

PROGRAMA DE PÓS-GRADUAÇÃO EM PRODUÇÃO VEGETAL

Ludimila Geiciane de Sá

**Inteligência Computacional Aplicada ao Estudo da Divergência e Fenotipagem
em Cultivares de Soja**

**Montes Claros
2018**

Ludimila Geiciane de Sá

Inteligência Computacional Aplicada ao Estudo da Divergência e Fenotipagem em Cultivares de Soja

Dissertação apresentada ao Programa de Pós-graduação em Produção Vegetal da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Produção Vegetal.

Orientador: Prof. Dr. Alcinei Místico Azevedo

Coorientador: Prof. Dr. Carlos Juliano Brant Albuquerque

Coorientador: Prof. Dr. Demerson Arruda Sanglard

Montes Claros
Fevereiro de 2018

S111i Sá, Ludimila Geiciane.
2018

Inteligência computacional aplicada ao estudo da divergência e fenotipagem em cultivares de soja / Ludimila Geiciane de Sá. Montes Claros, 2018.

56 f.: il.

Dissertação (Mestrado em Produção Vegetal) Universidade Federal de Minas Gerais / Instituto de Ciências Agrárias.

Orientador: Alcinei Místico Azevedo.

Banca examinadora: Carlos Juliano Brant Albuquerque, Cesar Fernandes Aquino, Silvia Nietsche.

Inclui referências: f. 21-28, 39-41, 53-55.

1. Redes neurais artificiais. 2. Melhoramento genético vegetal. 3. *Glycine max* (L.) Merrill. I. Azevedo, Alcinei Místico. II. Universidade Federal de Minas Gerais. Instituto de Ciências Agrárias. III. Título

CDU: 633.34

Ludimila Geiciane de Sá

Inteligência Computacional Aplicada ao Estudo da Divergência e Fenotipagem em Cultivares de Soja

Aprovada pela banca examinadora constituída pelos professores:

Prof. Carlos Juliano Brant Albuquerque - Coorientador
ICA / UFMG

Prof. Cesar Fernandes Aquino
UFOB

Prof.^a Silvia Nietsche
ICA / UFMG

Prof. Alcinei Místico Azevedo – Orientador
ICA / UFMG

Montes Claros, 15 de fevereiro de 2018.

Dedico aos meus pais, Sandra e Luiz, e às minhas irmãs,
Ludimara e Ludilene, por todo amor, apoio e incentivo.

AGRADECIMENTOS

A Deus, a fé que me move, por encher minha vida de bênçãos, iluminar meu caminho e nunca me deixar faltar força e coragem.

Aos meus pais, pelo exemplo de caráter, por serem o chão onde me apoio, pelo incentivo e pelo esforço na realização dos meus projetos.

Aos meus orientadores, Alcinei Místico Azevedo e Carlos Juliano Brant Albuquerque, pela confiança e pela paciência, por todos os ensinamentos e pela amizade construída. Foi muito engrandecedor ser orientada por vocês!

Ao Instituto de Ciências Agrárias da Universidade Federal de Minas Gerais, pela infraestrutura disponibilizada.

Aos colegas do Grupo de Pesquisa NEProTeC e GEEA, às alunas do Bic Júnior e aos amigos, pela ajuda impagável e por todos os bons momentos que partilhamos. Sem vocês, eu não conseguiria concluir esta etapa.

A todos os funcionários do ICA-UFMG, em especial, à Flávia Échila, pela prestatividade.

A todos os amigos e familiares que tornaram meus dias mais felizes e contribuíram para o meu crescimento pessoal.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela disponibilidade da bolsa.

Muito obrigada!

Inteligência Computacional Aplicada ao Estudo da Divergência e Fenotipagem em Cultivares de Soja

RESUMO

As redes neurais artificiais são modelos computacionais do cérebro humano que reconhecem padrões e regularidades dos dados e representam uma alternativa como aproximador universal de funções complexas. Podem ter desempenho superior aos modelos estatísticos convencionais, com vantagens de ser não paramétricos, não necessitarem de informações detalhadas sobre os processos físicos do sistema a ser modelado e tolerarem perda de dados. A inteligência computacional apresenta um grande potencial, já é amplamente consolidada nas áreas computacionais e torna-se uma abordagem interessante a ser utilizada na área de melhoramento de plantas. Desse modo, no primeiro artigo, objetivou-se estudar a dissimilaridade genética entre cultivares de soja, utilizando redes neurais do tipo SOM, e testar a eficiência da metodologia utilizada, por meio da análise discriminante de Anderson. Para selecionar a melhor topologia de rede, foram testadas diferentes arquiteturas de rede quanto à maior taxa de acerto médio, pela análise discriminante de Anderson, e o menor número de clusters vazios, efetuando-se 1000 treinamentos. Após, fizeram-se 1000 novos treinamentos de rede e, posteriormente, construiu-se uma matriz de dissimilaridade, observando-se a frequência com que os genótipos foram considerados como sendo de grupos distintos. A consistência do agrupamento foi verificada através da análise discriminante de Anderson. Seis grupos foram formados pelo método UPGMA, indicando a existência de variabilidade para as características avaliadas. Através da análise discriminante de Anderson, 100% dos acessos foram corretamente classificados. Logo, as redes SOM são eficientes para o agrupamento de linhagens de soja, mostrando a existência de dissimilaridade genética e a formação de seis agrupamentos entre as linhagens estudadas. O segundo artigo teve como objetivo auferir metodologias que permitam estimar a área foliar em soja, por meio do uso de redes neurais artificiais, considerando-se diferentes formatos foliares. Foram avaliadas trinta e seis cultivares, as quais foram separadas em três grupos, de acordo o formato foliar. Foram desenvolvidos perceptrons multicamadas, utilizando-se 300 folhas por grupo, destinando-se 70% para treinamento e 30% para a validação. Foram testadas também as medidas morfológicas mais importantes nesses estudos. As RNAs foram eficientes para estimar a área foliar em soja, com coeficientes de determinação próximos a 0,90. Apenas duas medidas foliares dos folíolos são suficientes para estimação da área foliar. A rede 4, treinada com folhas de todos os grupos, foi mais generalista e, conseqüentemente, é a mais indicada para a predição da área foliar em soja.

Palavras-chave: Redes neurais artificiais. Melhoramento genético vegetal. Mapas auto-organizáveis de Kohonen. Perceptrons multicamadas. *Glycine max* (L.) Merrill.

Computational Intelligence Applied to the Study of Divergence and Phenotyping in Soybean Cultivars

ABSTRACT

The artificial neural networks are computational models of the human brain that recognize patterns and regularities of the data and represent an alternative as a universal approximation of complex functions. They may outperform conventional statistical models, with the advantage of being non-parametric, not requiring detailed information on the physical processes of the system being modeled, and tolerating data loss. The computational intelligence has great potential, it is already widely consolidated in the computational areas and becomes an interesting approach to be used in the field of plant breeding. Thus, in the first article, the objective was to study the genetic dissimilarity between soybean cultivars, using SOM type neural networks, and to test the efficiency of the methodology used, through Anderson's discriminant analysis. In order to select the best network topology, different network architectures were tested for the highest average hit rate, Anderson discriminant analysis, and the lowest number of empty clusters, with 1000 trainings performed. After that, 1000 new net trainings were performed and, later, a matrix of dissimilarity was constructed, observing the frequency with which the genotypes were considered as being of different groups. The consistency of the clustering was verified through Anderson's discriminant analysis. Six groups were formed by the UPGMA method, indicating the existence of variability for the evaluated characteristics. Through Anderson's discriminant analysis, 100% of hits were correctly classified. Therefore, the SOM networks are efficient for the grouping of soybean strains, showing the existence of genetic dissimilarity and the formation of six clusters among the studied strains. The second article had as objective to obtain methodologies that allow estimating the leaf area in soybean, through the use of artificial neural networks, considering different leaf formats. Thirty-six cultivars were evaluated, which were divided into three groups, according the leaf format. Multilayer perceptrons were developed, using 300 sheets per group, with 70% for training and 30% for validation. The most important morphological measures were also tested in these studies. The RNAs were efficient to estimate the leaf area in soybean, with determination coefficients close to 0.90. Only two foliar measurements of leaflets are sufficient for estimation of leaf area. The network 4, trained with leaves of all groups, was more general and, consequently, is the most indicated for the prediction of the leaf area in soybean.

Keywords: Artificial Neural Networks. Genetic plant breeding. Kohonen Self-Organizing Maps. Multilayer Perceptrons. Glycinemax (L.) Merrill.

LISTA DE ILUSTRAÇÕES

ARTIGO 1 – MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN NO ESTUDO DA DISSIMILARIDADE GENÉTICA ENTRE CULTIVARES DE SOJA

- FIGURA 1- Porcentagem média de classificações corretas pela análise discriminante de Anderson (A) e porcentagem de clusters vazios, considerando 24 configurações de redes SOM (B)34
- FIGURA 2- Representação gráfica da matriz de dissimilaridade obtida pelos mapas auto-organizáveis de Kohonen.35
- FIGURA 3- Dendograma construído pelo método UPGMA a partir do complemento de uma matriz de coincidência de genótipos agrupados no mesmo cluster, em 1000 redes do tipo SOM, com duas linhas e três colunas (6 neurônios).36
- FIGURA 4- Dispersão gráfica da matriz de dissimilaridade em 1000 redes do tipo SOM, com duas linhas e 3 colunas (6 neurônios).38

ARTIGO 2 – ESTIMATIVA DA ÁREA FOLIAR DE SOJA POR MEIO DE REDES NEURAIIS ARTIFICIAIS

- FIGURA 1- Exemplo dos três formatos de folhas das cultivares de soja – lanceolada (a), triangular (b) e oval-pontiaguda (c) – utilizadas no trabalho.46
- FIGURA 2- Arquitetura da rede neural do tipo perceptron multicamadas, utilizada na estimação da área foliar em função do comprimento e largura de três folíolos.47
- FIGURA 3- Dispersão gráfica da área foliar estimada por redes neurais artificiais do tipo perceptron multicamadas, em função da área foliar observada da amostra de validação de folhas do grupo 1 (a), grupo 2 (b), grupo 3 (c) e para todos os grupos (d).50
- FIGURA 4- Gráficos de superfície resposta da área foliar predita (cm²), em função da largura e comprimento de folhas do grupo 1 (a), grupo 2 (b), grupo 3 (c) e para todos os grupos (c).51

LISTA DE TABELAS

ARTIGO 1 – MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN NO ESTUDO DA DISSIMILARIDADE GENÉTICA ENTRE CULTIVARES DE SOJA

Tabela 1 – Cultivares de soja (<i>Glycine max</i> (L.) Merr) avaliadas no Instituto de Ciências Agrárias da UFMG.....	32
Tabela 2 – Médias dos grupos para os caracteres quantitativos avaliados nas cultivares de soja.....	37
Tabela 3 – Distâncias médias dentro (diagonal principal) e entre (fora da diagonal) clusters, baseando-se na matriz de dissimilaridade obtida por redes neurais do tipo mapas auto-organizáveis de Kohonen.....	38
Tabela 4 – Porcentagem de classificações corretas, pela função discriminante de Anderson, para cada cluster formado pelo método UPGMA	39

ARTIGO 2 – ESTIMATIVA DA ÁREA FOLIAR DE SOJA POR MEIO DE REDES NEURAIIS ARTIFICIAIS

Tabela 1 – Cultivares de soja descrição quanto à empresa e ao grupo do formato foliar. Montes Claros, MG, 2017.....	45
Tabela 2 – Análise descritiva entre as características comprimento, largura e área foliar dos três folíolos de cada grupo em soja.....	48
Tabela 3 – Média do erro quadrático médio (EQM) e do coeficiente de determinação do ($R^2_{\text{médio}}$) para redes ajustadas, considerando diferentes arquiteturas de rede e importância relativa estimada pelo método Garson das medidas contempladas em cada avaliação.....	49
Tabela 4 – Coeficientes de determinação obtidos pelas quatro redes neurais selecionadas na predição da área foliar de plantas do grupo 1, 2, 3 e da soma destes (4)	51

SUMÁRIO

1	INTRODUÇÃO	11
2	OBJETIVOS	12
2.1	Objetivo geral.....	12
2.2	Objetivos específicos.....	12
3	REVISÃO DE LITERATURA	13
3.1	Aspectos gerais da cultura da soja.....	13
3.2	Área foliar	15
3.3	Aspectos do melhoramento genético da soja	15
3.4	Análise discriminante de Anderson	17
3.5	Redes neurais artificiais (RNAs).....	17
3.5.1	Perceptrons multicamadas	19
3.5.2	Mapas auto-organizáveis de Kohonen	20
3.6	REFERÊNCIAS	21
4	ARTIGOS	29
4.1	Artigo 1 - MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN NO ESTUDO DA DISSIMILARIDADE GENÉTICA ENTRE CULTIVARES DE SOJA.....	29
4.2	Artigo 2 - REDES NEURAI ARTIFICIAIS NA ESTIMATIVA DA ÁREA FOLIAR DE SOJA.....	422
5	CONSIDERAÇÕES FINAIS	56

1 INTRODUÇÃO

A soja possui um elevado potencial de uso em nível mundial e, embora haja muitas cultivares de soja de alto potencial produtivo disponível para comercialização, o melhoramento genético da cultura é contínuo. Dentre as principais características-alvo no melhoramento genético da soja estão a produtividade, a resistência a pragas e a doenças, a adaptação a diferentes condições de clima, solo e fotoperíodo (BUENO; MENDES; CARVALHO, 2006; VASCONCELOS *et al.*, 2012). Os programas de melhoramento vegetal utilizam de modelagens estatísticas para auxiliar na identificação de genótipos superiores. Para a utilização dessas metodologias estatísticas, várias pressuposições são feitas quanto à distribuição de probabilidades das variáveis estudadas, caracterizando-as como metodologias de análise paramétrica. Além disso, a maioria dessas técnicas assume natureza linear do fenômeno estudado, enquanto que na realidade isso pode não ser factível. Diferentemente dessas análises, a abordagem da inteligência computacional possui vantagens que podem possibilitar sucessos no melhoramento genético.

A inteligência computacional visa a simular, em máquinas, a capacidade de solucionar problemas e realizar tarefas, que são uma habilidade da inteligência natural do homem (NORVIG; RUSSELL, 2013). O aprendizado de máquinas é a área da inteligência computacional responsável por estudar formas de transferir o conhecimento às máquinas (COPPIN, 2010). Dentre as diversas técnicas baseadas no aprendizado de máquinas, as redes neurais artificiais (RNAs) demonstram ter potencial para auxiliar nas etapas de seleção de genótipos superiores e na recomendação de cultivares.

As RNAs são modelos computacionais do cérebro humano que reconhecem padrões e regularidades dos dados e representam uma alternativa como aproximador universal de funções complexas (GIANOLA *et al.*, 2011), seja lineares ou não. Consequentemente, podem ter desempenho superior aos modelos estatísticos convencionais, com vantagens de serem não paramétricas, não necessitarem de informações detalhadas sobre os processos físicos do sistema a ser modelado (SUDHEER *et al.*, 2003) e tolerarem perda de dados (SILVA *et al.*, 2014). Seu uso, associado a métodos de classificação como estudo da diversidade genética, é um caminho promissor (BARBOSA *et al.*, 2011).

Devido às suas vantagens práticas e econômicas, as RNAs têm sido aplicadas em muitas áreas da agricultura, como para prever a resposta do rendimento da cultura a diferentes parâmetros do solo (LIU *et al.*, 2005; DAIA *et al.*, 2011), agrupamento de genótipos (BARBOSA *et al.*, 2011), predição de evapotranspiração (DEHBOZORGI; SEPASKHAH, 2012) e previsão de produtividade das culturas (SINGH; PRAJNESHU, 2008; DAHIKAR; RODE, 2014). Modelos para estimativa de área foliar foram desenvolvidos para algumas culturas, como tomate (VAZQUEZ-CRUZ *et al.*, 2012), pimenta (AHMADIANMOGHADAM, 2012), milho (ODABAS *et al.*, 2013) e couve (AZEVEDO *et al.*, 2017), mas informações sobre o modelo de estimação da área foliar em soja ainda faltavam.

Dessa forma, a inteligência computacional é uma abordagem interessante a ser utilizada na área de melhoramento de plantas, tanto para diversidade genética quanto estimação de área foliar, uma vez que apresenta grande potencial e já é amplamente consolidada nas áreas computacionais.

2 OBJETIVOS

2.1 Objetivo geral

Estudar a dissimilaridade genética entre cultivares de soja, utilizando redes neurais do tipo *self-organizing map* (SOM), e testar a eficiência dessa metodologia, por meio da análise discriminante de Anderson.

Obter metodologias que permitam a estimação da área foliar em soja eficientemente por redes neurais artificiais (RNAs), considerando-se diferentes formatos foliares.

2.2 Objetivos específicos

- Avaliar a viabilidade das RNAs do tipo SOM como técnica para análise da diversidade genética em cultivares de soja;
- Formar grupos divergentes das cultivares em estudo, com base em dados quantitativos.
- Propor método para escolha de melhor topologia de rede;
- Verificar a eficiência da estimação da área foliar de soja por meio de redes neurais artificiais do tipo *multilayer perceptron* (MPL) e verificar a influência entre os formatos foliares;
- Identificar as variáveis explicativas com maior importância no processo de predição pelas RNAs na cultura de soja.

3 REVISÃO DE LITERATURA

3.1 Aspectos gerais da cultura da soja

A soja é a leguminosa mais produzida no mundo, devido ao seu elevado teor de óleo e proteínas encontradas em suas sementes (LAZZAROTTO; HIRAKURI, 2011).

No Brasil, a cultura da soja ganhou importância econômica a partir da década de setenta do século passado. As primeiras cultivares foram introduzidas a partir do Sul dos Estados Unidos, por exemplo, Bragg, Davis e Lee. Com a crescente importância dessa cultura, os melhoristas começaram a cruzar essas cultivares entre si e com outras fontes, obtendo as primeiras cultivares brasileiras, tais como Industrial, Santa Rosa e Campos Gerais (WYSMIERSKI; VELLO, 2013).

Conforme dados do Departamento de Agricultura dos Estados Unidos – USDA (2017), o país lidera a produção mundial de soja, seguido do Brasil e da Argentina. Esses três países juntos são responsáveis por mais de 80% da produção mundial de soja. Na safra 2016/2017, a área destinada para a cultura nos EUA foi de 33,482 milhões de hectares, produzindo 117,208 milhões de toneladas. No Brasil, a área plantada foi de 33,890 milhões de hectares, produzindo 113,923 milhões de toneladas para mesma safra (CONAB, 2017).

Segundo Carvalho *et al.* (2010), a soja possui grande variabilidade genética entre as cultivares quanto à resposta a fatores ambientais, proporcionando, assim, que uma mesma cultivar possa ter ótima adaptabilidade nas mais variadas regiões do Brasil. A cultura é, portanto, sensível a fatores ambientais como umidade, temperatura e principalmente o fotoperíodo, que podem influenciar diretamente no máximo potencial genético de planta. Isto é, para seu melhor crescimento e desenvolvimento, ela depende das diferentes épocas do ano e da escolha do genótipo para uma região (EMBRAPA, 2010).

A adaptação de cultivares de soja em novas fronteiras agrícolas, ao contrário de outras plantas que dependem, principalmente, das disponibilidades hídricas e térmicas, é controlada, preponderantemente, pelo fotoperíodo. Enquanto certas plantas possuem um fotoperíodo crítico, no qual se desenvolvem vegetativamente de maneira adequada, a soja floresce somente quando o fotoperíodo do ambiente de cultivo passa a ter valor inferior ao fotoperíodo crítico (MONDINE *et al.*, 2001), ou seja, a soja é uma planta de dias curtos, pois é sensível ao número de horas de ausência de luz no período de 24 horas. Garner e Allard (1920) foram uns dos primeiros a relatarem a importância do período luminoso no processo de indução floral na soja. Segundo eles, as cultivares têm uma faixa limitada de adaptação; visto que a duração do período escuro é em função da latitude, a adaptabilidade de cada cultivar varia à medida que é deslocada em direção ao sul ou ao norte. Portanto, cada variedade tem sua melhor adaptação em certa região.

A classificação do ciclo total das cultivares em superprecoce, precoce, semiprecoce e médio é válida dentro de cada faixa de adaptação. Fora de uma dada faixa, uma cultivar tardia no sul torna-se precoce no Brasil-Central (EMBRAPA, 2010). Portanto, esses grupos não são concordantes entre as cultivares e as diversas regiões de adaptação; dessa forma, uma mesma cultivar pode alcançar diferentes ciclos, conforme as condições de manejo e, principalmente, as condições edafoclimáticas entre regiões distintas (EMBRAPA, 2006).

O sistema de classificação da soja por grupos de maturação varia de zero a 10, ou seja, quanto maior é o seu número, mais próximo ao Equador será sua região de adaptação. No Brasil, os grupos indicados variam de 5.5 a 10. Para exemplificar, uma variedade 8.9 tem um ciclo mais longo que uma 8.7, e assim sucessivamente. Podemos dizer que, de forma geral, para cada aumento de número depois do ponto, teremos de 1,5 a 2 dias a mais de ciclo. Cada grupo de maturação se ajusta melhor em determinada faixa de latitude, em função de sua resposta ao fotoperíodo, variando de acordo com a quantidade de horas/luz a que é exposta. Quanto mais perto do Equador, na primavera e no verão, a quantidade de horas/luz é menor em relação às regiões mais ao Sul. Para a planta de soja, quanto menor a quantidade de luminosidade que ela recebe, mais rapidamente entrará na fase reprodutiva (florescimento), encurtando assim seu ciclo e reduzindo a altura das plantas. Então, ao localizarmos uma cultivar de grupo de maturação 9.0 no Sul do Brasil, esse material alongará seu ciclo de forma a comprometer totalmente sua produtividade (PENARIOL, 2000).

A soja é uma planta herbácea, anual, pertencente ao reino Plantae, classe Magnoliopsida, ordem Fabales, família Fabaceae (Leguminosae) e subfamília Faboideae (Papilionoideae). Autógama, possui $2n=40$ cromossomos (SEDIYAMA, 2009).

Possui sistema radicular pivotante, com a raiz principal bem desenvolvida e raízes secundárias em grande número, que se associam a bactérias *Rhizobium*, fixadoras de nitrogênio, formando nódulos (SEDIYAMA, 2009; EMBRAPA, 2014).

O seu caule é herbáceo, ereto e bastante ramificado. Sua altura varia de 0,60 m a 1,50 m (EMBRAPA, 2014). O desenvolvimento é dependente do tipo de crescimento, podendo sofrer influência das condições ambientais. Em cultivares de crescimento determinado e semideterminado, a gema terminal transforma-se em uma inflorescência terminal; em cultivares de crescimento indeterminado, o caule continua a crescer após o florescimento (MÜLLER, 1981; SEDIYAMA, 2009).

As plantas de soja, ao longo do seu desenvolvimento, possuem três tipos distintos de folhas: as cotiledonares, as unifolioladas e as trifolioladas, as quais possuem tamanhos, formatos e posicionamento diferentes (MÜLLER, 1981). As folhas são alternadas, compostas de três folíolos, seu comprimento pode variar de 0,5 a 12,5 cm (EMBRAPA, 2014). Segundo Müller (1981), os folíolos laterais são inseridos na raque através dos peciólulos, geralmente medindo menos de 1 cm. A inserção do folíolo terminal ocorre por meio do alongamento da raque; seu comprimento varia de 1 a 5 cm.

As flores da soja são completas, ou seja, formadas por cálice, corola, androceu e gineceu. Nascerem em racemos axilares variando de 2 a 35, do tipo papilionada, podendo ser brancas, amarelas ou violáceas, dependendo da cultivar (MÜLLER, 1981; SEDIYAMA *et al.*, 1996; SEDIYAMA, 2009).

Os frutos são do tipo vagem, achatados, curtos, pubescente e deiscente, geralmente possuem 2 a 7 cm de comprimento e podem chegar a 400 por planta. Cada vagem possui de 0 a 5 sementes, porém, a maioria das cultivares possui 2 a 3 sementes por vagem. A vagem possui variações quanto à forma, tamanho, cor do tegumento, intensidade do brilho e cor do hilo (MÜLLER, 1981; SEDIYAMA *et al.*, 1996; SEDIYAMA, 2009).

3.2 Área foliar

A área foliar é um dos principais parâmetros do crescimento vegetal, pois está relacionada com diversos processos fisiológicos da planta, tais como fotossíntese, respiração e transpiração (SILVA *et al.*, 2008; GHOREISHI *et al.*, 2012). Segundo Monteiro *et al.* (2005), a área foliar é um indicador de grande importância, sendo utilizada para investigar adaptação ecológica, competição com outras espécies e efeitos do manejo, além de ser usada para a determinação do índice de área foliar, que pode estimar a produtividade de um ecossistema vegetal, seu crescimento e o desenvolvimento das folhas.

Diversas são as formas de se medir a área foliar, porém muitas são inadequadas por serem destrutivas e/ou por dependerem de aparelhos disponíveis somente em laboratórios ou, ainda, por demandarem excessiva mão de obra para execução. É uma das características mais difíceis de serem mensuradas (BIANCO *et al.*, 2007).

Marshall (1968) classificou os métodos de estimação da área foliar em destrutivos e não destrutivos. Nos métodos destrutivos, é necessária a retirada da folha ou outras estruturas e, dependendo do número de plantas e da parcela experimental, muitas vezes não é possível. Nos métodos não destrutivos, as medidas são realizadas diretamente na planta, sem a exigência de retirada de estruturas, mantendo sua integridade e permitindo a continuidade das medições na mesma planta (ADAMI *et al.*, 2008).

Entre os métodos destrutivos empregados estão o método gravimétrico, dos discos foliares, os medidores automáticos, como os integradores de área foliar, e também a utilização de imagens digitais obtidas atrás de scanner. Segundo Odabas *et al.* (2013), esses são dispositivos muito caros e complexos para estudos simples. Entre os métodos não destrutivos, estão aqueles que estimam a área foliar através de medidas lineares das folhas, como o comprimento e largura, e através da equação de regressão linear entre as medidas lineares tomadas na folha (LIMA *et al.*, 2012; ODABAS *et al.*, 2013). A vantagem desse método é que a mensuração de largura e comprimento pode ser feito com alta precisão, sendo um método barato, rápido, menos laborioso e não destrutivo. Porém, o uso de regressão tem a desvantagem de considerar, em geral, natureza estritamente linear entre as variáveis. Além disso, no caso da soja, as folhas, com exceção da primeira folha, são trifoliadas. Assim, a necessidade de medir os três folíolos aumenta em três vezes o tempo para medir uma folha, comparada com espécies de folhas simples.

Uma alternativa é o uso das RNAs. A técnica tem capacidade de resolver sistemas não lineares e complexos (SILVA *et al.*, 2014). Diferentes pesquisadores mostraram que os modelos RNAs frequentemente dão melhores resultados, se comparados com os métodos tradicionais (MOOSAVI; SEPASKHA, 2012). Trabalhos com o uso da rede neural artificial para predição da área foliar demonstraram ser um método viável, rápido e preciso (ODABAS *et al.*, 2013; AZEVEDO *et al.*, 2017).

3.3 Aspectos do melhoramento genético da soja

As pesquisas sobre o melhoramento genético da soja no Brasil são recentes; as primeiras cultivares desenvolvidas no Brasil foram lançadas no ano de 1981. Desde então, os programas de melhoramento genético da cultura têm sido essenciais para atender à crescente demanda pelo grão e de

sementes. A geração de variabilidade genética e a ampliação da base genética, associadas aos métodos modernos de seleção dos melhores genótipos de uma população, têm permitido superar os patamares de produtividade década após década (COSTA *et al.*, 2004).

Os programas de melhoramento da soja selecionam linhagens mais produtivas, resistentes a pragas e doenças e adaptadas a diferentes condições edafoclimáticas, que permitam ser incorporadas a diferentes áreas. Essas características ajudam selecionar genótipos com rendimentos mais elevados, possibilitando maior rentabilidade para os produtores. (BUENO; MENDES; CARVALHO, 2006; VASCONCELOS *et al.*, 2012).

Em programas de melhoramento da cultura, os delineamentos como blocos ao acaso e *lattices* são uns dos mais utilizados. Podem-se citar como vantagens desses delineamentos a maior flexibilidade quanto ao número de tratamentos, a possibilidade de testar parte dos tratamentos sem repetições e, conseqüentemente, a economia de área, serviço e insumos. Entretanto, sua aplicação tem se deparado com uma limitação de ordem estatístico-experimental, que pode ser atribuída à dificuldade de se implementar uma análise estatística mais eficiente, com a rapidez operacional que o processo exige (DUARTE, 2000).

No melhoramento de plantas, tem-se exigido a utilização de técnicas multivariadas no estudo da divergência genética (ARAVIND, 2006; IQBAL *et al.*, 2010, ALMEIDA *et al.*, 2011; LIU, *et al.*, 2011; MALIK *et al.*, 2011; RIGON *et al.*, 2012; SALIMI *et al.*, 2012; VILELLA *et al.*, 2014) e na seleção de genótipos (VIANNA *et al.*, 2013; DALLASTRA *et al.*, 2014). Tais técnicas constituem-se em ferramentas de grande interesse no melhoramento genético, por conseguir unificar várias informações de um conjunto de caracteres (FONSECA *et al.*, 2006; CRUZ *et al.*, 2012).

Dentro das técnicas multivariadas, existem dois principais métodos de agrupamentos: os aglomerativos e os hierárquicos. O método não ponderado de agrupamento aos pares UPGMA (Unweighted Pair Group Method with Arithmetic Mean) tem a finalidade de reunir por critério de classificação os genitores em grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (KLOSTER *et al.*, 2011). Conhecendo-se a classificação dos grupos, podem-se estabelecer partições em maior ou menor número de classes (BOUROCHE; SAPORTA, 1981; ALEIXO *et al.*, 2007). O estabelecimento dos grupos de indivíduos permite verificar os parâmetros que evidenciam as características homogêneas dentro de um grupo e também as principais diferenças entre esses grupos (HAIR *et al.*, 2005).

As redes neurais artificiais são uma técnica de bioinformática com abordagem multivariada. A técnica vem sendo usada no contexto da agricultura de diversas formas, por exemplo, na identificação de estágios de desenvolvimento de pragas ou doenças (SILVA *et al.*, 2005), na estimativa da produtividade de diferentes culturas (PICOLI, 2007; LEAL *et al.*, 2015), na classificação das imagens de satélite (FRANÇA, 2010; SCHIMITH *et al.*, 2009; VIEIRA *et al.*, 2009; WATANABE *et al.*, 2009; CHAGAS *et al.*, 2009) e em agrupamento de dados (DUTRA *et al.*, 2009; BARBOSA *et al.*, 2011).

De acordo com Galvão *et al.* (1999), devido à sua estrutura não linear, as redes neurais artificiais podem capturar mais características complexas dos dados, o que nem sempre é possível com técnicas estatísticas tradicionais. Para Sudheer *et al.* (2003), a maior vantagem das redes neurais artificiais sobre os métodos convencionais é que elas não requerem informações detalhadas sobre os processos físicos do sistema a ser modelado. O uso de redes neurais associadas aos métodos de classificação é um

caminho promissor. Esses classificadores têm a vantagem de serem não paramétricos, exigindo amostras pequenas para treinamento (KAVZOGLU; MATHER, 2003), e toleram dados faltantes (BISHOP, 1995).

3.4 Análise discriminante de Anderson

Uma das técnicas de análise multivariada que permite alocar um novo indivíduo a uma das várias populações distintas, previamente conhecidas, é a análise discriminante. Segundo Cruz (2006), a análise discriminante de Anderson tem por finalidade conseguir a melhor discriminação entre os indivíduos, alocando-os em suas devidas populações, e permitir a classificação de novos materiais genéticos, de comportamento desconhecido, em populações já conhecidas. A utilização dessa técnica é simples e possui alta eficiência para uma ampla variedade de estruturas populacionais. A sua aplicação tem a finalidade de diminuir a probabilidade de uma classificação errônea, isto é, de se classificar um “indivíduo” em uma população, quando este na realidade pertence a outra (CRUZ *et al.*, 2012).

De acordo com Anderson (1958), quando se tem várias populações e necessita-se colocar um novo indivíduo em cada uma delas, é importante que os indivíduos estejam divididos em populações distintas, e que sejam estabelecidas as probabilidades “a priori” para as várias populações, pois existem casos em que a probabilidade de um determinado indivíduo pertencer a uma dada população pode ser muito distinta da dele pertencer a outra, e torna-se muito importante a experiência do pesquisador.

A partir dessas informações, é possível gerar funções, que são combinações lineares das características avaliadas, para obter a melhor discriminação entre os indivíduos e encontrar as populações a que pertencem. As funções também permitem a classificação de novos genótipos, de comportamento desconhecido, nas populações já conhecidas (SANT’ANNA, 2014).

A análise discriminante de Anderson tem sido utilizada para verificar a consistência dos agrupamentos propostos por outras técnicas, em acessos de espinheira-santa (MARIOT *et al.*, 2008) e na cultura do mamão (BARBOSA *et al.*, 2011). Em algumas culturas, a análise discriminante de Anderson teve por finalidade classificar genótipos e discriminar espécies (ASSIS *et al.*, 2003; ABREU *et al.*, 2004; SUDRÉ *et al.*, 2006; ZANINE *et al.*, 2007; MONTEIRO, 2008), e também foi utilizada para identificar e avaliar características de diferenciação das cultivares (NOGUEIRA *et al.*, 2008; LIMA JÚNIOR *et al.*, 2009).

3.5 Redes neurais artificiais (RNAs)

A inteligência computacional tem auxiliado no processo de tomada de decisão em diversas áreas da ciência, permitindo uma nova abordagem e demonstrando ter grande potencial para o melhoramento genético vegetal (VENTURA *et al.*, 2013).

Em muitos casos, o pesquisador dispõe de dados experimentais adequados à realização das análises biométricas para avaliação dos experimentos com a finalidade de discriminar indivíduos ou populações, entretanto, as análises biométricas nem sempre são capazes de produzir resultados satisfatórios, pois o modelo adotado e as estatísticas utilizadas podem ser insuficientes para descrever e caracterizar corretamente cada população. Assim, a realização de análises por meio de métodos

computacionais que sejam capazes de aprendizagem e generalização de todos os dados disponíveis, tolerantes a dados faltantes e não lineares, representa um grande progresso para os estudos estatísticos e para o melhoramento genético (SANT'ANNA, 2014).

Trata-se da inteligência artificial. Com o avanço da tecnologia, ela vem sendo cada vez mais utilizada e tem permitido a invenção de modelos de neurônios artificiais muito próximos dos neurônios biológicos, que conseguem se conectar em redes neurais artificiais (RNAs). Essas redes são capazes de análises de várias situações, aprendizagem, reconhecimento de padrões e generalização (BRAGA *et al.*, 2011).

A rede neural artificial é um modelo matemático baseado no funcionamento do sistema nervoso biológico, sendo composta por unidades chamadas de neurônios (CAMPOS *et al.*, 2016). Segundo Oikawa e Ishiki (2013), as RNAs podem ser definidas como uma ferramenta estatística, cujo princípio de funcionamento do modelo matemático é inspirado no funcionamento da estrutura neural de organismos inteligentes, que adquirem conhecimento através de experiência e, por meio do processamento de informações, geram uma saída (dados preditos), a partir de uma ou mais entradas apresentadas (preditores).

Os neurônios ou unidades de processamento calculam determinadas funções matemáticas, normalmente não lineares (ANOCHI, 2015). Podem ser distribuídos em uma ou mais camadas e interligados por um grande número de conexões, os pesos sinápticos, os quais armazenam o conhecimento representado no modelo e servem para avaliar a entrada recebida por cada neurônio da rede. Os dados de entrada e saída previamente conhecidos são apresentados várias vezes. Assim, a RNA aprende a relação entre elas, entrada e saída e, através de um algoritmo de minimização de erros, procura reduzir o erro quadrático médio em cada iteração do treinamento (DANTAS *et al.*, 2016).

A importância do uso de RNAs reside no seu poder de resolução de problemas do tipo não linear, como os problemas de classificação, por exemplo (LIMA *et al.* 2016). Nesse contexto, diversas pesquisas no meio agrícola e agropecuário têm sido feitas nos últimos anos utilizando RNAs (PEREIRA, 2013; SANGOI, 2014). Também têm sido usadas em substituição aos modelos de regressão, apresentando frequentemente superioridade na acurácia das estimativas de variáveis de interesse (GÖRGENS *et al.*, 2009; SILVA *et al.*, 2009).

Existem diferentes tipos de RNAs e diferentes maneiras de classificá-las. Uma das maneiras mais importantes é quanto à forma de aprendizado. O tipo de aprendizado depende da forma em que é efetuado o ajuste nos parâmetros (HAYKIN, 2009). Segundo Jang *et al.* (2012), as redes neurais podem ser classificadas como supervisionadas, em que há presença de realimentação explícita dos dados do mundo exterior, e em não supervisionadas, em que não ocorre essa realimentação dos dados.

No aprendizado não supervisionado, as redes neurais utilizam dados semelhantes para calcular as correlações entre eles e responder corretamente. Já nas redes neurais de aprendizado supervisionado, existe um agente externo indicando a resposta desejada, assinalando os acertos e erros para os padrões de entrada (BRAGA *et al.*, 2011).

No aprendizado não supervisionado, a própria rede se auto-organiza em relação às particularidades dos dados. Nessa forma de aprendizado, os pesos sinápticos e os limiares dos neurônios são ajustados pelo algoritmo de aprendizado, levando em consideração as características

internas da RNA (CARNEIRO, 2015). Por exemplo, mapa auto-organizável de Kohonen, redes de Hopfield e memória associativa bidirecional são redes com métodos de aprendizado não supervisionado.

A RNA com o tipo de aprendizado supervisionado necessita que sejam disponibilizadas amostras, com entradas e saídas desejadas, representativas do processo a ser mapeado. Portanto, os pesos sinápticos das RNAs são ajustados através de comparação feita pelo próprio algoritmo de aprendizagem, que supervisiona a diferença entre as respostas produzidas pela rede em relação às saídas desejadas. Isso faz com que as RNAs tenham a importante capacidade de generalização (CARNEIRO, 2015). Perceptrons, Adaline e Madaline, redes de função de base radial são algumas dentre as redes com métodos de aprendizado supervisionado.

Segundo Braga *et al.*, (2011), as RNAs são capazes de resolver problemas de diferentes naturezas, por exemplo, classificação, predição, agrupamento, entre outros. As RNAs também podem ser aplicadas em uma gama extensa de áreas.

Na área agrícola, as RNAs possibilitam a predição de produção de diferentes culturas (KAUL *et al.*, 2005; JI *et al.*, 2007; PICOLI, 2007; SOARES *et al.*, 2015), na identificação dos estágios iniciais do desenvolvimento de pragas ou doenças (SILVA *et al.*, 2005), na abundância de peixes (IGLESIAS *et al.*, 2006), na classificação das imagens de satélite para diversos fins (SCHIMITH *et al.*, 2009; VIEIRA *et al.*, 2009; WATANABE *et al.*, 2009; FRANÇA, 2010), na classificação de etanol (ROMAGNOLI *et al.*, 2016) e vinho (LIMA *et al.*, 2016), e área foliar (ODABAS *et al.*, 2013; AZEVEDO *et al.*, 2017; SHABANI *et al.*, 2017), dentre outros. No melhoramento genético as RNAs têm sido empregadas em estudos de diversidade genética (BARBOSA *et al.*, 2011), análises de adaptabilidade e estabilidade (BARROSO *et al.*, 2013; NASCIMENTO *et al.*, 2013) e predição de valor genético (SILVA *et al.*, 2014).

3.5.1 Percéptrons Multicamadas

Os Percéptrons Multicamadas (MLP) se caracterizam por possuir ao menos uma camada de pequenas unidades chamadas de Percéptrons. Seu treinamento é feito de forma supervisionada. Esse processo de treinamento baseia-se em um algoritmo conhecido como *backpropagation*, no qual o erro da análise é minimizado constantemente (TONSIG, 2000; FANUCCHI *et al.*, 2013).

Esse processo de treinamento utiliza pares de entrada e saída para, por meio de um mecanismo de correção de erros, ajustar os pesos e limiares da rede (BRAGA *et al.*, 2011). O treinamento constitui-se de duas fases: *forward* e *backward*. Na fase de *forward*, os sinais de uma amostra do conjunto de treinamento são inseridos nas entradas da rede e são propagados, camada a camada, até a produção das respectivas saídas. Em seguida, são calculados os erros entre as respostas produzidas pela rede e as respostas desejadas. Em função dos valores desses erros, o algoritmo aplica a fase *backward*, que executa alterações nos pesos sinápticos e limiares de todos os neurônios da rede. Desse modo, a aplicação sucessiva de ambas as fases faz com que os pesos sinápticos e limiares de neurônios se ajustem em cada iteração, de modo que haja uma gradativa diminuição da soma dos erros produzidos pelas respostas da rede em comparação às desejadas (SILVA; SPATTI; FLAUZINO, 2010).

3.5.2 Mapas auto-organizáveis de Kohonen

As redes neurais de Kohonen ou mapas auto-organizáveis (SOM) são modelos neurocomputacionais inicialmente desenvolvidos pelo professor Teuvo Kohonen (KOHONEN, 1990). São utilizadas principalmente para análises de agrupamentos e possuem a vantagem de trabalhar com dados não lineares. O modelo foi inspirado a partir da analogia com o cérebro humano; são algoritmos computacionais, representando, de maneira bastante elementar, o mecanismo de funcionamento cerebral (KOHONEN, 2001). As redes SOM possuem aprendizado competitivo e não supervisionado, em que a base dos algoritmos de adaptação dos pesos sinápticos é a competição global entre os neurônios.

A auto-organização da rede é realizada da seguinte forma: quando um vetor de entrada é apresentado, a rede identifica o neurônio que apresenta vetor peso mais semelhante à entrada apresentada. No decorrer do treinamento, a rede diminui a distância entre o vetor peso e o vetor de entrada, incluindo esse ajuste para a vizinhança do neurônio vencedor. Assim, permite que a rede construa um mapa topológico, em que neurônios que estão topologicamente próximos respondam, de forma semelhante, a padrões de entrada semelhantes (BRAGA *et al.*, 2011).

De acordo com Haykin (2009), a formação do mapa auto-organizável pode ser compreendida em três etapas: a competitiva, a cooperativa e a adaptativa.

Na fase competitiva, para cada padrão de entrada, os neurônios de saída competem entre si, mediante alguma função discriminante. O neurônio com o maior valor de função discriminante é denominado vencedor da competição. Entre as funções de distâncias utilizadas para quantificar a semelhança entre os vetores da rede, uma das mais empregadas é a distância Euclidiana (HAYKIN, 2009).

Na etapa cooperativa, o neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados, fornecendo, dessa forma, as bases para a cooperação entre os neurônios vizinhos (HAYKIN, 2009).

Na fase adaptativa, os neurônios excitados aumentam os valores de sua função discriminante em relação ao padrão de entrada, por meio de ajustes em seus pesos sinápticos. Conseqüentemente, a grade vai se tornando auto-organizável (HAYKIN, 2009).

3.6 REFERÊNCIAS

- ABREU, F. B.; LEAL, N. R.; RODRIGUES, R.; AMARAL JÚNIOR, A. T.; SILVA, D. J. H. Divergência genética entre acessos de feijão-de-vagem de hábito de crescimento indeterminado. **Horticultura Brasileira**, v. 22, n. 3, p. 547-552, 2004. Disponível em: <<https://bit.ly/2HHiO54>>. Acesso em: 30 mar. 2016.
- ADAMI, M.; HASTENREITER, F. A.; FLUMIGNAN, D. L.; FARIA, R. D. Estimativa de área de folíolos de soja usando imagens digitais e dimensões foliares. **Bragantia**, v. 67, n. 4, p. 1053-1058, 2008. Disponível em: <<https://bit.ly/2JI02h3>>. Acesso em: 30 set. 2017.
- AHMADIANMOGHADAM, H. Prediction of pepper (*Capsicum annum* L.) leaf area using group method of data handling- type neural networks. **International Journal of AgriScience**, v. 2, n. 11, p. 993-999, 2012.
- ALEIXO, S. S.; SOUZA, J. G.; FERRAUDO, A. S. Técnicas de análise multivariada na determinação de grupos homogêneos de produtores de leite. **Revista Brasileira de Zootecnia**, v. 36, n. 6, supl. 0, p. 2168-2175, 2007. Disponível em: <<https://bit.ly/2HpjmsU>>. Acesso em: 17 mar. 2017.
- ALMEIDA, R. D.; PELUZIO, J. M.; AFFÉRI, F. S. Divergência genética entre cultivares de soja, sob condições de várzea irrigada, no Sul do Estado Tocantins. **Revista Ciência Agronômica**, v. 42, n. 1, p. 108-115, 2011. Disponível em: <<https://bit.ly/2HTCOPF>>. Acesso em: 14 jun. 2016.
- ANDERSON, T. W. **An introduction to multivariate statistical analysis**. New York: John Wiley & Sons, 1958. 345 p.
- ANOCHI, J. A. Previsão climática de precipitação por redes neurais autoconfiguradas. 2015. 135 f. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2015. Disponível em: <<https://bit.ly/2qQypps>>. Acesso em: 25 fev. 2017.
- ARAVIND, G. **Genetic variability and diversity studies in soybean [Glycine max (L.) Merrill]**. 2006. 103 f. Thesis (Master in Genetics and Plant Breeding) – University of Agricultural Sciences, Dharwad, 2006. Disponível em: <<https://bit.ly/2vCCthA>>. Acesso em: 30 mar. 2016.
- ASSIS, G. M. L.; EUCLYDES, R. F.; CRUZ, C. D.; VALLE, C. B. Discriminação de espécies de *Brachiaria* baseada em diferentes grupos de caracteres morfológicos. **Revista Brasileira de Zootecnia**, v. 32, n. 3, p. 576-584, 2003. Disponível em: <<https://bit.ly/2Hm0L0O>>. Acesso em: 14 abr. 2017.
- AZEVEDO, A. M.; ANDRADE JÚNIOR, V. C.; SOUSA JÚNIOR, A. S.; SANTOS, A. A.; CRUZ, C. D.; PEREIRA, S. L.; OLIVEIRA, A. J. M. Eficiência da estimação da área foliar de couve por meio de redes neurais artificiais. **Horticultura Brasileira**, v. 35, n. 1, p. 14-19, 2017. Disponível em: <<https://bit.ly/2Jl40X4>>. Acesso em: 30 mar. 2016.
- BARBOSA, C. D.; VIANA A. P.; QUINTAL S. S. R.; PEREIRA, M. G. Artificial neural network analysis of genetic diversity in *Carica papaya* L. **Crop Breeding and Applied Biotechnology**, v. 11, p. 224-231, 2011. Disponível em: <<https://bit.ly/2qSILGI>>. Acesso em: 22 ago. 2017.
- BARROSO, L. M. A.; NASCIMENTO, M.; NASCIMENTO, A. C. C.; SILVA, F. F.; FERREIRA, R. P. Uso do método de EBERHART e RUSSELL como informação a priori para aplicação de redes neurais artificiais e análise discriminante visando a classificação de genótipos de alfafa quanto à adaptabilidade e estabilidade fenotípica. **Revista Brasileira de Biometria**, v. 31, n. 2, p. 176-188, 2013. Disponível em: <<https://bit.ly/2vK3cJ4>>. Acesso em: 30 out. 2017.
- BIANCO, S.; BIANCO, M. S.; PAVANI, M. C. M. D.; DUARTE, D. J. Estimativa da área foliar de *Ipomoea hederifolia* e *Ipomoea nil* Roth. usando dimensões lineares do limbo foliar. **Planta Daninha**, v. 25, n. 2, p. 325-329, 2007. Disponível em: <<https://bit.ly/2qVlxfx>>. Acesso em: 30 out. 2017.
- BISHOP, C. M. **Neural networks for pattern recognition**. Oxford: Clarendon Press, 1995. 482 p.
- BOUROCHE, J. M.; SAPORTA, L. **L'analyse des données**. Paris: Presses Universitaires de France, 1981. 127 p.

BRAGA, A. P., CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC, 2011.

BUENO, L. C. S.; MENDES, A. N. G.; CARVALHO, S. P. **Melhoramento genético de plantas: princípios e procedimentos**. 2. ed. Lavras: UFLA, 319 p, 2006.

CAMPOS, B. P. F.; SILVA, G. F.; BINOTI, D. H. B.; MENDONÇA, A. R.; LEITE, H. G. Predição da altura total de árvores em plantios de diferentes espécies por meio de redes neurais artificiais. **Pesquisa Florestal Brasileira**, v. 36, n. 88, p. 375-385, 2016. Disponível em: <<https://bit.ly/2HTKyBd>>. Acesso em: 30 set. 2017.

CARNEIRO, V. Q. Rede neural e lógica fuzzy aplicadas no melhoramento do feijoeiro. 2015. 91 f. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, 2015. Disponível em: <<https://bit.ly/2qTs8bD>>. Acesso em: 16 nov. 2017.

CARVALHO, E. R.; RESENDE, P. M.; OGOSHI, F. G. A.; BOTREL, E. P.; ALCANTRA, H. P.; SANTOS, J. P. Desempenho de cultivares de soja [*Glycine max* (L.) Merrill] em cultivo de verão no Sul de Minas Gerais. **Ciência e Agrotecnologia**, v. 34, n. 4, p. 892-899, 2010. Disponível em: <<https://bit.ly/2qT798Q>>. Acesso em: 17 nov. 2017.

CHAGAS, C. S.; VIEIRA, C. A. O.; FERNANDES FILHO, E. I.; CARVALHO JUNIOR, W. Utilização de redes neurais artificiais na classificação de níveis de degradação em pastagens. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n. 3, p. 319-327, 2009. Disponível em: <<https://bit.ly/2JkBXqE>>. Acesso em: 10 dez. 2017.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Perspectivas para a agropecuária: safra 2013/2014**. Brasília: CONAB, 2013. v. 1. Disponível em: <<https://bit.ly/2vGWDHq>>. Acesso em: 19 ago. 2017.

COPPIN, B. **Inteligência artificial**. Rio de Janeiro: LTC, 2010.

COSTA, M. M.; MAURO, A. O. D.; UNÊDA-TREVISOLI, S. H.; ARRIEL, N. H. C.; BÁRBARO, I. M.; MUNIZ, F. R. S. Ganho genético por diferentes critérios de seleção em populações segregantes de soja. **Pesquisa Agropecuária Brasileira**, v. 39, n. 11, p. 1095-1102, 2004. Disponível em: <<https://bit.ly/2HTPrU2>>. Acesso em: 16 nov. 2017.

CRUZ, C. D. **Programa genes: estatística experimental e matrizes**. Viçosa: Ed. UFV, 2006. 285 p.

_____; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: Ed. UFV, v. 2, 688 p, 2014.

_____; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 5 ed. Viçosa: Ed. UFV, 480 p. 2012.

_____; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 2011. 620 p.

DAIA, X., HUO, Z., WANG, H. Simulation for response of crop yield to soil moisture and salinity with artificial neural network. **Field Crops Research**, v. 121, p. 441-449, 2011. Disponível em: <<https://bit.ly/2x3bBs5>>. Acesso em: 2 out. 2017.

DALLASTRA, A.; UNÊDA-TREVISOLI, S. H.; FERRAUDO, A. S.; MAURO, A. O. Multivariate approach in the selection of superior soybean progeny which carry the RR gene. **Revista Ciência Agronômica**, v. 45, n. 3, p. 588-597, 2014. Disponível em: <<https://bit.ly/2HpAuPp>>. Acesso em: 29 out. 2017.

DANTAS, D.; LUZ, T. M. O.; SOUZA, M. J. H.; BARBOSA, G. P.; CUNHA, E. G. S. Uso de redes neurais artificiais na previsão da precipitação de períodos chuvosos. **Revista Espinhaço**, v. 5, n. 1, p. 11-18, 2016. Disponível em: <<https://bit.ly/2HJg6vG>>. Acesso em: 12 nov. 2017.

DAHIKAR, S. S.; RODE, S. V. Agricultural crop yield prediction using artificial neural network approach. **International journal of innovative research in electrical, electronics, instrumentation and control**

engineering, v. 2, n. 1, p. 683-686, 2014. Disponível em: <<https://bit.ly/2salbUn>>. Acesso em: 2 nov. 2017.

DEHBOZORGI, F.; SEPASKHAH, A. R. Comparison of artificial neural networks and prediction models for reference evapotranspiration estimation in a semi-arid region. **Archives of Agronomy and Soil Science** v. 58, n. 5, p. 477-499, 2012.

DUARTE, J. B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293 f. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2000. Disponível em: <<https://bit.ly/2F8C4TK>>. Acesso em: 14 set. 2017.

DUTRA, R. M. O.; SPERANDIO, M.; COELHO, J. **O método Ward de agrupamento de dados e sua aplicação em associação com os mapas auto-organizáveis de Kohonen**. 2009. Disponível em: <<https://bit.ly/2HJEY6l>>. Acesso em: 12 nov. 2017.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Tecnologias de produção de soja: região central do Brasil: 2009 e 2010. **Sistemas de Produção**, n. 13, p. 11-261, 2008. Disponível em: <<https://bit.ly/2qS66ql>>. Acesso em: 10 out. 2017.

_____. Tecnologias de produção de soja: região central do Brasil: 2014. **Sistemas de Produção**, n. 16, p. 11-265, 2013. Disponível em: <<https://bit.ly/2Hn6l7k>>. Acesso em: 4 dez. 2017.

_____; EMPRESA DE PESQUISA AGROPECUÁRIA DE MINAS GERAIS; FUNDAÇÃO TRIÂNGULO DE PESQUISA E DESENVOLVIMENTO. **Cultivares de soja: Minas Gerais e região central do Brasil: Safra 2010/2011**. Londrina: EMBRAPA Soja, 2010. 47 p. Disponível em: <<https://bit.ly/2HKkWZL>>. Acesso em: 10 out. 2017.

FANUCCHI, R. Z.; OLESKOVICZ, M.; BARBOSA, D. Análise da Detecção de Falhas de Alta Impedância Utilizando Redes Neurais Artificiais com Topologias Baseadas em Perceptron Multicamadas e Redes RBF. In: SBAI-2013, p. 1-6, 2013. Disponível em: <<https://bit.ly/2KMOxzB>>. Acesso em: 20 mar. 2017.

FONSECA, A. F. A.; SEDIYAMA, T.; CRUZ, C. D.; SAKAIYAMA, N. S.; FERRÃO, M. A. G.; FERRÃO, R. G.; BRAGANÇA, S. M. Divergência genética em café conilon. **Pesquisa Agropecuária Brasileira**, v. 41, n. 4, p. 599-605, 2006. Disponível em: <<https://bit.ly/2FaHNZn>>. Acesso em: 20 out. 2017.

FRANÇA, M. M. Análise do uso da terra no município de Viçosa-MG mediado por classificações supervisionadas com redes neurais artificiais e Maxver. **Revista Brasileira de Geografia Física**, v. 2, n. 3, p. 92-101, 2010. Disponível em: <<https://bit.ly/2HmpC91>>. Acesso em: 10 out. 2017.

GALVÃO, C. O. *et al.* **Sistemas inteligentes: aplicações a recursos hídricos e ciências ambientais**. Porto Alegre: UFRGS; ABRH, 1999. 246 p.

GARNER, W. W.; ALLARD, H. A. Effect of relative length of day and night and other factors of the environment on growth and reproduction in plants. **Journal of Agricultural Research**, v. 18, n. 11, p. 553-606, 1920. Disponível em: <<https://bit.ly/2K9Rh12>>. Acesso em: 12 out. 2017.

GHOEISHI, M.; HOSSINI, Y.; MAFTOON, M. Simple models for predicting leaf area of mango (*Mangifera indica* L.). **Journal of Biology and Earth Sciences**, v. 2, n. 2, p. 45-53, 2012. Disponível em: <<https://bit.ly/2HpljIH>>. Acesso em: 6 nov. 2017.

GIANOLA, D.; OKUT, H.; WEIGEL, K. A.; ROSA, G. J. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. **BMC Genetics**, v. 12, 87-101, 2011. Disponível em: <<https://bit.ly/2HUJexV>>. Acesso em: 10 out. 2017.

GÖRGENS, E. B.; LEITE, H. G.; GLERIANI, J. M.; SOARES, C. P. B.; CEOLIN, A. Influência da arquitetura na estimativa de volume de árvores individuais por meio de redes neurais artificiais. **Revista Árvore**, v. 38, n. 2, p. 289-295, 2014. Disponível em: <<https://bit.ly/2qT4n4p>>. Acesso em: 10 out. 2017.

HAIR, J. R.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005. 593 p.

HAYKIN, S. **Neural networks and learning machines**. 3rd. ed. New York: Prentice-Hall, 2009. 906 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Levantamento sistemático da produção agrícola. **Rio de Janeiro: IBGE**, v. 29, p. 8-84, 2015. Disponível em: <<https://bit.ly/2HslyzG>>. Acesso em: 18 out. 2017.

IQBAL, Z.; ARSHAD, M.; ASHRAF, M.; NAEEM, R.; MALIK, M. F.; WAHEED, A. Genetic divergence and correlation studies of soybean [*Glycine max* (L.) Merrill.] genotypes. **Pakistan Journal of Botanic**, v. 42, n. 2, p. 971-976, 2010. Disponível em: <<https://bit.ly/2HN2jVi>>. Acesso em: 10 nov. 2017.

JANG, J. S.; SUN, C. T.; MIZUTANI, E. **Neuro-fuzzy and soft computing**: a computational approach to learning and machine intelligence. Upper Saddle River: Prentice-Hall, 614 p, 2012.

Ji, B.; SUN, Y.; YANG, S.; WAN, J. Artificial neural networks for rice yield prediction in mountainous regions. **Journal of Agricultural Science**, v. 145, p. 249-261, 2007. Disponível em: <<https://bit.ly/2HUKtgz>>. Acesso em: 10 out. 2017.

KAUL, M.; HILL, R. L.; WALTHALL, C. Artificial neural networks for corn and soybean yield prediction. **Agricultural Systems**, v. 85, n. 1, p. 1-18, 2005. Disponível em: <<https://bit.ly/2K6xR6y>>. Acesso em: 24 out. 2017.

KAVZOGLU, T.; MATHER, P. M. The use of backpropagation artificial neural networks in land cover classification. **International Journal of Remote Sensing**, v. 24, n. 23, p. 4907-4938, 2003. Disponível em: <<https://bit.ly/2JhvrRC>>. Acesso em: 10 out. 2017.

KLOSTER, G. S.; BARELLI, M. A. A.; SILVA, C. R.; NEVES, L. G.; PAIVA SOBRINHO. S.; LUZ, P. B. Análise da divergência genética através de caracteres morfológicos em cultivares de feijoeiro. **Revista Brasileira de Ciências Agrárias**, v. 6, n. 3, p. 452-459, 2011. Disponível em: <<https://bit.ly/2Hm7snF>>. Acesso em: 8 jul. 2017.

KOHONEN, T. **Self-Organizing Maps**. Springer, 3a. ed. 2001.

_____. The self-organizing map. **Proceedings of the IEEE**, v. 78, n. 9, p. 1464-1480, 1990. Disponível em: <<https://bit.ly/2HMfBRC>>. Acesso em: 10 out. 2017.

LAZZAROTTO, J. J.; HIRAKURI, M. H. Evolução e perspectivas de desempenho econômico associadas com a produção de soja nos contextos mundial brasileiro. **Documentos**, n. 319, p. 11-67, 2011. Disponível em: <<https://bit.ly/2Jj4vRg>>. Acesso em: 14 set. 2017.

LEAL, A. J. F.; MIGUEL, E. P.; ROJO BAILO, F. H.; NEVES, D. C.; LEAL, U. A. S. Redes neurais artificiais na predição da produtividade de milho e definição de sítios de manejo diferenciado por meio de atributos do solo. **Bragantia**, v. 74, n. 4, p. 436-444, 2015. Disponível em: <<https://bit.ly/2Hs3R3C>>. Acesso em: 22 set. 2017.

LIMA JÚNIOR, L.; FERREIRA, R.; SILVA, J.; SANTOS, E.; LIRA JÚNIOR, M.; FERRAZ, I. Utilização de técnicas multivariadas na classificação de fases de crescimento de *Leucaena leucocephala* (LAM.) de Wit. **Floresta**, v. 39, n. 4, p. 921-935, 2009. Disponível em: <<https://bit.ly/2qZLxB>>. Acesso em: 17 out. 2017.

LIMA, M. A.; FERREIRA, G. G.; DINIZ, R. F. S.; OLIVEIRA, L. L. C.; FERREIRA, C. B. Uso de redes neurais artificiais (RNA) do tipo Multilayer Perceptrons (MLP) modificado com processamento estatístico em paralelo para estudo do problema de classificação da origem de vinho tinto. **Revista Brasileira de Agropecuária Sustentável**, v. 6, n. 2, p. 58-65, 2016. Disponível em: <<https://bit.ly/2F8JUwt>>. Acesso em: 14 set. 2017.

LIMA, M. F. P. D.; NASCIMENTO, P. G. M. L. D.; SILVA, M. G. O. D.; MESQUITA, H. C. D.; CUNHA, J. L. X. L. Comparação de métodos de área foliar em *Myrciaria cauliflora* (Mart.) O. Berg. **Agropecuária Científica no Semiárido**, v. 8, n. 1, p. 12-17, 2012. Disponível em: <<https://bit.ly/2Hnftt8>>. Acesso em: 5 ago. 2017.

- LIU, G., YANG, X., LI, M. An artificial neural network model for crop yield responding to soil parameters. **Lecture Notes in Computer Science**, p. 1017-1021, 2005.
- LIU, M.; ZHANG, M.; JIANG, W.; SUN, G.; ZHAO, H.; HU, S. Genetic diversity of Shaanxi soybean landraces based on agronomic traits and SSR markers. **African Journal of Biotechnology**, v. 10, n. 24, p. 4823-4837, 2011. Disponível em: <<https://bit.ly/2Fay7Ou>>. Acesso em: 14 nov. 2017.
- MALIK, M. F. A.; ASHRAF, M.; QURESHI, A. S.; KHAN, M. R. Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. **Pakistan Journal of Botanic**, v. 43, n. 2, p. 1249-1255, 2011. Disponível em: <<https://bit.ly/2qTmAyu>>. Acesso em: 14 set. 2017.
- MARIOT, M. P.; BARBIERI, R. L.; SINIGAGLIA, C.; RIBEIRO, M. V. Variabilidade em matrizes de acessos de espinheira-santa. **Ciência Rural**, v. 38, n. 2, p. 351-357, 2008. Disponível em: <<https://bit.ly/2qUzhsK>>. Acesso em: 3 dez. 2017.
- MARSHALL, J. K. Methods of leaf area measurement of large and small leaf samples. **Photosynthetica**, v. 2, p. 41-47, 1968.
- MONDINE, M. L.; VIEIRA, C. P.; CAMBRAIA, L. A. Época de semeadura: um importante fator que afeta a produtividade da cultura da soja. **Documentos**, n. 34, p. 9-16, 2001. Disponível em: <<https://bit.ly/2Jnkycl>>. Acesso em: 11 jun. 2017.
- MONTEIRO, E. R. **Identificação botânica e divergência genética em pimentas do gênero *Capsicum spp.*** 2008. 59 f. Dissertação (Mestrado em Agronomia) – Centro de Ciências Agrárias, Universidade Federal do Piauí, Teresina, 2008. Disponível em: <<https://bit.ly/2K7fwXj>>. Acesso em: 14 out. 2017.
- MONTEIRO, J. E. B. A.; SENTELHAS, P. C.; CHIAVEGATO, E. J.; GUISELINI, C.; SANTIAGO, A. V.; PRELA, A. Estimação da área foliar do algodoeiro por meio de dimensões e massa das folhas. **Bragantia**, v. 64, n. 1, p. 15-24, 2005. Disponível em: <<https://bit.ly/2qQY15A>>. Acesso em: 18 set. 2017.
- MOOSAVI, A. A.; SEPASKHA, A. Artificial neural networks for predicting unsaturated soil hydraulic characteristics at different applied tensions. **Archives of Agronomy and Soil Science**, v. 58, n. 2, p. 125-153, 2012.
- MÜLLER, L. Morfologia, anatomia e desenvolvimento. In: MIYASAKA, S.; MEDINA, J.C. (Ed.). **A soja no Brasil**. Campinas: ITAL, 1981. 1062 p.
- NASCIMENTO, M.; PETERNELLI, L. A.; CRUZ, C. D.; NASCIMENTO, A. C. C.; FERREIRA, R. P.; BHERING, L. L.; SALGADO, C. C. Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. **Crop Breeding and Applied Biotechnology**, v. 13, p. 152-156, 2013. Disponível em: <<https://bit.ly/2KaRJ8D>>. Acesso em: 18 set. 2017.
- NOGUEIRA, A. P. O.; SEDIYAMA, T.; CRUZ, C. D.; REIS, M. S.; PEREIRA, D. G.; JANGARELLI, M. Novas características para diferenciação de cultivares de soja pela análise discriminante. **Ciência Rural**, v. 38, n. 9, p. 2427-2433, 2008. Disponível em: <<https://bit.ly/2JhGZUY>>. Acesso em: 11 jul. 2017.
- NORVIG, P.; RUSSELL, S. **Inteligência artificial**. 3. ed. Rio de Janeiro: Campus, 2013. 988 p.
- ODABAS, M. S.; ERGUN, E.; ONER, F. Artificial neural network approach for the prediction of the corn (*Zea mays* L.) leaf area. **Bulgarian Journal of Agricultural Science**, v. 19, n. 4, p. 766-769, 2013. Disponível em: <<https://bit.ly/2HJ8MAf>>. Acesso em: 17 nov. 2017.
- OIKAWA, R. T.; ISHIKI, H. M. Modelos estatísticos e de redes neurais artificiais utilizados na predição de precipitação. **Periódico Eletrônico Fórum Ambiental da Alta Paulista**, v. 9, n. 8, p. 19-34, 2013.
- PENARIOL, A. Soja: cultivares no lugar certo. **Informações Agronômicas**, n. 90, p. 13-14, 2000. Disponível em: <<https://bit.ly/2qVcVqy>>. Acesso em: 13 ago. 2017.
- PEREIRA, L. A. A. **Classificação automática de áreas cafeeiras em imagens de satélite, utilizando redes neurais artificiais**. 2013. 83 f. Monografia (Graduação em Ciência da Computação) – Universidade Federal de Lavras, Lavras, 2013. Disponível em: <<https://bit.ly/2vKpwSS>>. Acesso em: 18 set. 2017.

PICOLI, M. C. A. **Estimativa da produtividade agrícola da cana-de-açúcar utilizando agregados de redes neurais artificiais**: estudo de caso usina Catanduva. 2007. 90 f. Dissertação (Mestrado) – Instituto Nacional de pesquisas Espaciais, São José dos Campos, 2007. Disponível em: <<https://bit.ly/2HYtk5K>>. Acesso em: 2 nov. 2017.

RIGON, J. P. G.; CAPUANI, S.; BRITO NETO, J. F.; ROSA, G. M.; WASTOWSKI, A. D.; RIGON, C. A. G. Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. **Revista Ceres**, v. 59, n. 2, p. 233-240, 2012. Disponível em: <<https://bit.ly/2qYTItE>>. Acesso em: 10 jun. 2017.

ROMAGNOLI, E.; SILVA, L. R. C.; ANGILELLI, K. G.; FERREIRA, B. A. D.; WALKOFF, A. R.; BORSATO, D. The use of multilayer perceptron artificial neural networks for the classification of ethanol samples by commercialization region. **Acta Scientiarum. Technology**, v. 38, n. 2, p. 227-232, 2016. Disponível em: <<https://bit.ly/2HTIzN1>>. Acesso em: 14 ago. 2017.

SALIMI, S.; LAHIJI, H. S.; ABADI, G. M.; SALIMI, S.; MORADI, S. Genetic diversity in soybean genotypes under drought stress condition using factor analysis and cluster analysis. **World Applied Sciences Journal**, v. 16, n. 4, p. 474-478, 2012. Disponível em: <<https://bit.ly/2HPsNFk>>. Acesso em: 11 ago. 2017.

SANGOI, L. F. **Análises do sistema produtivo suinícola da região central do Rio Grande do Sul**. 2014. 71 f. Tese (Doutorado em Zootecnia) – Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014. Disponível em: <<https://bit.ly/2HT8Lal>>. Acesso em: 11 mar. 2017.

SANT'ANNA, I. C. **Redes neurais artificiais na discriminação de populações de retrocruzamento com diferentes graus de similaridade**. 2014. 115 f. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, 2014. Disponível em: <<https://bit.ly/2HLZ7ZY>>. Acesso em: 19 jul. 2017.

SCHIMITH, R.; GLERIANI, J. M.; SOARES, V. P.; SILVA, E.; MARTINS, S. V.; PINTO, F. A. C. Generalização temporal de redes neurais artificiais na classificação espectro-temporal de culturas agrícolas. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14, 2009, Natal. **Anais...** Natal: INPE, 2009. p. 459-466. Disponível em: <<https://bit.ly/2JIKedU>>. Acesso em: 25 jun. 2017.

SEDIYAMA, T. *et al.* **Cultura da soja**: I parte. Viçosa: Ed. UFV, 1996. 96 p.

_____ (Ed.). **Tecnologias de produção e usos da soja**. Londrina: Mecenaz, 2009. 314 p.

SHABANI, A.; SEPASKHAH, A. R. Leaf area estimation by a simple and non-destructive method. **Iran Agricultural Research**, v. 36, n. 2, p. 101-104, 2017. Disponível em: <<https://bit.ly/2qSAYqJ>>. Acesso em: 12 nov. 2016.

SILVA, G. N.; TOMAZ, R. S.; SANT'ANNA, I. C.; NASCIMENTO, M.; BHERING, L. L.; CRUZ, C. D. Neural networks for predicting breeding values and genetic gains. **Scientia Agricola**, v. 71, n. 6, p. 494-498, 2014. Disponível em: <<https://bit.ly/2Jj4Xz4>>. Acesso em: 29 jan. 2017.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais**: para engenharia e ciências aplicadas. São Paulo: Artliber, 399 p, 2010.

SILVA, M. L. M.; BINOTE, D. H. B.; GLERIANE, J. M.; LEITE, H. G. Ajuste do modelo de Schumacher e Hall e aplicações de redes neurais artificiais para estimar volume de árvores de eucalipto. **Revista Árvore**, v. 33, n. 6, p. 1133-1139, 2009. Disponível em: <<https://bit.ly/2qReyXc>>. Acesso em: 11 dez. 2017.

SILVA, S. H. M. G.; MORAES, W.; MARQUE, E. S. H.; JORGE, L. A. C. Processamento digital de imagens para identificação da sigatoka negra. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 5.; SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO NO AGRONEGÓCIO COOPERATIVO, 2., 2005, Londrina. **Anais eletrônicos...** Londrina: SBI-Agro, 2005. 1 CD-ROM.

SILVA, S. H. M. G.; LIMA, J. D.; BENDINI, H. N.; NOMURA, E. S.; MORAES, W. S. Estimativa da área foliar do antúrio com o uso de funções de regressão. **Ciência Rural**, v. 38, n. 1, p. 243-246, 2008. Disponível em: <<https://bit.ly/2JofglA>>. Acesso em: 14 set. 2017.

SINGH, R.K., PRAJNESHU. Artificial neural network methodology for modelling and forecasting maize crop yield. **Agricultural Economics Research Review**, v. 21, p. 5-10, 2008. Disponível em: <<https://bit.ly/2s2flzq>>. Acesso em: 8 nov. 2017

SOARES, F. C.; ROBAINA, A. D.; PEITER, M. X.; RUSSI, J. L.; VIVAN, G. A. Redes neurais artificiais na estimativa da retenção de água do solo. **Ciência Rural**, v. 45, n. 11, p. 1987-1993, 2015. Disponível em: <<https://bit.ly/2Hn8EXN>>. Acesso em: 18 nov. 2017.

SUDHEER, K. P.; GOSAIN, A. K.; RAMASASTRI, K. S. Estimating actual evapotranspiration from limited climatic data using neural computing technique. **Journal of Irrigation and Drainage Engineering**, v. 129, n. 3, p. 214-218, 2003. Disponível em: <<https://bit.ly/2qWQAJf>>. Acesso em: 27 nov. 2016.

SUDRÉ, C. P.; CRUZ, C. D.; RODRIGUES, R.; RIVA, E. M.; AMARAL JUNIOR, A. T.; SILVA, D. J. H.; PEREIRA, T. N. S. Variáveis multicatóricas na determinação da divergência genética entre acessos de pimenta e pimentão. **Horticultura Brasileira**, v. 24, n. 1, p. 88-93, 2006. Disponível em: <<https://bit.ly/2Hrz5Yx>>. Acesso em: 30 mar. 2016.

TONSIG, S. L. **Redes Neurais Artificiais Multicamadas e o Algoritmo Backpropagation**. Campinas, 2000.

USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. World Agricultural Production. Disponível em: <<https://bit.ly/2lExnaT>>. Acesso em: 1 fev. 2018.

VASCONCELOS, E. S.; REIS, S. M.; SEDIYAMA, T. CRUZ, C. D. Estimativas de parâmetros genéticos da qualidade fisiológica de sementes de genótipos de soja produzidas em diferentes regiões de Minas Gerais. **Semina: Ciências Agrárias**, v. 33, n. 1, p. 65-76, 2012. Disponível em: <<https://bit.ly/2Hpsiyu>>. Acesso em: 28 mar. 2016.

VAZQUEZ-CRUZ, M. A.; LUNA-RUBIO, R.; CONTRERAS-MEDINA, L. M.; TORRES-PACHECO, I.; GUEVARA-GONZALEZ, R. G. Estimating the response of tomato (*Solanum lycopersicum*) leaf area to changes in climate and salicylic acid applications by means of artificial neural networks. **Biosystems Engineering**, v. 112, n. 4, p. 319-327, 2012. Disponível em: <<https://bit.ly/2lXZRfc>>. Acesso em: 2 mar. 2016.

VENTURA, R. V.; SILVA, M. A.; MEDEIROS, T. H.; DIONELLO, N. L.; MADALENA, F. E.; FRIDRICH, A. B.; VALENTE, B. D.; SANTOS, G. G.; FREITAS, L. S.; WENCESLAU, R. R.; FELIPE, V. P. S.; CORRÊA, G. S. S. Uso de redes neurais artificiais na predição de valores genéticos para peso aos 205 dias em bovinos da raça Tabapuã. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 64, n. 2, p. 411-418, 2013. Disponível em: <<https://bit.ly/2qVwVtc>>. Acesso em: 14 ago. 2016.

VIANNA, V. F.; UNÊDA-TREVISOLI, S. H.; DESIDÉRIO, J. A.; SANTIAGO, S.; CHARNAI, K., FERREIRA JUNIOR, J. A.; FERRAUDO, A. S.; MAURO, A. O. The multivariate approach and influence of characters in selecting superior soybean genotypes. **African Journal of Biotechnology**, v. 8, n. 30, p. 4162-4169, 2013. Disponível em: <<https://bit.ly/2KajlLo>>. Acesso em: 7 abr. 2017.

VIEIRA, T. G. C.; LACERDA, W. S.; BOTELHO, T. G. Mapeamento de áreas cafeeiras utilizando redes neurais artificiais: estudo de caso na região de Três Pontas, Minas Gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14, 2009, Natal. **Anais...** Natal: INPE, 2009. p. 7947-7954. Disponível em: <<https://bit.ly/2HUVJMM>>. Acesso em: 13 jun. 2017.

VILELLA, O. T.; UNÊDA-TREVISOLI, S. H.; SILVA, F. M.; BÁRBARO JUNIOR, L. S.; MAURO, A. O. Genetic divergence of roundup ready (RR) soybean cultivars estimated by phenotypic characteristics and molecular markers. **African Journal of Biotechnology**, v.13, n. 26, p. 2613-2625, 2014. Disponível em: <<https://bit.ly/2qTDddH>>. Acesso em: 28 jul. 2017.

WATANABE, F. S. Y.; IMAI, N. N.; SAMIZAVA, T. M.; ROCHA, P. C. Classificação da vegetação de áreas úmidas baseada em redes neurais artificiais: estudo de caso da planície fluvial do alto rio Paraná.

In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. **Anais...** Natal: INPE, 2009. p. 5515-5522. Disponível em: <<https://bit.ly/2HVDUdL>>. Acesso em: 21 ago. 2017.

WYSMIERSKI, P. T.; VELLO, N. A. The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. **Genetics and Molecular Biology**, v. 36, n. 4, p. 540-546, 2013. Disponível em: <<https://bit.ly/2HT3PK>>. Acesso em: 14 dez. 2016.

ZANINE, A. M.; DIAS, P. F.; SOUTO, S. M.; FERREIRA, D. J.; SANTOS, E. M.; PINTO, L. F. B. Uso de funções discriminantes para comparação de cultivares dos gêneros *Cynodon* e *Digitaria* quanto à produção de matéria seca e teores de macronutrientes. **Archivos Latinoamericanos de Producción Animal**, v. 15, n. 4, p. 159-164, 2007. Disponível em: <<https://bit.ly/2KaGP2D>>. Acesso em: 17 ago. 2017.

Artigos

4.1 Artigo 1 - MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN NO ESTUDO DA DISSIMILARIDADE GENÉTICA ENTRE CULTIVARES DE SOJA

Este artigo foi elaborado conforme normas do Programa de Pós-Graduação em Produção Vegetal.

RESUMO

Os métodos de agrupamento têm como finalidade agrupar indivíduos a partir de suas dissimilaridades, a fim de obter homogeneidade dentro dos grupos e heterogeneidade entre os grupos. Alguns trabalhos têm demonstrado que agrupamentos feitos por redes neurais artificiais (RNAs) do tipo mapas auto-organizáveis de Kohonen (SOM) podem ser mais eficientes que os métodos clássicos. Assim, objetivou-se estudar a dissimilaridade genética entre cultivares de soja, utilizando redes neurais do tipo SOM, e testar a eficiência da metodologia utilizada, por meio da análise discriminante de Anderson. O experimento foi conduzido em campo, no Instituto de Ciências Agrárias (ICA) da Universidade Federal de Minas Gerais (UFMG), *campus* regional de Montes Claros-MG, no período de fevereiro a julho de 2017. Utilizou-se o delineamento de látice simples (6x6), com duas repetições. Trinta e seis cultivares de soja foram avaliadas. Foram testadas diferentes arquiteturas de rede quanto à taxa de acerto médio, pela análise discriminante de Anderson, e o menor número de clusters vazios, efetuando-se 1000 treinamentos. Após selecionar-se a melhor topologia de rede, fizeram-se 1000 novos treinamentos de rede e, posteriormente, construiu-se uma matriz de dissimilaridade, observando-se a frequência com que os genótipos foram considerados como sendo de grupos distintos. Por meio da matriz de dissimilaridade obtida por RNAs, foi aplicado o método UPGMA para a obtenção de um dendograma. A consistência do agrupamento foi verificada através da análise discriminante de Anderson. Por meio das redes neurais artificiais do tipo SOM, foi constatada dissimilaridade genética entre os genótipos. Seis grupos foram formados pelo método UPGMA, indicando a existência de variabilidade para as características avaliadas. Através da análise discriminante de Anderson, 100% dos acessos foram corretamente classificados. Logo, as redes SOM são eficientes para o agrupamento de linhagens de soja, mostrando a existência de dissimilaridade genética e a formação de seis agrupamentos entre as linhagens estudadas. Os genótipos do cluster I e do cluster V apresentam potencial como genitores em programas de hibridação, podendo gerar grande variabilidade genética em suas progênes.

Palavras-chave: Análise multivariada. *Glycine max* (L.) Merrill. Melhoramento genético vegetal. Inteligência computacional.

4.1 Article 1 - SELF-ORGANIZABLE MAPS OF KOHONEN IN THE STUDY OF GENETIC DISSIMILARITY AMONG SOYBEAN CULTIVARS

This article was elaborate according to the norms of the Graduate Program in Plant Production.

ABSTRACT

The grouping methods aim to group individuals from their dissimilarities in order to obtain homogeneity within and heterogeneity between the groups. Some works have shown that clusters made by artificial neural networks (RNAs) of the Kohonen self-organizing maps (SOM) may be more efficient than classical methods. Thus, we aimed to study the genetic dissimilarity between soybean cultivars using SOM type neural networks, and to test the efficiency of the methodology used through the Anderson discriminant analysis. The experiment was conducted in the field of the Institute of Agricultural Sciences of the Federal University of Minas Gerais (UFMG), Montes Claros-MG regional campus, from February to July of 2017. The simple lattice design (6x6) was used, with two replications. Thirty-six soybean cultivars were evaluated. Different network architectures were tested for the mean hit rate by the Anderson discriminant analysis and the lowest number of empty clusters, with 1000 trainings performed. After selecting the best network topology, 1000 new network trainings were performed, and later a dissimilarity matrix was constructed considering the frequency in which the genotypes were considered to be of different groups. By means of the dissimilarity matrix obtained by RNAs, the UPGMA method was applied to obtain a dendogram. The clustering consistency was verified by Anderson's discriminant analysis. Through the artificial neural networks of the SOM type, genetic dissimilarity between the genotypes was observed. Six groups were formed by the UPGMA method, indicating the existence of variability for the evaluated characteristics. Through Anderson's discriminant analysis, 100% of hits were correctly classified. Therefore, the SOM networks are efficient for the grouping of soybean strains, showing the existence of genetic dissimilarity and the formation of six clusters among the studied strains. The genotypes of cluster I and cluster V present potential as parents in hybridization programs, and can generate great genetic variability in their progenies.

Keywords: Multivariate analysis. Glycine max (L.) Merrill. Genetic plant breeding. Computational intelligence.

Introdução

No melhoramento genético de plantas, uma forma de possibilitar a obtenção de progênes superiores é reunir informações sobre a superioridade agrônômica e a divergência genética dos genitores. Essas informações são valiosas e permitem realizar combinações entre genitores (CEOLIN *et al.*, 2007; CRUZ *et al.* 2011; ROSA *et al.*, 2017). Estudos de natureza preditiva orientam combinações híbridas, economizando tempo e recursos nos projetos de melhoramento. As diferenças morfológicas, fisiológicas e produtivas são levadas em consideração para quantificar as medidas de dissimilaridade, como base nas distâncias Euclidiana e Mahalanobis (CRUZ, 2006).

Estudos de dissimilaridade genética também podem ser feitos por redes neurais artificiais (BARBOSA *et al.*, 2011). As principais vantagens das redes neurais são a possibilidade de manipulação eficiente de grandes quantidades de dados e sua capacidade de generalização. Entretanto, as principais razões para seu uso são: as redes neurais não assumem nenhum tipo de distribuição dos dados *a priori*, diferentemente da abordagem estatística paramétrica tradicional, que assume que os dados possuem uma distribuição normal; a habilidade para manipular dados adquiridos de diferentes fontes e com diferentes níveis de precisão e ruídos (CHAGAS *et al.*, 2007).

Self-organizing map (SOM) é um tipo de rede neural artificial treinada por aprendizagem competitiva não supervisionada baseada em princípios de auto-organização de sistemas, que permite a representação de dados multidimensionais em espaços de dimensões menores (KOHONEN, 2001). O SOM foi efetivamente usado para executar muitas tarefas diferentes, que incluem agrupamento, análise exploratória de dados, pedidos topológicos, amostragem, extração de recursos e dados interpolação (LOBO, 2009).

Os mapas auto-organizáveis associados aos métodos de classificação são um caminho promissor. Em relação aos outros métodos de classificação, eles têm as vantagens de ser não paramétricos, exigindo amostras pequenas para treinamento (KAVZOGLU; MATHER, 2003), e tolerar dados faltantes (BISHOP, 1995).

Dessa forma, objetivou-se estudar a dissimilaridade genética entre cultivares de soja, utilizando redes neurais do tipo SOM, e testar a eficiência da metodologia utilizada, por meio da análise discriminante de Anderson.

Material e métodos

O trabalho foi conduzido em área experimental do Instituto de Ciências Agrárias (ICA), da Universidade Federal de Minas Gerais (UFMG), *campus* regional de Montes Claros (MG), no período de fevereiro a julho de 2017. A área experimental está localizada geograficamente entre os paralelos 16°51'00" latitude sul e entre os meridianos 44°55'00" oeste e altitude de 630 m.

O solo é, em sua maioria, classificado como Cambissolo (EMBRAPA, 1999). De acordo com a classificação de Köppen, o clima da região é do tipo Aw (tropical úmido), com inverno seco e verão chuvoso.

Foram estudadas 36 cultivares de soja (Tabela 1), utilizando-se o delineamento em látice simples, com duas repetições e aproximadamente 40 plantas por parcela, e foram selecionadas aleatoriamente

15 plantas para serem avaliadas. Foram utilizadas linhas de 5 m de comprimento, com espaçamento de 50 cm.

No processo de caracterização fenotípica, em cada uma das plantas, foram analisadas 11 caracteres quantitativos, sendo:

- a) Comprimento do hipocótilo (CH) – medido da superfície do solo até o nó cotiledonar, com utilização de paquímetro digital, no estágio vegetativo V2;
- b) Diâmetro do hipocótilo (DH) – foi obtida por meio de paquímetro digital, no estágio vegetativo V2;
- c) Comprimento dos cotilédones 1 e 2 (CC1/CC2) – medido a partir da inserção do cotilédone na haste principal até a sua extremidade, com utilização de paquímetro digital, no estágio vegetativo V2;
- d) Comprimento do epicótilo (CE) – medido a partir do nó cotiledonar até os nós das folhas unifolioladas, com utilização de paquímetro digital, no estágio vegetativo V3;
- e) Comprimento do pecíolo da primeira folha trifoliolada (CPFT) – medido a partir da inserção do pecíolo na haste principal até a inserção dos dois folíolos laterais da folha trifoliolada, com utilização de paquímetro digital, no estágio vegetativo V3;
- f) Comprimento da raque do folíolo central da primeira folha trifoliolada (CR) – medido a partir da junção dos dois folíolos laterais até a inserção do folíolo terminal, com utilização de paquímetro digital, no estágio vegetativo V3;
- g) Altura da planta (AP) – foi obtida pela distância do nível do solo até a extremidade apical da planta, com utilização de fita métrica, no estágio reprodutivo R8;
- h) Altura da inserção da primeira vagem (AIPV) – foi obtida pela distância do nível do solo até a primeira vagem da planta, com utilização de fita métrica, no estágio reprodutivo R8;
- i) Número total de vagens (NV) – foi contado de cada planta avaliada, e foram consideradas apenas vagens com sementes;
- j) Peso total de sementes (PS) – foram coletadas todas as sementes de cada planta e pesadas por meio de balança digital analítica.

Tabela 1 – Cultivares de soja (*Glycine max* (L.) Merrill) avaliadas em campo no Instituto de Ciências Agrárias da UFMG

(Continua)

Identificação da cultivar	Nome Comercial	Empresa
L1	97R21	DuPont Pioneer
L2	97R73	DuPont Pioneer
L3	98Y12	DuPont Pioneer
L4	98Y30	DuPont Pioneer
L5	99R03	DuPont Pioneer
L6	99R09	DuPont Pioneer
L7	AS3610IPRO	AGROESTE
L8	AS3730IPRO	AGROESTE
L9	BMXDESAFIO	BRASMAX
L10	BMXPontalPRO	BRASMAX
L11	BMXPotenciaRR	BRASMAX
L12	CD2720IPRO	Coodetec

Identificação da cultivar	Nome Comercial	Empresa
L13	CD2728IPRO	Coodetec
L14	CD2730IPRO	Coodetec
L15	CD2737RR	Coodetec
L16	CD2750IPRO	Coodetec
L17	CD2817IPRO	Coodetec
L18	DM6563RSFIPRO	DONMARIO
L19	DS5916IPRO	Dow AgroSciences
L20	M5947IPRO	Monsoy
L21	M6210IPRO	Monsoy
L22	M6410IPRO	Monsoy
L23	M7110IPRO	Monsoy
L25	M8210IPRO	Monsoy
L26	NA5909RG	Nidera
L27	NS5959IPRO	Nidera
L28	NS6906IPRO	Nidera
L29	NS6909IPRO	Nidera
L30	NS7000IPRO	Nidera
L31	NS7209IPRO	Nidera
L32	NS7300IPRO	Nidera
L33	NS7338IPRO	Nidera
L34	TMG7062IPRO	TMG
L35	RK6813RR	GDM
L36	RK7814IPRO	Monsoy

Fonte: Do autor, 2018.

Nota: As informações sobre as cultivares foram coletadas por Registro Nacional de Cultivares – RNC.

Para o estudo da dissimilaridade genética entre as cultivares, foi implementada a técnica de redes neurais artificiais do tipo SOM, com base no modelo de Kohonen (2001). A análise foi feita baseando-se em dados padronizados. Foram testadas diferentes arquiteturas de rede, variando o número de linhas (1 a 5) e de colunas (1 a 5), totalizando-se 24 configurações (excluindo-se a combinação com uma linha e uma coluna).

A fim de se selecionar a melhor arquitetura de rede, foram feitos 1000 treinamentos para cada uma das combinações, e para cada uma das combinações foi estimada a taxa de acerto médio, pela análise discriminante de Anderson, e o menor número de clusters vazios. Posteriormente, selecionou-se a melhor arquitetura de rede (maior taxa média de acerto e menor número de clusters vazios). A metodologia de análise discriminante de Anderson foi realizada conforme descrito por Cruz *et al.* (2014).

Após selecionar-se a melhor topologia de rede, fizeram-se 1000 novos treinamentos e, posteriormente, construiu-se uma matriz de dissimilaridade. Para a estimação da dissimilaridade entre as cultivares (duas a duas), estimou-se a frequência com que as linhagens foram alocadas em grupos distintos.

Posteriormente, foi aplicado o método UPGMA para a obtenção de um dendograma. O número de grupos estabelecidos no dendograma foi feito baseando-se no número de neurônios existentes na melhor topologia de rede. A consistência do agrupamento foi verificada por meio da análise discriminante de Anderson.

Para a complementação do estudo de agrupamento das cultivares, foi realizada a dispersão gráfica da matriz de dissimilaridade em projeção 3D, conforme Cruz (2006). Nesse procedimento, as medidas de dissimilaridade são convertidas em escores relativos a três variáveis (X, Y e Z) que, quando representadas em gráficos de dispersão, irão refletir, no espaço 3D, as distâncias originalmente obtidas a partir do espaço n-dimensional ($n =$ números de caracteres utilizados para obtenção das distâncias).

Todas as análises foram feitas no software R (THE R DEVELOPMENT CORE TEAM, 2013). Para o uso das redes SOM, utilizou-se o pacote RSNNS. Já para a obtenção do dendograma, utilizou-se a função `hclust`.

Resultados e discussão

A melhor arquitetura de rede foi encontrada utilizando-se três colunas e duas linhas (Figura 1). Nessa configuração, encontrou-se mais de 99% de acerto, pela análise discriminante de Anderson, e 0% de clusters vazios.

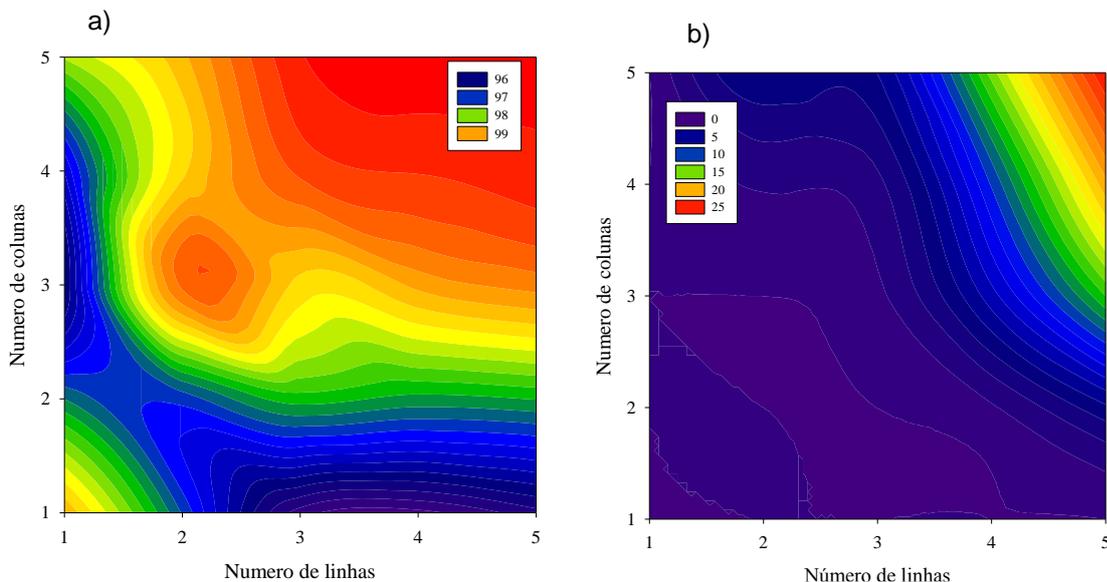


Figura 1. Porcentagem média de classificações corretas, pela análise discriminante de Anderson (a), e porcentagem de clusters vazios, considerando-se 24 configurações de redes SOM (b)
Fonte: Do autor, 2018.

Dessa forma, selecionando-se a melhor arquitetura de rede, têm-se então seis neurônios (duas linhas e três colunas), fixando-se conseqüentemente o número de clusters igual a seis. Kohonen (2001) ressalta que a determinação do número neurônios e dos parâmetros de aprendizagem é um processo empírico, baseado na experiência do usuário e em métodos de tentativa e erro. Diversos estudos utilizando a rede SOM definiram sua topologia por meio de tentativas ou de forma aleatória (CHRISTO e SOUZA, 2006; BIGNÉ, *et al.*, 2010; BARBOSA *et al.*, 2011; CHAUDHARY *et al.*, 2014; ALBÁN *et al.*, 2016). Portanto, essa metodologia para encontrar a melhor arquitetura de rede é muito importante, pois a cada vez que se utilizam redes do tipo SOM, podem-se ter resultados diferentes, já que as redes possuem pesos sinápticos aleatórios no início do treinamento (primeira iteração). Além disso, evita que a configuração (número de linhas e colunas) seja selecionada subjetivamente.

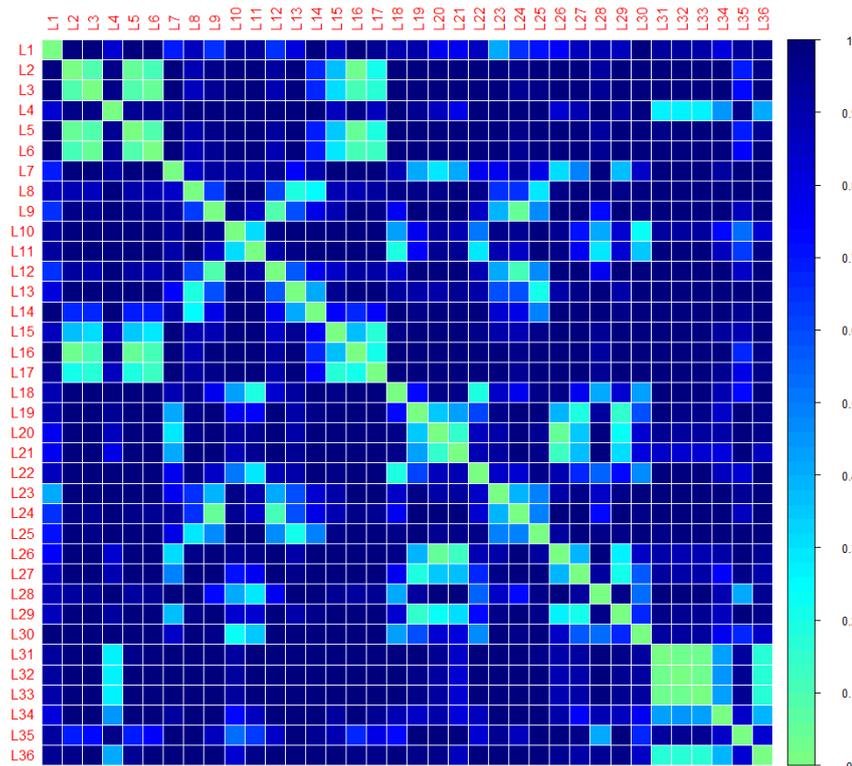


Figura 2. Representação gráfica da matriz de dissimilaridade obtida pelos mapas auto-organizáveis de Kohonen

Fonte: Do autor, 2018.

Após escolher a melhor arquitetura de rede, foi obtida a matriz de dissimilaridade, a partir dos mapas auto-organizáveis de Kohonen (Figura 2), na qual as cores mais claras indicam menor distância entre os genótipos, ou seja, são mais similares entre si, e as cores escuras indica que os genótipos são mais dissimilares (estão mais distantes). Observou-se que a maioria dos genótipos estão distantes, ou seja, são dissimilares, indicando presença de alta variabilidade genética. Dessa forma, quanto menor a distância, mais se pode ter indivíduos genéticos semelhantes ou parentais. Foram observados genótipos similares entre algumas cultivares da mesma empresa; por exemplo, as cultivares L2, L3, L5 e L6, da empresa DuPont Pioneer; L15, L16 e L17, da empresa Coodetec, e as cultivares L31, L32 e L33, da empresa Nidera. Não são, portanto, recomendáveis cruzamentos artificiais entre esses genótipos.

A quantificação da dissimilaridade genética existente entre os indivíduos permite informações sobre o grau de semelhança ou de diferença entre os genótipos, permitindo a formação dos grupos heteróticos pelos métodos de agrupamento, que são essenciais na escolha de genitores que possuam boa complementaridade gênica (CRUZ *et al.*, 2012). Logo, a avaliação da diversidade genética é de fundamental importância em programas de melhoramento e permite a otimização da seleção de genitores, de modo que as melhores combinações híbridas sejam preditas. Porém, além de dissimilares, é necessário que os genitores associem média elevada e variabilidade nas características que estejam sendo alvo do melhoramento genético (BARBOSA *et al.*, 2011).

Na figura 3 está representado o dendrograma obtido pelo método UPGMA, a partir da matriz de dissimilaridade dos genótipos de soja em estudo. Obteve-se um coeficiente de correlação cofenética de

0,95. Esse coeficiente é usado para avaliar a consistência do padrão de agrupamento e validar a capacidade do dendrograma em reproduzir as matrizes de dissimilaridade (CRUZ *et al.*, 2012).

Para análise do dendrograma, considerando-se o número pré-fixado de seis clusters, foi feito um corte em 60 % da distância.

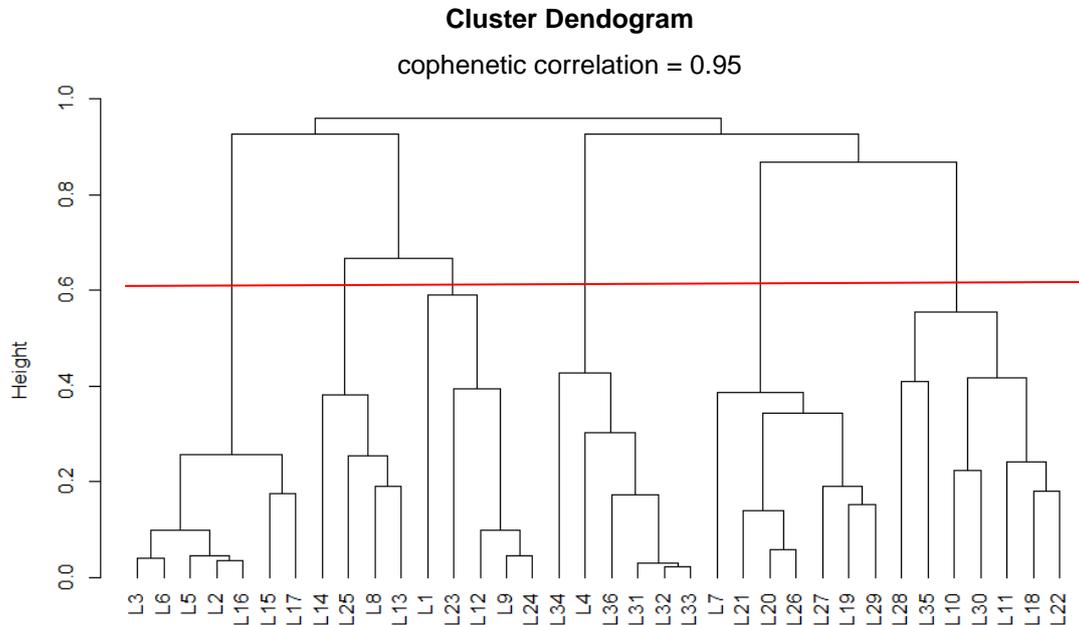


Figura 3. Dendrograma construído pelo método UPGMA, a partir do complemento de uma matriz de coincidência de genótipos agrupados no mesmo cluster, em 1000 redes do tipo SOM, com duas linhas e três colunas (6 neurônios).

Fonte: Do autor, 2018.

Foram formados seis grupos. O grupo 1 foi composto por sete cultivares das 36 estudadas. As cultivares pertencentes a esse grupo apresentaram, em geral, o menor número médio de vagens por planta (31,77) e o segundo menor peso médio das sementes (15,80 g). Porém, foi o grupo que apresentou maiores médias de altura da planta (73,12 cm), de altura da inserção da primeira vagem (19,50 cm) e de comprimento do hipocótilo (2,78 cm).

O grupo 2 foi composto por quatro cultivares. São as cultivares que produziram o maior número médio de vagens (47,58), maior peso sementes (25,19 g), maior diâmetro médio do hipocótilo (2,82 mm) e maior comprimento do epicótilo (46,06 mm).

O grupo 3, composto por cinco cultivares, é característico das cultivares que apresentaram número de vagens médio e peso médio das sementes com valores intermediários, 39,70 e 19,62 g, respectivamente. O grupo apresenta os maiores valores para o comprimento dos cotilédones (21,37 e 23,11 mm), para o comprimento da raque (7,77) e para o comprimento do pecíolo (27,93 cm).

O grupo 4 foi composto por seis cultivares. São característicos desse grupo os valores intermediários do número médio das vagens (43,64) e o segundo maior peso médio das sementes (23,56 g). O grupo apresentou menor diâmetro do hipocótilo (2,43 mm), menores comprimentos dos cotilédones (16,70 e 18,05 mm) e menor comprimento da raque (5,66 mm).

O grupo 5 foi composto por sete cultivares. É característico das cultivares que produzem o segundo maior número de vagens (45,62) e o terceiro maior peso médio das sementes (21,82 g). O

grupo apresentou menores valores médios das características comprimento do hipocótilo (2,28 cm), do epicótilo (36,44mm) e do pecíolo (20,01 mm).

O grupo 6 foi composto por sete cultivares. Este grupo caracterizou-se por apresentar o menor peso médio das sementes (15,20 g), a menor altura média da planta (50,86 cm) e a menor altura da inserção da primeira vagem (9,62 cm).

Tabela 2 – Médias dos grupos para os caracteres quantitativos avaliados nas cultivares de soja

Clusters	DH	CH	CC1	CC2	CE	CPFT	CR	AP	AIPV	NV	PS
I	2,62	2,78	19,98	21,43	40,96	23,59	6,43	73,12	19,50	31,77	15,80
II	2,82	2,44	21,26	22,57	46,06	24,92	5,82	66,29	13,23	47,58	25,19
III	2,74	2,62	21,37	23,11	42,02	27,93	7,77	61,76	12,03	39,70	19,62
IV	2,43	2,73	16,70	18,05	39,45	20,02	5,66	60,29	12,95	43,64	23,56
V	2,65	2,28	19,91	21,72	36,44	20,01	6,11	57,57	9,76	45,62	21,82
VI	2,67	2,39	20,04	21,90	37,34	26,24	6,57	50,86	9,62	34,54	15,20

DH: Diâmetro do hipocótilo (mm); CH: Comprimento do hipocótilo (cm); CC1: Comprimento dos cotilédones 1 (mm); CC2: Comprimento dos cotilédones 2 (mm); CE: Comprimento do epicótilo (mm); CPFT: Comprimento do pecíolo da primeira folha trifoliolada (mm); CR: Comprimento da raque do folíolo central da primeira folha trifoliolada (mm); AP: Altura da planta (cm); AIPV: Altura da inserção da primeira vagem (cm); NV: Número total de vagens; e PS: Peso total de sementes (g).

Fonte: Do autor, 2018.

Segundo VAL *et al.*, (2014), a mensuração de características agrônômicas da cultura, como altura de inserção da primeira vagem, altura da planta e número de vagens, são importantes, pois elas estão correlacionadas positivamente com a produtividade e permitem ao melhorista identificar e selecionar os melhores genótipos, por meio de caracteres de maior importância agrônômica. Sendo a produtividade uma das características mais importantes na seleção de genótipos, o cluster II apresentou ser o mais produtivo em relação às demais cultivares. É importante salientar que a produtividade é um caráter quantitativo, sendo influenciada fortemente pelas condições do ambiente, e, conseqüentemente, de herdabilidade baixa. Desse modo, observa-se um sistema de inter-relações sobre o rendimento, com influência de outro caráter, denominado efeito indireto (RIGON *et al.*, 2012).

A Tabela 3 apresenta os valores médios das distâncias entre os seis clusters formados. A distância inter-cluster foi máxima entre os clusters I e V (0,999). O cluster II foi mais distante do IV (0,994), o III do IV (0,977), e o VI do II (D = 0,981). A distância mínima inter-cluster foi observada entre os clusters II e III (0,667). Os resultados preliminares indicam que, em um futuro programa de melhoramento, os genótipos pertencentes aos clusters mais distantes poderiam ser usados nos programas de hibridação, para obter um amplo espectro de variação entre os segregantes. Quanto maior a divergência entre os genótipos, maior será a heterose dos híbridos em um programa de melhoramento, para desenvolver variedades de rendimentos superiores (BEKELE *et al.*, 2012). Dentro desse contexto, os genótipos do cluster I e do cluster V apresentam o potencial para serem selecionados como genitores em programas de hibridação.

Tabela 3 – Distâncias médias dentro (diagonal principal) e entre (fora da diagonal) clusters, baseando-se na matriz de dissimilaridade obtida por redes neurais do tipo mapas auto-organizáveis de Kohonen

Clusters	I	II	III	IV	V	VI
I	0.144	0.892	0.953	0.976	0.999	0.951
II	0.892	0.231	0.667	0.994	0.942	0.981
III	0.953	0.667	0.303	0.977	0.939	0.884
IV	0.976	0.994	0.977	0.219	0.914	0.943
V	0.999	0.942	0.939	0.914	0.256	0.870
VI	0.951	0.981	0.884	0.943	0.870	0.382

Fonte: Do autor, 2018.

A dissimilaridade genética entre os genótipos foi demonstrada com a formação dos seis grupos gerados pela rede neural SOM. Esses classificadores têm as vantagens de serem não paramétricos, necessitarem de pequenas amostras para treinamento (KAVZOGLU e MATHER, 2003), serem tolerantes a dados faltantes (BISHOP, 1995) e sua capacidade de agrupar dados a alta velocidade em relação às análises convencionais de agrupamento ser adequada para lidar com grandes conjuntos de dados (WANG *et al.*, 2009). Segundo Barbosa *et al.* (2011), a utilização das redes neurais artificiais como métodos de agrupamento é um caminho promissor. Budayan *et al.* (2009) concluíram que a rede SOM pode proporcionar resultados mais valiosos quando comparados com análises de agrupamentos tradicionais.

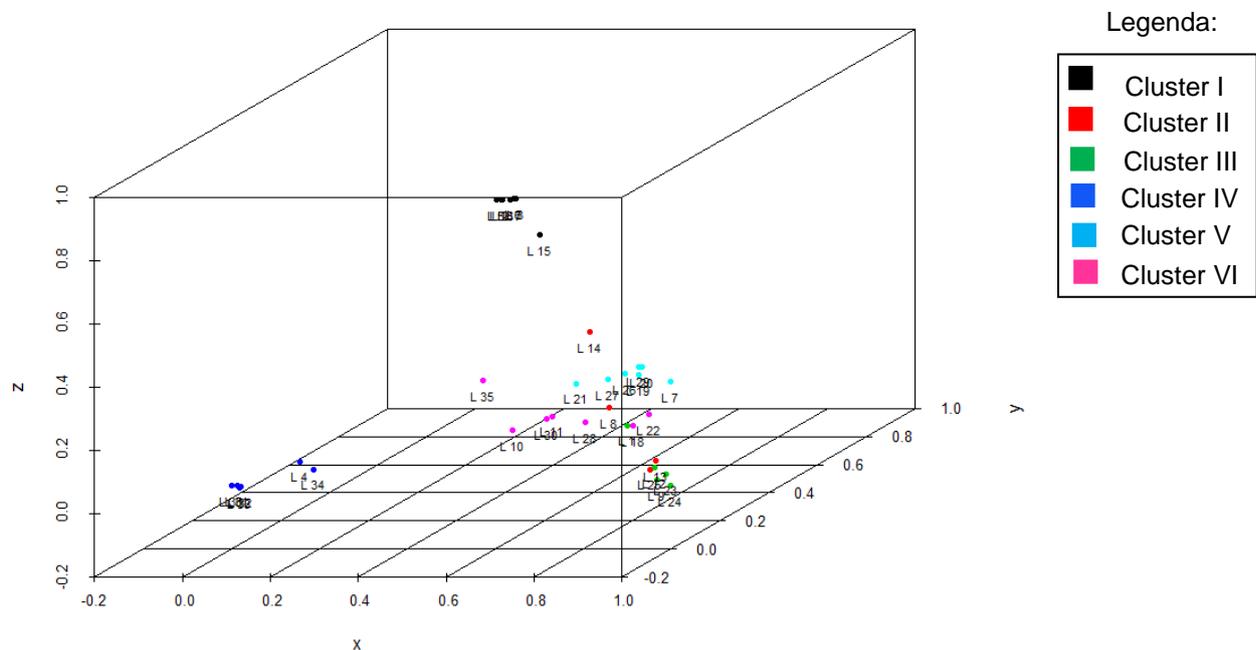


Figura 4. Dispersão gráfica da matriz de dissimilaridade em 1000 redes do tipo SOM, com duas linhas e três colunas (seis neurônios), dos seis agrupados

Fonte: Do autor, 2018.

A partir da dispersão gráfica da matriz de dissimilaridade (Figura 4), observou-se um comportamento similar dos seis grupos formados pelo método UPGMA. Entretanto, o cluster VI é o mais

disperso, ao contrário dos clusters I e IV. O cluster VI foi o que apresentou a maior distância (0.303), e os clusters I e IV, as menores – 0.144 e 0.219, respectivamente (Tabela 3).

Para a verificação da adequação dos agrupamentos obtidos pelo método de redes SOM, foi utilizada a análise discriminante de Anderson. De acordo os resultados fornecidos pelo método, os acessos foram 100% corretamente classificados nos clusters (tabela 4). Barbosa *et al.* (2011) afirmaram que sua utilização da análise discriminante de Anderson é considerada viável, como um método para verificar a consistência do agrupamento proposto pela técnica de redes neurais. Entretanto, não conseguiram 100% de classificações corretas para todos os clusters, possivelmente devido ao fato de terem utilizado uma topologia de rede de forma aleatória. Sudré *et al.* (2006) também afirmaram que essa técnica demonstrou ter um grande potencial de uso e ser uma ferramenta adicional para se conferir a correta classificação fornecida pelos vários métodos de análise multivariada.

Tabela 4 – Porcentagem de classificações corretas, pela função discriminante de Anderson, para cada cluster formado pelo método UPGMA.

Clusters	1	2	3	4	5	6
I	100	0	0	0	0	0
II	0	100	0	0	0	0
III	0	0	100	0	0	0
IV	0	0	0	100	0	0
V	0	0	0	0	100	0
VI	0	0	0	0	0	100

Fonte: Do autor, 2018.

Conclusão

As redes auto-organizáveis de Kohonen (SOM) são eficientes na avaliação da diversidade genética das cultivares de soja, de acordo com a análise discriminante de Anderson.

Há dissimilaridade genética entre os genótipos, com formação de 6 clusters. Os genótipos do cluster I são mais distantes dos genótipos do cluster V.

Os genótipos do cluster I e do cluster V apresentam potencial como genitores em programas de hibridação, podendo gerar grande variabilidade genética em suas progênes.

Referências

- ALBÁN, H.; CABRERA, J.; ACHIPIZ, O.; BASTIDAS, J. Application of Kohonen maps for the prioritization of market areas: a practical approximation. **Revista EIA**, v. 13, n. 25, p.157-169, 2016. Disponível em: <<https://bit.ly/2HmTePk>>. Acesso em: 19 out. 2017.
- BARBOSA, C. D.; VIANA A. P.; QUINTAL S. S. R.; PEREIRA M. G. P. Artificial neural network analysis of genetic diversity in *Carica papaya* L. **Crop Breeding and Applied Biotechnology**, v. 11, p. 224-231, 2011. Disponível em: <<https://bit.ly/2HVTpSU>>. Acesso em: 22 out. 2017.
- BEKELE, A.; ALEMAW, G.; ZELEKE, H. Genetic divergence among soybean (*Glycine max* (L) Merrill) introductions in Ethiopia based on agronomic traits. **Jornal of Biology Agriculture and Healthcare**, v. 2, n. 6, p. 66-12, 2012. Disponível em: <<https://bit.ly/2JksBuV>>. Acesso em: 29 out. 2017.
- BIGNÉ, E. ; ALDAS-MANZANO, J.; KÜSTER, I.; VILA, N. Mature market segmentation: a comparison of artificial neural networks and traditional methods. **Neural Computing and Application**, v. 19, p. 1-11, 2010. Disponível em: <<https://bit.ly/2qWOFV1>>. Acesso em: 6 out. 2017.
- BISHOP, C. M. **Neural networks for pattern recognition**. New York: Claredon Press, 1995. 482 p.
- BUDAYAN, C.; DIKMEN, I.; BIRGONUL, M. T. Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. **Expert Systems with Applications**, v. 36, n. 9, p. 11772-11781, 2009. Disponível em: <<https://bit.ly/2F99Bx7>>. Acesso em: 13 out. 2017.
- CEOLIN, A. C. G.; GONÇALVES-VIDIGAL, M. C.; VIDIGAL FILHO, P. S.; KVITSCHAL, M. V.; GONELA, A.; SCAPIM, C. A. Genetic divergence of the common bean (*Phaseolus vulgaris* L.) group Carioca using morpho-agronomic traits by multivariate analysis. **Hereditas**, v. 144, n. 1, p.1-9, 2007. Disponível em: <<https://bit.ly/2HWUgmr>>. Acesso em: 23 out. 2017.
- CHAGAS, C. S.; FERNANDES FILHO E. I.; VIEIRA C. A. O.; CARVALHO JÚNIOR W. Utilização de redes neurais artificiais para predição de classes de solo em uma bacia hidrográfica no Domínio de Mar de Morros. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007, Florianópolis. **Anais...** Florianópolis: INPE, 2007. p. 2421-2428, 2007. Disponível em: <<https://bit.ly/2qXIsJk>>. Acesso em: 2 nov. 2017.
- _____; VIEIRA, C. A. O.; FERNANDES FILHO, E. I.; CARVALHO JUNIOR, W. Utilização de redes neurais artificiais na classificação de níveis de degradação em pastagens. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n. 3, p. 319-327, 2009. Disponível em: <<https://bit.ly/2vHTalJ>>. Acesso em: 2 nov. 2017.
- CHAUDHARY, V.; BHATIA, R. S.; AHLAWAT, A. K. A novel self-organizing map (SOM) learning algorithm with nearest and farthest neurons. **Alexandria University Journal**, v. 53, n. 4, p. 827-831, 2014. Disponível em: <<https://bit.ly/2HGhLT1>>. Acesso em: 5 nov. 2017.
- CHRISTO, E. S.; SOUZA, R. C. Uma abordagem estatística para a previsão de potência reativa em sistemas elétricos. **Pesquisa Operacional**, v. 26, n. 2, p. 361-381, 2006. Disponível em: <<https://bit.ly/2Fds8bo>>. Acesso em: 9 nov. 2017.
- CRUZ C. D. **Programa genes**: biometria. Viçosa: Ed. UFV, 2006. 382 p.
- _____; REGAZZI A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed. rev. Viçosa: Ed. UFV, 2001. v. 2, 390 p.
- _____. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. rev. e ampl. Viçosa: Ed. UFV, 2014. v. 2, 390 p.
- _____; CARNEIRO; P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 4. ed.. Viçosa: Ed. UFV, 2012. v. 2.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Sistema brasileiro de classificação de solos**. Rio de Janeiro: EMBRAPA/CNPS, 1999. 412 p.

KAVZOGLU, T.; MATHER, P. M. The use of backpropagation artificial neural networks in land cover classification. **International Journal of Remote Sensing**, v. 24, n. 23, p. 4907-4938, 2003.

KOHONEN, T. **Self-organizing maps**. 3rd. ed. [Berlin]: Springer-Verlag, 2001. 502 p.

LOBO, V. J. A. S. Application of self-organizing maps to the maritime environment. In: POPOVICH, V. V.; SCHRENK, M.; KOROLENKO, K. V. (Ed.). **Information fusion and geographic information systems**. Berlin: Springer-Verlag, 2009. p.19-36.

THE R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013.

RIGON, J. P. G.; CAPUANI, S.; BRITO NETO, J. F.; ROSA, G. M.; WASTOWSKI, A. D.; RIGON, C. A. G. Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. **Revista Ceres**, v. 59, n. 2, p. 233-240, 2012. Disponível em: <<https://bit.ly/2qYTItE>>. Acesso em: 9 dez. 2017.

ROSA, D. P.; NOBRE D. A. C.; OLIVEIRA D. S.; SILVA F. C. S.; BEZERRA A. R. G.; SILVA A. F.; SEDIYAMA, T. Genetic diversity in soybean seed quality under different storage conditions. **Semina: Ciências Agrárias**, v. 38, n. 1, p. 57-72, 2017. Disponível em: <<https://bit.ly/2HpHqvM>>. Acesso em: 8 dez. 2017.

SUDRÉ, C. P.; CRUZ, C. D.; RODRIGUES, R.; RIVA, E. M.; AMARAL JUNIOR, A. T.; SILVA, D. J. H.; PEREIRA, T. N. S. Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. **Horticultura Brasileira**, v. 24, n. 1, p. 88-93, 2006. Disponível em: <<https://bit.ly/2Hrz5Yx>>. Acesso em: 8 dez. 2017.

VAL, B. H. P.; FERREIRA JÚNIOR, J. A.; BIZARI, E. H.; MAURO, A. O.; TREVISOLI, S. H. Diversidade genética de genótipos de soja por meio de caracteres agromorfológicos. **Ciência e Tecnologia**, v. 6, n. 1, p. 72-83, 2014. Disponível em: <<https://bit.ly/2HmkmCi>>. Acesso em: 28 nov. 2017.

WANG, D.; LARDER, B.; REVELL, A.; MONTANER, J.; HARRIGAN, R.; WOLF, F.; LANGE, J.; WEGNER, S.; RUIZ, L.; PEREZ-ELIAS, M. J.; EMERY, S.; GATELL, J.; MONFORTE, A. A.; TORTI, C.; ZAZZI, M.; LANE, C. A comparison of three computational modeling methods for the prediction of virological response to combination HIV therapy. **Artificial Intelligence in Medicine**, v. 47, n. 1, p. 63-74, 2009. Disponível em: <<https://bit.ly/2K8H66k>>. Acesso em: 29 out. 2017.

4.2 Artigo 2 - REDES NEURAIS ARTIFICIAIS NA ESTIMATIVA DA ÁREA FOLIAR DE SOJA

Este artigo foi elaborado conforme normas do Programa de Pós-Graduação em Produção Vegetal.

RESUMO

A área foliar é um dos parâmetros fisiológicos mais utilizados na análise de crescimento das plantas, dando suporte à interpretação de fatores de produção e dos a ele relacionados. Sua estimativa é de grande importância, pois medidas diretas são difíceis e imprecisas, sendo necessários equipamentos caros e muita mão de obra. No presente estudo, buscou-se obter metodologias que permitam estimar a área foliar em soja, por meio do uso de redes neurais artificiais, considerando diferentes formatos foliares. O experimento foi conduzido em campo no Instituto de Ciências Agrárias (ICA) da Universidade Federal de Minas Gerais (UFMG), *campus* regional de Montes Claros-MG, no período de fevereiro a julho de 2017. Foram avaliadas trinta e seis cultivares, as quais foram separadas em três grupos, de acordo o formato foliar. Foram desenvolvidos perceptrons multicamadas, utilizando-se 300 folhas por grupo, destinando-se 70% para treinamento e 30% para a validação. Foram testadas também as medidas morfológicas mais importantes nesses estudos. As RNAs foram eficientes para estimar a área foliar em soja, com coeficientes de determinação próximos a 0,90. Apenas duas medidas foliares dos folíolos são suficientes para estimação da área foliar. A rede 4, treinada com folhas de todos os grupos, foi mais generalista e, conseqüentemente, é a mais indicada para a predição da área foliar em soja.

Palavras-chave: *Glycine max* (L.) Merrill. Perceptrons multicamadas. Inteligência computacional.

4.2 Article 2 - ARTIFICIAL NEURAL NETWORKS IN THE ESTIMATE OF THE SOYBEAN FOLIAR AREA

This article was elaborate according to the norms of the Graduate Program in Plant Production.

ABSTRACT

The leaf area is one of the most used physiological parameters in the analysis of plant growth, supporting the interpretation of related factors of production. Their estimation is of great importance, since direct measures are difficult and imprecise, requiring expensive equipment and a lot of manpower. In the present study, we aimed to obtain methodologies to estimate the leaf area in soybean, through the use of artificial neural networks, considering different leaf formats. The experiment was conducted in the field of the Institute of Agricultural Sciences of the Federal University of Minas Gerais (UFMG), regional campus of Montes Claros-MG from February to July of 2017. Thirty-six cultivars were evaluated, which were separated into three groups, according to the foliar format. Multilayer perceptrons were developed, using 300 sheets per group, with 70% for training and 30% for validation. The most important morphological measures were also tested in these studies. The RNAs were efficient to estimate the leaf area in soybean, with determination coefficients close to 0.90. Only two foliar measurements of leaflets are sufficient for estimation of leaf area. The network 4, trained with leaves of all the groups, was more generalist and consequently is the most indicated for the prediction of the leaf area in soybean.

Keywords: Glycine max (L.) Merrill. Multilayer perceptrons. Computational intelligence.

Introdução

A determinação da área foliar é fundamental para estudar a eficiência fotossintética das plantas, determinar danos bióticos e abióticos na cultura, analisar seu crescimento e por ser uma característica importante na estimação da produtividade das culturas (SETIYONO *et al.*, 2008).

Em diversas espécies vegetais, o comprimento e a largura do limbo foliar têm sido utilizados para a estimação da área foliar, em frutíferas (FALLOVO *et al.*, 2008; MENDOZA-DE GYVES *et al.*, 2008; DEMIRSOY e LANG 2010; MAZZINI *et al.*, 2010), olerícolas (PEKSEN, 2007; RIVERA *et al.*, 2007; TSIALTAS e MASLARIS, 2008; KANDIANNAN *et al.*, 2009; OLFATI *et al.*, 2010), culturas ornamentais (ROUPHAEL *et al.*, 2007; 2010; FASCELLA *et al.*, 2009; GIUFFRIDA *et al.*, 2011) e outras plantas.

Para a cultura da soja, o principal método para estimação da área foliar é por meio de modelos matemáticos, a partir de dimensões lineares da folha (WIERSMA; BAILEY, 1975; ADAMI *et al.*, 2007; BAKHSHANDEH *et al.*, 2011; RICHTER *et al.*, 2014). As cultivares de soja possuem folhas de tamanhos e formatos diferentes (MÜLLER, 1981), o que pode afetar na estimação da área foliar. Além disso, a presença de três folíolos pode dificultar a avaliação.

A área foliar pode ser medida por meios destrutivos ou não destrutivos. Muitos métodos foram desenvolvidos para facilitar a medição da área foliar. No entanto, esses métodos, incluindo os de discos foliares, impressão em linha, scanners de mesa ou portáteis, fotografando ou usando um convencional planímetro, requerem a excisão de parte das plantas. Dessa maneira, não é possível fazer medidas sucessivas da mesma folha (FALLOVO, *et al.*, 2008). Vale ressaltar que alguns métodos demandam equipamentos de alto valor, são métodos destrutivos e exigem muito tempo e mão de obra. Um dos desafios para medir a área foliar é a utilização de um método de baixo custo, rápido, confiável e não destrutivo.

Um dos métodos para estimar a área foliar é realizado a partir dos dados do comprimento e da largura do limbo foliar por modelos matemáticos de regressão, utilizando abordagem linear ou não linear (ROUPHAEL *et al.*, 2006; SILVA *et al.*, 2008). Recentemente, a técnica de redes neurais artificiais (RNAs) tem sido citada como uma importante alternativa na estimativa e previsão de várias características. Diferentes pesquisadores mostraram que as RNAs frequentemente dão melhores resultados do que os métodos tradicionais (MOOSAVI e SEPASKHA, 2012).

A rede neural artificial é aplicada para vários assuntos na ciência agrícola, tais como: previsão de produção de colheitas (KAUL *et al.*, 2005, DAI *et al.*, 2011; MOOSAVIZADEH-MOJARAD e SEPASKHAH, 2012;), previsão do peso do fruto (SOARES *et al.*, 2013; NAROUI RAD *et al.*, 2015), previsão de evapotranspiração (TRAJKOVIC, 2005; DEHBOZORGI e SEPASKHAH, 2012), estimação dos parâmetros do solo (MOOSAVI e SEPASKHA, 2012), entre outros. Essa técnica apresenta a vantagem de ser não paramétrica, tolerar a perda de dados e não necessitar de informações detalhadas sobre o sistema a ser modelado (SILVA *et al.*, 2014).

Dessa forma, buscou-se obter metodologias que permitam a estimação da área foliar em soja eficientemente por redes neurais artificiais, considerando-se diferentes formatos foliares.

Material e métodos

O experimento foi conduzido em área experimental do Instituto de Ciências Agrárias (ICA) da Universidade Federal de Minas Gerais, *campus* regional de Montes Claros-MG, no período de fevereiro a julho de 2017, localizada geograficamente entre os paralelos 16°51'00" latitude sul e entre os meridianos 44°55'00" oeste e altitude de 630 m.

O solo é, em sua maioria, classificado como Cambissolo (EMBRAPA, 1999). De acordo com a classificação de Köppen, o clima da região é do tipo Aw (tropical úmido), com inverno seco e verão chuvoso.

Foram utilizadas 36 cultivares comerciais de soja (Tabela 1). As cultivares foram pré-selecionadas por representarem diferentes formato de folha (Figura 1). Foi utilizado delineamento em látice simples (6x6), com duas repetições e aproximadamente 40 plantas por parcela.

Tabela 1 – Cultivares de soja, descrição quanto à empresa e ao grupo do formato foliar, Montes Claros, MG, 2017

Grupo	Cultivar	Empresa
1	CD2728IPRO	Coodetec
1	NS7209IPRO	Nidera
1	RK6813RR	GDM
1	RK7814IPRO	Monsoy
2	97R21	DuPont Pioneer
2	97R73	DuPont Pioneer
2	98Y12	DuPont Pioneer
2	99R03	DuPont Pioneer
2	99R09	DuPont Pioneer
2	BMXDESAFIO	BRASMAX
2	BMXPotenciaRR	BRASMAX
2	CD2730IPRO	Coodetec
2	CD2737RR	Coodetec
2	CD2750IPRO	Coodetec
2	CD2817IPRO	Coodetec
2	DM6563RSFIPRO	DONMARIO
2	DS5916IPRO	DowAgroSciences
2	M5947IPRO	Monsoy
2	M6210IPRO	Monsoy
2	M8210IPRO	Monsoy
2	NS6906IPRO	Nidera
2	NS6909IPRO	Nidera
3	98Y30	DuPont Pioneer
3	AS3610IPRO	AGROESTE
3	AS3730IPRO	AGROESTE
3	BMXPontaIPRO	BRASMAX
3	CD2720IPRO	Coodetec
3	M6410IPRO	Monsoy
3	M7110IPRO	Monsoy
3	NA5909RG	Nidera
3	NS5959IPRO	Nidera
3	NS7000IPRO	Nidera
3	NS7300IPRO	Nidera
3	NS7338IPRO	Nidera
3	TMG7062IPRO	TMG

Fonte: Do autor, 2018.

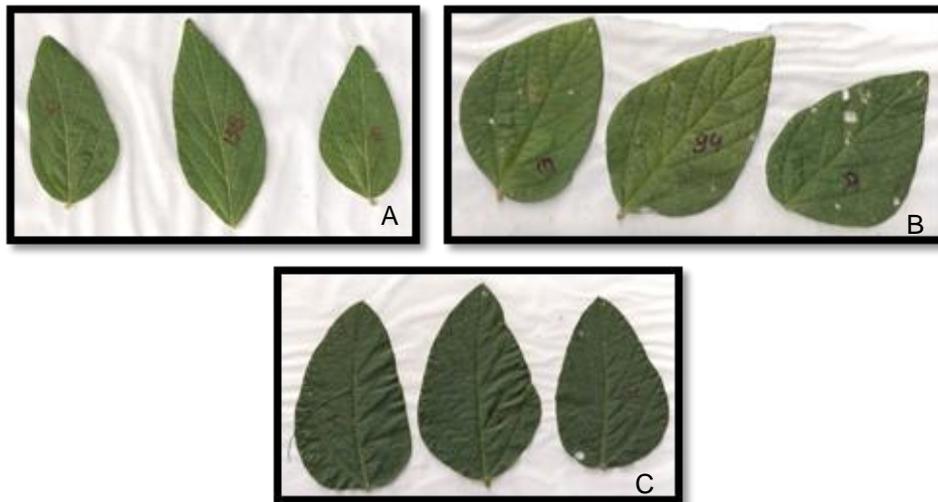


Figura 1. Exemplo dos três formatos de folhas das cultivares de soja – lanceolada (a), triangular (b) e oval-pontiaguda (c) – utilizadas no trabalho.
Fonte: BRASIL, 2009.

As folhas foram coletadas na mesma época, de forma aleatória na planta, para que se pudessem ter folhas de diferentes tamanhos. Em cada folíolo foi avaliado o comprimento e a largura (medido com régua graduada em centímetros), e a área foliar foi estimada em cm^2 com o auxílio do software Image Pro Plus (cada folha foi escaneada, utilizando um Scanner HP Photosmart C4480).

Para a predição da área foliar, utilizaram-se as redes MLP (*multilayer perceptron*), com o auxílio do pacote RSNNS do software R. No processo de treinamento da MLP, foram utilizados, como variáveis de entrada, os dados de comprimento e largura de cada folíolo, e como saída desejada, a área foliar observada. Foram utilizadas informações de 300 folhas de cada grupo estudado, totalizando 900 folhas. Desse conjunto de dados, foram destinadas as informações de 70% das folhas para treinamento, e 30% para validação.

Para garantir a melhor eficiência no treinamento das redes, tanto os dados de entrada quanto os de saída foram padronizados para o intervalo entre 0 e 1 pela seguinte equação: $V_n = [1 + (V_{obs} - V_{max})] / (V_{max} - V_{min})$, em que: V_n é o valor normalizado; V_{obs} , o valor observado; V_{max} é o valor máximo da amostra; e V_{min} , o valor mínimo da amostra. Os valores máximos e mínimos encontrados para cada uma das variáveis são apresentados na Tabela 2. Para a padronização, utilizou-se a função *normalizeData* do pacote RSNNS (BERGMEIR; BENÍTEZ, 2012).

O número máximo de épocas de treinamento foi arbitrado como 500. Foram feitos 100 treinamentos para a arquitetura de rede com seis neurônios na camada intermediária (figura 2). Foi obtido o diagrama com a arquitetura da rede, utilizando a função *plotnet* (pacote NeuralNetTools). Também foram obtidas a média do erro quadrático médio ($EQM_{\text{médio}}$) e do coeficiente de determinação ($R^2_{\text{médio}}$) para redes ajustadas, considerando-se diferentes arquiteturas de rede e importância relativa das medidas contempladas em cada avaliação estimada pelo método de Garson (1991), pela função *garson* do pacote *NeuralNetTools*.

Após escolher a melhor arquitetura de rede para cada grupo, foram feitos 100 novos treinamentos com cada uma das quatro redes, para os quatro grupos (o quarto grupo é composto dos três grupos pré-

definidos de acordo com os tipos de formatos foliares) e foram obtidos novos coeficientes de determinação.

Para visualizar a eficiência das quatro redes neurais selecionadas na predição da área foliar dos quatro grupos, foi feita a dispersão gráfica da área foliar estimada em função da observada para a amostra de validação. Além disso, foram obtidos os coeficientes de determinação como indicativo da qualidade do ajuste. Na obtenção dos dados para plotar os gráficos de superfície resposta da área foliar em função do comprimento e largura dos folíolos, foram utilizadas as funções *expand.grid* e *predict* (pacote *RSNNS*). É válido lembrar que os valores preditos pelas RNAs são normalizados e, conseqüentemente, devem ser desnormalizados para a obtenção da escala desejada (cm²). Para a desnormalização utilizou-se a função *denormalizeData* (pacote *RSNNS*), a qual considera a expressão $V_{dn} = V_{max} + (V_n - 1) * (V_{max} - V_{min})$ em que: V_{dn} é o valor desnormalizado; V_n , o valor normalizado; V_{min} , é o valor mínimo da amostra e V_{max} , o valor máximo da amostra. A partir dos dados preditos pela rede, foram feitos gráficos de superfície resposta, com o auxílio do software Sigma Plot v.11.

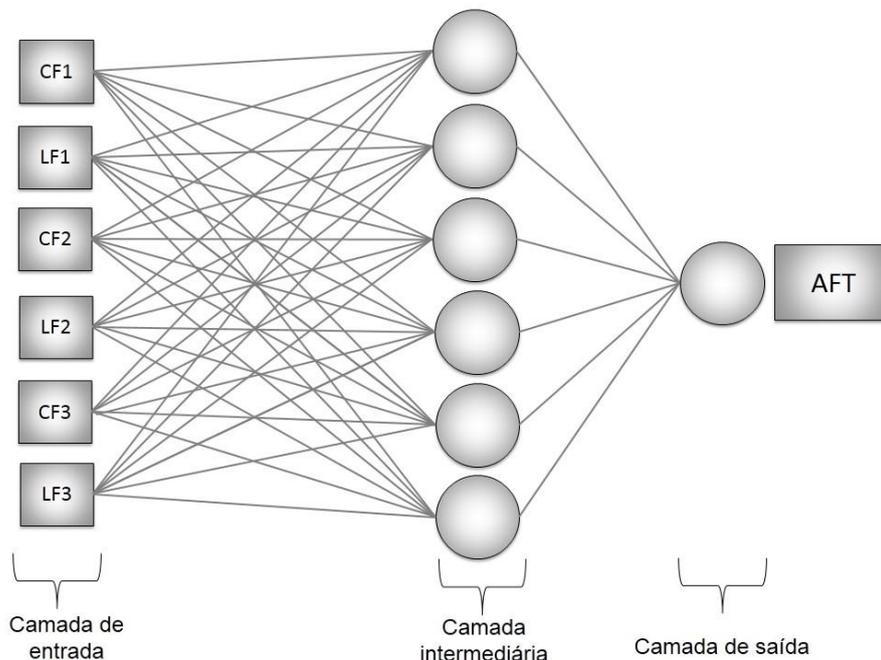


Figura 2. Arquitetura da rede neural do tipo perceptrons multicamadas, utilizada na estimaco da rea foliar em funo do comprimento e largura de trs fololos

Legenda: CF1: comprimento do fololo central; LF1: largura do fololo central; CF2: comprimento do fololo esquerdo; LF2: largura do fololo esquerdo, CF3: comprimento do fololo direito; LF3: largura do fololo direito; AFT: rea foliar total.

Fonte: Do autor, 2018.

Resultados e discusso

A Tabela 2 apresenta os valores dos coeficientes de variao dos fololos de acordo com sua posio na folha, fololo central, esquerdo e direito de cada grupo. Observando-se os valores dos coeficientes de variao obtidos, notou-se que os valores no variaram muito entre os grupos. Devido  escolha aleatria das folhas em diferentes posies na planta e tamanho, foi verificada uma grande variao para a rea foliar para o primeiro, segundo e terceiro grupos, com coeficientes de variao de 39,38%, 41,70% e 40,02%, respectivamente. A eficincia e a generalizao na predio da rea foliar

pelas RNAs dependem da variação do formato e tamanho da folha na amostra utilizada no treinamento (FALLOVO *et al.*, 2008; WANG; ZHANG, 2012).

Tabela 2 – Análise descritiva entre as características comprimento, largura e área foliar dos três folíolos de cada grupo em soja

Grupos	Parâmetros	CF1	LF1	CF2	LF2	CF3	LF3	Área foliar
1	Máximo (mm)	147.000	105.000	203.000	94.000	132.000	81.000	210.877
	Média (mm)	91.781	49.140	78.050	46.543	76.852	46.194	87.392
	Mínimo (mm)	30.000	26.000	37.000	19.000	30.000	21.000	26.435
	Desvio padrão (mm)	17.352	11.855	17.617	11.673	16.223	11.330	34.420
	CV (%)	18.906	24.125	22.572	25.081	21.110	24.527	39.386
2	Máximo (mm)	155.000	96.000	142.000	102.000	142.000	100.000	286.225
	Média (mm)	95.170	60.210	83.057	57.969	83.098	57.884	112.639
	Mínimo (mm)	51.000	27.000	36.000	26.000	40.000	24.000	29.6503
	Desvio padrão (mm)	20.076	12.659	20.067	14.129	19.864	13.959	46.973
	CV (%)	21.095	21.025	24.161	24.373	23.905	24.115	41.702
3	Máximo (mm)	140.000	93.000	124.000	85.000	132.000	110.000	216.887
	Média (mm)	89.209	50.707	75.378	47.852	74.546	47.384	88.750
	Mínimo (mm)	52.000	21.000	36.000	24.000	32.000	20.000	30.969
	Desvio padrão (mm)	18.443	10.915	17.737	11.840	18.086	12.453	35.522
	CV (%)	20.674	21.525	23.530	24.744	24.262	26.282	40.024

CF1: comprimento do folíolo central; LF1: largura do folíolo central; CF2: comprimento do folíolo esquerdo; LF2: largura do folíolo esquerdo; CF3: comprimento do folíolo direito; LF3: largura do folíolo direito.
Fonte: Do autor, 2018.

Neste trabalho, foram utilizados o comprimento e a largura do limbo de cada folíolo para cada grupo como informações da camada de entrada (variáveis explicativas), e a área foliar, observada na camada de saída (variável dependente). Saber qual variável explicativa tem maior importância no processo de predição pelas RNAs é uma informação muito importante e útil, podendo reduzir o número de medições. No presente estudo, foi aplicado o método Garson (1991), no qual em cada configuração de rede foi excluída a variável menos importante, até chegar na configuração de rede 5, com as duas variáveis mais importantes para a estimação da área foliar. Assim, indicou-se como mais importante no grupo 1, o comprimento do folíolo esquerdo e a largura do folíolo direito; no 2 grupo, o comprimento do folíolo central e a largura do folíolo esquerdo; no grupo 3, o comprimento do folíolo direito e a largura do folíolo esquerdo; e para o grupo 4, as variáveis mais importantes foram iguais para o grupo 3 (Tabela 3).

Tabela 3 – Média do erro quadrático médio (EQM) e do coeficiente de determinação ($R^2_{\text{médio}}$) para redes ajustadas considerando diferentes arquiteturas de rede e importância relativa estimada pelo método Garson das variáveis contempladas em cada avaliação

Grupos	Config.	Importância relativa						R^2 médio	EQM médio
		CF1	LF1	CF2	LF2	CF3	LF3		
1	1	0.149±0.006	0.126±0.007	0.235±0.011	0.133±0.007	0.177±0.008	0.179±0.007	0.923	0.003
	2	0.170±0.009	-	0.250±0.014	0.173±0.007	0.197±0.008	0.21±0.0085	0.920	0.003
	3	-	-	0.328±0.017	0.185±0.008	0.269±0.013	0.218±0.010	0.916	0.003
	4	-	-	0.432±0.017	-	0.262±0.013	0.306±0.011	0.915	0.003
	5	-	-	0.582±0.019	-	-	0.418±0.019	0.885	0.004
2	1	0.158±0.007	0.15±0.007	0.162±0.007	0.190±0.007	0.152±0.008	0.187±0.007	0.938	0.002
	2	0.173±0.008	-	0.174±0.008	0.241±0.008	0.172±0.008	0.240±0.008	0.937	0.002
	3	0.226±0.010	-	0.219±0.011	0.250±0.011	-	0.306±0.010	0.939	0.002
	4	0.328±0.011	-	-	0.361±0.012	-	0.311±0.012	0.940	0.002
	5	0.397±0.014	-	-	0.603±0.014	-	-	0.929	0.003
3	1	0.150±0.007	0.164±0.007	0.119±0.006	0.236±0.009	0.188±0.009	0.143±0.008	0.858	0.006
	2	0.168±0.008	0.192±0.008	-	0.268±0.010	0.218±0.009	0.154±0.008	0.859	0.006
	3	0.186±0.008	0.237±0.010	-	0.307±0.010	0.271±0.010	-	0.859	0.006
	4	-	0.246±0.012	-	0.330±0.010	0.424±0.011	-	0.864	0.005
	5	-	-	-	0.509±0.013	0.491±0.013	-	0.861	0.006
4 (1, 2 e 3)	1	0.168±0.008	0.141±0.007	0.153±0.007	0.191±0.008	0.182±0.009	0.164±0.009	0.922	0.002
	2	0.184±0.009	-	0.167±0.009	0.245±0.010	0.207±0.011	0.198±0.010	0.921	0.002
	3	0.242±0.012	-	-	0.32±0.0130	0.22±0.011	0.219±0.011	0.923	0.002
	4	0.261±0.015	-	-	0.426±0.015	0.314±0.012	-	0.920	0.002
	5	-	-	-	0.522±0.016	0.478±0.016	-	0.919	0.002

CF1: comprimento do fólolo central; LF1: largura do fólolo central; CF2: comprimento do fólolo esquerdo; LF2: largura do fólolo esquerdo; CF3: comprimento do fólolo direito; LF3: largura do fólolo direito.

Fonte: Do autor, 2018.

Nota: Valores seguidos do símbolo \pm referem-se aos desvios para a obtenção dos intervalos de confiança obtidos pelo teste t ao nível de 5% de significância.

Os coeficientes de determinação tiveram valores elevados (R^2 entre 0,85-0,94), indicando que a estimativa da área foliar em soja pelo método de rede neural do tipo perceptrons multicamadas foi eficiente para todas as cultivares de soja utilizadas no experimento, independentemente do grupo. O EQM expressa a magnitude do erro para as redes ajustadas, de modo que quanto mais próximo de zero, melhor é a rede.

Devido à similaridade dos resultados da predição da área foliar de cada grupo individualmente (grupos 1, 2 e 3) com o resultado de todos os grupos simultaneamente (grupo 4), não se justifica o ajuste da rede separadamente para cada grupo; pode-se utilizar uma configuração geral para mais de um formato foliar. Stoppani *et al.* (2003) relataram, em seus estudos, a precisão das previsões da variação da forma da folha entre cultivares, na cultura do brócolis. Shabani *et al.* (2017) concluíram, em estudo, que a aplicação de RNAs é eficiente para estimar área foliar de diferentes espécies de plantas, enquanto que na aplicação de outros métodos deve ser preparada uma equação específica para cada tipo de planta. De acordo Richter *et al.* (2014), isso é importante, pois a cada ano são lançadas novas cultivares, mas, como o formato das folhas das cultivares tende a ser similar, não será necessário estimar uma configuração de rede com tanta frequência, o que se alinha aos resultados encontrados por Bakhshandeh *et al.* (2011).

Este tipo de estudo ganha maior importância quando é avaliado um maior número de características, principalmente quando se torna vantajosa a exclusão de características para a redução de mão de obra e esforço computacional (PALIWAL; KUMAR, 2011). Com exceção da primeira folha, as folhas da soja são trifoliadas, aumentando assim, em três vezes, o tempo e o trabalho para se medir uma folha, comparada com culturas de folhas simples. No entanto, os resultados deste trabalho indicam,

como outros realizados para cultura da soja (BAKHSHANDEH *et al.*, 2011; RICHTER *et al.*, 2014), que duas medidas são suficientes para se ter boa estimativa da área dos três folíolos por RNAs.

Para a melhor visualização da eficiência de predição da área foliar pelas RNAs, são apresentados, na figura 3, os valores observados em função dos valores estimados da área foliar para os dados de cada grupo, e todos os dados das amostras de validação. A eficiência da predição pelas RNAs depende da variação do formato da folha (comprimento e largura) e dos materiais genéticos utilizados no treinamento (WANG; ZHANG, 2012; FALLOVO *et al.*, 2008). Assim, a utilização do grande número de cultivares de soja estudado neste trabalho contribuiu para que a rede treinada apresente um caráter generalista para vários formatos e tamanhos de folha.

Para o grupo 1, foi obtido um coeficiente de determinação 0,9374; para o grupo 2, 0,9737; para o grupo 3, 0,9395; e para todos os grupos simultaneamente, 0,959. Esse resultado indicou que mais de 95% da informação da área foliar observada foi explicada pela área foliar estimada por meio das redes perceptrons multicamadas. A eficiência foi semelhante a Bakhshandeh *et al.* (2011) e Richter *et al.* (2014), que também obtiveram valores de R^2 superiores a 0,95, utilizando a regressão linear.

A RNA aplicada em outras espécies para estimativa da área foliar também foi eficiente. Para cultura da pimenta, com coeficiente de determinação de 98% (AHMADIAN, 2012); para a cultura do milho, que também acharam coeficiente de determinação de 98% (ODABAS *et al.*, 2013); e para cultura de couve, com coeficiente de determinação de 96% (AZEVEDO *et al.*, 2017).

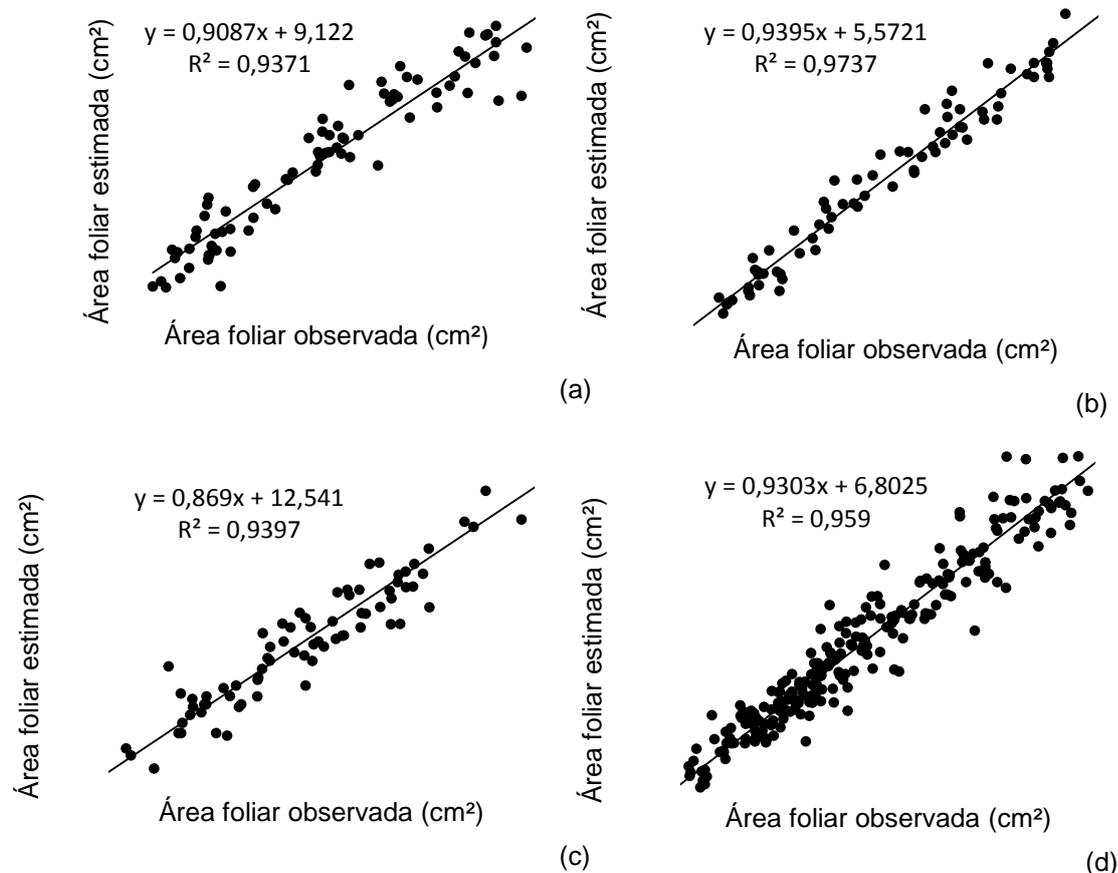


Figura 3. Dispersão gráfica da área foliar estimada por redes neurais artificiais do tipo perceptrons multicamadas, em função da área foliar observada da amostra de validação de folhas do grupo 1 (a), grupo 2 (b), grupo 3 (c) e para todos os grupos (d).

Fonte: Do autor, 2018.

A Tabela 4 apresenta os valores do coeficiente de determinação resultante das quatro redes neurais utilizadas na predição da área foliar, de acordo com cada grupo. Observando-se os valores dos coeficientes de determinação obtidos, notou-se uma similaridade da rede 4 em relação às demais redes e sua eficiência para todos os grupos, evidenciando ser a rede mais generalista.

Tabela 4 – Coeficientes de determinação obtidos pelas quatro redes neurais selecionadas na predição da área foliar de plantas do grupo 1, 2, 3 e da soma destes (4)

Rede utilizada	Amostras do grupo			
	1	2	3	4
Rede 1	0.937	0.921	0.860	0.912
Rede 2	0.913	0.973	0.852	0.916
Rede 3	0.871	0.916	0.939	0.908
Rede 4	0.911	0.935	0.862	0.959

Fonte: Do autor, 2018.

Para a melhor aplicabilidade deste trabalho, é apresentada, na figura 4, a área foliar predita por RNAs em função de diferentes valores de comprimento e largura do limbo para cada grupo e, de forma geral, para todos os formatos foliares. A partir dessas figuras, é possível que pesquisadores consigam prever a área foliar em soja, conhecendo apenas o comprimento e largura de um folíolo.

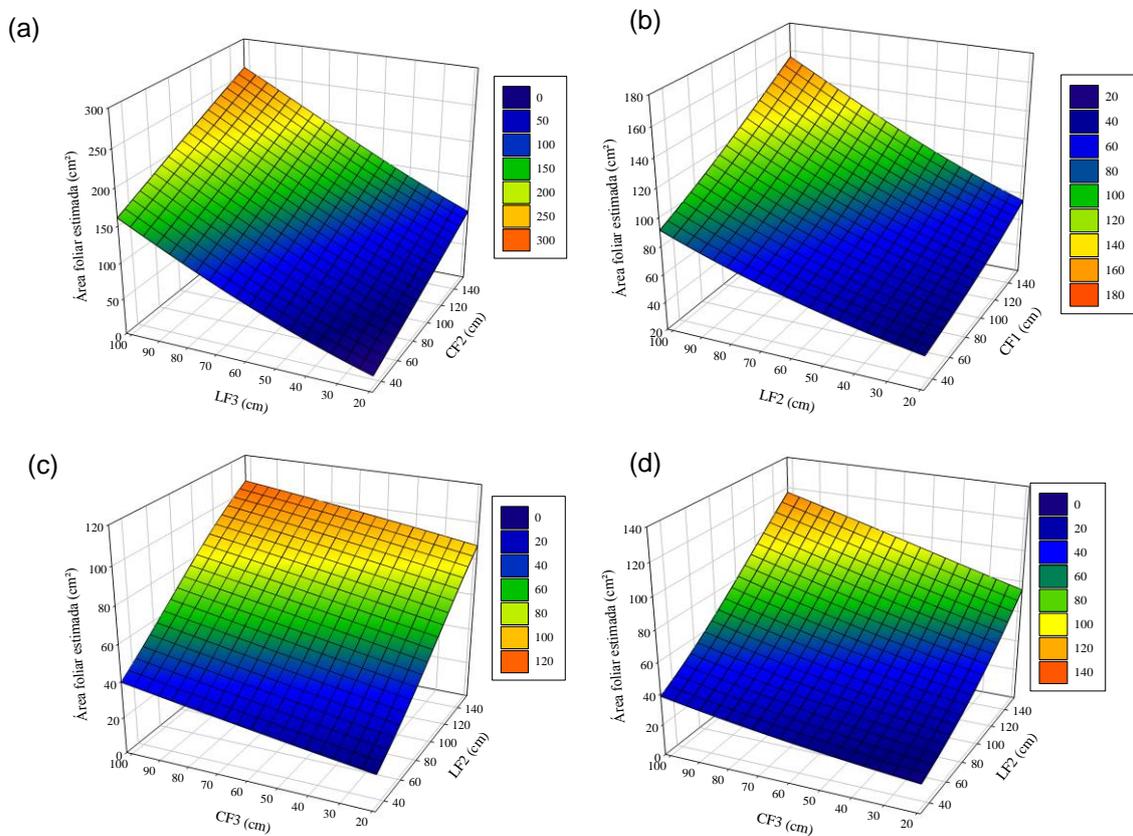


Figura 4. Gráficos de superfície resposta da área foliar predita (cm²) em função da largura e comprimento de folhas do grupo 1 (a), grupo 2 (b), grupo 3 (c) e para todos os grupos (c)
Fonte: Do autor, 2018.

Conclusão

As redes neurais artificiais do tipo perceptrons multicamadas são eficientes na predição da área foliar nas cultivares de soja avaliadas. Apenas duas medidas foliares dos folíolos são suficientes para estimar a área foliar. A rede 4, treinada com folíolos de todos os grupos, foi mais generalista e, conseqüentemente, é a mais indicada para a predição da área foliar em soja.

Referências

- ADAMI, M.; HASTENREITER, F. A.; FLUMIGNAN, L.; FARIA, R. T. Estimativa de área foliar de soja usando imagens digitais e dimensões foliares. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13, 2007, Florianópolis. **Anais...** Florianópolis: INPE, 2007. p. 9-14. Disponível em: <<https://bit.ly/2K7pOq9>>. Acesso em: 29 out. 2017.
- AHMADIANMOGHADAM, H. Prediction of pepper (*Capsicum annuum*) leaf area using group method of data handling-type neural networks. **International Journal of AgriScience**, v. 2, n. 11, p. 993-999, 2012. Disponível em: <<https://bit.ly/2HP4QOM>>. Acesso em: 29 out. 2017.
- AZEVEDO, A. M.; ANDRADE JÚNIOR, V. C.; SOUSA JÚNIOR, A. S.; SANTOS, A. A.; CRUZ, C. D.; PEREIRA, S. L.; OLIVEIRA, A. J. M. Eficiência da estimação da área foliar de couve por meio de redes neurais artificiais. **Horticultura Brasileira**, v. 35, n. 1, p. 14-19, 2017. Disponível em: <<https://bit.ly/2Jl40X4>>. Acesso em: 30 mar. 2016.
- BAKHSHANDEH, E.; KAMKAR, B.; TSIALTAS, J.T. Application of linear models for estimation of leaf area in soybean [*Glycine max* (L.) **Photosynthetica**, v. 49, n. 3, p. 405-416, 2011. Disponível em: <<https://bit.ly/2Fdl3qi>>. Acesso em: 13 nov. 2017.
- BERGMEIR, C.; BENÍTEZ, M. J. Neural networks in R using the Stuttgart Neural Network Simulator: RSNNS. **Journal of Statistical Software**, v. 46, n. 7, p. 1-26, 2012. Disponível em: Merr]. <<https://bit.ly/2HUak3n>>. Acesso em: 16 out. 2017.
- DEMIRSOY, H., LANG, G. A. Short communication: validation of a leaf area estimation model for sweet cherry. **Spanish Journal Agricultural Research**, v. 8, n. 3, p. 830-832, 2010. Disponível em: <<https://bit.ly/2qTZbNp>>. Acesso em: 29 out. 2017.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Sistema brasileiro de classificação de** Disponível em: <<https://bit.ly/2s5q14Y>>. Acesso em: 6 abr. 2017.
- DAI, X.; HUO, Z.; WANG, H. Simulation for response of crop yield to soil moisture and salisolos. Rio de Janeiro: EMBRAPA/CNPQ, 2011. 412 p.
- FALLOVO, C.; CRISTOFORI, V.; GYVES, E. M.; RIVERA, C. M.; REA, R.; FANASCA, S. Leaf area estimation model for small fruits from linear measurements. **HortScience**, v. 43, n. 7, p. 2263-2267, 2008. Disponível em: <<https://bit.ly/2HWqCNY>>. Acesso em: 20 nov. 2017.
- FASCELLA, G.; MAGGIORE, P.; ZIZZO, G.; COLLA, G.; ROUPHAEL, Y. A simple and low-cost method for leaf area measurement in *Euphorbia x lomi* Thai hybrids. **Advances in Horticultural Science**, v. 23, n. 1, p. 57-60, 2009. Disponível em: <<https://bit.ly/2HtMOOF>>. Acesso em: 29 out. 2017.
- GARSON, G. D. Interpreting neural-network connection weights. **Journal AI Expert**, v. 6, p. 47-51, 1991.
- GIUFFRIDA, F.; ROUPHAEL, Y.; TOSCANO, S.; SCUDERI, D.; ROMANO, D.; RIVERA, C. M.; COLLA, G.; LEONARDI, C. A simple model for nondestructive leaf area estimation in bedding plants. **Photosynthetica**, v. 49, n. 3, p. 380-388, 2011. Disponível em: <<https://bit.ly/2HW0N0D>>. Acesso em: 29 out. 2017.
- KANDIANNAN, K.; PARTHASARATHY, U.; KRISHNAMURTHY, K.; THANKAMANI, C.; SRINIVASAN, V. Modeling individual leaf area of ginger (*Zingiber officinale* Roscoe) using leaf length and width. **Scientia Horticulturae**, v. 120, n. 4, p. 532-537, 2009. Disponível em: <<https://bit.ly/2vHWG5B>>. Acesso em: 21 dez. 2017.
- KAUL, M.; HILL, R.L.; WALTHALL, C. Artificial neural networks for corn and soybean yield prediction. **Agricultural Systems**, v. 85, n. 1, p. 1-18, 2005. Disponível em: <<https://bit.ly/2K6xR6y>>. Acesso em: 20 dez. 2017.
- MAZZINI, R. B.; RIBEIRO, R. V.; PIO, R. M. A simple and nondestructive model for individual leaf area estimation in citrus. **Fruits**, v. 65, n. 5, p. 269-275, 2010. Disponível em: <<https://bit.ly/2KcmJp4>>. Acesso em: 14 nov. 2017.

MENDOZA-DE GYVES, E.; CRISTOFORI, V.; FALLOVO, C.; ROUPHAEL, Y.; BIGNAMI, C. Accurate and rapid technique for leaf area measurement in medlar (*Mespilus germanica* L.). **Advances in Horticultural Science**, v. 22, n. 3, p. 223-226, 2008. Disponível em: <<https://bit.ly/2Hpx2Z7>>. Acesso em: 29 out. 2017.

MOOSAVI, A. A.; SEPASKHAH, A. Artificial neural networks for predicting unsaturated soil hydraulic characteristics at different applied tensions. **Archives of Agronomy and Soil Science**, v. 58, n. 2, p. 125-153, 2012.

MOOSAVIZADEH-MOJARAD, R.; SEPASKHAH, A. R. Comparison between rice grain yield predictions using artificial neural networks and a very simple model under different levels of water and nitrogen application. **Archives of Agronomy and Soil Science**, v. 58, n. 11, p. 1271-1282, 2012.

MÜLLER, L. Morfologia, anatomia e desenvolvimento. In: MIYASAKA, S.; MEDINA, J.C. (Ed.). **A soja no Brasil**. Campinas: ITAL, 1981. 1062 p.

ODABAS, M. S.; ERGUN, E.; ONER, F. Artificial neural network approach for the prediction of the corn (*Zea mays* L.) leaf area. **Bulgarian Journal of Agricultural Science**, v. 19, n. 4, p. 766-769, 2013. Disponível em: <<https://bit.ly/2HJ8MAf>>. Acesso em: 17 nov. 2017.

OLFATI, J. A.; PEYVAST, G.; SHABANI, H.; NOSRATIE-RAD, Z. An estimation of individual leaf area in cabbage and broccoli using non-destructive methods. **Journal of Agricultural Science and Technology**, v. 12, p. 627-632, 2010. Disponível em: <<https://bit.ly/2Hohzs0>>. Acesso em: 17 nov. 2017.

PALIWAL, M.; KUMAR, U. A. Assessing the contribution of variables in feed forward neural network. **Applied Soft Computing**, v.11, p. 3690-3696, 2011. Disponível em: <<https://bit.ly/2sFVkv7>>. Acesso em: 2 set. 2017.

PEKSEN, E. Non-destructive leaf area estimation model for faba bean (*Vicia faba* L.). **Scientia Horticulturae**, v.113, n. 4, p. 322-328, 2007. Disponível em: <<https://bit.ly/2HrRsfZ>>. Acesso em: 20 set. 2017.

RAD, M. R. N.; KOOHKAN, S. H.; FANAEI, H. R.; RAD, M. R. P. Application of artificial neural networks to predict the final fruit weight and random forest to select important variables in native population of melon (*Cucumis melo* L.). **Scientia Horticulturae**, v. 181, p. 108-112, 2015. Disponível em: <<https://bit.ly/2qUcMEem>>. Acesso em: 3 jan. 2017.

RICHTER, G. L.; ZANON, A. J.; STRECK, N. A.; GUEDES, J. V. C.; KRÄULICH, B.; ROCHA, T. S. M.; WINCK, J. E. M.; CERA, J. C. Estimating leaf area of modern soybean cultivars by a non-destructive method. **Bragantia**, v. 73, n. 4, p. 416-425, 2014. Disponível em: <<https://bit.ly/2qVXKNO>>. Acesso em: 29 out. 2017.

RIVERA, C. M.; ROUPHAEL, Y.; CARDARELLI, M.; COLLA, G. A simple and accurate equation for estimating individual leaf area eggplant from linear measurements. **European Journal of Horticultural Science**, v. 70, n. 5, p. 228-230, 2007. Disponível em: <<https://bit.ly/2HnkxwR>>. Acesso em: 4 dez. 2017.

ROUPHAEL, Y.; COLLA, G.; FANASCA, S.; KARAM, F. Leaf area estimation of sunflower leaves from simple linear measurements. **Photosynthetica**, v. 45, p. 306-308, 2007. Disponível em: <>. Acesso em: 29 out. 2017.

ROUPHAEL, Y.; RIVERA, C. M.; CARDARELLI, M.; FANASCA, S.; COLLA, G. Leaf area estimation from linear measurements in zucchini plants of different ages. **Journal of Horticultural Science and Biotechnology**, v. 81, n. 2, p. 238-241, 2006.

SETIYONO, T. D.; WEISS, A.; SPECHT, J. E.; CASSMAN, K. G.; DOBERMANN, A. Leaf area index simulation in soybean grown under near-optimal conditions. **Field Crops Research**, v. 108, p.82-92, 2008. Disponível em: <<https://bit.ly/2HVbOz2>>. Acesso em: 2 dez. 2017.

SHABANI, A.; GHAFFARY, K. A.; SEPASKHAH, A. R.; KAMGAR-HAGHIGHI, A. A. Using the artificial neural network to estimate leaf area. **Scientia Horticulturae**, v. 216, p. 103-110, 2017. Disponível em: <<https://bit.ly/2HGClmj>>. Acesso em: 29 out. 2017.

- SILVA, G. N.; TOMAZ, R. S.; SANT'ANNA, I. C.; NASCIMENTO, M.; BHERING, L. L.; CRUZ, C.D. Neural networks for predicting breeding values and genetic gains. **Scientia Agricola**, v. 71, n. 6, p. 494-498, 2014. Disponível em: <<https://bit.ly/2Jqc48R>>. Acesso em: 8 nov. 2017.
- SILVA, S. H. M. G.; LIMA, J. D.; BENDINI, H. N.; NOMURA, E. S.; MORAES, W. S. Estimativa da área foliar do antúrio com o uso de funções de regressão. **Ciência Rural**, v. 38, n. 1, p. 243-246, 2008. Disponível em: <<https://bit.ly/2JqJKmD>>. Acesso em: 5 dez. 2017.
- SOARES, J. D. R.; PASQUAL, M.; LACERDA, W. S.; SILVA, S. O.; DONATO, S. L. R. Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants. **Scientia Horticulturae**, v. 155, p. 24-29, 2013. Disponível em: <<https://bit.ly/2Feb84W>>. Acesso em: 22 nov. 2017.
- STOPPANI, M. I.; WOLF, R.; FRANCESCANGELI, N.; MARTÍ, H. R. A nondestructive and rapid method for estimating leaf area of broccoli. **Advances in Horticultural Science**, v. 17, n. 3, p. 173-175, 2003. Disponível em: <<https://bit.ly/2vKqHSu>>. Acesso em: 14 set. 2017.
- TRAJKOVIC, S. Temperature-based approaches for estimating reference evapotranspiration. **Journal of Irrigation and Drainage Engineering**, v. 131, n. 4, p. 316-323, 2005. Disponível em: <<https://bit.ly/2HWTKEO>>. Acesso em: 29 out. 2017.
- TSIALTAS, J. T.; MASLARIS, N. Leaf area prediction model for sugar beet (*Beta vulgaris* L.) cultivars. **Photosynthetica**, v. 46, n. 2, p. 291-293, 2008. Disponível em: <<https://bit.ly/2HIY2IH>>. Acesso em: 20 dez. 2017.
- WANG, Z.; ZHANG, L. Leaf shape alters the coefficients of leaf area estimation models for *Saussurea stoliczkai* in central Tibet. **Photosynthetica**, v. 50, n. 3, p. 337-342, 2012. Disponível em: <<https://bit.ly/2HoIMXD>>. Acesso em: 5 nov. 2017.
- WIERSMA, J. V.; BAILEY, T. B. Estimation of leaf et, trifoliolate and total leaf area of soybean. **Agronomy Journal**, v. 67, p. 26-30, 1975.

4 CONSIDERAÇÕES FINAIS

No decorrer desta dissertação, foram estudados aspectos relativos à utilização e à análise das redes neurais artificiais para estudos de agrupamento e predição de dados.

No primeiro artigo, mostrou-se a eficiência das RNAs para estudar a dissimilaridade genética entre 36 cultivares de soja, com a formação de seis grupos divergentes, com base em dados quantitativos, e confirmou-se essa eficiência por meio da análise discriminante de Anderson. Apresentou-se, ainda, uma metodologia para escolha da melhor topologia da rede neural, que se mostrou eficiente. Por não existir nenhuma metodologia para escolha da topologia de rede, essas técnicas se fazem muito importantes, podendo ser aprofundadas em futuros estudos. Essas técnicas podem ser valiosas no melhoramento – por exemplo, na seleção de acessos em bancos de germoplasma ou para avaliar um número muito grande de populações.

No segundo artigo, obtivemos metodologias para a estimação da área foliar em soja eficientemente pelas RNAs, considerando-se diferentes formatos foliares. Verificamos a eficiência na estimação e identificamos as variáveis explicativas com maior importância no processo de predição pelas RNAs em cultura de soja. Essa técnica pode gerar grandes benefícios em estudos de fenotipagem, agregando rapidez, facilidade e eficiência nos estudos, podendo também ser usada para outras estimativas importantes, como área de raízes, resistência a doenças, entre outros.