo que o resultado final é a fusão das respostas de várias redes e é superior àquele alcançável apenas por uma única RNA. Um dos métodos de implementar o comitê de máquinas é a média de ensemble, que consiste em combinar linearmente as respostas de um certo número de redes treinadas através de uma média. O uso de um ensemble diminui o erro final devido a *bias*, pois há um ajuste excessivo

proposital de cada rede e também devido a variância pelo uso de diferentes condições iniciais de treinamento de cada rede (HAYKIN, 2001). Neste trabalho foi utilizada uma média de *ensemble* com 50 redes, sendo que para cada rede um novo conjunto de pesos iniciais e de exemplos de treinamento era gerado aleatoriamente.

2.7 Otimização dos parâmetros da rede

Antes da escolha da melhor rede neural para o problema proposto, alguns parâmetros foram otimizados através do simplex sequencial (Tabela 2). A otimização foi realizada visando minimizar o erro quadrado médio para o conjunto de amostras utilizadas para teste (BONA *et al.*, 2011). O princípio do simplex básico foi deslocar uma figura regular, por exemplo um triângulo equilátero sobre uma superfície, quando duas variáveis estavam sendo consideradas (SPLENDLEY *et al.*, 1962). Os parâmetros escolhidos para a otimização estão representados na Tabela 2. O algoritmo utilizado nesse trabalho está descrito em maiores detalhes em Gao e Han (2012). A otimização segue até que o valor do erro quadrado médio varie apenas dentro da tolerância estabelecida que foi de 0,001 ou pela avaliação gráfica que tem por finalidade auxiliar na visualização da otimização, que está representada como uma suavização na variação das respostas e variáveis independentes.

Tabela 2 - Parâmetros otimizados através do simplex sequencial.

Rede	Parâmetros	Variação
Perceptron de Múltiplas Camadas	Quantidade de neurônios na camada oculta	5 a 15
	Função de pré-processamento das entradas	Mínimax ou Autoescalamento
	Função de ativação utilizada na camada oculta	Logística ou Tangente hiperbólica sigmoide
	Função de ativação utilizada na camada de saída	Logística, Tangente hiperbólica sigmoide e Linear
	Quantidade de variáveis independentes	5 a 20

3 Resultados e Discussão

Um espectro FTIR típico obtido e tratado para o genótipo IPR 105 cultivado em Cornélio Procópio está representado na Figura 3. Através da análise do espectro destaca-se a banda em 1744 cm⁻¹ resultante da deformação axial simétrica do C=O de ésteres de ácidos graxos (WANG *et al.*, 2011).

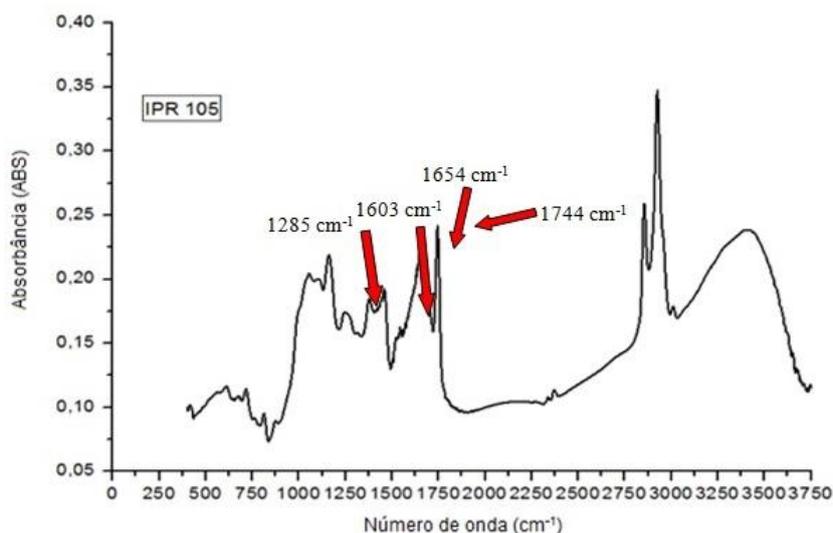


Figura 3 - Espectro obtido para o genótipo IPR 105 cultivado em Cornélio Procópio..

A região em torno de 1285 cm^{-1} está relacionada com os ácidos clorogênicos (BRIANDET *et al.*, 1996). As regiões de 1200 cm^{-1} a 900 cm^{-1} e de 1500 cm^{-1} a 1400 cm^{-1} estão relacionadas ao conteúdo de carboidratos (BRIANDET *et al.*, 1996; RIBEIRO *et al.*, 2010).

3.1 Perceptron de múltiplas camadas (MLP)

Todas as redes utilizadas possuem uma camada oculta. Após a otimização dos parâmetros selecionados (Tabela 2), foram criadas mil redes para cada faixa de trabalho no FTIR e tipo de tratamento do espectro (Tabelas 3 e 4). A melhor rede de cada tipo foi selecionada pelo menor erro quadrado médio e maior porcentagem de classificação para as amostras de teste.

Foi utilizado também a média de *ensemble* com 50 redes (HAYKIN, 2001). Os resultados obtidos para os perceptron de múltiplas camadas propostos para a classificação geográfica e genotípica de café arábica estão representados nas Tabelas 3 e 4. É importante salientar que as condições de cada problema influenciam na escolha dos parâmetros ótimos obtidos para cada rede (SUKTHOMYA; TANNOCK, 2005).

A partir de uma análise da Tabela 3, é possível concluir que, o perceptron de múltiplas camadas que obteve menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica por região de cultivo foi a Rede 4, que foi treinada com os espectros originais (sem derivá-los) e com o número de onda entre $1900\text{-}800\text{ cm}^{-1}$, uma faixa espectral mais específica que contém as impressões bioquímicas (*fingerprints*) para discriminação de diferentes amostras de cafés (BRIANDET *et al.*, 1996; LYMAN *et al.*, 2003; WANG *et al.*, 2009).

Para a segmentação genotípica (Tabela 4) o MLP que obteve menor erro quadrado médio e maior porcentagem de classificação correta (Rede 11) foi o que utilizou a primeira derivada do espectro e, assim como para a classificação geográfica, com a faixa espectral entre $1900\text{-}800\text{ cm}^{-1}$. Os espectros variam muito pouco por se tratarem de uma mesma variedade de café, por isso, a primeira derivada do espectro obteve melhores resultados, pois ela acentua as diferenças espectrais removendo ruídos em uma mesma região espectral e/ou definindo bandas que se encontram sobrepostas (COSTA FILHO; POPPI, 2002). Através da análise da performance dos MLPs nas Tabelas 3 e 4 fica evidente que a classificação genotípica é um problema mais difícil de ser modelado. Os MLPs desenvolvidos apresentam melhor desempenho quando comparados ao SIMCA (*Soft Independent Modelling of Class Analogies*) que obteve 83,33% de acerto para a classificação geográfica e 55,56% para a classificação genotípica (SATO *et al.*, 2012). Devido ao SIMCA, ser um método linear de classificação que se baseia na classificação de uma nova amostra a partir de sua projeção nas componentes principais de cada classe (BRANDEN; HUBERT, 2005), era esperado um desempenho superior dos MLPs desenvolvidos, já que o MLP realiza o mapeamento de relações não lineares entre múltiplas variáveis de entrada e saída.

Destaca-se também que para a classificação das amostras por região as redes com função de ativação da camada oculta do tipo tangente hiperbólica sigmoide obtiveram um erro quadrado médio menor e uma porcentagem de classificação superior aquelas com função de ativação do tipo função logística, provando que para a segmentação geográfica a tangente hiperbólica sigmoide é a função de ativação adequada. Para a classificação dos genótipos a combinação da função de ativação do tipo tangente hiperbólica sigmoide tanto na camada oculta quanto na camada de saída apresentou melhores resultados, sendo esta combinação adequada para este outro problema.

Com relação à normalização, ficou evidente que o método minimax foi superior ao autoescalonamento, devido as redes que apresentaram melhores resultados (Rede 4 e Rede 11) apresentarem o mesmo método de normalização.

O número de variáveis independentes utilizadas na camada de entrada de neurônios também foi otimizado através do simplex sequencial. A média das componentes principais utilizadas foi de 19 para a classificação geográfica e 18 para a classificação genotípica. Tais valores encontram-se próximos do que era esperado, pois qualquer componente principal deve explicar a variância de pelo menos uma variável se a mesma há de ser mantida para interpretação. Cada

variável contribui com um valor 1 do autovalor total. Logo, apenas os fatores que apresentam autovalor maior que 1 são considerados significantes (HAIR *et al.*, 2005). Para este estudo obteve-se 19 componentes principais que apresentaram autovalores maiores que 1, comprovando assim, a capacidade que a ACP possui para compactar os dados e a capacidade de otimização do simplex sequencial.

A partir de uma análise das Tabelas 3 e 4, observa-se uma redução no erro quadrado médio e um aumento na porcentagem de classificação correta quando são comparadas as respostas da melhor rede e da média de ensemble.

Tabela 3 - Resultados obtidos para os perceptron de múltiplas camadas propostos para a classificação geográfica de café arábica.

	Número de onda					
	3750-750 cm ⁻¹			1900-800 cm ⁻¹		
	Rede 1	Rede 2	Rede 3	Rede 4	Rede 5	Rede 6
Espectros	Original	1ª Derivada	2ª Derivada	Original	1ª Derivada	2ª Derivada
Neurônios (camada oculta)	4	8	10	2	2	10
Função de pré-processamento (entradas)^a	MM	MM	AE	MM	MM	MM
Função de Ativação (Camada oculta)^b	FL	THS	THS	THS	FL	THS
Função de Ativação (Saída)^b	THS	THS	L	THS	THS	THS
Componentes utilizadas	20	17	20	20	20	18
Desempenho das melhores redes obtidas para os dados de treinamento						
Erro quadrado médio (Ensemble)	0,0014	0,0016	0,0330	0,0015	0,0031	0,0161
Erro quadrado médio (Melhor Rede)	0,0002	0,0066	0,0463	0,0377	0,0000	0,0196
%Classificação correta (Ensemble)	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
%Classificação correta (Melhor Rede)	100,00%	98,61%	94,44%	95,83%	100,00%	93,06%
Desempenho das melhores redes obtidas para os dados de teste						
Erro quadrado médio (Ensemble)	0,0250	0,0209	0,0792	0,0064	0,0240	0,0650
Erro quadrado médio (Melhor Rede)	0,2147	0,0406	0,1042	0,0353	0,0000	0,1337
%Classificação correta (Ensemble)	94,44%	100,00%	88,89%	100,00%	94,44%	88,89%
%Classificação correta (Melhor Rede)	55,56%	88,89%	72,22%	100,00%	83,33%	93,06%

NOTAS: ^a Função de pré-processamento: MM (Minimax) e AE (Autoescalonamento).

^b Função de ativação: FL (Função Logística), THS (Tangente Hiperbólica Sigmoidal) e L (Linear)

Tabela 4 - Resultados obtidos os perceptron de múltiplas camadas propostos para a classificação genotípica de café arábica.

	Número de onda					
	3750-750 cm ⁻¹			1900-800 cm ⁻¹		
	Rede 7	Rede 8	Rede 9	Rede 10	Rede 11	Rede 12
Espectros	Original	1ª Derivada	2ª Derivada	Original	1ª Derivada	2ª Derivada
Neurônios (camada oculta)	9	6	5	3	6	8
Função de pré-processamento (entradas)^a	AE	MM	MM	MM	MM	AE
Função de Ativação (Camada oculta)^b	THS	FL	THS	FL	THS	THS
Função de Ativação (Saída)^b	L	L	L	L	THS	THS
Componentes utilizadas	20	19	20	19	16	13
Desempenho das melhores redes obtidas para os dados de treinamento						
Erro quadrado médio (Ensemble)	0,0563	0,0305	0,0698	0,0280	0,0251	0,0235
Erro quadrado médio (Melhor Rede)	0,0811	0,0667	0,0171	0,0944	0,0213	0,0229
%Classificação correta (Ensemble)	100,00%	98,61%	97,22%	98,61%	100,00%	100,00%
%Classificação correta (Melhor Rede)	86,11%	91,67%	95,83%	79,17%	95,83%	94,44%
Desempenho das melhores redes obtidas para os dados de teste						
Erro quadrado médio (Ensemble)	0,1381	0,0961	0,1365	0,0862	0,0835	0,1316
Erro quadrado médio (Melhor Rede)	0,1766	0,1557	0,9482	0,1092	0,0986	0,2476
%Classificação correta (Ensemble)	61,11%	77,78%	55,56%	72,22%	83,33%	61,11%
%Classificação correta (Melhor Rede)	55,56%	72,22%	50,00%	72,22%	77,78%	50,00%

NOTAS: ^a Função de pré-processamento: MM (Minimax) e AE (Autoescalonamento).

^b Função de ativação: FL (Função Logística), TH (Tangente Hiperbólica Sigmoidal) e L (Linear).

No trabalho realizado por Bona *et al.* (2011) foram observados resultados semelhantes, comprovando assim, que o uso da resposta do ensemble melhora o desempenho e aumenta a precisão da rede que se torna estável e convergente, duas qualidades importantes em um modelo neural (HAYKIN, 2001).

A Rede 4 obteve 100% de classificação correta (Tabela 3), ou seja, o classificador neural obtido classificou todas as amostras corretamente apesar das cidades estudadas estarem geograficamente próximas. A maioria dos trabalhos realizados para classificação de cafés utilizam amostras geograficamente muito distantes, de espécies diferentes e com diferentes graus de torra (KEMSLEY *et al.*, 1995; BRIANDET *et al.*, 1996; LYMAN *et al.*, 2003; WANG *et al.*, 2009). Neste trabalho, além das amostras serem de uma

mesma espécie de café verde (arábica), os cafés são de regiões geograficamente muito próximas, sendo aproximadamente 195 km a distância máxima entre as cidades (Figura 1), comprovando assim, a eficácia do modelo neural obtido.

A classificação por genótipos obteve porcentagens de classificação correta inferiores (Tabela 4) por se tratar de um problema mais complexo como já comentado anteriormente. A Tabela 5 apresenta a tabela de confusão com os resultados de reconhecimento e classificação genotípica da Rede 11 para as amostras de teste. Esta rede, que alcançou melhores resultados para a classificação por genótipo, obteve 77,78% de classificação correta (classificação incorreta de 4 amostras), duas amostras (IPR 105 e IPR 106) foram confundidas com IPR 99, uma amostra IA 59 foi

confundida com IPR 106 e uma amostra IPR 99 foi confundida com IPR 105. Novamente ressalta-se que estes resultados demonstram que a classificação genotípica é muito mais complexa do que a segmentação geográfica.

Tabela 5 - Tabela de confusão com os resultados de reconhecimento e classificação genotípica da Rede 11 para as amostras de teste.

Classe Prevista	Classe Observada			
	IPR 105	IPR 106	IPR 99	IA 59
IPR 105	3	0	1	0
IPR 106	0	3	0	1
IPR 99	1	1	3	0
IA 59	0	0	0	5

A Rede 4 apresenta 20 neurônios na camada de entrada (componentes utilizadas), 2 neurônios na camada oculta e 4 na camada de saída. A Rede 11 apresenta 16 neurônios na camada de entrada, 6 neurônios na camada oculta e 4 na camada de saída. Devido ao elevado número de neurônios da Rede 11, ela apresenta um número muito alto de pesos sinápticos (130). Este número de pesos sinápticos deve ser menor que o número de exemplos utilizados no treinamento da rede neural (72 exemplos) para que ela possa aprender de maneira confiável, ou seja, devem haver graus de liberdade suficiente para que a aprendizagem da rede seja segura, o que não se observa na Rede 11. Portanto, para melhorar o desempenho da classificação genotípica, os autores sugerem a utilização de outro tipo de RNA com menos parâmetros livres.

4 Conclusão

Os perceptrons de múltiplas camadas otimizados foram capazes de classificar geograficamente as amostras de café arábica. Entretanto, para a classificação genotípica o desempenho não foi totalmente satisfatório. Além disso, o perceptron de múltiplas camadas desenvolvido para classificação genotípica apresenta um número muito elevado de pesos sinápticos, demandando assim, um número alto de graus de liberdade (amostras) para que a rede neural possa aprender de maneira confiável. Confirmando a sugestão da literatura consultada, as redes que utilizaram a faixa de 1900-800 cm^{-1} do espectro FTIR obtiveram menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica tanto por região de cultivo quanto por genótipo. Também ficou evidente que o uso da média de ensemble aumentou o desempenho e reduziu a variabilidade das respostas das redes. O uso do simplex sequencial contornou a dificuldade para se determinar o melhor conjunto de parâmetros das redes além de realizar a maximização

do desempenho e a minimização do tamanho das redes de uma maneira heurística. Assim, a combinação das redes MLP e a espectroscopia de infravermelho com transformada de Fourier (FTIR) mostrou ser uma ferramenta eficiente e rápida para o reconhecimento e classificação geográfica das amostras de café arábica.

5 Referências

- ABIC. **Indicadores da indústria de café no Brasil – 2012**. 2012. Disponível em: <<http://www.abic.com.br/publique/cgi/cgilua.exe/sys/start.htm?sid=61#1389>>. 2012. Acesso em: 26 ago. 2012.
- BISHOP, C. M. **Neural Networks for Pattern Recognition**. Oxford: Oxford University: 1995. 482p.
- BONA, E. et al. Optimized Neural Network for Instant Coffee Classification through an Electronic Nose. **International Journal of Food Engineering**, v. 7, 2011.
- BORSATO, D. et al. Application of artificial neural networks in the geographical identification of coffee samples. **Eur Food Res Technol**, v. 233, p. 533-543, 2011.
- BRANDEN, K. V.; HUBERT, M. Robust classification in high dimensions based on the SIMCA Method. **Chemometrics and Intelligent Laboratory Systems**, v. 79, p. 10-21, 2005.
- BRASIL, Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa n° 8**, de 11 de junho de 2003, que aprova o Regulamento Técnico de Identidade e de Qualidade para a Classificação do Café Beneficiado Grão Cru. Disponível em: <http://www.claspar.pr.gov.br/arquivos/File/pdf/cafebenef008_03.pdf>. Acesso em: 09 nov. 2011.
- BRIANDET, R.; KEMSLEY, E. K.; WILSON, R. H. Approaches to Adulteration Detection in Instant Coffees using Infrared Spectroscopy and Chemometrics. **J. Sci. Food Agric**, v. 71, p. 359-366, 1996.
- BRIANDET, R.; KEMSLEY, E. K.; WILSON, R. H. Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. **J. Agric. Food Chem.**, v. 44, p. 170-174, 1996.
- CECAFÉ. **Dados estatísticos**. Disponível em: <<http://www.cncafe.com.br/ba/file/CECAFÉ%20-%20Resumo%20das%20Exportacoes%20de%20Café%20M ARCO%202012.pdf>>. 2012. Acesso em: 19 out. 2012.
- COSTA FILHO, P. A. D.; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio: determinação simultânea de glicose, maltose e frutose. **Química Nova**, v. 25, p. 46-52, 2002.
- DOWNEY, G. et al. Near- and Mid-Infrared Spectroscopies in Food Authentication: Coffee Varietal Identification. **J. Agric. Food Chem.**, v. 45, p. 4357-4361, 1997.
- FARAH, A. **Coffee as a speciality and functional beverage**. Cambridge: Woodhead Publishing Limited: 2009. p.370-395.
- GALÃO, O. F. et al. Artificial Neural Networks in the Classification and Identification of Soybean Cultivars by Planting Region. **J. Braz. Chem. Soc.**, v. 22, p. 142-147, 2011.

GAO, F.; HAN, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. **Comput Optim Appl**, v. 51, p. 259–277, 2012.

HAIR, J. F. et al. **Análise Fatorial em: Análise Multivariada de Dados**. 5ªed. Porto Alegre: Bookman, 2005. 38p.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2ª edição. Porto Alegre: Bookman, 2001. 900p.

HUANG, Y.; KANGAS, L. J.; RASCO, B. A. Applications of artificial neural networks (ANNs) in food science. **Critical Reviews in Food Science and Nutrition**, v. 47, p. 13, 2007.

KAROUI, R.; DOWNEY, G.; BLECKER, C. Mid-Infrared Spectroscopy Coupled with Chemometrics: A Tool for the Analysis of Intact Food Systems and the Exploration of Their Molecular Structure-Quality Relationships - A Review. **Chemical Reviews**, v. 110, p. 6144-6167, 2010.

KEMSLEY, E. K.; RUAULT, S.; WILSON, R. H. Discrimination between *Coffea arabica* and *Coffea canephora* variant robusta beans using infrared spectroscopy. **Food Chemistry**, v. 54, p. 321-327, 1995.

LUCIA, S. M. D.; MINIM, L. A. **Redes Neurais Artificiais: Fundamentos e Aplicações**. 2006. p.173-194.

LYMAN, D. J., BENCK, R., DELL, S., MERLE, S., MURRAY-WIJELATH, J. FTIR-ATR Analysis of brewed coffee: effect of roasting conditions. **Journal of Agriculture and Food Chemistry**, v. 51, 3268-3273, 2003.

MARINI, F. Artificial neural networks in foodstuff analyses: Trends and perspectives A review. **Analytica Chimica Acta**, v. 635, p. 121–131, 2009.

PÉREZ-MAGARIÑO, S., ORTEGA-HERAS, M., GONZÁLEZ-SAN JOSÉ, M. L., BOGER, Z. Comparative study of artificial neural network and multivariate methods to classify Spanish DO rose wines. **Talanta**, v. 62, n. 5, p. 983-990, 2004.

RIBEIRO, J. S.; SALVA, T. J.; FERREIRA, M. M. C. Chemometric studies for quality control of processed brazilian coffees using drifts. **Journal of Food Quality**, v. 33, p. 212-227, 2010.

RUBAYIZA, A. B.; MEURENS, M. Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy. **Journal of Agriculture and Food Chemistry**, v. 53, p. 4654-4659, 2005.

SATO, H. P.; BONA, E.; LINK, J. V.; LEMES, A. G. Viabilidade de uso do FTIR e análise multivariada para a segmentação de amostras de café verde. **XVII Seminário de Iniciação Científica e Tecnológica da UTFPR - SICITE**. Curitiba: 8 p. 2012.

SINDICAFÉ. Dados Estatísticos. Disponível em: <<http://sindicafe.com.br/plus/modulos/conteudo/?tac=cafe-no-mundo>>. 2011. Acesso em: 24 nov. 2012.

SPLENDLEY, W.; HIMSWORTH, F. R.; HEXT, G. R. **Sequential application of simplex designs in optimization and evolutionary operation**. **Technometrics**, v. 4, p. 441-461, 1962.

SUKTHOMYA, W.; TANNOCK, J. The optimization of neural network parameters using Taguchi's design of experiments approach: an application in manufacturing process modeling. **Neural Computation & Application**, v. 14, p. 7, 2005.

SÁ, J. P. M. D. **Applied Statistics Using SPSS, STATISTICA, MATLAB and R**. 2nd ed. New York: 2007.

WANG, J. et al. Fourier Transform Infrared Spectroscopy for Kona Coffee authentication. **Journal of food science**, v. 74, p. 385-389, 2009.

WANG, N.; FU, Y.; LIM, L.-T. Feasibility study on chemometric discrimination of roasted arabica coffees by solvent extraction and Fourier Transform Infrared Spectroscopy. **J. Agric. Food Chem.**, v. 59, p. 3220-3226, 2011.