

## **Predição de produtividade do mamoeiro (*Carica Papaya* L.) por Redes Neurais Artificiais**

**Prediction of productivity of papaya (*Carica Papaya* L.) by Artificial Neural Networks**

**Predicción de la productividad de la papaya (*Carica Papaya* L.) por Redes Neuronales Artificiales**

Recebido: 16/09/2021 | Revisado: 26/09/2021 | Aceito: 30/09/2021 | Publicado: 02/10/2021

### **Leila Verena da Conceição**

ORCID: <https://orcid.org/0000-0002-1557-0159>  
Universidade Federal do Recôncavo da Bahia, Brasil  
E-mail: [verenaleila@gmail.com](mailto:verenaleila@gmail.com)

### **Cristina Ferreira Nepomuceno**

ORCID: <https://orcid.org/0000-0002-8242-1620>  
Universidade Federal do Recôncavo da Bahia, Brasil  
E-mail: [cfnbio@gmail.com](mailto:cfnbio@gmail.com)

### **Bruno Vinícius Castro Guimarães**

ORCID: <https://orcid.org/0000-0003-2585-8794>  
Instituto Federal do Amazonas, Brasil  
E-mail: [bvinicius20@yahoo.com.br](mailto:bvinicius20@yahoo.com.br)

### **Sebastião de Oliveira e Silva**

ORCID: <https://orcid.org/0000-0002-1276-5233>  
Universidade Federal do Recôncavo da Bahia, Brasil  
E-mail: [ssilva3000@gmail.com](mailto:ssilva3000@gmail.com)

### **Resumo**

A estimativa da produtividade do mamoeiro é importante para o planejamento de produtores rurais, tornando-se uma ferramenta eficiente e estratégica na tomada de decisões da produção agrícola, sobretudo no que diz respeito ao planejamento de pós colheita, armazenamento e transporte. O objetivo desse estudo foi selecionar modelos de Redes Neurais Artificiais com maior capacidade preditiva da colheita para cultura do mamoeiro CNPMF-L78. Foram mensuradas 352 plantas em uma área útil de 2.112 m<sup>2</sup>, com os descritores de planta e os caracteres relacionados à qualidade do fruto. Os dados foram submetidos a análise por Redes Neurais Artificiais, utilizando-se o software R. Os melhores ajustes para a estimativa da produtividade do mamoeiro foram considerados três camadas das RNAs: entrada, intermediária e a camada de saída, testando-se a composição com um a 10 neurônios na camada intermediária. Esses modelos apresentaram os menores erros quadráticos médios, o que corresponde a maior proximidade entre os dados preditos e os reais, e, por conseguinte, maior eficiência das redes na predição da produtividade. Pelo coeficiente de determinação, verificaram-se os melhores ajustes para caracteres reprodutivos  $R^2 = 0,99$ , os caracteres vegetativos proporcionaram  $R^2 = 0,54$ , enquanto que os caracteres de qualidade e rendimento viabilizou  $R^2 = 0,24$ . Previsões de produtividade para o mamoeiro CNPMF-L78 foram obtidas com alta eficiência para os caracteres reprodutivos por meio de redes neurais artificiais.

**Palavras-chave:** Inteligência artificial; Mamão; Planejamento rural; Predição.

### **Abstract**

The estimation of the productivity of papaya is important for the planning of rural producers, becoming an efficient and strategic tool in the decision making of agricultural production, especially with regard to post-harvest planning, storage and transportation. The objective of this study was to select Artificial Neural Network models with greater predictive capacity of the harvest for the papaya crop CNPMF-L78. A total of 352 plants were measured in a useful area of 2,112 m<sup>2</sup>, with the plant descriptors and the characters related to fruit quality. The data were submitted to artificial neural network analysis using the R software. The best fits for estimating the productivity of papaya were considered for three layers of ANNs: input, intermediate, and output layers, testing the composition with one to 10 neurons in the intermediate layer. These models presented the lowest mean square errors, which corresponds to greater proximity between predicted and actual data, and therefore greater efficiency of the networks in predicting yields. By the coefficient of determination, the best fits were verified for reproductive characters  $R^2 = 0.99$ , vegetative characters provided  $R^2 = 0.54$ , while quality and yield characters enabled  $R^2 = 0.24$ . Productivity predictions for CNPMF-L78 papaya were obtained with high efficiency for reproductive characters by means of artificial neural networks.

**Keywords:** Artificial intelligence; Papaya; Rural planning; Prediction.

## Resumen

La estimación de la productividad de la papaya es importante para la planificación de los productores rurales, convirtiéndose en una herramienta eficiente y estratégica en la toma de decisiones en la producción agrícola, especialmente en lo que respecta a la planificación de la poscosecha, el almacenamiento y el transporte. El objetivo de este estudio fue seleccionar los modelos de Redes Neuronales Artificiales con mayor capacidad de predicción de la cosecha para el cultivo de papaya CNPMF-L78. Se midieron 352 plantas en una superficie útil de 2.112 m<sup>2</sup>, con los descriptores de la planta y los caracteres relacionados con la calidad del fruto. Los datos se sometieron a un análisis de redes neuronales artificiales utilizando el software R. Los mejores ajustes para estimar la productividad de la papaya se consideraron en tres capas de la RNA: de entrada, intermedia y de salida, probando la composición con una a 10 neuronas en la capa intermedia. Estos modelos presentaron los errores cuadrados medios más bajos, lo que corresponde a la mayor proximidad entre los datos predichos y los reales, y por tanto a una mayor eficacia de las redes en la predicción de la productividad. Por el coeficiente de determinación, los mejores ajustes se encontraron para los caracteres reproductivos  $R^2 = 0,99$ , los caracteres vegetativos proporcionaron  $R^2 = 0,54$ , mientras que los caracteres de calidad y rendimiento permitieron  $R^2 = 0,24$ . Las predicciones de productividad para la papaya CNPMF-L78 se obtuvieron con alta eficiencia para los caracteres reproductivos utilizando redes neuronales artificiales.

**Palabras clave:** Inteligencia artificial; Papaya; Planificación rural; Predicción.

## 1. Introdução

A estimativa precisa do rendimento é um dos propósitos buscado em cultivos de mamoeiro devido ao alto valor econômico de sua produção. O mamão (*Carica papaya* L.) tem grande importância para a agricultura e é amplamente cultivado nas regiões tropicais e subtropicais do mundo (Mishra, Varun, Hans & Saxena, 2021). Seu desempenho na agricultura brasileira, demanda novas técnicas, que busque eficiência em ganhos de produção. Todavia um dos problemas mais importantes da agricultura é a falta de previsão do rendimento da safra (Pan, Pant, Singh, Singh & Pant, 2021).

Impasse esse, que tem sido alvo em estudos nos programas de melhoramento com o desenvolvimento de variedades de alta produtividade e práticas que permitam reduzir o uso de defensivos e o custo de fertilizantes, priorizando a redução do custo de produção (Empresa Brasileira de Pesquisa Agropecuária [Embrapa], 2020). Contudo, mesmo com os avanços nos programas de melhoramento, é de extrema necessidade técnicas mais robustas que possam fazer a predição de produtividade de forma mais eficiente possível.

Nesse sentido, técnicas de produção se alicerçam, cada vez mais, no aperfeiçoamento do uso dos insumos e nas metodologias destinadas a estimar a produção, de forma a melhorar a produtividade e a rentabilidade. O novo mercado agrícola digital tem como vantagem a digitalização, as plataformas digitais amplificam benefícios para vários setores agrícolas, auxiliando na melhor tomada de decisão (Merladete, 2021). Entretanto, o rendimento da agricultura depende de vários fatores, como condições climáticas, ambientais, de manejo e tratos culturais, o que dificulta a exata estimativa de produtividade (Embrapa, 2020).

Na direção em otimizar as atividades agrícolas, driblando os percalços associados à predição da produtividade, modelagens em Redes Neurais Artificiais têm se mostrado promissor. Pela relevância e necessidade de se conhecer a produtividade previamente, diversas culturas já desenvolveram modelos preditores da produtividade, como a cultura do milho (Soares, Robaina, Peiter & Russi, 2015), palma forrageira (Guimarães, Donato, Azevedo, Aspiazú & Silva Jr, 2018), de trigo (Mamann et al., 2019), soja na região do Matopiba (Santos, 2020), bananeiras 'Prata-Anã' e 'BRS Platina' (Guimarães, Donato, Aspiazú & Azevedo, 2021).

Todavia, o mamoeiro carece de uma ferramenta precisa, que proponha menos tempo de análise, para que o produtor tenha maior prazo para o planejamento, para assim assegurar um bom empreendimento. Dessa forma, os modelos heurísticos, ou seja, a Ciência da inteligência Computacional traz uma proposta que viabiliza os dados biométricos e propõem resultados mais rápidos, com menor margem de erro. Auxiliando o profissional da área quanto à tomada de decisão em uma produção já prevista.

Nesse sentido, este estudo teve por propósito selecionar modelos de Redes Neurais Artificiais com maior capacidade

preditiva da colheita para cultura do mamoeiro CNPMF-L78.

## 2. Metodologia

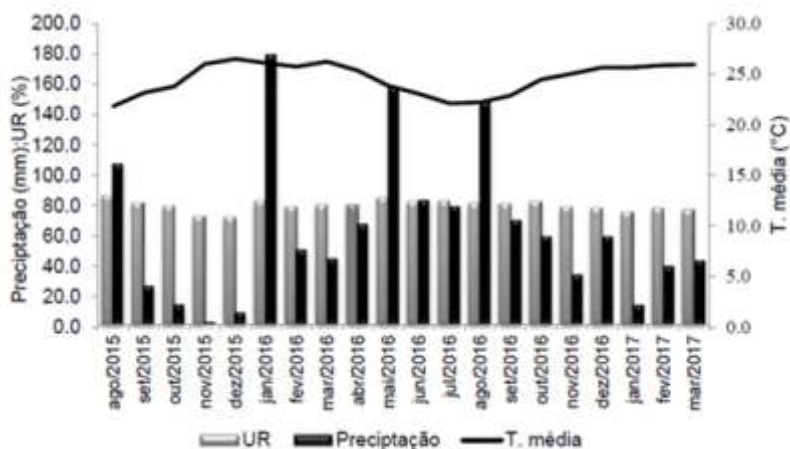
A abordagem deste estudo é voltada à análise quantitativa de natureza básica, e o procedimento de pesquisa experimental utilizado foi com base na proposta metodológica estabelecida por Gil (2017). Esta pesquisa foi conduzida em campo com base na experimentação homogênea ou ensaio de uniformidade, em que todas as plantas, denominadas como unidades básicas receberam o mesmo tratamento. Adicionalmente, a metodologia em questão foi similar às propostas de Guimarães et al. (2018 e 2021) com a cultura da palma forrageira e da banana, respectivamente.

### 2.1 Implantação e condução da cultura do mamoeiro (Linhagem CNPMF-L78).

O estudo foi conduzido na Embrapa Mandioca e Fruticultura, no período de agosto de 2015 a março de 2017, em Cruz das Almas (BA), situado nas coordenadas geográficas: 12°40'39" latitude sul e 39°06'23" longitude oeste de Greenwich e altitude de 226 m, com temperatura média de 24,5 °C, umidade relativa de 82% e precipitação média anual de 1.197 mm (Instituto Brasileiro de Geografia e Estatística [IBGE], 2020).

As características climáticas médias registradas no período do experimento (Figura 1) foram obtidas no Posto Meteorológico da, localizado próximo à área experimental.

**Figura 1.** Características climáticas médias registradas entre agosto de 2015 e março de 2017, na área experimental da Embrapa Mandioca e Fruticultura em Cruz das Alma- BA.



Fonte: Estação Meteorológica da Embrapa Mandioca e Fruticultura (2020).

O solo da área experimental é do tipo Latossolo Amarelo distrófico típico (Santos et al., 2018), que corresponde a um Ferralsol (IUSS WORKING GROUP WRB, [WRB], 2015) ou um Oxisol (Smith, 2014). Para a correção e adubação do solo, seguiram-se as recomendações para a cultura do mamoeiro (Oliveira, Souza, & Coelho, 2010), sendo utilizadas na semeadura três sementes por saco plástico preenchidos com terra vegetal.

As sementes utilizadas no plantio foram da Linhagem CNPMF-L78, grupo Solo, provenientes do Banco Ativo de Germoplasma de Mamão (BAG-Mamão) da Embrapa Mandioca e Fruticultura.

Foram semeadas três sementes por saco plástico de polietileno com dimensões de 7,0 cm x 18,5 cm x 0,006 cm, correspondentes à largura, à altura e à espessura, respectivamente. Preenchidos com substrato composto por solo, areia e esterco de curral curtido, na proporção de 3:1:1.

O desbaste foi efetuado aos 15 dias após a emergência das plantas, de forma a manter uma planta por saco plástico. E

após 40 dias foram transplantadas para área experimental com adubação orgânica (esterco), com três plantas por cova e no início do florescimento das plantas foi mantida apenas uma planta a que apresentava inflorescências hermafroditas.

A área plantada foi constituída por 18 fileiras com 24 plantas cada e consideradas como útil as 16 fileiras centrais, com 22 plantas por fileira, no espaçamento de 3 m x 2 m, com o total de 352 plantas e área útil de 2.112 m<sup>2</sup>, não avaliando a primeira e a última planta de cada fileira.

## 2.2 Características morfoagronômicas avaliadas

Foram avaliados 17 caracteres agrônômicos, três relacionados às características vegetativas da planta, sete relacionados as reprodutivas e sete as de qualidade dos frutos, descritas a seguir:

### 2.2.1 Características vegetativas

a. Altura da planta (AP6); (AP12); (AP18): distância entre a superfície do solo até o ponto de inserção da folha mais nova, localizado no ápice caulinar, avaliada aos seis, 12 e 18 meses de idade, expressa em centímetros (cm);

b. Altura de inserção dos primeiros frutos (AIPF): altura de inserção da primeira flor funcional (que irá produzir frutos), no início da produção, de oito a 10 meses após o plantio, a partir da superfície do solo, contígua ao colo da planta, até o ponto de inserção do primeiro fruto, expressa em centímetros (cm);

c. Diâmetro do caule (DC6); (DC12); (DC18): medido a 20 cm acima do nível do solo, avaliado aos seis, 12 e 18 meses de idade;

### 2.2.2 Caracteres reprodutivos

a. Precocidade (PREC): avaliada mediante indicação da data da primeira colheita de frutos;

b. Número de frutos comerciais por planta (NFC9); (NFC14): contagem de frutos comerciais presentes em cada planta aos nove e 14 meses após o plantio;

c. Número de frutos comerciais totais (NFCT): soma de frutos comerciais presentes em cada planta aos nove e 14 meses após o plantio;

d. Número de frutos deformados por planta (NFD9); (NFD14): contagem de frutos carpelóides, pentândricos e bananiformes presentes em cada planta, aos nove e 14 meses após o plantio;

e. Número de nós sem frutos (NNSF9); (NNSF14): avaliado aos nove e 14 meses após o plantio, na região denominada usualmente por “pescoço”, mediante contagem do número de nós que não produziram frutos;

f. Produtividade (PROD): Estimada pela multiplicação do número de frutos comerciais por planta (NFC) pelo peso médio do fruto por planta, considerando-se o espaçamento do plantio;

### 2.2.3 Caracteres de qualidade do fruto

a. Comprimento de fruto (CF): expresso em cm, com o auxílio de um paquímetro de madeira mediu-se o comprimento da base ao ápice do fruto;

b. Diâmetro de fruto (DF): expresso em centímetros cm, com o auxílio de um paquímetro de madeira realizou-se a medição na parte de maior diâmetro do fruto;

c. Massa de fruto (PF): expresso em gramas, pesou-se os frutos colhidos por planta em uma balança analítica. Os frutos foram colhidos no estágio 2 de maturação, ou seja, ¼ maduro, com até 25% da casca amarela;

d. Teor de sólidos solúveis (SS): expresso em o Brix (°Bx), obtido com auxílio de refratômetro digital portátil modelo r2mini Reicher;

e. Firmeza do fruto (FF): expresso em kg cm<sup>-2</sup>, determinada em frutos maduros íntegros na região central com auxílio de um eletrômetro, a partir de 03 (três) leituras;

f. Diâmetro da cavidade interna do fruto (DCI): expresso em cm, mediu-se o diâmetro da cavidade na parte central do fruto. No caso de cavidades em formato de estrela, as medidas foram tomadas de uma extremidade a outra de maior de distância;

g. Espessura da polpa (EP): expresso em cm, com o auxílio de um paquímetro tomou-se a medição da espessura da polpa de maior tamanho após o corte transversal do fruto.

#### 2.2.4 Modelagem dos dados Redes Neurais Artificiais-RNA

Os dados foram avaliados por Redes Neurais Artificiais (RNA) no software R. Para o treinamento das RNAs tanto os dados das variáveis predictoras (AP6; AP12; AP18; AIPF; DC6; DC12; DC18; PREC; NFC9; NFC14; NFCT; NFD9; NFD14; NNSF9; NNSF14; PROD; CF; DF; PF; SS; FF; DCI; EP) quanto os da variável predita (Produtividade) foram normalizados no intervalo entre 0 e 1 pela função *normalizeData* do pacote RSNNS (Bergmeir & Benítez, 2012).

Na análise por RNAs, 80% dos dados (282 plantas) foram destinadas ao treinamento da rede, e 20% para a análise de validação (70 plantas). O conjunto de dados foi selecionado de modo aleatório para as duas amostras. O modelo neural empregado foi do tipo perceptron de multicamadas (MLP - *Multi-Layer-Perceptron*) com o auxílio dos pacotes Keras e tensorflow no software R. Para a estimativa da produtividade do mamoeiro foram considerados três camadas das RNAs: entrada, intermediária e a camada de saída, testando-se a composição com um a 10 neurônios na camada intermediária.

Foi considerada a função de ativação “sigmoide” para a camada intermediária e de saída. Na compilação do modelo foi utilizado o otimizador “rmsprop”, função de perda “mse” e taxa de aprendizado de 0,1. No treinamento foi arbitrado 500 épocas de treinamento, com as funções de ativação para as camadas intermediárias e de saída ajustadas como logística e linear, respectivamente. Posteriormente, a melhor arquitetura de rede, selecionada pelo menor erro quadrático médio na predição da amostra de validação, foi treinada por 1000 vezes.

Para avaliar a capacidade preditiva do modelo, os dados da produtividade predita e a observada na amostra de validação foram submetidos à análise de regressão. Assim, pela associação entre as variáveis preditas e observadas, testou-se a significância do coeficiente angular da reta pelo teste t. Dessa forma, se coeficiente angular da reta não se diferir de 1 e o coeficiente de determinação for alto, assume-se eficiência no processo de predição. O coeficiente de determinação do modelo foi determinado por (Equação 1).

$$R^2 = \left( \frac{\sum_{i=1}^n (Xobs_i - \overline{Xobs})(Xpred_i - \overline{Xpred})}{\sqrt{\sum_{i=1}^n (Xobs_i - \overline{Xobs})^2 (Xpred_i - \overline{Xpred})^2}} \right)^2 \quad (\text{Eq.1})$$

O erro quadrático médio (EQM) foi estabelecido pela média entre diferença quadrática entre os valores observados e os previstos pelos modelos, em que o menor valor indicou o melhor ajuste do modelo. EQM para este estudo é definido pela seguinte expressão (Equação 2).

$$EQM = \sum_i^n \frac{(Xobs_i - Xpred_i)^2}{n} \quad (\text{Eq. 2})$$

Em que:

$Xobs_i$ : i-ésimo valor observado para a variável dependente;

$\overline{X_{obs}}$ : média do valor observado para a variável dependente;

$X_{pred_i}$ : i-ésimo valor predito para a variável dependente;

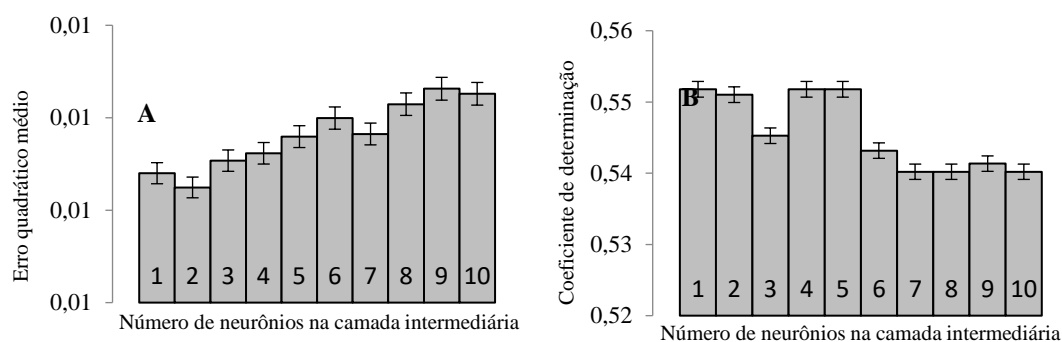
$\overline{X_{pred}}$ : média do valor predito para a variável dependente.

### 3. Resultados e Discussão

#### 3.1 Modelagem dos dados de RNA para as características vegetativas

O desempenho das arquiteturas de rede foi medido pelo erro quadrático médio (EQM) e os resultados mostram boa consistência na análise dos dados, como retratados na Figura 2, com a expressão do número de neurônios das camadas intermediárias, EQM (Figura 2A) e o coeficiente de determinação ( $R^2$ ) (Figura 2B). Após intenso treinamento para se obter uma rede adequada (Soares, Pasqual, Lacerda, Silva & Donato, 2014), ficou estabelecido a melhor arquitetura da rede a que possibilitou menor erro quadrático médio na predição da amostra de validação.

**Figura 2.** Estimativas do erro quadrático médio-EQM (A) e coeficiente de determinação- $R^2$  (B) obtidos, considerando diferentes números de neurônios na camada intermediária. Os desvios referem-se aos limites inferiores e superiores ao nível de 95% confiança obtida por bootstrap com 10.000 simulações.

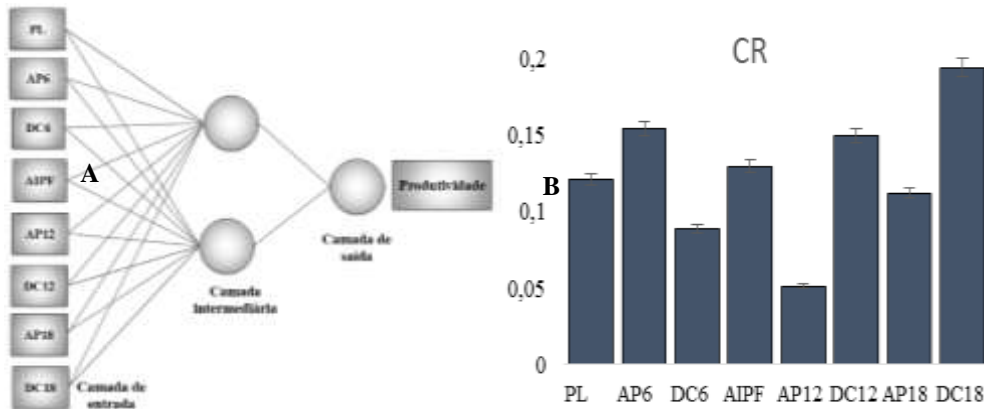


Fonte: Autores (2021).

Assim, foram testados modelos de RNA para prever a produtividade do mamoeiro CNPMF-L78 por meio dos caracteres vegetativos. Ao analisar o coeficiente de determinação ( $R^2$ ), resultados razoáveis foram averiguados com dois neurônios na camada intermediária, em que o ajuste dos dados possibilitou prever a produtividade em 55% (Figura 2B).

Essa estrutura de rede foi selecionada com base no menor valor do EQM (Azevedo et al., 2019) associado ao  $R^2$ . Na Figura 3A é possível verificar a topologia da melhor adequação da rede com os caracteres vegetativos, ao passo que a Figura 3B mostra a contribuição relativa dos caracteres com maior importância para expressar a produtividade.

**Figura 3.** Topologia da melhor rede ajustada (A) e contribuição relativa – CR (B), obtida pelo método de Garson (1991), dos parâmetros agrônômicos, (AP- altura da planta, DC- diâmetro do caule, avaliada aos seis, 12 e 18 meses de idade; AIPF- altura de inserção da primeira flor funcional) apresentada na camada de entrada e diâmetro do caule avaliado aos seis, 12 e 18 meses para a predição da produtividade mamoeiro CNPMF-L78 por redes neurais artificiais. Os desvios referem-se aos intervalos de confiança de 95%, obtidos por bootstrap BCa com 10.000 simulações.

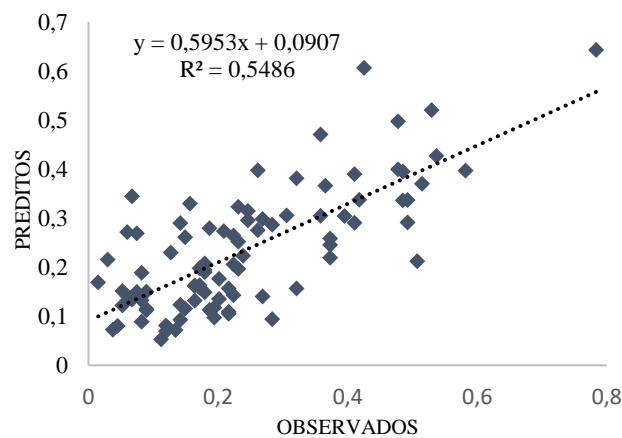


Fonte: Autores (2021).

Pela importância relativa das características preditivas, ponderadas pelo método de Garson (1991), o diâmetro do caule avaliado aos 18 meses de idade (Figura 3B) foi o caractere mais relevante com contribuição relativa de 19% no rendimento do mamoeiro. Dentre os caracteres vegetativos, a altura da planta e diâmetro do caule avaliados aos 6 meses de idade apresentaram as menores contribuições relativas com 5,12% e 8,87%, nesta ordem, para estimativa da produtividade (Figura 3B).

Para testar a eficiência do modelo de predição, a produtividade estimada foi comparada com a produtividade observada por meio das variáveis de entrada destinadas à amostra de validação ajustadas por redes neurais com valores de  $R^2$  de 0,55 (Figura 4).

**Figura 4.** Representação gráfica da qualidade de predição da produtividade do mamoeiro CNPMF-L78 por redes neurais artificiais para a amostra de validação considerando todos os caracteres avaliados.



Fonte: Autores (2021).

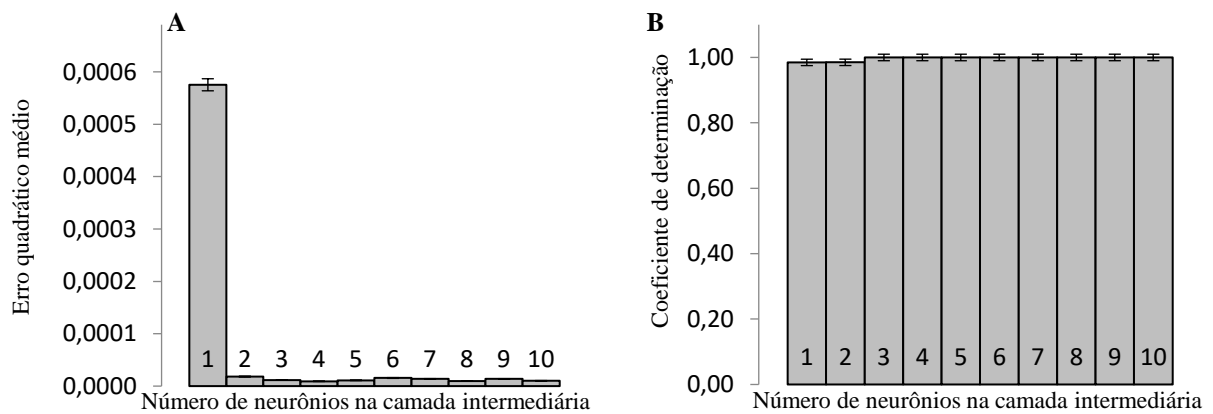
Uma boa consistência dos dados é muito importante para fins de predição de rendimentos em plantios comerciais, pois podem ser utilizados para alimentar modelos matemáticos de predição do rendimento, gerando informações úteis para

aperfeiçoar diversas atividades agrícolas, como a colheita, pós colheita, embalagem e comercialização (Aggelopoulou et al., 2011). A seleção dos modelos e a comprovação de sua eficiência para generalização da RNA ocorreu em resposta ao  $R^2$  e a não significância ( $p > 0,05$ ) do teste t para a hipótese nula da inclinação da reta ( $H_0: b = 1$ ). Tonando possível o melhor entendimento do processo modelado (Azevedo et al., 2019).

### 3.2 Modelagem dos dados de RNA para as características reprodutivas

O desempenho das arquiteturas de rede para modelagem dos caracteres de produção do fruto: PREC (precocidade); NFC (número de frutos comerciais por planta); NFCT (número de frutos comerciais totais); NFD (número de frutos deformados por planta); NNSF (número de nós sem frutos) foi medido pelo erro quadrático médio (EQM) (Figura 5). Os números de neurônios das camadas intermediárias são expressos no EQM (Figura 5A) e amostra de treinamento e o coeficiente de determinação ( $R^2$ ) (Figura 5B).

**Figura 5.** Estimativas do erro quadrático médio-EQM (A) e coeficiente de determinação- $R^2$  (B) obtidos considerando diferentes números de neurônios na camada intermediária. Os desvios referem-se aos limites inferiores e superiores ao nível de 95% confiança obtidos por bootstrap com 10.000 simulações.



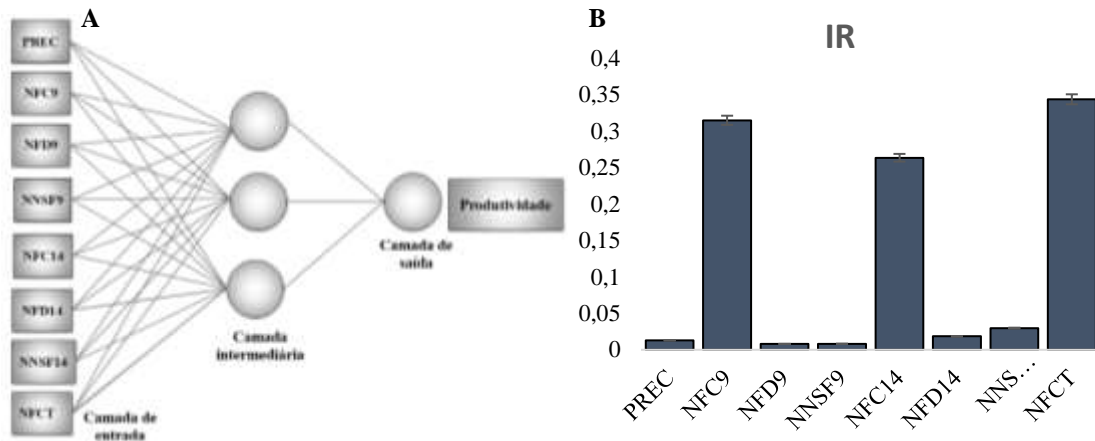
Fonte: Autores (2021).

Resultados significativos foram obtidos quando se testou três neurônios na camada intermediária (Figura 5B), em que o ajuste apresentou os menores erros do quadrado médio (EQM) (Figura 5A) associados aos maiores valores do coeficiente de determinação ( $R^2$ ) (Figura 5B).

Os menores valores de EQM representam a maior proximidade entre os dados previstos e reais e, portanto, a maior eficiência das redes na previsão do rendimento. O  $R^2$ , expressou resultados satisfatórios quando ponderados para o número de três neurônios na camada intermediária, com ajuste dos dados em 99,98% (Figura 5B). A estrutura da rede com três neurônios na camada intermediária é apresentada na Figura 6A, com a estratificação da contribuição relativa expressa pela Figura 6B.



**Figura 6.** Topologia da melhor rede ajustada (A) e contribuição relativa (B), obtida pelo método de Garson (1991), dos parâmetros agrônômicos (PREC-precocidade; NFC- número de frutos comerciais por planta (avaliado aos 9 e 14 meses); NFCT- número de frutos comerciais totais (soma); NFD- número de frutos deformados por planta e NNSF-número de nós sem frutos) apresentados na camada de entrada para a predição da produtividade mamoeiro CNPMF-L78 por redes neurais artificiais. Os desvios referem-se aos intervalos de confiança de 95%, obtidos por bootstrap BCa com 10.000 simulações.

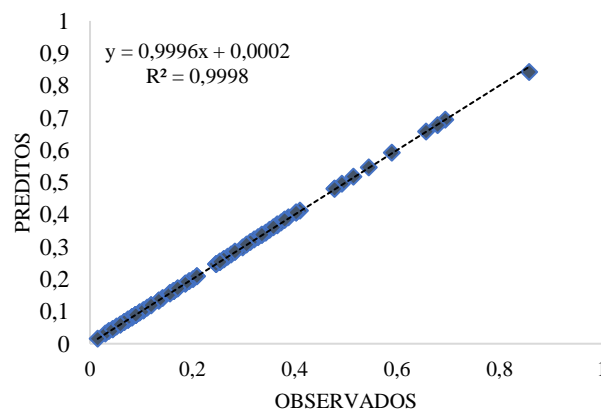


Fonte: Autores (2021).

As características preditivas com maior importância no processo de predição pelas RNA's foram ligadas ao número de frutos comerciais totais e o número de frutos comerciais aos nove meses de idade com 34,38 e 31,50% de contribuição relativa, respectivamente (Figuras 6B). Já os caracteres de menores contribuições preditivas foram o número de frutos deformados por planta e número de nós sem frutos, avaliados aos 9 meses, com as respectivas contribuições de 0,8 e 0,81% (Figura 6B).

Estudos dessa natureza têm importância por avaliar um maior número de características, principalmente quando se torna rentável a exclusão de características para a redução de mão-de-obra (Paliwal & Kumar, 2011). Para testar a eficiência do modelo de predição, a produtividade estimada foi comparada com a produtividade observada por meio das variáveis de entrada destinadas à amostra de validação. Assim, o modelo de RNA obteve sucesso devido à similaridade entre os dados observados e previstos, testados com valores de  $R^2$  de alta significância 0,99 (Figura 7).

**Figura 7.** Representação gráfica da qualidade de predição da produtividade do mamoeiro CNPMF-L78 por redes neurais artificiais para a amostra de validação considerando todos os caracteres avaliados.



Fonte: Autores (2021).

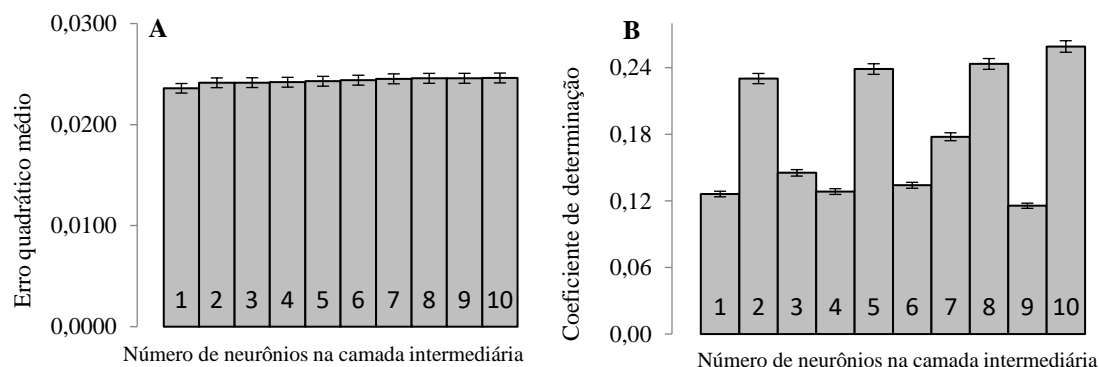
Ao selecionar modelo de predição, deve-se considerar a melhor relação entre o número de amostras de treinamento e o número de conexões intermediárias, sendo que este último deve ser superior a dois, pois favorece o menor erro relativo médio de validação (Soares, Robaina, Peiter & Russi, 2015; Guimarães et al., 2021). Similarmente por RNA, alcançou-se a estimativa da produção de frutos de pimenta com a arquitetura 8:10:1 precisão  $R^2 = 0,97$ , obtendo com o caractere número de frutos por planta maior determinação da produtividade dos frutos (Gholipour & Nadali, 2019).

Nestes modelos, foram consideradas características de fácil mensuração em campo, e que podem ser medidas no momento de inserção dos primeiros frutos comerciais, no caso deste estudo, cerca de nove meses após o plantio Donato et al. (2015), o que facilita o agendamento da colheita e a previsão de entrega pelo produtor (Guimarães et al., 2021).

### 3.3 Modelagem dos dados de RNA para as características de qualidade e rendimento do fruto

A modelagem de Redes Neurais Artificiais em mamoeiro CNPMF-L78, também foi empregada para medir os caracteres de qualidade e rendimento do fruto (CF) comprimento de fruto; (DF) diâmetro de fruto; (PF) peso de fruto; (FF) firmeza do fruto; (DCI) diâmetro da cavidade interna do fruto; (EP) espessura da polpa; (SS) sólidos solúveis. O funcionamento das arquiteturas de rede foi medido pelo erro quadrático médio (EQM) (Figuras 8A) e pelo coeficiente de determinação ( $R^2$ ) (Figuras 8B).

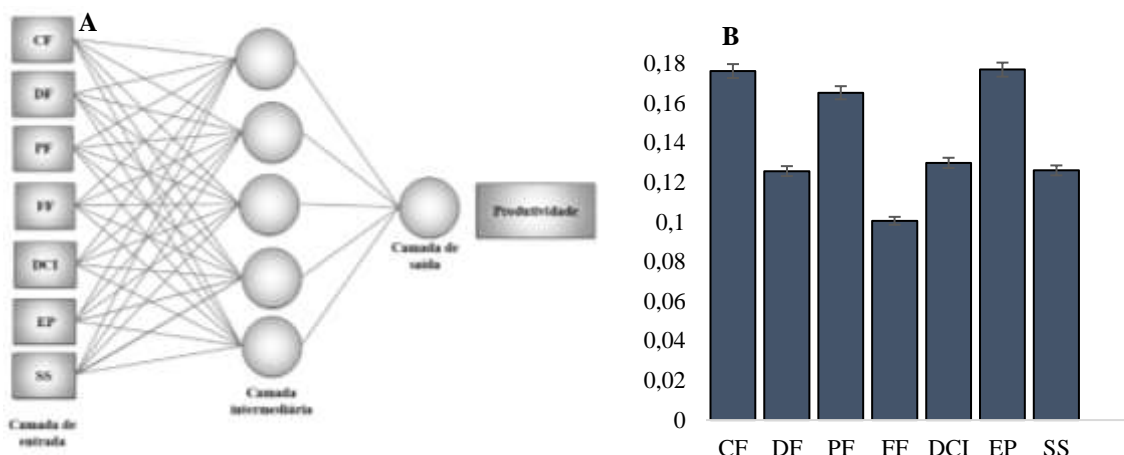
**Figura 8.** Estimativas do erro quadrático médio-EQM (A) e coeficiente de determinação-  $R^2$  (B) obtidos considerando diferentes números de neurônios na camada intermediária. Os desvios referem-se aos limites inferiores e superiores ao nível de 95% confiança obtidos por bootstrap com 10.000 simulações.



Fonte: Autores (2021).

Para variáveis relacionadas à qualidade e rendimento do fruto, constatou-se que o ajuste da rede com cinco neurônios na camada intermediária apresentou os menores erros do quadrado médio (EQM) (Figura 8A) associados aos maiores valores do coeficiente de determinação ( $R^2$ ) (Figura 8B). Contudo, foram encontrados resultados insatisfatórios, mesmo para o melhor ajuste, com o número médio de cinco neurônios na camada intermediária, o que proporcionou apenas 23,87% da eficiência preditiva (Figura 8B). A estrutura de redes com cinco neurônios na camada intermediária é apresentada na Figura 9.

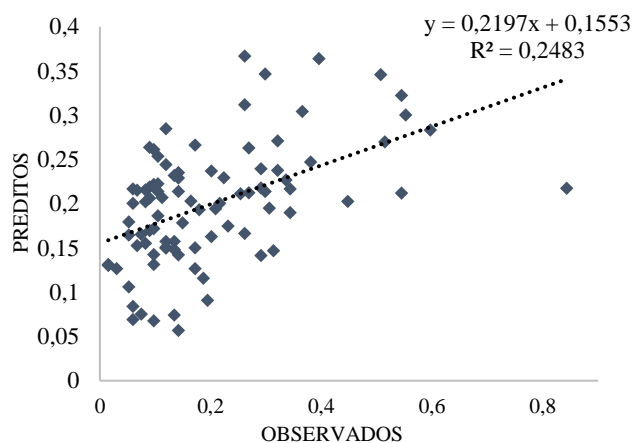
**Figura 9.** Topologia da melhor rede ajustada (A) e contribuição relativa (B), obtida pelo método de Garson (1991), dos parâmetros agrônômicos (CF- comprimento de fruto; DF- diâmetro de fruto; PF- peso de fruto; FF- firmeza do fruto; DCI- diâmetro da cavidade interna do fruto; EP- espessura da polpa e SS- sólidos solúveis) apresentados na camada de entrada para a predição da produtividade mamoeiro CNPMF-L78 por redes neurais artificiais. Os desvios referem-se aos intervalos de confiança de 95%, obtidos por bootstrap BCa com 10.000 simulações.



Fonte: Autores (2021).

Quanto à importância relativa das variáveis respostas, verificou-se que a variável de qualidade do fruto espessura da polpa teve maior importância (Figura 9B), com contribuição relativa de 17,68%. A maior expressão do parâmetro EP corrobora o moderado coeficiente de correlação com a produtividade do mamoeiro. Já a característica firmeza do fruto teve a menor contribuição relativa (10%). Foram ajustadas redes neurais com  $R^2$  de 24% para a amostra de validação (Figura 10).

**Figura 10:** Representação gráfica da qualidade de predição da produtividade do mamoeiro CNPMF-L78 por redes neurais artificiais para a amostra de validação considerando todos os caracteres avaliados.



Fonte: Autores (2021).

Mesmo com baixos valores de precisão ligados aos caracteres vegetativos e de qualidade de rendimento foi possível evidenciar medidas de importância relativa para cada variável. Destacando a utilidade das Redes Neurais Artificiais, como meio de avaliação a ser utilizada em identificação de variáveis com maior e menor contribuição, além disso tais variáveis podem proporcionar o reconhecimento de genótipos superiores, subsidiando estudos em melhoramento genético.

Estudos com alto desempenho da RNA na obtenção das variáveis de interesse agrônomo, têm sido difundidos em diversos ecossistemas (Gemici, Yucedag, Ozel & Imren, 2019; Vitor, Diniz, Morgante, Antônio, & Oliveira, 2019; Niedbala, 2019; Heidari, Rezaei & Rohani, 2020; Liu, Ma, Wang, Lu, & Lin, 2021). Dentre os motivos associados à eficiência dos modelos RNA, existe o menor erro preditivo, o que converte na maior eficiência preditiva do processo.

Ressalta-se no presente estudo que as variáveis associadas aos descritores de rendimento apresentaram maior poder de predição da produtividade do mamoeiro em detrimento das variáveis relacionadas às características vegetativas e reprodutivas do fruto. O que sinaliza para a construção do planejamento agrícola, a utilização de modelos com as características ligadas ao rendimento.

#### 4. Conclusão

As previsões de produtividade do mamoeiro CNPMF-L78 foram obtidas com por meio de redes neurais artificiais do tipo Perceptron Multicamadas.

Os resultados deste estudo comprovam a aplicação dos modelos de RNA's, na previsibilidade da produtividade em mamoeiro CNPMF-L78, tornando-se uma ferramenta eficiente e estratégica na tomada de decisões da produção agrícola, sobretudo no que diz respeito ao planejamento de pós colheita, armazenamento e transporte.

Os caracteres morfoagronômicos com maior contribuição relativa na previsão do rendimento do mamoeiro CNPMF-L78 foram associados ao número de frutos comerciais avaliado aos nove e quatorze meses de idade e o número de frutos comerciais totais.

A possibilidade de um planejamento agrícola através de uma tecnologia confiável e precisa, é um significativo avanço, por possibilitar o desenvolvimento de um sistema especialista em estima de produtividade, através de dados morfoagronômicos, demonstrando a praticidade em executar a técnica, além da vantagem de poder mensurar os dados diretamente no campo, sem necessidade de análise destrutiva. Modelagem computacional em estudos preditivos, demonstra ser uma ferramenta promissora para obtenção de ganhos na agricultura e em estudos futuros no melhoramento genético vegetal.

#### Agradecimentos

Os autores agradecem o apoio financeiro concedido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- CAPES pelo financiamento para o desenvolvimento deste estudo. Os autores também agradecem a Embrapa Mandioca e Fruticultura.

#### Referências

- Aggelopoulou, A. D., Bochtis, D., Fountas, S., Swain, K. C., Gemtos, T. A., & Nanos, G. D. (2011). Previsão de rendimento em pomares de maçã com base no processamento de imagens. *Agricultura de precisão*, 12 (3), 448-456.
- Azevedo, A. M. de, Silveira, V. A. da, Oliveira, C. M., Pedrosa, C. E., Lemos, V. T., Valadares, N. R., & Guimarães, A. G. (2019). Predição da área foliar em acerola por redes neurais e regressão múltipla. *Revista Agrária Acadêmica*, 2 (3), 96-105.
- Bergmeir, C. N., & Benítez Sánchez, J. M., (2012). Redes neurais em R usando o simulador de redes neurais de Stuttgart: RSNNs. *American Statistical Association*, 46 (7), 1-26.
- Donato, S. L. R., Arantes, A. D. M., Coelho, E., & Rodrigues, M. G. V. (2015). Considerações ecofisiológicas e estratégias de manejo da bananeira. In *Embrapa Mandioca e Fruticultura-Artigo em anais de congresso (ALICE)*. In: Simpósio Brasileiro Sobre Bananicultura, 8, 2015, Montes Claros. Palestras e resumos... Belo Horizonte: Epamig, 2015. 1, 45-110.
- Empresa Brasileira de Pesquisa Agropecuária (2020). *VII Plano Diretor da Embrapa: 2020–2030 / Embrapa*, (1), 31. autor.
- Estação Meteorológica da Embrapa Mandioca e Fruticultura. (2020). *Série Documentos (CNPMF)*. Recuperado de: <https://www.embrapa.br/busca-de-eventos/-/evento/150188/estacao-meteorologica-da-embrapa-mandioca-e-fruticultura#sec-locais>.
- Garson, G. D. (1991). Interpretando pesos de conexão de redes neurais. *International Journal of Artificial Intelligence and Expert Systems*, 6 (1), 47-51.

- Gemici, E., Yucedag, C., Ozel, H. B., & Imren, E. (2019). Predição da produção de cones em pomar de sementes clonais de pinheiro negro da Anatólia com rede neural artificial. *Ecologia Aplicada e Pesquisa Ambiental*, 17 (2), 2267-2273.
- Gil, C. A. (2017). Como Elaborar Projetos de Pesquisa. 6, 192. Atlas.
- Gholipour, M., & Nadali, F. (2019). Previsão da produção de frutos de pimenta por meio de rede neural artificial. *Scientia Horticulturae*, 250, 249-253.
- Guimarães, B. V. C., Donato, S. L. R., Aspiazú, I., & Azevedo, A. M. (2021). Predição da produtividade de bananeiras 'Prata-Anã' e 'BRS Platina' por redes neurais artificiais. *Pesquisa Agropecuária Tropical*, 51, e66008-e66008.
- Guimarães, B. V., Donato, S. L., Azevedo, A. M., Aspiazú, I., & Junior, S. (2018). Predição da produtividade da palma forrageira 'Gigante' por caracteres morfológicos e redes neurais artificiais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 22, 315-319.
- Heidari, P., Rezaei, M., & Rohani, A. (2020). Soft computing-based approach on prediction promising pistachio seedling base on leaf characteristics. *Scientia Horticulturae*, 274, 109647.
- Cidades, I. B. G. E. (2020). Conheça cidades e Estados do Brasil. *É o sistema agregador de informações do IBGE sobre os municípios e Estados do Brasil. Recuperado de <https://cidades.ibge.gov.br>*.
- Wrb, I. W. G. (2015). World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps. *World Soil resources reports*, (106), 192.
- Liu, L. W., Ma, X., Wang, Y. M., Lu, C. T., & Lin, W. S. (2021). Usando algoritmos de inteligência artificial para prever a taxa de crescimento do arroz (*Oryza sativa* L.) para agricultura de precisão. *Computadores e Eletrônicos na Agricultura*, 187, 106286.
- Mamann, A. T. W., da Silva, J. A. G., Binelo, M. O., Scremin, O. B., Kraisig, A. R., Carvalho, I. R., & Argenta, C. V. (2019). Inteligência artificial simulando a produtividade de grãos durante o desenvolvimento do trigo considerando indicadores biológicos e ambientais. *Journal of Agricultural Studies*, 7 (3), 197-212.
- Merladete, A. (2021). Digitalização da cadeia agrícola já é realidade Brasil afora. *Agrolink*. Recuperado de: [https://www.agrolink.com.br/noticias/digitalizacao-da-cadeia-agricola-ja-e-realidade-brasil-afora\\_454406.html](https://www.agrolink.com.br/noticias/digitalizacao-da-cadeia-agricola-ja-e-realidade-brasil-afora_454406.html).
- Mishra, R., Varun, P., Hans, A. L., & Saxena, S. (2021). Análise da diversidade de begomovírus que infectam o mamão e seus mecanismos de resistência. Em *Plant Virus-Host Interaction* (pp. 507-524). Academic Press.
- Niedbala, G. (2019). Modelo simples baseado em rede neural artificial para previsão antecipada e simulação de safra de colza de inverno. *Journal of Integrative Agriculture*, 18 (1) 54-61.
- Oliveira, A. M. G., Souza, L. D., & Coelho, E. F. (2010). Recomendações de calagem e adubação para mamoeiro. Embrapa Mandioca e Fruticultura. *Comunicado técnico 139*, ISSN- 1809- 502X.
- Paliwal, M., & Kumar, U. A. (2011). Avaliando a contribuição das variáveis na rede neural feedforward. *Applied Soft Computing*, 11 (4), 3690-3696.
- Pant, J., Pant, R. P., Singh, M. K., Singh, D. P., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*.
- Team, R. C. (2018). R: A language and environment for statistical computing; 2018.
- Santos, H. G., Jacomine, P. K. T., Dos Anjos, L. H. C., Oliveira, V. A., Lumbreras, J. F., Coelho, M. R., & Cunha, T. J. F. (2018). *Sistema brasileiro de classificação de solos*. Brasília, DF: Embrapa, 2018.
- Santos, J. F. S. (2020). Produção de soja, desigualdades no campo e mudanças climáticas na região do Matopiba. *DRd-Desenvolvimento Regional em debate*, 10, 535-561.
- Soares, F. C., Robaina, A. D., Peiter, M. X., & Russi, J. L. (2015). Predição da produtividade da cultura do milho utilizando rede neural artificial. *Ciência Rural*, 45, 1987-1993.
- Soares, J. D. R., Pasqual, M., Lacerda, W. S., Silva, S. O., & Donato, S. L. R. (2014). Comparação de técnicas utilizadas na predição da produtividade em bananeira. *Scientia Horticulturae*, 167 (1), 84-90.
- Smith, D. W. (2014). Soil survey staff: keys to soil taxonomy. *Washington: Natural Resources Conservation Service*.
- Vitor, A.B., Diniz, R. P., Morgante, C. V., Antônio, R. P., & de Oliveira, E. J. (2019). Modelos precoces de predição da produtividade de raízes de mandioca em diferentes regimes hídricos. *Field Crops Research*, 239, 149-158.