

UNIVERSIDADE ESTADUAL DE GOIÁS
CÂMPUS CENTRAL SEDE ANÁPOLIS - CET
PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM ENGENHARIA AGRÍCOLA
GABRIELLA ANDREZZA MEIRELES CAMPOS

**CLASSIFICAÇÃO DO ESTÁDIO DE MATURAÇÃO DE CAGAITAS UTILIZANDO
ESTATÍSTICA MULTIVARIADA E REDES NEURAIS ARTIFICIAIS**

ANÁPOLIS-GO

2021

GABRIELLA ANDREZZA MEIRELES CAMPOS

**CLASSIFICAÇÃO DO ESTÁDIO DE MATURAÇÃO DE CAGAITAS UTILIZANDO
ESTATÍSTICA MULTIVARIADA E REDES NEURAIS ARTIFICIAIS**

Dissertação apresentada à Universidade Estadual de Goiás - UEG, Câmpus Central Sede Anápolis – CET, como parte das exigências do Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola, para obtenção do título de Mestre em Engenharia Agrícola.

Área de concentração CNPQ: Inteligência Artificial e Processamento e Armazenamento de Produtos Agrícolas

Orientador: Prof. Dr. Francisco Ramos de Melo

Co-orientador: Prof. Dr. André José de Campos

ANÁPOLIS-GO

2021



TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL (BDTD)

Na qualidade de titular dos direitos de autor, autorizo a Universidade Estadual de Goiás a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UEG), regulamentada pela Resolução, **CsA n.1087/2019** sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

Dados do autor (a)

Nome Completo: Gabriella Andreza Meireles Campos

E-mail: g_abicampos@hotmail.com

Dados do trabalho

Título: **CLASSIFICAÇÃO DO ESTÁDIO DE MATURAÇÃO DE CAGAITAS UTILIZANDO ESTATÍSTICA MULTIVARIADA E REDES NEURAIS ARTIFICIAIS**

Data da Defesa: 16 / 12 / 2021

Tipo

() Tese (X) Dissertação

Curso/Programa: Programa de Pós-Graduação *Stricto sensu* em Engenharia Agrícola

Concorda com a liberação documento

[X] SIM

[] NÃO¹

Anápolis , 14 / 02 / 2021

Local

Data

Assinatura do autor (a)

Assinatura do orientador (a)

1 Casos de impedimento:

- Período de embargo é de um ano a partir da data de defesa
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da Dissertação/Tese em livro.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UEG
com os dados fornecidos pelo(a) autor(a).

MC198 MEIRELES CAMPOS, GABRIELLA ANDREZZA
c CLASSIFICAÇÃO DO ESTÁDIO DE MATURAÇÃO DE CAGAITAS
UTILIZANDO ESTATÍSTICA MULTIVARIADA E REDES NEURAIS
ARTIFICIAIS / GABRIELLA ANDREZZA MEIRELES CAMPOS;
orientador FRANCISCO RAMOS DE MELO; co-orientador ANDRÉ
JOSÉ DE CAMPOS. -- Anápolis, 2021.
57 p.

Dissertação (Mestrado - Programa de Pós-Graduação
Mestrado Acadêmico em Engenharia Agrícola) -- Câmpus
Central - Sede: Anápolis - CET, Universidade Estadual
de Goiás, 2021.

1. Processamento digital de imagens. 2. Análise de
componentes principais. 3. Análise de agrupamentos. 4.
Eugenia dysenterica DC. 5. Visão computacional. I.
RAMOS DE MELO, FRANCISCO, orient. II. JOSÉ DE CAMPOS,
ANDRÉ, co-orient. III. Título.

GABRIELLA ANDREZZA MEIRELES CAMPOS

**CLASSIFICAÇÃO DO ESTÁDIO DE MATURAÇÃO DE CAGAITAS UTILIZANDO
ESTATÍSTICA MULTIVARIADA E REDES NEURAIS ARTIFICIAIS**

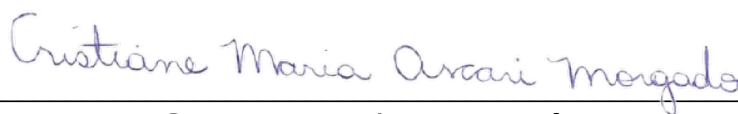
Dissertação defendida ao Programa de Pós-Graduação
Stricto Sensu em Engenharia Agrícola, Câmpus Central
Sede Anápolis – CET, Universidade Estadual de Goiás,
para obtenção do título de mestre.

Aprovada: Anápolis, 16 de Dezembro de 2021.

Banca examinadora



Prof. Dr. Francisco Ramos de Melo
Universidade Estadual de Goiás



Dra. Cristiane Maria Ascari Morgado
Universidade Estadual de Goiás



Prof. Dr. Luiz Fernando Batista Loja
Instituto Federal de Goiás

ANÁPOLIS – GO
2021

AGRADECIMENTOS

A Deus, por me conceder forças, sabedoria e fé, me permitindo chegar até aqui.

Aos meus pais Filinto Pereira Campos e Míriam Luciana Meireles da Costa, pela base, sustento e fortaleza. Sou muito grata por todo amor e cuidado. Amo muito vocês!

Neste trabalho, em especial ao meu pai, por me acompanhar nos experimentos, buscar de madrugada no laboratório e me apoiar dia após dia.

Aos meus irmãos André Felipe de O. Campos (*in memorian*), Kaique Everton Felipe de O. Campos, Matheus Eduardo M. Campos, Maryah Eduarda R. Campos, Manuella R. Campos, Eloah Luiza R. Campos, que são minha motivação e sinônimo de amor.

Aos meus avós Inês Maria Campos, Marilene Aparecida Meireles e Getulino Meireles, por todo incentivo, carinho e cuidado. Vocês são minha preciosidade!

Ao meu padrasto João Wilian Xavier da Costa por acreditar no meu potencial e me incentivar a ser melhor. Obrigada tio Wilian.

À minha madrastra Danyelle Rodrigues Geraldo por todo cuidado e carinho.

Ao meu orientador Prof. Dr. Francisco Ramos de Melo, por todos os ensinamentos, pela paciência e confiança. Sou muito grata pelo privilégio da sua orientação. Muito obrigada Professor!

Ao meu co-orientador Prof. Dr. André José de Campos, por mais uma vez me auxiliar, ensinar e apoiar. Aprendi muito com o senhor em todos esses anos. Muito obrigada Professor!

À Profa. Dra. Eli Regina B. de Souza e ao Sr. Natal, por nos receber tão bem e me aconselhar no experimento. Muito obrigada!

Às minhas amigas de longa data, que foram os meus maiores presentes da UEG, Amanda Cristina Elias Lourenço, Maressa Silva Moreira Carvalho e Maria Rita Teixeira da Silva. Sou muito grata por ter conhecido vocês. Obrigada pela parceria, conselhos e paciência.

A quem se tornou um grande amigo e me auxiliou muito nesse período, Frank Freire Capuchinho. Sou muito grata por toda ajuda. Você é excepcional. Muito obrigada!

À Sielly Lobo Pereira por me auxiliar nas análises de laboratório. Muito obrigada!

Ao Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola da Universidade Estadual de Goiás, pela oportunidade. Foram anos de muito aprendizado.

A Capes pelo auxílio financeiro para a realização desta pesquisa.

RESUMO

O presente trabalho teve como objetivo avaliar a classificação de diferentes estádios de maturação da cagaita (*Eugenia dysenterica* DC.) utilizando-se a estatística multivariada e redes neurais artificiais. Foram coletadas 120 amostras, possuindo 3 frutos cada amostra, a cada 3 dias em 5 diferentes dias após antese (DAA) (28, 31, 34, 37 e 40 DAA), totalizando 600 amostras. Para a análise utilizando estatística multivariada foram avaliadas características, como: peso, diâmetro longitudinal, diâmetro transversal, área, coloração (luminosidade, °hue e croma), firmeza da polpa e da casca, sólidos solúveis (SS), potencial hidrogeniônico (pH), acidez titulável (AT) e índice de maturação (IM). Os dados obtidos foram padronizados e submetidos a análise de componentes principais (ACP) e análise de agrupamento (*Clusters*), e para a validação dos métodos estabelecidos, calculou-se o coeficiente de correlação cofenética (CCC). Na classificação do estágio de maturação de cagaita utilizando redes neurais artificiais, foram obtidas imagens de cada fruto de cagaita em cada dia após a antese analisado, totalizando 600 imagens. As imagens foram convertidas para tons de cinzas e segmentadas. Posteriormente foram extraídas informações dos modelos de cor (RGB, HSV, L, °hue e croma) e os atributos geométricos (comprimento e área) das imagens dos frutos. Os parâmetros extraídos das imagens foram utilizados como entradas das redes neurais artificiais. As saídas foram estabelecidas base aos dias após antese analisados, em que, cada um passou a corresponder a uma classe (1, 2, 3, 4 e 5), respectivamente. Foram treinadas 20 redes, alterando, em cada uma delas, a quantidade de nodos na camada oculta (3 a 12 nodos) e as funções de otimização (adam e nadam). Foi calculada a precisão e perda (erro) de cada rede com o intuito de determinar qual foi o modelo que apresentou melhor desempenho no treinamento. Para avaliar o desempenho do melhor modelo foram calculados 4 parâmetros estatísticos, sendo eles: acurácia, precisão, recall e F-score. Para a primeira etapa da análise, por meio da análise de componentes principais, os CPs 1, 2 e 3 explicaram 98,44% da variabilidade dos dados. A análise de agrupamento (*Clusters*) explicou o estágio de maturação das cagaitas por meio do comportamento das variáveis físicas, físico-químicas e químicas em diferentes estádios de maturação, dividindo esses dias em 3 grupos (grupo 1, grupo 2 e grupo 3), onde no grupo 1 encontram-se 28 e 31 DAA, o grupo 2, 40 DAA e o grupo 3, 34 e 37 DAA. As análises de estatística multivariada apresentaram um coeficiente de correlação cofenética de 0,8234. Já para a segunda parte, utilizando as RNAs, o modelo da rede que apresentou a melhor precisão, foi o com 9 nodos na camada oculta e otimizador “adam”, mostrando a precisão de 57,353%. No relatório das métricas de

classificação da rede para determinação de diferentes estádios de maturação, a maior acurácia, precisão e R-score é da Classe 1, 97,1%, 75% e 75%, respectivamente e o melhor recall da Classe 2 (88,9%). Verificando o cálculo do coeficiente de correlação cofenética, quanto à adequação do agrupamento e a precisão de definição do melhor modelo de RNAs, foi possível concluir que o melhor método para a classificação de diferentes estádios de maturação de cagaita neste trabalho foi o de estatística multivariada, embora a classificação por meio das redes neurais artificiais também tenha sido eficiente.

Palavras-chave: Análise de agrupamentos. Análise de componentes principais. *Eugenia dysenterica* DC. Processamento digital de imagens. Visão computacional.

ABSTRACT

The present work aimed to evaluate the classification of different maturation stages of cagaita (*Dysenteric eugenics* DC.) using multivariate statistics and artificial neural networks. 120 samples were collected, with 3 fruits each sample, every 3 days on 5 different days after anthesis (DAA) (28, 31, 34, 37 and 40 DAA), totaling 600 samples. For the analysis using multivariate statistics, characteristics were evaluated, such as: weight, longitudinal diameter, transversal diameter, area, color (luminosity, °hue and chroma), pulp and skin firmness, soluble solids (SS), hydrogenion potential (pH), titratable acidity (AT) and maturation index (IM). The data obtained were standardized and submitted to principal components analysis (PCA) and cluster analysis (Clusters), and to validate the established methods, the cophenetic correlation coefficient (CCC) was calculated. In the classification of the cagaita maturation stage using artificial neural networks, images of each cagaita fruit were obtained on each day after the anthesis analyzed, totaling 600 images. The images were converted to grayscale and segmented. Subsequently, information was extracted from the color models (RGB, HSV, L, °hue and chroma) and the geometric attributes (length and area) of the fruit images. The parameters extracted from the images were used as inputs to the artificial neural networks. The outputs were established based on the days after anthesis and analyzed, in which each one came to correspond to a class (1, 2, 3, 4 and 5), respectively. 20 networks were trained, changing, in each of them, the number of nodes in the hidden layer (3 to 12 nodes) and the optimization functions (adam and nadam). The precision and loss (error) of each network were calculated in order to determine which model presented the best performance in the training. To evaluate the performance of the best model, 4 statistical parameters were calculated, namely: accuracy, precision, recall and F-score. For the first stage of the analysis, through principal component analysis, PCs 1, 2 and 3 explained 98,44% of the data variability. Cluster analysis (Clusters) explained the maturation stage of cagaitas through the behavior of physical, physicochemical and chemical variables at different stages of maturation, dividing these days into 3 groups (group 1, group 2 and group 3), where in group 1 we find if 28 and 31 DAA, group 2, 40 DAA and group 3, 34 and 37 DAA. Multivariate statistical analyzes showed a cophenetic correlation coefficient of 0,8234. For the second part, using the ANNs, the network model that presented the best precision was the one with 9 nodes in the hidden layer and “adam” optimizer, showing the accuracy of 57,353%. In the network classification metrics report for determining different stages of maturation, the highest accuracy, precision and R-score are from Class 1, 97,1%, 75%

and 75%, respectively, and the best recall from Class 2 (88,9%). Checking the calculation of the cophenetic correlation coefficient, regarding the adequacy of the grouping and the precision of definition of the best model of ANNs, it was possible to conclude that the best method for the classification of different stages of maturation of cagaita in this work was the multivariate statistics, although the classification through artificial neural networks was also efficient.

Keywords: Cluster analysis. Principal component analysis. Dysenteric eugenics DC. Digital image processing. Computer vision.

LISTA DE FIGURAS

Figura 1 - Cagaiteiras localizadas na coleção na Escola de Agronomia da Universidade Federal de Goiás.	13
Figura 2 - Cagaiteira com flores.	14
Figura 3 - Frutos da cagaiteira.	15
Figura 4 - Cagaitas acondicionadas em embalagem de polietileno de baixa densidade (PEBD), com fecho tipo zip.	22
Figura 5 - Cagaitas (a) imersas em água potável; e (b) secando em temperatura ambiente.	23
Figura 6 - Etapas de processamento digital de imagens utilizadas para a classificação de diferentes estádios de maturação da cagaita.	27
Figura 7 - Imagens dos primeiros frutos de cada dia após antese. Imagens originais (a); convertidas em tons de cinza (b); e segmentadas (c). 28 (a1, b1, c1), 31 (a2, b2, c2), 34 (a3, b3, c3), 37 (a4, b4, c4) e 40 DAA (a5, b5, c5).	28
Figura 8 - Dendrograma da análise de agrupamento das cagaitas em diferentes dias de antese (DAA).	35
Figura 9 - Análise dos componentes principais das características físicas, físico-químicas e químicas das cagaitas em diferentes estádios de maturação. Em que: L = Luminosidade; DL = Diâmetro Longitudinal; DT = Diâmetro Transversal; AT = Acidez Titulável; SS = Sólidos Solúveis; FP = Firmeza da Polpa; FC = Firmeza da Casca e IM = Índice de Maturação.	38
Figura 10 - Rede Multi Layer Perceptron (MLP) do melhor modelo.	40
Figura 11 - Matriz confusão para a previsão de 5 diferentes estádios de maturação de cagaita, com otimizador “adam” e 9 neurônios na camada oculta.	41

LISTA DE TABELAS

Tabela 1 - Componentes principais (CP's), autovalores, variância explicada (%) e variância total (%) das características físicas, físico-químicas e químicas das cagaitas em diferentes dias após antese (28, 31, 34, 37 e 40 DAA).....	32
Tabela 2 - Correlação das características físicas, físico-químicas e químicas das cagaitas em diferentes dias após antese (28, 31, 34, 37 e 40 DAA), com as 3 primeiras componentes principais.	33
Tabela 3 - Perda e precisão dos dados de teste para diferentes quantidades de neurônios na camada oculta.	39
Tabela 4 - Matriz confusão binária para a previsão de 5 diferentes estádios de maturação de cagaita, com otimizador “adam” e 9 neurônios na camada oculta. Em que: C1 = Classe 1, C2 = Classe 2, C3 = Classe 3, C4 = Classe 4 e C5 = Classe 5.	43
Tabela 5 - Relatório de classificação da rede para classificação de 5 diferentes estádios de maturação de cagaita.	43

SUMÁRIO

1 INTRODUÇÃO	11
2 REVISÃO BIBLIOGRÁFICA	13
2.1 Caracterização da cagaita.....	13
2.2 Diferentes estádios de maturação de frutos	16
2.2.1 Classificação dos estádios de maturação utilizando características físicas e químicas ...	16
2.2.2 Classificação dos estádios de maturação utilizando imagens digitais	17
2.3 Contextualizando a estatística multivariada	18
2.3.1 Estatística multivariada na classificação de frutos	19
2.4 Contextualizando as redes neurais artificiais	19
2.4.1 Redes neurais na classificação de frutos.....	21
3 MATERIAL E MÉTODOS	21
3.1 O local e suas características	21
3.2 Coleta das amostras.....	22
3.3 Delineamento experimental	23
3.4 Variáveis analisadas	23
3.4.1 Peso dos frutos.....	23
3.4.2 Diâmetro (transversal e longitudinal)	23
3.4.3 Área.....	24
3.4.4 Coloração (Luminosidade, °hue e croma)	24
3.4.5 Firmeza (casca e polpa)	25
3.4.6 Sólidos solúveis.....	25
3.4.7 Potencial hidrogeniônico	25
3.4.8 Acidez titulável.....	25
3.4.9 Índice de maturação.....	25
3.5 Estatística Multivariada.....	26
3.5.1 Análise de componentes principais (ACP).....	26
3.5.2 Análise de agrupamento (<i>Cluster</i>).....	26
3.6 Etapas sistemáticas para o processamento digital de imagens.....	26
3.6.1 Aquisição das imagens e calibração do sistema	27
3.6.2 Segmentação das imagens	27
3.6.3 Extração dos atributos de cor	28

3.6.4 Extração de atributos geométricos.....	29
3.7 Classificação dos estádios de maturação das cagaitas por meio de redes neurais artificiais.....	29
3.7.1 Separação dos conjuntos (treino, validação e teste)	29
3.7.2 Construindo, compilando e ajustando o modelo	30
3.7.3 Processo de avaliação do sistema de classificação	30
3.7.4 Parâmetros estatísticos para a avaliação do desempenho	30
3.8 Especificação do software e da máquina	31
4 RESULTADOS E DISCUSSÃO	32
4.1 Estatística Multivariada.....	32
4.1.1 Análise de componentes principais e análise de agrupamento das características físicas, físico-químicas e químicas das cagaitas em diferentes dias após a antese.....	32
4.2 Redes Neurais Artificiais	39
4.2.1 Determinação do modelo com melhor otimizador, número de lotes e neurônios	39
4.2.2 Fase de confirmação da rede de classificação de diferentes estádios de maturação da cagaita.....	41
4.3 Comparação de métodos (Estatística Multivariada e Redes Neurais Artificiais)	44
5 CONCLUSÃO.....	45
REFERÊNCIAS BIBLIOGRÁFICAS	46

1 INTRODUÇÃO

O Cerrado é o segundo maior bioma brasileiro e apresenta grande diversidade de frutos. Esses frutos são uma importante fonte de alimento de interesse tanto para a alimentação *in natura* quanto para a industrialização (REIS e SCHMIELE, 2019; MOTA, 2013). No entanto, as frutas nativas do Cerrado são pouco encontradas na literatura, embora sejam muito consumidas na região. A divulgação de estudos sobre essas frutas pode atribuir valor comercial, já que as mesmas são desvalorizadas (ALMEIDA et al., 2018). Dentre essas frutas, têm-se a cagaita.

A cagaita é uma fruta nativa do Cerrado de alta perecibilidade. Seu desenvolvimento e amadurecimento dos frutos ocorrem em um curto período de 30 a 40 dias, apresentando a coloração da casca verde quando imaturos e amarelo pálido quando maduros (BRAZ et al., 2019; MARTINOTTO et al., 2008; SOUZA et al., 2008). Os frutos possuem comportamento climatérico, demonstrando o aumento da atividade respiratória de forma rápida, em uma etapa do seu ciclo, tendendo ao rápido amadurecimento dos frutos, estando eles na planta ou fora dela, o que pode induzir, a deterioração dos frutos (CHITARRA e CHITARRA, 2005).

A alternativa para se evitar o desperdício dos frutos é conhecendo seus diferentes estádios de maturação, entendendo suas características ao longo destes estádios e determinando a colheita no estágio correto, já que a colheita no ponto adequado de maturação é o principal determinante para a qualidade das frutas (SILVA, 2021; CEAGESP, 2016). Além disso, o conhecimento das etapas de desenvolvimento dos frutos pode auxiliar na determinação das práticas culturais, revelando os períodos críticos no desenvolvimento dos frutos que possibilitam a produção com qualidade (SALOMÃO et al., 2006).

Entendendo a complexidade das alterações fisiológicas que os frutos sofrem ao longo dos estádios de maturação, uma sugestão de avaliação é a determinação com base nas características físico-químicas, sendo que estas são responsáveis pelas mudanças sensoriais e físicas dos frutos (LOBO e PAULL, 2017; STEINGASS et al., 2015). De acordo com Coelho et al. (2010), ao longo do amadurecimento dos frutos é possível observar mudanças em suas características, como a coloração, além das alterações químicas como sólidos solúveis (SS), potencial hidrogeniônico (pH), acidez titulável (AT), índice de maturação (relação SS/AT), rendimento em suco, vitamina C, entre outros, sendo que esses parâmetros são um indicativo para o ponto de colheita.

Neste sentido, visando verificar a interação entre as variáveis e seus comportamentos ao longo de determinados dias, um método muito utilizado é a estatística multivariada. A análise multivariada é usada para avaliar a interação da variância dentro dos grupos de dados obtidos, possibilitando estabelecer as variáveis de maior influência no processo de maturação dos frutos (SILVA et al., 2021; SANTOS, 2011).

Moreno et al. (2021), avaliando as mudanças físico-químicas e fisiológicas durante o amadurecimento da banana (*Musaceae*) cultivada na Colômbia, utilizou estatística multivariada, mais precisamente a análise de componentes principais (ACP), para determinar quais variáveis foram responsáveis pela maior variância dos dados. Silva et al. (2021) também utilizou análise multivariada de componentes principais, em seu estudo sobre a comparação de parâmetros de qualidade obtidos por meio de métodos destrutivos e ópticos na classificação de tomates. Lawson et al. (2019), avaliando a caracterização de mangas do Sudeste Asiático (*Mangifera indica* L) de acordo com seus atributos físico-químicos, utilizou análise multivariada para caracterizar o processo de amadurecimento.

Um outro método que tem sido muito utilizado nos últimos anos, visando a não destruição, bem como evitar erros causados por classificações não confiáveis de frutos é o de processamento digital de imagens combinada com redes neurais artificiais. Essa tecnologia tende a garantir a consistência do nível de maturação da fruta, possibilitando a diminuição das reclamações dos consumidores. A visão computacional permite a classificação do grau de maturação dos frutos com o auxílio de computadores (SIDEHABI et al., 2018).

No processo de classificação artificial de frutos, têm-se utilizado modelos que permitem que seja estabelecido o relacionamento entre a entrada de padrões de classificação (entradas), análise/processamento dessa informação e convergência para uma saída de classificação definida (saída). A rede neural tem como função aprender a reconhecer padrões de entrada e definir a saída segundo classes definidas (RAMOS, 2003).

Kheiralipour e Pormah (2017) em seu estudo, apresentando novos recursos para classificação de pepino utilizando técnica de processamento de imagem e redes neurais artificiais apresentaram dois novos recursos para distinguir duas formas do pepino e após avaliação de diferentes modelos de redes neurais artificiais, o melhor modelo atingiu a acurácia de 97,1%. Mazen e Nashat (2019) avaliando a classificação de bananas usando uma rede neural artificial identificaram que o modelo que foi proposto atingiu uma alta acurácia e se mostrou melhor do que as outras técnicas utilizadas.

Diante do exposto, o presente trabalho teve como objetivo avaliar diferentes estádios de maturação de cagaitas (*Eugenia dysenterica* DC.) utilizando-se a estatística multivariada e redes neurais artificiais.

2 REVISÃO BIBLIOGRÁFICA

2.1 Caracterização da cagaita

A cagaita (*Eugenia dysenterica* DC.) pertence à família das Myrtaceae, sendo esta a mesma família da jabuticaba, pitanga, goiaba, gariobas, araçá, entre outras e é uma das espécies nativas do Cerrado podendo ser encontrada, principalmente, nos estados de: Goiás, Tocantins, Bahia, Mato Grosso, Minas Gerais, Distrito Federal, São Paulo, Ceará, Pernambuco, Maranhão e Piauí (PERFEITO et al., 2017; DAZA et al., 2016).

A árvore de cagaita (Figura 1) apresenta grande adaptação a solos muito pobres, ambientes com escassez de chuva e resistência ao fogo. Pode chegar até 10 m de altura, apesar de sua média estar entre 4 a 8 m, seu tronco possui muitos sulcos, com presença de cortiça (SOUZA et al., 2013).

Figura 1 - Cagaiteiras localizadas na coleção na Escola de Agronomia da Universidade Federal de Goiás.



Fonte: Próprio autor (2020).

Sua expansão ocorre de forma sexuada, sendo possível a ocorrência tanto de autofecundação quanto de fecundação cruzada (polinização) e esse processo geralmente ocorre no período da manhã (SILVA, 1999; PROENÇA e GIBBS, 1994).

O surgimento de flores (Figura 2) normalmente ocorre entre o mês de julho e agosto, neste mesmo período é comum o surgimento de folhas novas (BRITO et al., 2003). O surgimento de novas folhas ocorre antes da queda de todas as folhas senescentes, no período das primeiras chuvas (CAMILO et al., 2013).

Figura 2 - Cagaiteira com flores.



Fonte: Próprio autor (2020).

A produção dos frutos ocorre de forma abundante e é fonte de alimento para diversos animais silvestres. Um dos principais responsáveis pela dispersão das sementes são os morcegos, que levam o fruto para se alimentar em diversos pontos diferentes em torno da planta mãe (CHAVES e TELLES, 2018).

A frutificação ocorre nos períodos de setembro a novembro (BRITO et al., 2003). Seus frutos (Figura 3) possuem formato globoso e casca com coloração amarela, quando maduros. Esses frutos possuem de uma a três sementes e estas são revestidas com a polpa que possui o sabor levemente ácido (SILVA, 2020). Pesam entre 14 a 20 g, com o comprimento de

3 a 4 cm, diâmetro de 3 a 5 cm e média de 3 a 4 grumos (NAVES et al. 2002). Os frutos podem ser consumidos tanto *in natura* quanto processados em forma de sorvetes, bebidas ou geleias. Estas diferentes formas têm se tornando meio de renda para pequenos agricultores do Cerrado (ALMEIDA JÚNIOR, 2014).

Figura 3 - Frutos da cagaiteira.



Fonte: Prof. Dr. Eli Regina B. de Souza (2020).

Tanto seu nome científico (*Eugenia dysenterica* DC.) quanto seu nome vulgar (cagaita) remete à propriedade laxativa que os frutos da espécie possuem. Este fato ocorre quando o fruto é consumido maduro em excesso, ou quando estão fermentados, sendo que, quando consumido perto de amadurecer, o mesmo não causa dano algum (MORAIS, 2020; CHAVES e TELLES, 2010).

Por apresentar alta densidade e possuir boa produção, é um fruto que possui grande potencial em sistemas de produção agrícola (SCARIOT e RIBEIRO, 2015), além de alto potencial para industrialização visto que os frutos apresentam baixo valor energético, baixo teor de lipídios, carboidratos, altos teores de vitamina C, contribuindo beneficemente para saúde (ALVARENGA et al., 2019). Também é um fruto que apresenta alto teor de ferro e grande capacidade antioxidante (NASCIMENTO et al., 2020).

A cagaita é um fruto que apresenta comportamento climatérico, possuindo altas taxas de respiração e produção de etileno, aumentando parâmetros como os sólidos solúveis e reduzindo a firmeza da fruta no seu período de amadurecimento (FERREIRA et al., 2019).

Os frutos podem sofrer alterações de suas características físicas de uma planta para outra, o que, para a industrialização poderia ser necessário uma seleção dos frutos, visando a uniformização. Existe uma relação positiva entre a altura e comprimento dos frutos com a massa, significando que quanto maior os frutos, maior a massa. Entretanto, esta relação não

está atrelada ao número de sementes, pois nem sempre o tamanho e massa dos frutos estará relacionado ao número de sementes. O tamanho dos frutos também não possui relação com a coloração (CAMILO et al., 2014).

Por possuir grande potencial para consumo *in natura* e processamento industrial, em suas propriedades, as cagaitas podem ser empregadas em muitas aplicações, podendo ser utilizado desde os seus frutos, sementes, folhas (para chás) e madeiras. Porém, poucos estudos são atribuídos ao conhecimento dessas propriedades e por esse motivo se faz necessário e importante a realização de estudos para o melhor conhecimento dessa espécie (MORAIS, 2020). A maior parte dos estudos voltados à cagaita estão no Brasil, sendo possível observar o grande investimento das instituições de ensino para impulsionar o conhecimento da espécie (SILVA-RODRIGUES, 2021).

2.2 Diferentes estádios de maturação de frutos

A maturação dos frutos é a fase em que os mesmos passam por um processo de desenvolvimento, onde ocorre uma sequência de transformações, sendo elas fisiológicas, bioquímicas e estruturais. Os frutos saem de um estágio incompleto (verde) para atingir a fase adequada para o consumo (maduro) (CHITARRA e CHITARRA, 2005).

Uma série de fatores influenciam o prolongamento da vida útil dos frutos, sendo que, a colheita dos frutos no estágio de maturação adequado é uma das mais utilizadas, já que esse fator é o que determina o potencial de conservação pós-colheita. Quando os frutos são colhidos antes do tempo (verdes), possuem baixa qualidade sensorial e são mais suscetíveis a desidratação e mudanças fisiológicas. Em contrapartida, quando os frutos são colhidos muito maduros, entram rapidamente em senescência (DREHMER e AMARANTE; 2008; WILLS et al., 1998).

Portanto, estudos sobre a determinação do estágio de maturação de um determinado fruto pode dar suporte a futuros estudos, com o intuito de aumentar seu consumo, aproveitamento, diminuir o desperdício e fortalecer a comercialização (MELO et al., 2015).

2.2.1 Classificação dos estádios de maturação utilizando características físicas e químicas

Uma das principais formas de determinar os diferentes estádios de maturação dos frutos de forma precisa é por meio do conhecimento do comportamento físico e químico dos frutos. Essas características podem indicar se os frutos estão apropriados para o consumo *in natura* e para a industrialização (RAMPAZZO et al. 2018).

Para a determinação do tempo ideal de colheita dos frutos em sua maturação fisiológica, é necessário verificar se ocorreu alterações químicas no fruto, como o aumento na taxa de respiração, no teor de sólidos solúveis e no pH (VIANNA-SILVA et al., 2008). Um outro fator muito utilizado para essa verificação é a relação entre sólidos solúveis (SS) e acidez titulável (AT), já que esses parâmetros estão associados ao sabor do fruto. Valores elevados indicam que os frutos se encontram mais doces do que ácidos (AROUCHA, ET AL. 2012).

Farias et al. (2020) verificaram a ocorrência de redução da firmeza de araçá, em quase a metade, de um estágio para outro. Aumento dos teores de sólidos solúveis, do índice de maturação e bem como também dos açúcares solúveis. As características físicas, químicas e sensoriais são responsáveis por tornar os frutos mais atrativos para os consumidores e essas características são desenvolvidas no período de amadurecimento dos frutos (CHAVES NETO et al., 2020).

Frutos com comportamento climatérico possuem vida pós-colheita mais curta e apresentam alterações físicas e químicas elevadas durante o amadurecimento. Isso ocorre devido ao aumento da taxa respiratória e produção de etileno, alterando as características físico-químicas (NERIS et al., 2018). Boas práticas, como armazenamento, podem prolongar a vida pós-colheita, no entanto, são necessários estudos buscando fornecer alternativas mais acessíveis de maturação para pequenos produtores e, também, para os consumidores (SANTOS, et al. 2019).

2.2.2 Classificação dos estádios de maturação utilizando imagens digitais

A maioria dos métodos para determinação dos diferentes estádios de maturação de frutos são destrutivos, o que gera prejuízos. Por essa razão faz-se necessária a aplicação de técnicas não destrutivas para a classificação da maturação de frutos (COSTA, 2019). Utilizar processamento digital de imagens para a seleção e classificação de frutas, tem um futuro promissor, podendo ser um grande aliado de indústrias (SANTIAGO et al., 2019).

Os sistemas por classificação de imagens além de possuírem benefícios de serem autocalibráveis, imunes a aspectos intrínsecos à natureza humana como estresse, emoções, fadiga e julgamentos subjetivos, é um método exato que avalia individualmente cada fruto. A visão de máquinas aliada à inteligência artificial já é uma técnica que tem sido bem utilizada na classificação de frutos, visando obter melhor precisão na padronização por meio das imagens (SANCHES E LINO, 2010; COSTA, 2018).

A seleção das frutas acontece, na maioria das vezes, baseando-se em características visuais e essas seleções acontecem manualmente. O processamento e análise digital de imagens, são recomendáveis para esse caso, onde possam ser evitados danos causados pelo contato, bem como erros que possam ocorrer por meio das seleções humanas. As pesquisas utilizando essas técnicas se mostram apropriadas para obter informações sobre os frutos sem que tenha necessidade de contato físico, bem como tornando o processo de classificação muito mais rápido (SILVA, 2015).

2.3 Contextualizando a estatística multivariada

A análise multivariada é o estudo de técnicas estatísticas que analisa as informações com o intuito de simplificar a estrutura de dados, verificando a possibilidade de relacionamento entre as variáveis analisadas, visando sintetizar e facilitar a interpretação. As variáveis são medidas simultaneamente em uma mesma unidade experimental, em cada elemento amostral (SANTOS e SOUZA, 2018; MINGOTI, 2013). Os motivos aos quais as análises multivariadas conduzem são: a redução de dados ou simplificação estrutural, ordenação e agrupamento, investigação da independência entre as variáveis e predição (FERREIRA, 1996).

De acordo com Manly (1986), a estatística multivariada busca resposta para como as variáveis se relacionam dentro de cada grupo, bem como se há diferença significativa quanto aos valores médios das variáveis, se os grupos apresentam quantidades similares de variação para as variáveis e, havendo diferença entre os dois grupos quanto a distribuição de variáveis, se há a possibilidade de construir alguma função que separe esses dois grupos. Para responder essas questões, várias técnicas foram desenvolvidas, buscando ser útil para cada objetivo proposto. As técnicas mais utilizadas para o desenvolvimento dessas metodologias são a análise de componentes principais, análise fatorial e análise de agrupamentos (MINGOTI, 2013).

No passado, a única forma de analisar as variáveis, era utilizando estatística univariada. No entanto, para problemas aos quais dependem de muitas variáveis, o mais adequado e necessário é conhecer a totalidade das informações que são fornecidas por todo o conjunto das variáveis e a relação entre elas (VICINI, 2005). A estatística multivariada tem como intuito identificar as relações entre as variáveis no lugar de confirmar hipóteses de relacionamentos, como é típico da estatística univariada (LOESCH e HOELTGEBAUM, 2017).

O desenvolvimento da estatística multivariada teve um grande impulso na metade do século passado. Porém, devido à grande complexidade dos cálculos matemáticos envolvendo

operações com matrizes de altas ordens, este método só se popularizou no século atual. Isso se deve aos diversos avanços tecnológicos que tem surgido para ajudar a aplicar as técnicas multivariadas. E entre os mais importantes estão os desenvolvimentos de software. Pacotes computacionais nos possibilitam a analisar inúmeros dados complexos com uma relativa facilidade. A aplicação deste método tem sido muito comum em áreas como ciências biológicas, que abrange biologia, ecologia, ciências da saúde e ciências agronômicas (SONG, 2013; HAIR et al., 2009).

A complexidade de compreensão do assunto por boa parte dos usuários tem contribuído para a limitação da aplicação dos métodos multivariados e principalmente, reduzindo a qualidade da exploração dos resultados (FERREIRA, 1996). No entanto, as técnicas de análise multivariada têm-se apresentado como instrumentos poderosos para a análise de dados (DUARTE, 1997).

2.3.1 Estatística multivariada na classificação de frutos

A análise de estatística multivariada vem sendo amplamente aplicada na classificação de frutos. Um exemplo é a utilização de técnicas da estatística multivariada visando classificar polpas de frutas de acordo com as propriedades físicas e difusividade térmica. A estatística multivariada oferece ferramentas para o entendimento das propriedades físicas e químicas, permitindo traçar padrões de comportamento para diversas matrizes (ALMEIDA et al., 2020; SEMPREBON JUNIOR, 2018).

Visando a identificação de variáveis associadas a uma característica do fruto e no intuito de determinar parâmetros de qualidade, a melhor alternativa é a aplicação de técnicas de análise multivariada, como análise de componentes principais (ACP) e análise de agrupamento (*cluster*) (MIGUEL et al., 2011). Portanto, o emprego desses métodos apresenta grande importância na análise de vários frutos.

2.4 Contextualizando as redes neurais artificiais

As redes neurais artificiais baseiam-se no mecanismo de comportamento de um cérebro humano, simulados por meio de um hardware e/ou software, que por sua vez são capazes de adquirir conhecimentos a cada experiência. As redes neurais artificiais são organizadas em camadas (entrada, oculta e saída) ou fileiras, possuindo muitas interconexões, tendo como elemento base de aprendizado, os pesos, sendo este o ponto de calibração da rede (LIMA e MINUSSI, 2011).

Uma rede neural artificial funciona por meio da composição de camadas e nodos. As camadas são referentes a localização em que se encontram na rede e cada qual possui uma função. Os nodos (também chamados de neurônios ou perceptrons) são comportados nas camadas, carregando as informações e sendo também pontos de conexão de uma camada e outra. A rede neural artificial é formada por conexões de nós entre as camadas (SAS, 2020).

As entradas das RNAs podem ser conectadas em muitos neurônios, o que resulta em uma determinada quantidade de saídas, a depender do que é esperado como resposta. A conexão desses neurônios tem como função transformar o sinal da saída de um neurônio em um sinal da entrada, de outro. São as diferentes possibilidades de conexões que geram diversas estruturas ou arquiteturas diferentes, estando entre elas a Rede Feedforward de 1 camada, Rede Feedforward Multicamadas, Rede Recorrente ou Realimentada e Rede GRNN (FURTADO, 2019).

Pesquisas nessa área tem como foco principal fazer com que computadores aprendam a resolver problemas que são considerados simples para o ponto de vista humano, pois mesmo sendo problemas simples, quando executados manualmente, geralmente exigem o ato de repetição, tornando o processo mais passível à ocorrência de erro, isto em razão da percepção ou até mesmo cansaço de quem o executa e um outro problema é que em alguns casos podem até colocar em risco o executor. Obter uma boa execução desses processos podem beneficiar a economia e segurança. A inteligência artificial tem beneficiado vários setores, automatizando as tarefas (VINUESA et al., 2020).

No treinamento de uma rede neural artificial são utilizados diferentes tipos de aprendizado, podendo ser: supervisionado, não supervisionado e por reforço, em que, os dois primeiros têm como direção se há existência ou não de um supervisor, ou seja, se os dados já são ou não anteriormente divididos em classes. Já o de reforço, o algoritmo ganha pontos quando atinge resultados coerentes com a realidade ou perde pontos quando erra (MATOS et al. 2020; BARCA et al., 2005)

As RNAs podem ser aplicadas na solução de diversos tipos de classes de problemas, sendo elas, o reconhecimento e classificação de padrões, processamento digital de imagens, processamento de sinais, identificação e controle de sistemas, visão computacional, robótica, análise do mercado financeiro, entre outros (PEDRYCZ e GOMIDE, 2007).

As redes possuem a capacidade, juntamente com um algoritmo, de aprender com base em treinamento da rede com dados agrupados, possibilitando encontrar padrões para classificar de forma correta uma base de dados. A cada treinamento a rede é melhorada

continuamente, com influência nas mudanças dos parâmetros que influenciam no aprendizado da mesma, como a taxa de aprendizado, número de ciclos (épocas) e a quantidade de neurônios na camada intermediária (oculta). Ajustes nesses parâmetros ditam o comportamento da eficácia da rede e são muito importantes para que ocorra um bom treino e consequentemente uma boa validação da capacidade de generalização das redes neurais artificiais (FIGUEIREDO e CAMPOS, 2019).

Embora seja um assunto com ampla abordagem, todas se assemelham a um modelo da atividade cerebral, envolvida em exercer a tarefa proposta ou de simular comportamentos e funções de um neurônio biológico, por meio de um modelo matemático. Essa perspectiva reflete diretamente em pesquisas acadêmicas sobre o assunto, quanto em suas aplicações em um problema específico (SILVA D. E., 2021; YACIM e BOSHOFF, 2018).

2.4.1 Redes neurais na classificação de frutos

Geralmente a classificação de frutos e a aplicação de redes neurais estão relacionadas à maturidade de frutos, pois se uma rede neural é capaz de classificar o quão maduro os consumidores gostam de seus frutos, é possível então, associar esse aspecto com conjuntos de modelos matemáticos com capacidade de predição do tempo necessário até que eles cheguem a esse ponto de maturidade (THAI e SHEWFELT, 1991). Ramos (2003) apresentou a possibilidade de uso de redes neurais artificiais no desenvolvimento de modelos de classificação de frutos por meio de vetores de padrões.

Uma outra aplicação eficaz das RNAs para classificação é a de determinação de preços de frutas ao longo do tempo. As redes podem fornecer ótimas previsões, com influência positiva na estabilidade ou mesmo na queda de preços dos produtos (SOUZA et al., 2014).

Trabalhos utilizando redes neurais artificiais para classificação de frutas tendem a possibilitar inovação para que essas redes sejam utilizadas por supermercados e consumidores, visando auxiliar em tempo real a detecção das características do produto, sem que ocorra a necessidade de intervenção humana (ROSA, 2019).

3 MATERIAL E MÉTODOS

3.1 O local e suas características

As cagaitas foram colhidas na área experimental da Escola de Agronomia, Engenharia de Alimentos e Engenharia Florestal, da Universidade Federal de Goiás, localizada

no município de Goiânia/GO. O município está situado em latitude 16° 36' 16.920" Sul, longitude 49° 15' 44.280" Oeste e altitude de 749 metros.

Segundo a classificação Köppen, para Goiânia/GO, o clima característico é Aw, tropical, caracterizada por apresentar duas estações bem definidas, um verão (quente e chuvoso) e um inverno (ameno e seco) (VASQUES, 2021; CARDOSO et al., 2014).

3.2 Coleta das amostras

Foram selecionados e marcados aleatoriamente cinco ramos em diferentes árvores, com lãs de diferentes cores, sendo elas: azul escuro, rosa, azul claro, vermelho e lilás, por ocasião da antese. As marcações dos ramos foram realizadas em sete árvores, as quais ocorreram antese de forma uniforme.

Os frutos foram colhidos manualmente, entre os dias 11 e 23 de outubro de 2020 e foram acondicionados em sacos de polietileno de baixa densidade (PEBD), com fecho tipo zip, de modo que fosse evitado que as amostras perdessem a umidade por transpiração (Figura 4). Foram colhidos o total de 120 cagaitas (*Eugenia desynerica* DC.), para cada estágio de maturação totalizando 600 frutos.

Figura 4 - Cagaitas acondicionadas em embalagem de polietileno de baixa densidade (PEBD), com fecho tipo zip.

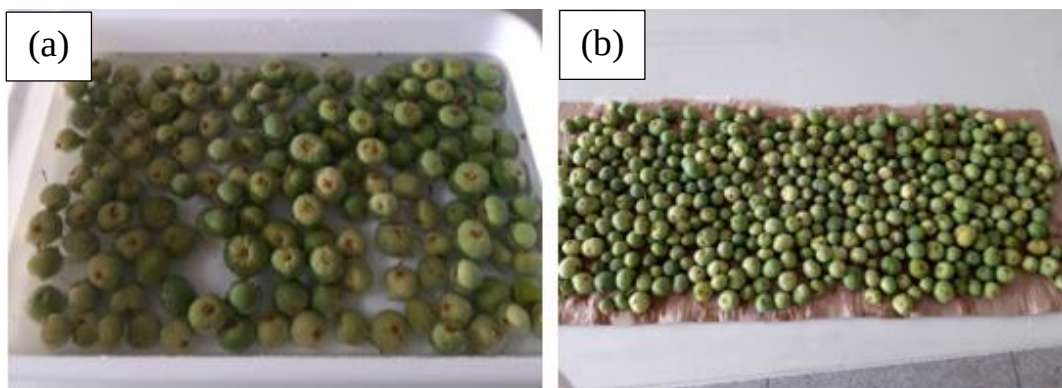


Fonte: Próprio autor (2020).

Posteriormente, os frutos foram transportados para o Laboratório de Pós-Colheita, do curso de Engenharia Agrícola da Universidade Estadual de Goiás, Campus Central - Henrique Santillo, localizado no município de Anápolis/GO. Os frutos foram lavados em

imersão na água potável (Figura 5a), secados em temperatura ambiente (Figura 5b) e selecionados com o intuito de determinar um padrão de cor da casca, evitar integridade física, ausência de danos mecânicos e patogênicos, visando a uniformização dos lotes.

Figura 5 - Cagaitas (a) imersas em água potável; e (b) secando em temperatura ambiente.



Fonte: Próprio autor (2020).

3.3 Delineamento experimental

O delineamento experimental utilizado foi o inteiramente casualizado (DIC), com um fator - dias após antese (DAA), com 40 repetições, sendo três frutos por repetição. Os frutos foram colhidos a partir do 28 DAA, com o intervalo de três dias, totalizando 5 tratamentos (28, 31, 34, 37 e 40 DAA) e foram analisados no mesmo dia da colheita respectiva.

Nesse período, foram realizadas as análises físicas, físico-químicas e químicas, como: peso, diâmetro (transversal e longitudinal), área, coloração (luminosidade, °hue e croma), firmeza (casca e polpa), sólidos solúveis (SS), potencial hidrogeniônico (pH), acidez titulável (AT) e índice de maturação (IM).

3.4 Variáveis analisadas

3.4.1 Peso dos frutos

Os pesos dos frutos foram determinados utilizando a balança de precisão shimadzu modelo BL3200H. e os resultados foram expressos em gramas (g).

3.4.2 Diâmetro (transversal e longitudinal)

O diâmetro longitudinal (DL) foi medido no sentido horizontal do fruto e o diâmetro transversal (DT) foi medido da região de inserção do pedúnculo à sua região oposta, sentido vertical, ambos utilizando o paquímetro digital Stainless Hardened 150 mm, com 3 medições

para DL e 3 medições para DT, por fruto. Os resultados foram as médias das medições expressas em milímetros (mm).

3.4.3 Área

A obtenção da área das cagaitas foi realizada por meio da Equação (1), e expresso em milímetros quadrados (mm²):

$$\text{Área} = 4.\pi.r^2 \quad (1)$$

Em que:

r = raio.

3.4.4 Coloração (Luminosidade, °hue e croma)

Foram analisados os parâmetros: Luminosidade (L) referente ao aspecto visual do fruto no que diz respeito ao quão claro (100) e quão escuro (0) se encontra o fruto, ângulo de tonalidade cromática (°hue) e a saturação da cor (croma).

Por meio do aparelho colorímetro da marca CR-400 (Konica Minolta) foram obtidas, por meio de refletância, informações do parâmetro Luminosidade (L), das coordenadas a* relacionada à intensidade de verde (-a) e vermelho (+a) e b* relacionada à intensidade de azul (-b) e amarelo (+b). Cada fruto foi avaliado, permitindo a determinação do °hue e croma, por meio das Equações (2) e (3), respectivamente:

$$^{\circ} Hue = 180 + \arctang \left(\frac{b^*}{a^*} \right) \quad (2)$$

$$Croma = \sqrt{a^{*2} + b^{*2}} \quad (3)$$

Em que:

a* = coordenada cromática obtida no colorímetro CR-400;

b* = coordenada cromática obtida no colorímetro CR-400.

3.4.5 Firmeza (casca e polpa)

Foi realizada com o uso do aparelho *Texture Analyser* CT3, da marca Brookfield, na velocidade de penetração de $7,0 \text{ mm s}^{-1}$ a uma profundidade de 2 mm. Os valores foram expressos em Newton (N). Foram realizadas 2 medições, por fruto, em lados opostos no sentido longitudinal.

3.4.6 Sólidos solúveis

Cada fruto foi submetido a maceração, com o auxílio de um almofariz e pistilo. Posteriormente, com o auxílio de uma pipeta, foi procedida a leitura no refratômetro mini digital Reichert, em °Brix, conforme recomendação do AOAC (2012).

3.4.7 Potencial hidrogeniônico

Foi obtido por meio do aparelho pHmetro portátil modelo K39 – 0014P da marca Kasvi, sendo realizado para cada amostra, conforme técnica descrita por AOAC (2012).

3.4.8 Acidez titulável

Para análise de acidez titulável foi pesado 5 g de cada fruto macerado e em seguida acrescentou-se 95 ml de água destilada, homogeneizando a mistura, e, posteriormente, utilizou-se NaOH $0,1 \text{ mol L}^{-1}$, para determinação da % de ácido cítrico. O ponto de viragem foi observado visualmente, com a alteração da coloração de cada amostra, seguindo a recomendação do AOAC (2012). O percentual de ácido cítrico foi calculado pela Equação (4):

$$AT \text{ (Ácido cítrico \%)} = \frac{V \times FC \times 0,1 \times 6,404}{Pa} \quad (4)$$

Em que:

V = Volume de NaOH usado na titulação (mL);

FC = Fator de Correção da solução de NaOH;

Pa = Peso da amostra adicionado no erlenmeyer (g);

0,1 = Concentração da solução de NaOH (mol L^{-1});

6,404 = Equivalente do ácido cítrico na amostra (g)

3.4.9 Índice de maturação

Foi obtido por meio da relação entre teor de sólidos solúveis e acidez titulável.

3.5 Estatística Multivariada

3.5.1 Análise de componentes principais (ACP)

Os dados foram padronizados, devido a não possuir unidades de medidas iguais, com o intuito de se evitar erros nos resultados. Posteriormente, foram calculadas as correlações dos grupos de variáveis, visando identificar se as mesmas se correlacionam entre si. Baseando-se na matriz de correlação, os dados foram submetidos a análise de componentes principais (ACP) onde foram analisadas as relações entre as características físicas, físico-químicas e químicas da cagaita com o estágio de maturação.

Foram determinadas as quantidades de componentes principais a serem mantidos no estudo por meio dos autovalores, utilizando o critério sugerido por Kaiser (1958), em que os componentes principais que permanecerão devem ser maiores que 1. Geralmente são esses os componentes que resumem uma variância explicada em até 70%. Posteriormente, calculou-se o círculo de correlação unitário das características físicas, físico-químicas e químicas dos frutos, para verificar quais variáveis possuem maior influência mediante cada componente em estudo.

3.5.2 Análise de agrupamento (*Cluster*)

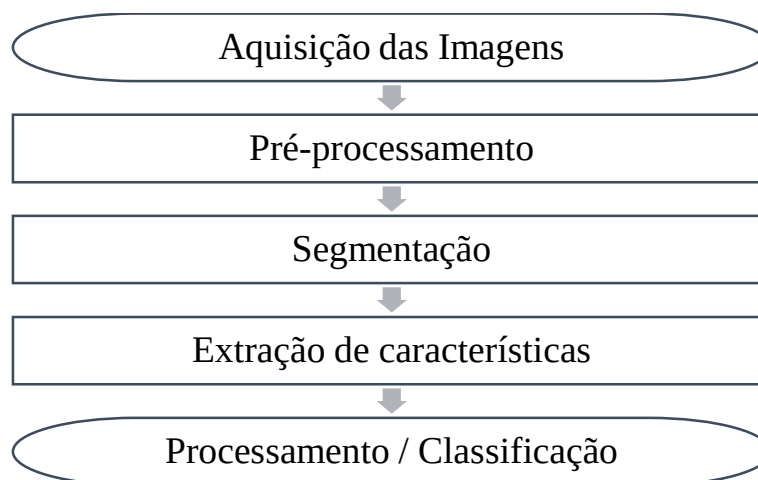
Foi realizada a análise de agrupamento com a finalidade de avaliar os diferentes estádios de maturação das cagaitas baseando-se nas características extraídas em diferentes dias após a antese (DAA), onde os dados utilizados foram os dados padronizados na primeira análise, visando evitar erros. Adotou-se, como medida de dissimilaridade, o método de distância *Canberra*, para calcular a distância entre cada ponto amostral.

O método de agrupamento utilizado foi o método de Ward (1963), também conhecido como método da mínima variância, visando a produção de grupos heterogêneos entre si, e mais homogêneos internamente. Para a validação dos métodos estabelecidos para a análise de agrupamento, calculou-se o coeficiente de correlação cofenética (CCC), com a contribuição da função *hclust* (GAMA, 2020). De acordo com Bussab et al. (1990) são considerados aceitáveis os agrupamentos que obtiverem o coeficiente de correlação cofenética (CCC) a partir de 0,80, sendo esse o critério adotado para definir os métodos adequados neste trabalho.

3.6 Etapas sistemáticas para o processamento digital de imagens

O sistema de extração de informações para a definição de diferentes estádios de maturação da cagaita por meio de processamento digital de imagens foi composto por etapas vitais (Figura 6)

Figura 6 - Etapas de processamento digital de imagens utilizadas para a classificação de diferentes estádios de maturação da cagaita.



Fonte: Próprio autor (2021).

3.6.1 Aquisição das imagens e calibração do sistema

As imagens digitais das cagaitas foram extraídas nos dias após antese (DAA) específico de cada tratamento para as características físicas, físico-químicas e químicas, no laboratório. Inicialmente as cagaitas foram colocadas sobre um tecido horizontal de fundo branco e as imagens foram adquiridas utilizando uma câmera fotográfica Sony, modelo Cyber-Shot DSC-H400, de 20.1MP, sob iluminação ambiente, tomando-se o cuidado para que não ocorresse sombras sobre a área da imagem. A câmera ficou situada a 10 cm acima da bandeja, sendo fotografado um fruto para cada imagem. Foram obtidas 120 imagens por dia após antese (28, 31, 34, 37 e 40 DAA), totalizando 600 imagens. As imagens capturadas foram armazenadas no formato JPG com resolução de 500x375 pixels.

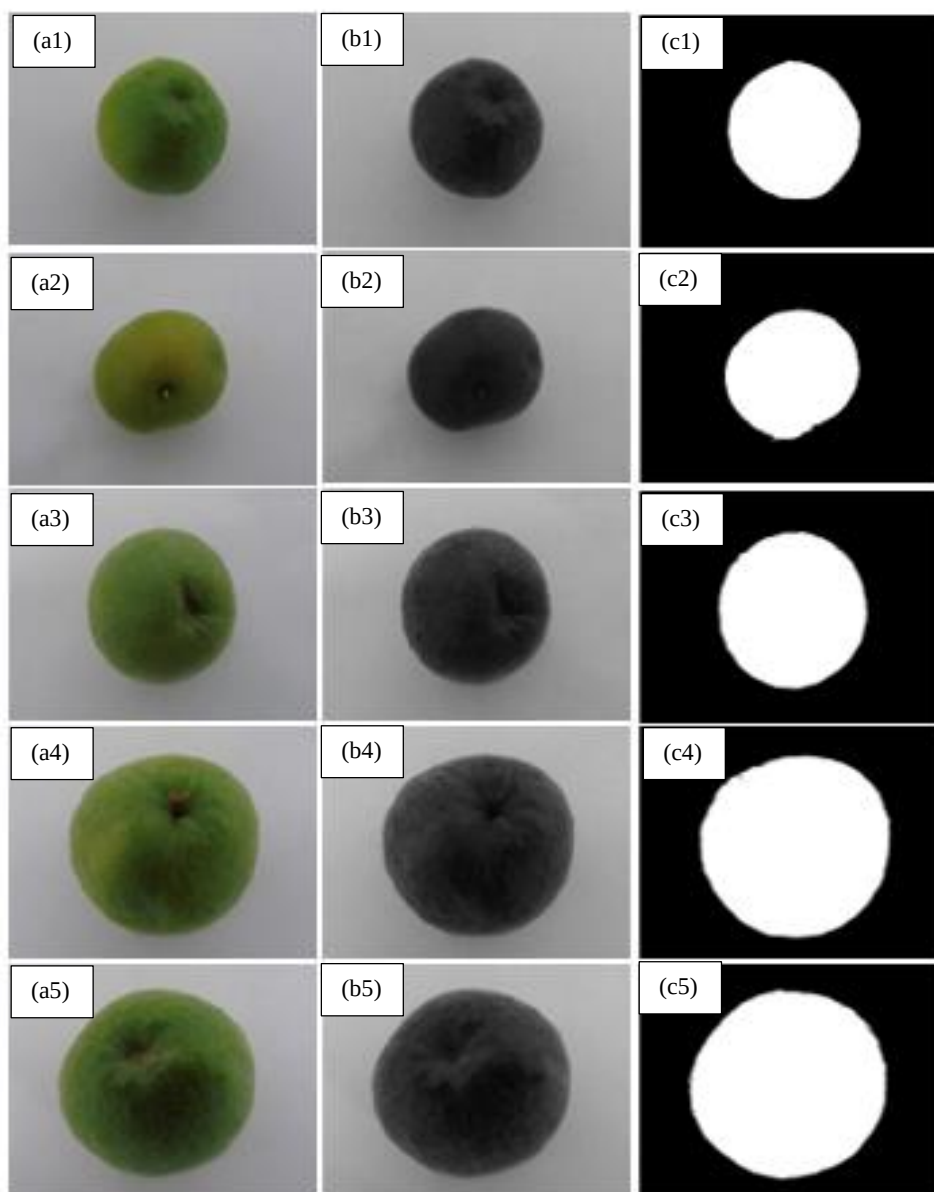
3.6.2 Segmentação das imagens

Inicialmente as imagens originais (Figura 7a) foram alteradas para o modo cinza (Figura 7b), onde foram convertidas para apenas uma camada (RGB). A camada “blue” foi a escolhida, sendo esta a que melhor ajustou para a obtenção de segmentações mais precisas das imagens, mesmo a cor predominante sendo verde. Foi executada também a correção gama (com expoente 1.0) em que o contraste das imagens foi alterado para melhor identificar o objeto (fruto) e o fundo das imagens, tornando os frutos mais escuros (MATTOS, 2017).

Realizou-se a segmentação das imagens (Figura 7c) com o intuito de extrair e identificar as regiões de interesse. O método utilizado foi o de Otsu (1979), onde as imagens

iniciais foram convertidas em duas classes (primeiro plano e segundo plano) visando minimizar as interações. Posteriormente, a partir da parte extraída, foi feita a caracterização para análise e classificação (AZARMDEL et al., 2020; RIBEIRO, 2016).

Figura 7 - Imagens dos primeiros frutos de cada dia após antese. Imagens originais (a); convertidas em tons de cinza (b); e segmentadas (c). 28 (a1, b1, c1), 31 (a2, b2, c2), 34 (a3, b3, c3), 37 (a4, b4, c4) e 40 DAA (a5, b5, c5).



Fonte: Próprio autor (2021).

3.6.3 Extração dos atributos de cor

Para esse trabalho, realizou-se a extração dos canais RGB (*Red*, *Green* e *Blue*) e posteriormente foram calculados seus respectivos valores médios. Os valores foram convertidos

para o HSV, em que H refere-se ao matiz que varia entre 0° e 360°, o S a saturação que varia no intervalo de 0 e 1 e V o value que refere-se ao brilho, variando no intervalo de 0 e 1. Os canais RGB também foram convertidos para as cores primárias XYZ e por meio da conversão das cores primárias, foram obtidos os parâmetros CIELab, em que o L é referente a Luminosidade com variação de 0 a 100, e duas gamas de cor a e b que estão associadas à cromaticidade que vão do verde ao vermelho e do azul ao amarelo, respectivamente, que variam de 180 a -180 (COSTA, 2015). Foram determinados °hue e croma, por meio das Equações (2) e (3), respectivamente, que estão no item 3.4.4.

3.6.4 Extração de atributos geométricos

Outro parâmetro muito importante para determinação de diferentes estádios de maturação de frutos são suas características referentes ao tamanho. Nesse trabalho foi realizada a estimação da área dos frutos, por meio das imagens segmentadas, em que a função da área possui as mesmas condições da Equação (1), no item 3.4.3 e por meio dela foi calculado o comprimento do fruto, utilizando a Equação (5):

$$\text{Comprimento} = \frac{\text{Área}}{2 \cdot \pi} \quad (5)$$

3.7 Classificação dos estádios de maturação das cagaitas por meio de redes neurais artificiais

Na etapa de classificação foram definidas cinco classes (1, 2, 3, 4 e 5), sendo elas correspondentes ao comportamento dos frutos nos cinco dias após antese avaliados (28, 31, 34, 37 e 40 DAA).

Para classificar as cagaitas nas diferentes classes, foram treinadas 20 redes com os parâmetros derivados dos modelos de cores e dos atributos geométricos extraídos por meio das imagens, sendo que, os parâmetros dos modelos de cores e os atributos geométricos (R, G, B, H, S, V, L, °hue, croma, área e comprimento) foram considerados como as variáveis de entrada da rede e as diferentes classes (1, 2, 3, 4 e 5) baseadas nos diferentes dias após antese, as variáveis de saída.

3.7.1 Separação dos conjuntos (treino, validação e teste)

Segundo Abu-Mostafa et al. (2012), amostras utilizadas para o treinamento da rede precisam ser divididas em treino, teste e validação, sendo que a proporção dessa divisão é essencial para um treinamento com resultados satisfatórios.

Para a execução da rede, 58,667% de todos os dados de entrada foram considerados de treinamento, sendo que esses dados foram utilizados para estimar os parâmetros desconhecidos da rede (pesos e tendências), 30% de validação em que os dados são subdivididos dos dados de treinamento e essa subdivisão foi utilizada como critério de parada do treinamento, totalizando 532 dados para o treinamento e validação e por fim, 11,334% de teste, totalizando 68 dados que foram utilizados com o intuito de verificar a eficiência da rede aplicada às condições reais (RODRIGUES, 2021).

3.7.2 Construindo, compilando e ajustando o modelo

Na camada oculta para cada rede, foram testados um conjunto de até 10 neurônios, variando de 3 a 12 neurônios. A função de ativação que melhor se ajustou aos dados foi a ‘relu’ e para a camada de saída a função ‘exponencial’. Após configurar a arquitetura do modelo, o mesmo foi compilado para que fosse ajustado o melhor desempenho dos dados. Nesta etapa de compilação, foram avaliadas duas diferentes funções de otimização “adam” e “nadam”.

Para cada aplicação da função, o modelo foi ajustado com 3000 interações e 5 lotes, visando identificar qual apresentou resultados mais satisfatórios. Esses parâmetros foram definidos visando melhorar a eficiência das redes e aumentar a precisão.

O melhor desempenho da rede ocorreu nos ciclos de treinamento (época) 2400, momento este em que as curvas de aprendizagem se encontraram. Este momento é conhecido como ponto de parada antecipada, que de acordo com Melo (2012), é o ponto que determina a partir de quando a rede irá ajustar em relação aos dados de classificação, sugerindo que a rede está se ajustando ao ruído contido nos dados de treinamento. Neste mesmo momento os pesos finais foram salvos, visando evitar a ocorrência de ajustes excessivos. Posteriormente, foi verificada a predição da rede neural artificial com novos valores.

3.7.3 Processo de avaliação do sistema de classificação

Para a análise avaliativa, bem como para a verificação do poder classificatório, foram utilizadas as amostras do conjunto teste, aplicando o método da matriz confusão.

3.7.4 Parâmetros estatísticos para a avaliação do desempenho

Para avaliar o desempenho do modelo, foram calculados 4 parâmetros estatísticos, visando fornecer uma visão mais detalhada do classificador. As métricas calculadas foram a Acurácia, Precisão, Recall e F-Score, utilizando as Equações (6), (7), (8) e (9), respectivamente:

$$Acurácia (\%) = \frac{VP + VN}{VP + FP + VN + FN} \quad (6)$$

$$Precisão = \frac{VP}{VP + FP} \quad (7)$$

$$Recall = \frac{VP}{VP + FN} \quad (8)$$

$$F - Score = 2 * \frac{Precisão * Recall}{Precisão + Recall} \quad (9)$$

Em que:

VP = Verdadeiro Positivo;

VN = Verdadeiro Negativo;

FP = Falso Positivo; e

FN = Falso Negativo.

3.8 Especificação do software e da máquina

Para as análises de estatística multivariada (componentes principais e análise de agrupamento), de processamento digital de imagens (modelo de cores e atributos geométricos) e redes neurais artificiais, foi utilizado o software RStudio versão 2021.09.0 Build 351 para Windows.

O computador utilizado possui processador Intel(R) Core (TM) i3, CPU @ 2.00GHz, sistema operacional de 64 bits.

4 RESULTADOS E DISCUSSÃO

4.1 Estatística Multivariada

4.1.1 Análise de componentes principais e análise de agrupamento das características físicas, físico-químicas e químicas das cagaitas em diferentes dias após a antese

Os resultados obtidos por meio da análise dos componentes principais (CP's) aplicados as características físicas, físico-químicas e químicas em diferentes dias após antese estão expostos na Tabela 1. Cada variável em estudo (13 variáveis) gerou um componente principal (CP). No entanto, na Tabela 1 são apresentados apenas 4, tendo em vista que os demais apresentaram autovalores bem próximos de zero, o que significa que esses componentes (5 a 13) pouco se explicam em relação ao estágio de maturação baseado nas variáveis em diferentes dias após antese.

Além dos componentes principais e seus respectivos autovalores, são também apresentadas as variâncias explicadas e as variâncias totais. Seguindo o critério de Kaiser (1958), os três primeiros componentes são os que melhor explicam a variação no conjunto dos dados. Apresentando autovalores maiores que 1, sendo eles: 7,57; 4,11; 1,12, onde CP1 explica 58,21%, CP2 31,65% e CP3 8,58%, representando um total de 98,44% da variabilidade dos dados.

Tabela 1 - Componentes principais (CP's), autovalores, variância explicada (%) e variância total (%) das características físicas, físico-químicas e químicas das cagaitas em diferentes dias após antese (28, 31, 34, 37 e 40 DAA).

Componente Principal (CP)	Autovalor	Variância Explicada (%)	Variância Total (%)
CP 1	7,57	58,21	58,21
CP 2	4,11	31,65	89,86
CP 3	1,12	8,58	98,44
CP 4	0,20	1,56	100,00

Fonte: Próprio autor (2021).

Na Tabela 2 é exposta a correlação das variáveis físicas, físico-químicas e químicas das cagaitas, com as três primeiras componentes principais (CP 1, CP 2 e CP 3), onde é possível verificar que a maioria das variáveis apresentaram melhor correlação nas duas primeiras componentes, com exceção do potencial hidrogeniônico que foi o único presente no componente principal 3.

Na componente principal 1, há correlação em 8 das 13 variáveis em estudo, em que se correlaciona positivamente o diâmetro longitudinal, diâmetro transversal, área, peso, °hue e negativamente a firmeza da casca, firmeza da polpa e sólidos solúveis, isso quer dizer que quanto maiores e mais pesados os frutos se encontram, menos consistentes e com menor teor de açúcar os mesmos estarão.

Tabela 2 - Correlação das características físicas, físico-químicas e químicas das cagaitas em diferentes dias após antese (28, 31, 34, 37 e 40 DAA), com as 3 primeiras componentes principais.

Variável	CP 1	CP 2	CP3
Diâmetro Longitudinal	0,9746	-0,0247	0,1965
Diâmetro Transversal	0,9931	-0,0385	0,0919
Área	0,9682	-0,0604	0,2064
Peso	0,9446	-0,2161	0,2286
Firmeza da Casca	-0,9975	-0,0207	-0,0489
Firmeza da Polpa	-0,9719	-0,1804	-0,1507
Luminosidade	0,5135	0,8123	-0,2693
°Hue	0,7801	-0,5157	-0,3126
Croma	0,2703	0,9256	-0,2522
Sólidos Solúveis	-0,854	-0,3886	0,1788
pH	-0,3356	0,4256	0,8245
Acidez Titulável	0,0718	-0,9956	0,0001
Índice de Maturação	-0,2638	0,9609	0,0539

Números em negrito apresentam variáveis com melhor representação para a variância observada. Fonte: Próprio autor (2021).

Rodrigues et al. (2021) avaliando compostos bioativos, atividade antioxidante e características físicas e químicas de frutos da cagaiteira nativos do cerrado mineiro, verificaram que quanto maior o conteúdo de sólidos solúveis, maior será o rendimento do produto final, sendo, portanto, desejável, diferente do encontrado neste trabalho. Os autores também constataram que a acidez e o pH são parâmetros de grande relevância que dependem do grau de maturação do fruto, fato este também contrário ao deste trabalho, tendo em vista que as variáveis acidez e pH não apresentaram relação, sendo que a acidez é explicada no componente principal 2 e o potencial hidrogeniônico é melhor explicado no componente principal 3.

Na segunda componente principal ocorre correlação em 4 das variáveis em estudo, sendo elas positivamente a luminosidade, croma e índice de maturação e negativamente a acidez titulável. A redução da acidez da cagaita está relacionada com a respiração do fruto, isso quer dizer que quanto mais maduros os frutos se encontram, menor é a acidez dos mesmos (TERÁN-ORTIZ et al., 2013), corroborando com este trabalho, em que no decorrer dos dias analisados, os frutos apresentaram redução da acidez.

O índice de maturação, de acordo com Lattuada et al. (2020), é um dos principais indicativos para avaliar se os frutos estão maduros e também, um dos principais parâmetros para se avaliar o flavor. Quanto maior essa relação se encontra, mais maduros os frutos estão. Foi possível observar que nesse trabalho ocorreu esse aumento, indicando que os frutos se apresentaram mais maduros ao longo dos diferentes estádios de maturação.

Quanto a coloração, os frutos ao longo dos diferentes estádios de maturação apresentaram aumento no brilho e as cores mais vívidas. A coloração da casca dos frutos pode apresentar alterações ao longo da maturação, sendo que este fato se deve a degradação da clorofila ou revelação de pigmentos, como carotenoides e antocianinas (TUCKER, 1993).

O dendrograma da análise de agrupamento (Figura 8) foi gerado baseando-se nas características físicas, físico-químicas e químicas das cagaitas em diferentes dias após antese, com o intuito de explicar o estágio de maturação. O critério de dissimilaridade adotado foi com base na distância de *Canberra* e no método de agrupamento proposto por Ward. O valor calculado de coeficiente de correlação cofenética está dentro do recomendando pela literatura, sendo que neste trabalho foi $r = 0,8234$, mostrando assim a confiabilidade do dendrograma.

Por meio deste dendrograma, os diferentes estádios de maturação das cagaitas foram separados por 3 grupos onde neles estão expostos como as características físicas, físico-químicas e químicas das cagaitas se comportaram nos diferentes dias de análise, sendo que nos dias 28 e 31 após antese (grupo 1) os frutos apresentaram características semelhantes. Silva (2016), em seu estudo sobre a fisiologia da cagaita, verificou que entre o 27° e 31° DAA, foi detectada uma etapa de maturação dos frutos, com tendência à estabilização da massa fresca do fruto, endurecimento e lignificação do endocarpo.

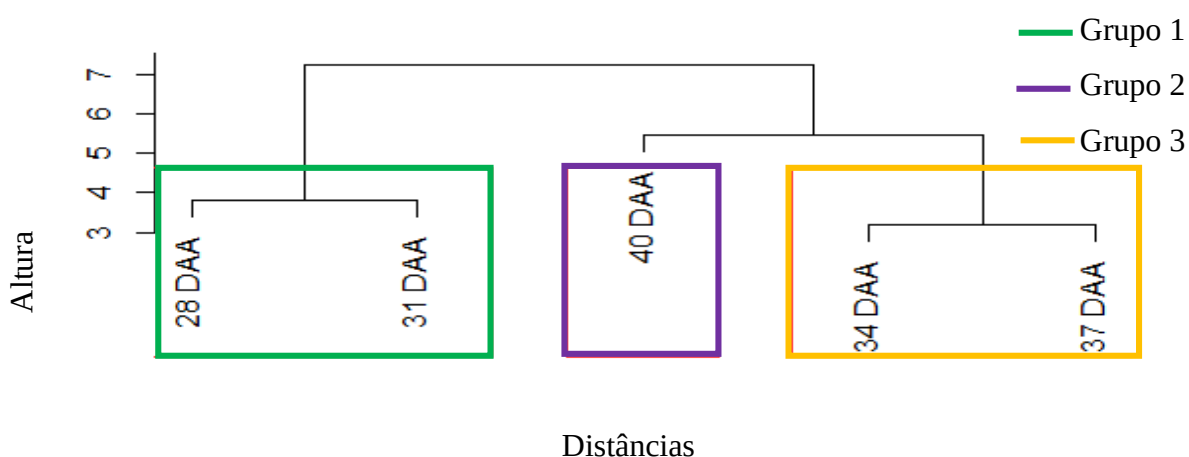
O grupo 2 refere-se ao 40 DAA, em que a maior parte dos parâmetros analisados apresentaram valores elevados, sendo que este fato é distinto ao esperado, pois essa elevação de valores dos parâmetros geralmente ocorre nos dias iniciais dos estádios de maturação. O último estágio de maturação analisado apresentou características mais semelhantes aos dois primeiros estádios, sendo que, o adequado para uma definição mais precisa desses resultados

seria que os dois estádios que antecederam (34 e 37 DAA) ao último, nos dias analisados, apresentassem características mais próximas aos dois primeiros estádios. Tokairin (2017) avaliando a caracterização e conservação pós-colheita de cambuci, observou que mesmo que estivessem no último dia de armazenamento, ao qual a tendência dos frutos era estar em estágio de senescência, os mesmos ainda apresentavam características adequadas e agradáveis a aceitação do consumidor, corroborando com esse trabalho, já que no último estágio, os frutos se quer apresentaram-se próximos a senescência.

No grupo 3, os frutos comportaram semelhantemente e de forma satisfatória. Os parâmetros que apresentaram maior importância dentro desse grupo possuem tendência a se elevar no decorrer dos estádios de maturação e é um dos principais indícios de que os frutos estão mais próximos do amadurecimento, justificando o fato dos 34 e 37 DAA estar no último grupo.

Diante ao exposto, é possível verificar que por meio do dendrograma pode-se obter respostas quanto ao estágio de maturação das cagaitas, embora o comportamento dos frutos no dia 40 tenha sido mais semelhante as fases iniciais do estágio de maturação, o decorrer dos dias mostrou o desenvolvimento do fruto quanto ao seu tamanho, peso, coloração, nível de açúcar, consistência dos frutos, entre outras características, sendo que estas características estão diretamente associadas ao tempo de vida útil da cagaita.

Figura 8 - Dendrograma da análise de agrupamento das cagaitas em diferentes dias de antese (DAA).



Fonte: Próprio autor (2021).

De acordo com a análise de componentes principais (Figura 9) é possível verificar que quanto maior o comprimento do vetor (seta), maior a importância dele para o componente em que ele está localizado, sendo possível observar que todas as variáveis nesse estudo são importantes nos componentes em que se encontram, com exceção do potencial hidrogeniônico. As variáveis que apresentaram maior importância, considerado o critério do tamanho dos vetores, são acidez titulável, índice de maturação e firmeza da polpa.

As variáveis que se encontram no eixo horizontal, possuem maior relevância ao componente 1, enquanto as que se encontram no eixo vertical, possuem maior relevância ao componente 2, as que não se encontram em nenhum dos dois eixos, fazendo ângulo de aproximadamente 45°, são relevantes para ambos os componentes.

As variáveis de firmeza e sólidos solúveis possuem relação entre si, onde de acordo com o exposto, a consistência está diretamente ligada ao nível de açúcar dos frutos. Ou seja, quanto mais consistentes, mais doces os frutos se encontraram. E as mesmas são relevantes tanto para CP1 e CP2, negativamente. Essas variáveis foram mais representativas nos 28 e 31 dias após antese, em que os frutos se encontravam mais firmes e doces. Essas características (firmeza de casca, firmeza de polpa e sólidos solúveis), também associadas ao pH e ao índice de maturação, se encontram no grupo 1. De acordo com Ordóñez et al. (2004), os frutos tendem a aumentar seu grau de doçura ao longo da maturação do fruto, o que não ocorreu neste trabalho, tendo em vista que os frutos mais doces foram analisados nos primeiros dois dias de análise.

Ribeiro (2011), considerou os frutos de cagaita como ácidos, baseando-se no pH (2,8) e na acidez titulável (8,5); verificando os sólidos solúveis (14,2 °Brix) o mesmo autor verificou também que os frutos não são muito doces.

Garcia (2017), verificou que no decorrer dos dias ocorreu aumento dos teores de pectinas solúveis e total, reforçando com os dados de firmeza do fruto, que se apresentaram firmes até o segundo dia de armazenamento e após esse dia o parâmetro diminuiu, ou seja, essas duas características tendem a se relacionar de forma inversa ao apresentado neste trabalho. Araújo et al. (2015), avaliando diferentes frutos do Cerrado, verificou que a cagaita obteve maior teor de sólidos solúveis, comparado aos demais frutos avaliados, que foi 8,6 (°Brix), valor superior a todos os dias analisados neste trabalho, em que o dia 28 após antese foi o que apresentou valores maiores, com valor médio de 8,42 (°Brix).

O índice de maturação encontra-se próximo ao pH, sendo que este não apresentou, quanto ao seu vetor, relevância neste estudo, quando comparado as outras 12 variáveis. Silva et al. (2021), analisando a avaliação biométrica e físico-química e estudo do perfil químico da

cagaita, verificaram que para o pH não foi constatada correlação significativa, corroborando com o exposto neste trabalho. De acordo com Lima et al. (2015), o pH é um parâmetro muito importante, isto porque está relacionado ao flavor bem como a verificação do estágio de maturação das frutas.

O índice de maturação ficou próximo a completar um ângulo de 180° com a acidez titulável, o que é coerente com as informações das características que ambas apresentam, tendo em vista que o índice de maturação é a relação dos sólidos solúveis com a acidez titulável. O índice de maturação também apresenta relevância para ambas as componentes, sendo que esta relevância é negativa para CP1 e positiva para CP2.

O diâmetro longitudinal, diâmetro transversal e área por se encontrarem próximos, demonstram forte correlação entre si e a firmeza da casca, por se encontrar localizada a aproximadamente 180° das demais, demonstrando que quanto maior os frutos, menor será a consistência dos mesmos. Essas variáveis são mais relevantes para o componente principal 1.

Silva (2016), observou aumento significativo do diâmetro transversal das cagaitas desde a antese até o seu amadurecimento, atingindo em média 38 mm no 37° dia. Este aumento, também foi verificado no diâmetro longitudinal (36 mm) até o 23° dia após antese (DAA), chegando ao fim com um decréscimo, atingindo seu tamanho limite (30 mm). Neste trabalho houve aumento significativo tanto do diâmetro longitudinal, quanto transversal dos frutos até o último estágio de maturação (40 DAA).

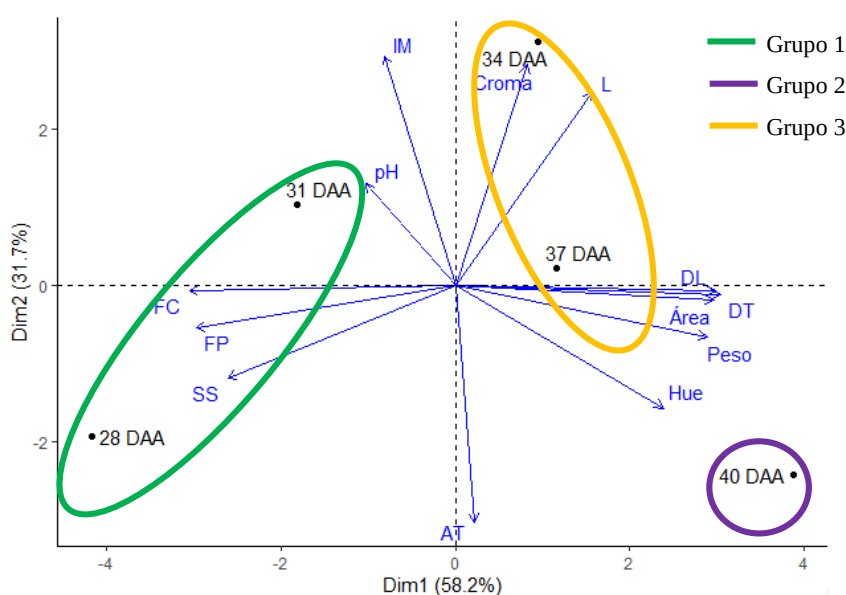
As variáveis peso e °hue estão próximas tanto entre si, quanto dos diâmetros (longitudinal e transversal), mostrando que as mesmas também possuem correlação entre si, sendo que quando uma destas características aumentam, as outras terão tendência em aumentar também. As variáveis peso e °hue são relevantes tanto para CP1 positivamente, quanto para CP2 negativamente. O último dia de análise (40 DAA).

Paz et al. (2020), avaliando a caracterização biométrica de frutos e sementes de cagaita (*Eugenia dysenterica* DC.) na região sul do Piauí, Brasil, identificou frutos com o peso variando entre 2,89 a 11,87 (g) e diâmetro longitudinal e transversal variando entre 11 a 21 (mm) e 13 a 24 (mm). Neste trabalho, a média dos pesos variaram dos diferentes dias após antese variaram entre 8,53 a 18,73 (g), sendo que nos primeiros dois dias (28 e 31 DAA) apresentaram valores dentro do encontrado no trabalho citado e os três últimos dias (34, 37 e 40 DAA) apresentaram-se inferiores. O diâmetro longitudinal e transversal apresentaram valores entre 26,35 a 34,47 (mm) e 20,77 a 26,69 (mm), respectivamente.

Os parâmetros de coloração: croma e luminosidade, estão próximos entre si, o que explica que quanto mais luminoso o fruto se encontra, maior será a tonalidade cromática do mesmo e essas variáveis apresentaram relevância positiva tanto para CP1 quanto para o CP2. Esses parâmetros foram bem representados nos dias 34 e 37 DAA, grupo 3, ou seja, durante esse período os frutos apresentaram-se mais claros e com maior ângulo de cromaticidade, principalmente para o 34 DAA. Almeida (2020), verificou em seu trabalho que a coloração da casca é um dos parâmetros mais adequados para a determinação de diferentes estádios de maturação de cagaitas.

De acordo com Vilas Boas et al. (2012), o avanço dos dias após antese possibilita verificar os comportamentos das variáveis físicas e químicas, em diferentes estádios de maturação dos frutos. Os frutos colhidos em diferentes dias após a antese apresentam um bom padrão de uniformidade quanto a influência das características físicas e químicas dos frutos sobre os estádios de maturação (VIANNA-SILVA et al., 2008). Bem como neste trabalho, onde a extração das características físicas, físico-químicas e químicas das cagaitas em diferentes dias após antese, possibilitou compreender e estabelecer padrões para as cagaitas, possibilitando assim a classificação de estágio de maturação, por meio de grupos bem distintos, conseguindo determinar claramente as fases do fruto.

Figura 9 - Análise dos componentes principais das características físicas, físico-químicas e químicas das cagaitas em diferentes estádios de maturação. Em que: L = Luminosidade; DL = Diâmetro Longitudinal; DT = Diâmetro Transversal; AT = Acidez Titulável; SS = Sólidos Solúveis; FP = Firmeza da Polpa; FC = Firmeza da Casca e IM = Índice de Maturação.



Fonte: Próprio autor (2021).

4.2 Redes Neurais Artificiais

Os resultados de classificação e o desempenho do melhor modelo para a determinação de diferentes estádios de maturação de cagaitas são expostos a seguir.

4.2.1 Determinação do modelo com melhor otimizador e neurônios

Os otimizadores, na etapa de compilação e a quantidade de neurônios, na etapa de treinamento, bem como o número de épocas e de lotes, possibilitaram determinar os melhores resultados para todos os modelos de redes. Para cada modelo treinado foram obtidos parâmetros de perda e precisão dos dados de teste (Tabela 3). As perdas ficaram entre 1,44 e 1,30 e quanto a precisão, é possível observar que o maior valor obtido é 57,35% e o menor é 30,88%.

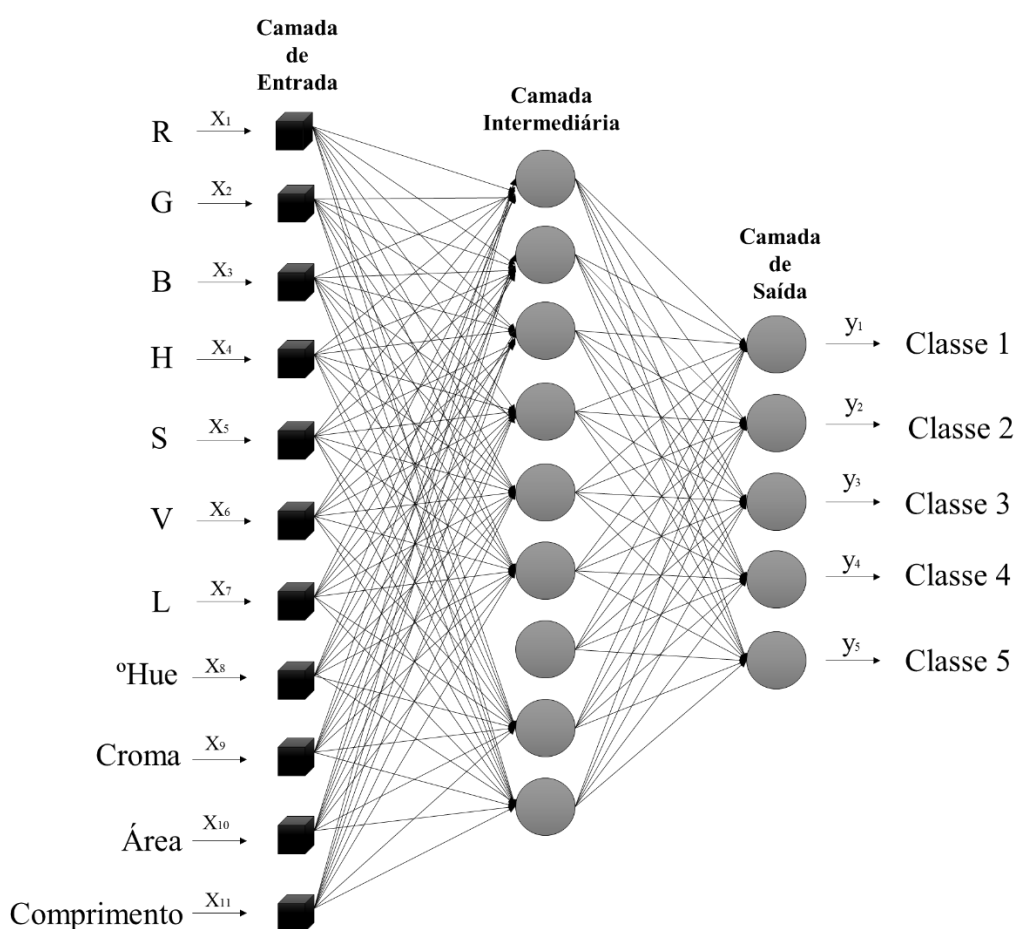
Tabela 3 - Perda e precisão dos dados de teste para diferentes quantidades de neurônios na camada oculta.

Otimizadores	Número de Neurônios	Perda	Precisão (%)
Adam	3	1,340	47,059
	4	1,323	44,118
	5	1,323	41,176
	6	1,398	39,706
	7	1,361	39,706
	8	1,339	47,059
	9	1,324	57,353
	10	1,402	45,588
	11	1,301	52,941
	12	1,451	39,706
Nadam	3	1,394	35,294
	4	1,383	30,882
	5	1,365	44,118
	6	1,374	45,882
	7	1,332	50,000
	8	1,303	41,176
	9	1,366	42,647
	10	1,380	51,471
	11	1,442	51,471
	12	1,290	52,941

Em negrito, o modelo final. Fonte: Próprio Autor (2021).

A rede (Figura 10) que apresentou a maior precisão foi a com o otimizador “adam”, 9 neurônios na camada oculta. E embora a perda dessa rede não tenha sido a menor, sendo a menor: 1,30 com o otimizador “nadam”, 4 neurônios na camada oculta, de acordo com Camargo e Sentelhas (1997) o critério mais adequado para determinar o melhor modelo é apresentado em termos de exatidão e precisão. Sendo assim, o melhor modelo deste trabalho, foi determinado base ao maior valor de precisão dos dados de teste, apresentando (57,35%), sendo este o modelo final.

Figura 10 - Rede Multi Layer Perceptron (MLP) do melhor modelo.



Fonte: Próprio Autor (2021).

Comparando o desempenho de cada rede, individualmente, dos dois otimizadores utilizados, quanto a precisão, é possível verificar que para o adam as redes com 3, 4, 8, 9 e 11 neurônios na camada oculta e para o nadam com 5, 6, 7, 10 e 12 neurônios na camada oculta, apresentaram melhores resultados. Observando assim, que para problemas de classificação semelhantes, pode-se utilizar qualquer um dos dois otimizadores. Ambos podem ser usados no

lugar do procedimento clássico de descida gradiente estocástico, para atualizar os pesos das redes de forma iterativa, baseando nos dados de treinamento (FALCÃO et al., 2019).

4.2.2 Fase de confirmação da rede de classificação de diferentes estádios de maturação da cagaita

A matriz de confusão do melhor modelo é apresentada na Figura 11. Na matriz de confusão é apresentado o desempenho do modelo dos dados de teste. As classes reais dos estádios de maturação do fruto de cagaita são apresentadas na horizontal e as previstas na vertical. Nela é possível verificar que, dos dados provenientes de 68 imagens (modelos de cores e atributos geométricos), sendo esse o número de informações resultante dos dados de teste para o treinamento das redes, nenhuma classe foi prevista totalmente correta. Todas as reais classes (na horizontal), em algum momento, apresentaram a predição de outras determinadas classes. Nota-se também que de uma classe para a outra, ocorreu diferença na quantidade de dados. Foram classificados 4 frutos na classe 1, 14 frutos na classe 2, 13 frutos na classe 3, 18 frutos na classe 4 e 15 frutos na classe 5. Do qual apenas 3, 8, 9, 8 e 11 desses frutos, correspondem as suas reais classes, respectivamente.

Figura 11 - Matriz confusão para a previsão de 5 diferentes estádios de maturação de cagaita, com otimizador “adam” e 9 neurônios na camada oculta.

		Classe Prevista				
		CLASSE 1	CLASSE 2	CLASSE 3	CLASSE 4	CLASSE 5
Classe Real	CLASSE 1	3	0	0	1	0
	CLASSE 2	0	8	0	0	1
	CLASSE 3	1	2	9	4	3
	CLASSE 4	0	2	4	8	0
	CLASSE 5	0	2	4	5	11

Fonte: Próprio autor (2021).

Apesar de um pouco mais da metade dos dados terem sido classificados de forma correta, a quantidade de dados que foram classificados indevidamente também foram consideravelmente alta. Teixeira (2018), aplicando RNA à automação de balanças comerciais de hortifrúti, verificou que a base de dados pequena pode ter influenciado em resultados ruins, isso devido a terem acarretado em um treinamento mal ajustado. O que justifica o ocorrido neste trabalho, já que a quantidade de dados influencia e nesse estudo a quantidade foi baixa, devido as condições do experimento. A aquisição de mais imagens a cada dia de análise demandaria o tempo necessário para a execução das demais análises. Um outro fator que influenciou nas condições dos dados aplicados a rede, foram os frutos ao longo dos dias de análise não terem sofrido tanta alteração quanto a coloração, influenciando diretamente nos parâmetros de modelos de cores. Sendo possível verificar uma alteração mais relevante apenas em seus tamanhos, bem como é possível observar na Figura 7.

Na Tabela 4 é apresentada a matriz confusão binária de cada classe, referente aos diferentes dias após antese, para a classificação de diferentes estádios de maturação. A matriz binária auxilia na visualização dos parâmetros (de cada classe): VP - Verdadeiro Positivo, que diz respeito aos frutos que foram classificados corretamente em suas devidas classes e o mesmo corresponde a diagonal principal da matriz confusão (Figura 11), VN - Verdadeiro Negativo, que é referente ao número de dados verdadeiramente reprovados para uma determinada classe, FP - Falso Positivo, que representa os dados que foram classificados de forma incorreta em uma determinada classe e FN - Falso Negativo, que refere-se aos dados que foram indevidamente rejeitados em uma determinada classe. Esses parâmetros são necessários para os cálculos de algumas métricas que auxiliarão nos resultados de desempenho do modelo selecionado para análise. De acordo com Saedi et al. (2021), para uma boa classificação, os valores máximos devem estar na diagonal principal, ou seja, os verdadeiros positivos e os zeros máximos fora da diagonal. Neste trabalho, os maiores valores apresentaram-se na diagonal principal, no entanto, devido a uma considerável ocorrência de falsas classificações, ocorreram poucos zeros fora da diagonal.

Tabela 4 - Matriz confusão binária para a previsão de 5 diferentes estádios de maturação de cagaita, com otimizador “adam” e 9 neurônios na camada oculta. Em que: C1 = Classe 1, C2 = Classe 2, C3 = Classe 3, C4 = Classe 4 e C5 = Classe 5.

		PREDITO			
		Classe		Não Classe	
ATUAL	Classe	VP =	3	C1	1
			8	C2	1
			9	C3	10
			8	C4	6
			11	C5	11
	Não Classe	FP =	1	C1	63
			6	C2	53
			8	C3	41
			10	C4	44
			4	C5	42

Fonte: Próprio autor (2021).

O relatório de classificação (Tabela 5) tem por finalidade apresentar as principais métricas de avaliação de cada classe da rede. Esta etapa ocorre para que seja possível comparar os resultados das classes. De acordo com Xie et al. (2021), a seleção destas métricas desempenham um papel significativo na avaliação do modelo. Quanto a acurácia das classes, todas apresentaram um valor significativamente bom, sendo que a classe 1 foi a que obteve o melhor resultado (97,1%). Isso significa que a classificação de cagaitas nesse estágio apresenta alto índice de estar correta e a taxa de dados classificados incorretamente é mínima. Já a classe que apresentou a menor acurácia no presente trabalho, foi a classe 3, com uma precisão geral de 73,5%. De acordo com Sabzi et al. (2018), a baixa precisão geral (acurácia) comprova a complexidade do problema em estudo, isso ocorre devido aos dados em análise possuírem valores de características muito próximos.

Tabela 5 - Relatório de classificação da rede para classificação de 5 diferentes estádios de maturação de cagaita.

	Acurácia	Precisão	Recall	F-Score	
	(%)	(%)	(%)	(%)	Apoio
Classe 1	97,1	75,0	75,0	75,0	4
Classe 2	89,7	57,1	88,9	69,6	9
Classe 3	73,5	52,9	47,4	50,0	19
Classe 4	76,5	44,4	57,1	50,0	14
Classe 5	77,9	7,33	50,0	59,5	22

Fonte: Próprio autor (2021).

Júnior Piedad et al. (2018), avaliando a Classificação pós-colheita de banana (*Musa acuminata*) usando aprendizado de máquina baseado em camadas, observou que todos o F-Scores apresentaram valores próximos aos valores apresentados de acurácia. Mostrando que a acurácia e a recuperação dos dados são próximas e que a métrica de acurácia para classificação pode ser utilizada como a mais adequada para avaliação. No presente trabalho, os valores de F-Score apresentaram-se muito inferiores ao de acurácia, com diferença entre ambos de até 26,5% (classe 3).

A precisão deste trabalho apresentou-se de forma mais efetiva na classe 1 (75%). A precisão mostra a capacidade de uma determinada classe em recusar um determinado dado positivo, quando o mesmo é negativo.

Quanto a métrica de recall, verificou-se no presente trabalho que a classe que apresentou melhor valor foi a classe 2 (88,9%). Isso quer dizer que foram classificadas em suas devidas classes, tudo que ocorreu positivamente na classe 2. Isso significa que boa parte dos frutos da classe 2, foram classificados de forma correta quando se tratavam da classe 2.

O apoio tem como intuito demonstrar quantos frutos de cada classe foram classificados em quaisquer que sejam as classes, seja ela a real classe daquele respectivo fruto ou se foi classificado de forma incorreta em outras classes. Sendo que, a classe real que apresentou a maior quantidade de frutos classificados, seja ele por quaisquer outras classes ou até por ela mesma foi a classe 5.

É possível verificar que há possibilidade de se classificar as cagaitas em diferentes estádios de maturação utilizando redes neurais artificiais, de forma efetiva, desde que alguns detalhes do experimento, sendo eles no campo e até a ajustes da rede, sejam adaptados de forma correta. Como por exemplo, a quantidade de imagens, a influência da coloração ou de outros fatores a serem avaliados. Ajustando principalmente o número de dados, aumentará a possibilidade de melhor aprendizagem da rede e assim será possível obter bons resultados.

4.3 Comparação de métodos (Estatística Multivariada e Redes Neurais Artificiais)

Base ao coeficiente de correlação cofenética determinado por meio da estatística multivariada e a precisão estabelecida por redes neurais artificiais, é possível verificar qual foi o método mais eficiente para determinar os diferentes estádios de maturação das cagaitas. Como mencionado anteriormente, o coeficiente de correlação cofenética é de $r = 0,8234$ e a precisão do melhor modelo é de 0,5735. Verificando-se assim que, o método mais eficiente para as condições do experimento, foi o de estatística multivariada. Este fato é atribuído as

características físicas, físico-químicas e químicas dos frutos terem apresentado maiores variações do que os parâmetros extraídos por meio do processamento digital de imagens e também devido ao fato da quantidade de dados para rede ter sido inferior ao que recomenda outros estudos. Estudos estes que chegam a trabalhar com até 3 vezes mais o valor trabalhado neste estudo.

5 CONCLUSÃO

Portanto, por meio da estatística multivariada e das redes neurais artificiais é possível classificar os diferentes estádios de maturação da cagaita.

Na análise de componentes principais (ACP's), 98,44% da variabilidade dos dados foi explicada pelo CP1, CP2 e CP3, sendo que, a maioria das variáveis (8 de 13 variáveis) foram explicadas pelo CP1. A análise de agrupamento gerou 3 grupos, em que o grupo 1 corresponde aos dois primeiros estádios de maturação que foram determinados por meio dos 28 e 31 DAA. No grupo 2, a tendência da maior parte das variáveis foi aumentar e esse grupo foi melhor representado pelo 40 DAA e o grupo 3 representou os 34 e 37 DAA. O cálculo do coeficiente de correlação cofenética para a validação do dendrograma foi de $r = 0,8234$.

Quanto as redes neurais artificiais, foram treinados 20 modelos, em que o melhor modelo foi determinado pela maior precisão (57,353%). No relatório das métricas a maior acurácia, precisão e R-score foram apresentados na Classe 1 (97,1%, 75% e 75%, respectivamente) e o melhor recall foi da Classe 2 (88,9%). As duas primeiras classes foram as que melhor classificaram as cagaita nos diferentes estádios de maturação.

Verificando o coeficiente de correlação cofenética (0,8234) da estatística multivariada e a precisão de determinação do melhor modelo da rede (0,57353) é possível concluir que para as condições deste estudo o método mais eficaz para a classificação foi a estatística multivariada.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABU-MOSTAFA, Y. S.; MALIK, M. I.; HSUAN-TIEN, L. **Learning from data**. v. 4. New York, AMLBook, 2012.
- ALMEIDA, A. B de.; SILVA, A. K. C. ; LODETE, A. R. ; EGEA, M. B. ; LIMA, M. C. P. M. ; SILVA, F. G. Assessment of chemical and bioactive properties of native fruits from the Brazilian Cerrado. **Nutrition & Food Science**, ISSN: 0034-6659, 2018.
- ALMEIDA, C. R de., OLIVEIRA, A. F. S de.; PEREIRA, S de. F. P.; FREITAS, K. H. G. ; SANTOS, D. C dos. ; COSTA, H. C da. ; SILVA, T de. M e.; ROCHA, R. M. Estatística multivariada aplicada na avaliação dos parâmetros químicos e físico-químicos da água engarrafada do município de Belém-PA. **Brazilian Journal of Development**, v. 6, n. 6, p.40002-40014, 2020.
- ALMEIDA, M. J de. **Cultivo consorciado de cagaiteira (*Eugenia dysenterica* DC): Influência na qualidade tecnológica dos frutos**. 2020. 53 f. Dissertação (Mestrado em Agroquímica) - Instituto Federal de Educação, Ciência e Tecnologia Goiano - Câmpus Rio Verde, Rio Verde, 2020.
- ALMEIDA JÚNIOR, E. B.; CHAVES, L. J.; SOARES, T. N. Caracterização genética de uma coleção de germoplasma de cagaiteira, uma espécie nativa do cerrado. **Bragantia**, Campinas, v. 73, n. 3, p. 246-252, 2014.
- ALVARENGA, A. M. L.; SILVA, R. E. V.; COSTA, P. M. C.; PINTO, D. M.; RODRIGUES, L. J. Physical and Chemical Parameters of “Cagaita” (*Eugenia dysenterica*) at Different Maturity Stages. **Journal of Applied Life Sciences International**, v. 22, n. 1, p. 1-9, 2019.
- ARAÚJO, F. P de. ; SOARES, M. F. ; SILVA, D. X da. **Investigação de compostos orgânicos dos frutos do cerrado tocantinense**. 11º Seminário de Iniciação Científica da Universidade Federal de Tocantins - Câmpus de Araguaína, Gurupi e Palmas, 2015.
- AROUCHA, E. M. M.; SOUZA, C. S. M de.; SOUZA, A. E. D de.; FERREIRA, R. M de. A.; AROUCHA FILHO, J. C. Qualidade pós-colheita da cajarana em diferentes estádios de maturação durante armazenamento refrigerado. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 34, n. 2, p. 391-399, 2012.
- ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS. **Official methods of analysis**. 19th Edition, Washington: AOAC, 2012.
- AZARMDEL, H.; JAHANBAKHSHI, A.; MOHTASEBI, S. S.; ROSADO, A. "Evaluation of image processing technique as an expert system in mulberry fruit grading based on ripeness level using artificial neural networks (ANNs) and support vector machine (SVM). **Postharvest Biology and Technology**, New York, v. 166, ISSN 111201, 2020.
- BARCA, M. C. S.; SILVEIRA, T. R. S.; MAGINI, M. Treinamento de Redes Neurais Artificiais: O Algoritmo Backpropagation. **IX Encontro Latino Americano de Iniciação Científica**. 2005.

- BRAZ, A.J.; NASCENTE, L.P.; CORRÊA, N.C.; ROCHA, R.A.; SOUZA, E.R.B.; SIQUEIRA, A.P.S. Influence of coverage based on biopolymers on the maturation of cagaita (*Eugenia dysenterica* DC.). **Revista de Agricultura Neotropical**, Cassilândia, v. 7, n. 2, p. 62-65, 2020.
- BRITO, M. A.; PEREIRA, E. B. C.; PEREIRA, A. V.; RIBEIRO, J. F. **Cagaita: biologia e manejo**. Planaltina. DF: Embrapa Cerrados. 80p, 2003.
- BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 1990. 105p.
- CAMARGO, A. P; SENTELHAS, P. C. Avaliação do desempenho de diferentes métodos de estimativas da evapotranspiração potencial no Estado de São Paulo, Brasil. **Revista Brasileira de Agrometeorologia**, Santa Maria, v.5, n.1, p.89–97, 1997.
- CAMILO, Y. M. V.; SOUZA, E. R. B.; VERA, R.; NAVES, R. V. Caracterização de frutos e seleção de progênies de cagaiteiras (*Eugenia dysenterica* DC.). **Científica**, v. 42 n.1, p. 1-10, 2014.
- CAMILO Y. M. V., SOUZA E.R.B., VERA R., NAVES R.V. Fenologia, produção e precocidade de plantas de *Eugenia dysenterica* visando melhoramento genético. **Revista de Ciências Agrárias**, Lisboa, v. 36, n. 2, p. 192-198, 2013.
- CARDOSO, M. R. D.; MARCUZZO, F. F. N.; BARROS, J. R. Classificação climática de Köppen-Geiger para o estado de Goiás e o Distrito Federal. **ACTA Geográfica**, Boa Vista, v.8, n.16, p. 40-55, 2014.
- CHAVES, L. J.; TELLES, M. P de. C. **Cagaita. Futas Nativas da Região Centro-Oeste do Brasil**. Embrapa Informações tecnológicas, Brasília, p. 120-134, 2010.
- CHAVES, L. J.; TELLES, M. P de. C. **Espécies nativas da flora brasileira de valor econômico atual ou potencial: plantas para o futuro: Região Centro-Oeste. *Eugenia dysenterica* - Cagaita**. Embrapa Informações tecnológicas, Brasília, p. 216-223, 2018.
- CHAVES NETO, J. R.; SILVA, S. de M.; DANTAS, R. L. Atributos de qualidade, compostos bioativos e atividade antioxidante de frutos de uvaieira durante a maturação. **Agrarian**, Dourados, v. 13, n. 49, p. 296-308, 2020.
- CHITARRA, M.I.F.; CHITARRA, A.B. 2005. **Pós-colheita de frutas e hortaliças: fisiologia e manuseio**, 2.ed. UFLA, Lavras.
- COELHO, A. A.; CENCI, S. A.; RESENDE, E. D de. Qualidade do suco de maracujá-amarelo em diferentes pontos de colheita e após o amadurecimento. **Ciências agrotecnológicas**, Lavras, v. 34, n. 3, p. 722-729, 2010.
- COMPANHIA DE ENTREPOSTOS E ARMAZÉNS GERAIS DE SÃO PAULO - CEAGESP. **A medida da doçura das frutas**. São Paulo: Centro de qualidade, pesquisa e desenvolvimento. Cartilha técnica 08, 2016. 17 p.

COSTA, D. C da. **Uso de imagens digitais na avaliação da qualidade pós-colheita de frutas de caroço**. 2018. 100 f. Dissertação (Mestrado em Tecnologia da Produção Agrícola). Agricultura Tropical e Subtropical - Instituto Agrônomo, Campinas, 2018.

COSTA, J. D de. S. **Métodos não destrutivos na avaliação da maturação de mangas ‘palmer’ e simulação de danos decorrentes do transporte**. 2019. 109 f. Tese (Doutorado em Engenharia Agrícola). Centro de Tecnologia e Recursos Naturais - Universidade Estadual de Campina Grande, Campina Grande, 2019.

COSTA, A. G. **Identificação da maturação de frutos da macaúba por meio de características óticas**. 2015. 126 f. Tese (Doutorado em Engenharia Agrícola). Universidade Federal de Viçosa. Viçosa, 2015.

DAZA, L.D.; FUJITA, A.; FÁVARO-TRINDADE, C.S., et al. Effect of spray drying conditions on the physical properties of Cagaita (*Eugenia dysenterica* DC.) fruit extracts. **Food and bioproducts processing**, Monterrey, v.97, n.1, p.20-29, 2016.

DREHMER, A.M.F.; AMARANTE, C.V.T. Conservação pós-colheita de frutos de araçá-vermelho em função do estágio de maturação e temperatura de armazenamento. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 30, n. 2, p. 322-326, 2008.

DUARTE, J. B. **Princípios e utilização de técnicas multivariadas no melhoramento de plantas**. 1997. 69 f. Monografia (Conclusão de Disciplina em Genética e Melhoramento de Espécies Alógamas), Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1997.

FALCÃO, J. V. R.; MOREIRA, V de. A.; SANTOS, F. A. O.; RAMOS, C de A. Redes neurais deep learning com tensorflow. **RE3C - Revista Eletrônica Científica de Ciência da Computação**, v. 14, n. 1, [s.l], 2019. Disponível em: <https://revistas.unifenas.br/index.php/RE3C/article/view/232https://revistas.unifenas.br/index.php/RE3C/article/view/232>. Acesso em: 14 nov. 2021.

FARIAS, E. T do. R.; SANTOS, A. F dos.; LOPES, M. F.; BEZERRA, J. M.; SILVA, F. V. G da. Componentes biométricos, físico-químicos, bioativos e antioxidantes em araçás em diferentes estágios de maturação. **Research, Society and Development**, v. 9, n.8, e331985800, 2020.

FERREIRA, D. C. M.; MOLINA, G.; PELISSARI, F. M. Effect of Edible Coating from Cassava Starch and Babassu Flour (*Orbignya phalerata*) on Brazilian Cerrado Fruits Quality. **Food and Bioprocess Technology**, Dublin, v. 13, [s.n], p. 172–179, 2019.

FERREIRA, D. F. **Análise Multivariada**. Lavras: UFLA, 1996. 200 p. Universidade Federal de Lavras - Departamento de Ciências Exatas. 1996.

FIGUEIREDO, Y. F. C de.; CAMPOS, L. M. L de. **Aplicação de Técnicas de Mineração de Dados em problemas de Classificação**. XVIII Simpósio de Excelência em Gestão e Tecnologia. XVIII SEGTe, 2019.

FURTADO, M. I. V. **Redes Neurais Artificiais: Uma Abordagem Para Sala de Aula**. Ponta Grossa: Editora Atena, 2019.

GAMA, A. T da. **Desempenho agrônômico, divergência genética, fenotipagem de alta eficiência e qualidade de sementes de variedades crioulas de feijão-fava cultivadas no Semiárido Norte Mineiro**. 2020. 87 f. Tese (Doutorado em Produção Vegetal) - Instituto de Ciências Agrárias, Universidade Federal de Minas Gerais, Montes Claros, 2020.

GARCIA, L. G. C. **Desenvolvimento fisiológico e conservação pós-colheita de jabuticaba**. 2017. 97 f. Tese (Doutorado em Agronomia) - Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2017.

HAIR, J. F.; BLACK, W.; BABIN, B.; ANDERSON, R. E.; TATHAM, R. L. **Análise Multivariada de dados**. Editora Bookman, Porto Alegre, 6ª ed., 2009. 688 p.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, Boston, v. 23, n. 3, p. 187- 200, 1958.

KHEIRALIPOUR, K.; PORMAH, A. Introducing new shape features for classification of cucumber fruit based on image processing technique and artificial neural networks. **Journal of Food Process Engineering**, e12558, 2017.

LATTUADA, D. S.; BARROS, N.; HAGEMANN, A. SOUZA, V. D de. Caracterização físico-química e desenvolvimento pós-colheita de jabuticabas (*Plinia peruviana* e *P. cauliflora*). **Série Botânica**, Porto Alegre, 75: e2020015, 2020.

LAWSON, T.; LYCETT, G. W.; ALI, A.; CHIN, C. F. Characterization of southeast asia mangoes (*Mangifera indica* L) according to their physicochemical attributes. **Scientia Horticulturae**, Dresden, v. 243, [s.n], p. 189-196, 2019.

LIMA, F. P dos. A.; MINUSSI, C. R. Usando uma rede neural de Kohonen para reconhecimento de padrões de som. **Omnia Exatas**, v.4, n.2, p.19-30, 2011.

LIMA, T. L. S.; CAVALCANTE, C. L.; SOUSA, D. G.; SILVA, P. H. A.; SOBRINHO, L. G. A. Avaliação da composição físico-química de polpas de frutas comercializadas em cinco cidades do Alto Sertão paraibano. **Revista Verde**, Cataguases, v.10, n.2, p. 49-55, 2015.

LOBO, M. G.; PAULL, R. E. **Handbook of Pineapple Technology: Production, Postharvest Science, Processing and Nutrition**: 1. Ed. Hoboken: Wiley-Blackwell, 2017.

LOESCH, C.; HOELTGEBAUM, M. **Métodos Estatísticos Multivariados**. 1.Ed. Rio de Janeiro: Saraiva Educação S.A., 2017.

MANLY, B. F. J. **Multivariate statistical methods**, a primer. New York: Chapman Hall, 159 p., 1986.

MARTINOTTO, C.; PAIVA, R.; SOARES, F.P.; SANTOS, B.R.; NOGUEIRA, R.C. 2008. Cagaiteira (*Eugenia dysenterica* DC.). Universidade Federal de Lavras, Lavras.

MATOS, W. L. N.; LOPES, C. H. G.; PORTAL, L. M. M.; SILVA, O. F. **REDES NEURAIIS ARTIFICIAIS**: fundamentos e aplicações em Engenharia Elétrica. VII Encontro

Regional dos Grupos do Programa de Educação Tutorial da região Norte - VII NORTEPET, UNIR, 2020.

MATTOS, A do. P. **Determinação da severidade de doenças de plantas por meio de imagens digitais**. 2017. 63 f. Trabalho de Conclusão de Curso (Bacharel em Agronomia) - Campus de Curitibanos, Universidade Federal de Santa Catarina, Curitibanos, 2017.

MAZEN, F. M. A.; NASHAT, A. A. Ripeness Classification of Bananas Using an Artificial Neural Network. **Arabian Journal for Science and Engineering**, Dhahran v. 44, [s.n], p. 6901–6910, 2019.

MELO, F. R de. **Modelo Neural por padrões proximais de aprendizagem para automação personalizada de conteúdos didáticos**. 2012. 173 p. Tese (Doutorado em Ciências), Faculdade de Engenharia Elétrica - Universidade Federal de Uberlândia, Uberlândia, 2012.

MELO, R.S.; SILVA, S.M.; LIMA, R.P.; SOUSA, A.S.B.; DANTAS, R.L.; DANTAS, A.L. Qualidade física e físico-químicas de frutos de mandacaru (*Cereus jamacaru* P. DC.) colhidos na região do Curimataú Paraibano. Congresso Brasileiro de Processamento mínimo e Pós-colheita de frutas, flores e hortaliças, 2015, 001.

MIGUEL, A. C. A.; DURIGAN, J. F.; FERRAUDO, A. S. Técnicas de análise multivariada na avaliação de injúrias pelo frio em mangas cv. Tommy Atkins. **Revista Brasileira de Fruticultura**, Jaboticabal , v.e. 371-376, 2011.

MINGOTI, S.A.; **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. 3ª Edição. Belo Horizonte, Editora UFMG. 2013, 297p.

MORAIS, M de. S. **Caracterização bioquímica e biofísica da fração proteica do fruto da cagaita (*Eugenia Dysenterica*)**. 2020. 57 p. Dissertação (Mestrado em Química) - Intituto de Química, Universidade de Brasília, Brasília, 2020.

MORENO, J. L.; TRAN, T.; CANTERO-TUBILLA, B.; LÓPEZ-LÓPEZ, K.; Physicochemical and physiological changes during the ripening of Banana (*Musaceae*) fruit grown in Colombia. **International Journal of Food Science and Technology**, Oxford, v. 56, [s.n], p. 1171-1183, 2021.

MOTA, E. E. S. **Caracterização fenotípica e variação genética quantitativa em *Dipteryx alata* Vog**. 2013. 82 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2013.

NASCIMENTO, A. L. A. A.; BRANDI, I. V.; DURÃES, C. A. F.; LIMA, J. P de.; SOARES, S. B.; MESQUITA, B. M. A de. C. Chemical characterization and antioxidant potential of native fruits of the Cerrado of northern Minas Gerais. **Brazilian Journal of Food Technology**, Campinas, 23, e2019296, 2020.

NAVES, R. V.; BORGES, J. D.; CHAVES, L. J. A cagaiteira. **Revista Brasileira de Fruticultura**, v. 24, n. 2, p. 0, 2002.

NERIS, T. S.; SILVA, N. S.; LOSS, S. R. A.; CARVALHO, W. P.; GUEDES, S. F. Avaliação físico-química da casca da banana (*Musasp.*) *in natura* e desidratada em diferentes estádios de maturação. **Ciência e Sustentabilidade**, Juazeiro do Norte, v. 4, n.1, p. 5-21, 2018.

ORDÓÑEZ, J. A. Tecnologia de Alimentos: Componentes dos Alimentos e Processos, 1ª Edição. Porto Alegre, Editora Artmed, 2004.

OTSU, N. A Threshold Selection Method from Gray-Level Histogram. **IEEE Transactions on System Man Cybernetics**, v. 9, n. 1, p. 62-66, 1979.

PAZ, F. L. B da.; MATIAS, A. O.; SANTOS, M. F dos.; NETO, W. V de. A.; COSTA, M. F. Caracterização biométrica de frutos e sementes de cagaita (*Eugenia dysenterica* DC.) na região sul do Piauí, Brasil. **Diversitas Journal**, v. 5, n. 3, p. 1432-1441, 2020.

PEDRYCZ, W.; GOMIDE, F. **Fuzzy systems engineering: toward human-centric computing**. [S.l.]: John Wiley & Sons, p. 34, 2007.

PERFEITO, D.G.A.; CORRÊA, I.M.; PEIXOTO, N. Elaboração de bebida com extrato hidrossolúvel de soja saborizada com frutos do cerrado. **Journal of Neotropical Agriculture**, Mato Grosso do Sul, v. 4, n. 1, p. 21-27, 2017.

PROENÇA, C.E.B.; GIBBS, P.E. Reproductive biology of eight sympatric Myrtaceae from Central Brazil. **New Phytologist**, v. 126, [s.n], p. 343-354, 1994.

JÚNIOR PIEDAD, E.; LARADA, J. I.; POJAS, G. J.; FERRER, L. V. V. Postharvest classification of banana (*Musa acuminata*) using tier-based machine learning. **Postharvest Biology and Technology**, New York, v. 145, [s.n], p. 93-100, 2018.

RAMOS, J. P. S. Redes neurais artificiais na classificação de frutos: cenário bidimensional. **Ciências Agrotécnicas**, Lavras, v. 27, n.2, p. 356-362, 2003.

RAMPAZZO, M. C.; NUNES, R de. C. S.; DUTRA, F. V.; LEITE, G. S.; CARDOSO, A. D. Características físico-químicas de frutos de noni physiochemical characteristics of noni fruits. In: III Congresso Internacional das Ciências Agrárias COINTER - PDVAGRO, 2018.

REIS, A. F.; SCHIMIELE, M. Características e potencialidades dos frutos do Cerrado na indústria de alimentos. **Brazilian Journal of Food Technology**, Campinas, v. 22, e20171502, 2019.

Ribeiro, E.M.G. **Atividade antioxidante e polifenóis totais do fruto de cagaita (*Eugenia dysenterica* DC.) com e sem casca**. Dissertação (Mestrado em Farmácia). 2011. 77f. Universidade Federal do Rio de Janeiro, Faculdade de Farmácia. Rio de Janeiro, RJ.

RIBEIRO, S. S. **Métodos computacionais aplicados a agricultura: Python & Weka**. Rio de Janeiro: Nova Terra, 2016. 122p.

RODRIGUES, B. P. **Reamostragem em redes neurais com aplicação a dados espaciais**. 2021, 30 f. Dissertação (Mestrado em Agricultura e Informações Geoespaciais), Programa de

Pós-Graduação em Agricultura e Informações - Universidade Federal de Uberlândia, Uberlândia, 2021.

RODRIGUES, C. G.; CORREIA, V. T da. V. C.; BERNADES, N. R.; OLIVEIRA, D. B de.; MATRANGOLO, W. J. R.; SILVA, E. C da.; CARLOS, L da, A. **Compostos bioativos, atividade antioxidante e características físicas e químicas de frutos de cagaiteira nativos do cerrado mineiro: Ensino, Pesquisa e Extensão em Engenharia de Alimentos**, 2021. p. 288-302.

ROSA, A. L da. **Classificação de Imagens de Frutas utilizando Aprendizado de Máquina**. 2019. 74 f. Trabalho de Conclusão de Curso (Engenharia Eletrônica), Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2019.

RStudio (2021). RStudio: Desenvolvimento integrado para R. RStudio, PBC, Boston, MA. Disponível em: <<http://www.rstudio.com/>>.

SABZI, S.; ABBASPOUR-GILANDEHA, Y.; GARCÍA-MATEOS, G. A new approach for visual identification of orange varieties using neural networks and metaheuristic algorithms. **Information Processing in Agriculture**, Pequim, v. 5, n. 1, p. 162-172, 2018.

SAEDI, S. I.; KHOSRAVI, H. A deep neural network approach towards real-time on-branch fruit recognition for precision horticulture. **Expert Systems with Applications**, Louisiana, v. 159, ISSN 113594, 2020.

SALOMÃO, L. C. C.; SIQUEIRA, D. L. DE; PEREIRA, M. E. C. Desenvolvimento do fruto da lichieira (*Litchi chinensis* Sonn.) ‘Bengal’. **Revista Brasileira de Fruticultura, Jaboticabal**, v. 28, n.1, p. 11-13, 2006.

SANCHES, J.; LINO, A.C.L. **Uso de imagem digital para seleção e classificação de frutas e hortaliças**. 2010. Artigo em Hipertexto. Disponível em: <http://www.infobibos.com/Artigos/2010_1/imagem/index.htm>. Acesso em: 02 dez. 2021.

SANTIAGO, W. E. E.; LOPES, G. D.; FACETO, A. D.; LOBO JÚNIOR, A. R.; CALDEIRA, R. F. Classificação de laranjas por meio de processamento digital de imagens. **Acta Iguazu**, Cascavel, v.8, n.2, p. 49-58, 2019.

SANTOS, E. F. N.; SOUSA, I. F. Análise estatística multivariada da precipitação do Estado de Sergipe através dos fatores e agrupamentos. **Revista Brasileira de Climatologia**, v. 23, n. 14, p. 205-222, 2018.

SANTOS, M da. S. **Impacto do processamento sobre as características físico-químicas, reológicas e funcionais de frutos da gabirobeira (*Campomanesia xanthocarpa* Berg)**. 2011. 148 f. Tese (Doutorado em Tecnologia de Alimentos). Universidade Federal do Paraná - Setor de Tecnologia, Curitiba, 2011.

SANTOS, W. W. V.; SILVA, K. R de. O.; BARBOSA, R. C.; OLIVEIRA, J. B de.; SILVA, J. S. A da.; MEDEIROS, E. V de. Efeito de diferentes métodos de maturação sobre a qualidade da banana prata. **Diversitas Journal**, Santana do Ipanema, v. 4, n. 3, p. 1092-1104, 2019.

SAS. **Redes Neurais**: O que são e qual sua importância?. 2020. Disponível em: <https://www.sas.com/pt_br/insights/analytics/neural-networks.html>. Acesso em: 02 dez. 2021.

SCARIOT, A.; RIBEIRO, J. F. **Boas Práticas de Manejo para o Extrativismo Sustentável da Cagaita**. Embrapa Recursos Genéticos e Biotecnologia, Brasília, 71 p., 2015.

SEMPREBON JUNIOR, E. J. **Emprego da análise multivariada na classificação de polpas de frutas de acordo com as propriedades físicas e a difusividade térmica**. 2018. 34 f. Trabalho de Conclusão de Curso (Bacharel em Engenharia de Alimentos), Departamento Acadêmico de Alimentos, Universidade Tecnológica Federal do Paraná, Campo Mourão, 2018.

SIDEHABI, S. W.; SUYUTI, A.; ARENI, I. A.; NURTANIO, I. Classification on Passion Fruit's Ripeness using K-Means Clustering and Artificial Neural Network. International Conference on Information and Communications Technology (ICOIACT), p. 304-309, 2018.

SILVA, C. S da. **Determinação do ponto de colheita da manga 'ubá' (*Mangifera indica* L. var. Ubá) pela análise de suas propriedades físico-químicas**. 2021. 64 f. Dissertação (Mestrado em Engenharia Agrícola). Universidade Federal de Viçosa, Viçosa, 2021.

SILVA, E. D. **Redes neurais artificiais e processamento de linguagem Natural aplicados à previsão do minicontrato futuro do índice Ibovespa**. 2021. 115 f. Dissertação (Mestrado em Engenharia Elétrica), Centro Universitário Fei, São Bernado do Campo, 2021.

SILVA, M. M. M da.; SILVA, E. P.; GARCIA, L. G. C.; ASQUIERI, E. R.; VILAS BOAS, E. V. B.; SILVA, A. P. G da.; XIAO, J.; DAMIANI, C. Bioactive Compounds and Nutritional Value of Cagaita (*Eugenia dysenterica*) during its Physiological Development. **eFood**, v. 1, n. 4, p. 288–297, 2020.

SILVA, M. M. M da. **Estudo do desenvolvimento fisiológico da cagaita (*Eugenia dysenterica*)**. 2016. 120 f. Dissertação (Mestrado em Agronomia). Escola de Agronomia - Universidade Federal de Goiás, Goiânia, 2016.

SILVA, P. R da.; MEDONÇA, H de. O. P.; NOGUEIRA, L. A.; PAULA, A. C. C. F. F de.; SANTOS, M. N. S.; AUGUSTI, R.; ARAÚJO, R. L. B de.; SILVA, M. R.; REINA, L. D. C. B.; MELO, J. O. F. Avaliação biométrica e físico-química e estudo do perfil químico da *Eugenia dysenterica*. **Ciências Agrárias: o avanço da ciência no Brasil**, Belo Horizonte, v.1, [s.n], p. 367-388, 2021.

SILVA-RODRIGUES, P. L.; AMORIM, G. C.; ANDRADE, I. E. P. C.; CUNHA, V. A.; FIGUEIREDO, L. H. M.; GROSSI-DE-AS, M. F. Monitoramento Tecnológico da Planta Cagaita (*Eugenia dysenterica*) e Aplicações Biotecnológicas Potenciais. **Cadernos de Prospecção**, Salvador, v. 14, n. 4, p. 1248-1264, 2021.

SILVA, R. S. M. **Caracterização de sub-populações de Cagaita (*Eugenia dysenterica* DC.) da região sudeste do estado de Goiás, Brasil**. 1999. 107 f. Dissertação (Mestrado) – Universidade Federal de Goiás, Goiânia, 1999.

SILVA, R. V da. **Identificação de defeitos superficiais em tomates (*Solanum lycopersicum*) utilizando técnicas de processamento de imagens**. 2015. 56 f. Dissertação (Mestrado em Máquinas Agrícolas), Faculdade de Engenharia Agrícola - Universidade Estadual de Campinas, Campinas, 2015.

SILVA, T. R da.; COSTA, A. G.; PAES, J. L.; OLIVEIRA, M. M de.; PINTO, F de. A de. C. Comparando os parâmetros de qualidade obtidos usando destrutivos e ópticos métodos de classificação de tomates. **Revista Ciência Agronômica**, Fortaleza, v. 52, n. 3, e20207475, 2021.

SONG, F. N. **Técnicas de análise multivariada com aplicações a dados de natureza biológica**. 2013. 67 f. Trabalho de Conclusão de Curso (Bacharel em Ciências Biológicas), Instituto de Biociência - Universidade Estadual Paulista, Rio Claro, 2013.

SOUZA, C. C de.; REIS NETO, J. F.; ARAUJO, L. O.; NUNES, H. S.; CONCEIÇÃO, L. R. Utilização de redes neurais artificiais para previsão de preços de frutas e hortaliças. **Desenvolvimento e Gestão**, v.17, n.3, p.259-272, 2014.

SOUZA, E. R. B de.; NAVES, R. V.; BORGES, J. D.; VERA, R.; FERNANDES, E. P.; SILVA, L. B e.; TRINDADE, M da. G. Fenologia de cagaiteira (*Eugenia dysenterica* DC.) no Estado de Goiás. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 30, n. 4, p. 1009-1014, 2008.

SOUZA, E. R. B de.; NAVES, R. V.; OLIVEIRA, M. F. início da produção de frutos de cagaiteira (*Eugenia dysenterica* DC) implantada em Goiânia, Goiás. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 35, n. 3, p. 906-909, 2013.

STEINGASS, C.B.; JUTZI, M.; MÜLLER, J.; CARLE, R. Ripening-dependent metabolic changes in the volatiles of pineapple (*Ananas comosus* (L.) Merr.) fruit: II. Multivariate statistical profiling of pineapple aroma compounds based on comprehensive two-dimensional gas chromatography-mass spectrometry. **Analytical and Bioanalytical Chemistry**, [s.l.], v. 407, n. 9, p. 2609-2624, 2015.

TEIXEIRA, I. R da. S. **Deep learning aplicado à automação de balanças comerciais de hortifrutis**. 2018. 41 f. Trabalho de Conclusão de Curso (Bacharel em Ciências da Computação). Instituto de Ciências Exatas e da Terra - Universidade Federal de Mato Grosso, Barra do Garças, 2018.

TERÁN-ORTIZ, G. P.; SILVA, M. G da.; LIMA, H. C de.; PACIULLI, S de. O. D.; SILVA, V. A. Avaliação pós-colheita de cagaíta (*Eugenia dysenterica*). In : Congresso Brasileiro de Processamento de Frutas e Hortaliças, 3. 2013. Ilheus : [s.n], 2013.

TOKAIRIN, T de. O. **Caracterização e conservação pós-colheita de cambuci, fruto nativo da Mata Atlântica**. 2017. 134 f. Tese (Doutorado em Ciências). Escola Superior de Agricultura 'Luiz de Queiroz' - Universidade de São Paulo, Piracicaba, 2017.

THAI, C. N.; SHEWFELT, R. L. Modeling sensory color quality of tomato and peach: neural networks and statistical regression. **Transactions of the ASAE**, Saint Joseph, v. 34, n. 3, p. 950-954, 1991.

TUCKER, G.A. Introduction. In: SEYMOR, G.B.; TAYLOR, J.E.; TUCKER, G.A. **Biochemistry of fruit ripening**. New York: Chapman & Hall, 1993. p. 2-51.

VASQUES, H. S. **Tempo meteorológico e ritmo climático: efeitos sobre a morbidade e a mortalidade causadas por doenças cardiovasculares no município de Goiânia – GO**. 217 f. Tese (Doutorado em Geografia) - Instituto de Estudos Socioambientais - Universidade Federal de Goiás, Goiânia, 2021.

VIANNA-SILVA, T. RESENDE, E. D de. ; VIANA, A. P. ; PEREIRA, S. M de. F. ; CARLOS, L de. A. ; VITORAZI, L. Qualidade do suco de maracujá-amarelo em diferentes épocas de colheita. **Ciência e Tecnologia de Alimentos**, Campinas, v. 28, n. 3, p. 545-550, 2008.

VICINI, L. **Análise multivariada da teoria à prática**. 2005. 215 f. Trabalho de Conclusão de Curso (Licenciatura em Matemática), Universidade Federal de Santa Maria, Santa Maria, 2005.

VILAS BOAS, B. M.; ALVES, A de. P.; ALVES, J. A.; RODRIGUES, L. J.; ALVES, T. C.; VILAS BOAS, E. V de. B. Caracterização física, química e bioquímica do mesocarpo interno de frutos do pequizeiro colhidos em diferentes estádios de desenvolvimento. **Ciência Rural**, Santa Maria, v.43, n.12, p.2285-2290, 2013.

VINUESA, R.; AZIZPOUR, H.; LEITE, I.; BALAAM, M.; DIGNUM, V.; DOMISCH, S.; FELLÄNDER, A.; LANGHANS, S. D.; TEGMARK, M.; NERINI, F. F. The role of artificial intelligence in achieving the Sustainable Development Goals. **Nature Communications**, v. 11, n. 1, p. 233, dez. 2020.

WARD, J. H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, New Haven, v. 58, [s.n.], p. 236 – 244. 1963.

WILLS, R.H.; McGLASSON, W.B.; GRAHAM, D.; JOYCE, D. **Postharvest**: an introduction to the physiology and handling of fruit, vegetables and ornamentals. 4th ed. New York: CAB International, 1998. 262 p.

XIE, W.; WEI, S.; ZHENG, Z.; JIANG, Y.; YANG, D. Recognition of Defective Carrots Based on Deep Learning and Transfer Learning. **Food and Bioprocess Technology**, Dublin, v. 14, [s.n], p. 1361-1374, 2021.

YACIM, J.; BOSHOF, D. Impact of Artificial Neural Networks Training Algorithms on Accurate Prediction of Property Values. **Journal of Real Estate Research**, v. 40, p. 375–418, 2018.