



UNIVERSIDADE ESTADUAL DO SUDOESTE DA BAHIA – UESB
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E
CIÊNCIA DE ALIMENTOS
Área de Concentração: Engenharia de Alimentos



IDENTIFICAÇÃO DA QUALIDADE E ORIGEM DE AMÊNDOAS DE
CACAU PRODUZIDAS NO ESTADO DA BAHIA USANDO
QUIMIOMETRIA

Autor: Acsa Santos Batista

Orientador: Prof. DSc. Leandro Soares Santos

ITAPETINGA - BA
FEVEREIRO 2018

ACSA SANTOS BATISTA

**IDENTIFICAÇÃO DA QUALIDADE E ORIGEM DE AMÊNDOAS DE
CACAU PRODUZIDAS NO ESTADO DA BAHIA USANDO
QUIMIOMETRIA**

Dissertação apresentada como requisito para obtenção do grau de Mestre em Engenharia e Ciência de Alimentos, área de concentração em Engenharia de Alimentos, da Universidade Estadual do Sudoeste da Bahia - UESB.

Orientador: Prof. DSc. Leandro Soares Santos

Co-orientadores: Prof. DSc. Ivan Oliveira Pereira

Prof^a. DSc^a. Silmara Carvalho

**ITAPETINGA - BA
FEVEREIRO 2018**

633.74 Batista, Acsa Santos

B336i Identificação da qualidade e origem de amêndoas de cacau produzidas no estado da Bahia usando quimiometria. / Acsa Santos Batista. - Itapetinga: UESB, 2018.
75p.

Dissertação apresentada como parte das exigências para obtenção do Título de Mestre em Engenharia e Ciência de Alimentos, no Programa de Pós-Graduação em Engenharia e Ciência de Alimentos da Universidade Estadual do Sudoeste da Bahia. Sob a orientação do Prof. D.Sc. Leandro Soares Santos e coorientação do Prof. D.Sc. Ivan Oliveira Pereira e Silmara Carvalho.

1. Amêndoas de Cacau - Quimiometria. 2. Amêndoas de cacau - Qualidade. 3. Amêndoas de cacau – Origem. I. Universidade Estadual do Sudoeste da Bahia. Programa de Pós-Graduação em Engenharia e Ciência de Alimentos. II. Santos, Leandro Soares. III. Pereira, Ivan Oliveira. IV. Carvalho, Silmara. V. Título.

CDD(21): **633.74**

Catálogo na fonte:

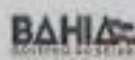
Adalice Gustavo da Silva – CRB/5-535
Bibliotecária – UESB – Campus de Itapetinga-BA

Índice Sistemático para Desdobramento por Assunto:

1. Amêndoas de cacau - Quimiometria
2. Amêndoas de cacau - Qualidade
3. Amêndoas de cacau – Origem



Universidade Estadual do Sudoeste da Bahia
Recredenciada pelo decreto estadual nº 15.825 de 04.07.2016
Programa de Pós-Graduação em Engenharia e Ciência de Alimentos
PPG-ECAL / UESB



Áreas de Concentração: Engenharia de Alimentos
Ciência de Alimentos

DECLARAÇÃO DE APROVAÇÃO

Título: IDENTIFICAÇÃO DA QUALIDADE E ORIGEM DE AMÊNDOAS DE CACAU PRODUZIDAS NO ESTADO DA BAHIA USANDO QUIMIOMETRIA.

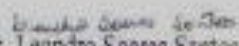
Autor (a): ACSA SANTOS BATISTA


Orientador (a): Prof. Dr. Leandro Soares Santos

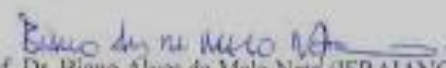
Coorientador (a): Prof. Dr. Ivan Oliveira Pereira

Prof.^a Dr.^a Silmara Almeida de Carvalho

Aprovada como parte das exigências para obtenção do Título de **MESTRE EM ENGENHARIA E CIÊNCIA DE ALIMENTOS, ÁREA DE CONCENTRAÇÃO: ENGENHARIA DE ALIMENTOS**, pela Banca Examinadora.


Prof. Dr. Leandro Soares Santos (UESB)


Prof. Dr. Sérgio de Souza Castro (UESB)


Prof. Dr. Bianco Alves de Melo Neto (IFBAIANO/ URUÇUCA)

Itapetinga-BA, 28 de fevereiro de 2018.

*Independemente das circunstâncias,
devemos ser sempre humildes, recatados e
despidos de orgulho.*

Dalai Lama

A Deus

Aos meu pais, minha base

A minha irmã

Ao meu namorado Antonio

DEDICO!

AGRADECIMENTOS

Agradeço primeiramente à Deus, pela vida, pela saúde, pela fé, pela coragem, pela proteção em todos os momentos.

À minha mãe, pelo enorme cuidado, preocupação, amor e grande exemplo de conquistas.

Ao meu pai, pela alegria, simplicidade e amor transmitidos. Sempre essencial.

À minha irmã e meu cunhado pelo carinho.

À minha família, que sempre acreditou em mim, incentivando a prosseguir e alcançar o alvo.

Ao meu amor Antonio Carlos que sempre esteve ao meu lado, me apoiando, acreditando no alcance das vitórias, pelo companheirismo e compreensão!

À Universidade Estadual do Sudoeste da Bahia (UESB), em especial o LABMESQ pela oportunidade de desenvolver este trabalho.

Aos alunos de Iniciação Científica Júlia, Antonia e Esdras e a amiga desde a graduação, Josane pelo apoio e incentivo de sempre.

Ao IF Baiano - *campus* Uruçuca-BA e a CEPLAC pelo grande apoio desde o início do projeto.

Ao meu querido orientador, Leandro Soares, pela amizade, orientação, ensinamentos e pela contribuição intelectual que sempre foram fundamentais.

Ao meu co-orientador, professor DSc. Ivan Pereira pela disponibilidade, pelo direcionamento e pela indispensável contribuição para esta pesquisa.

Aos membros da banca examinadora por aceitarem fazer parte da avaliação deste trabalho.

Aos colegas de Mestrado, em especial Meline, Fabiola, Polyany, Josiane, Mateus e Karine pelas trocas de conhecimentos e por tornarem esses anos mais divertidos e a Daniele pela companhia nas madrugadas e na multivariada, dividindo sofrimentos e alegrias. Foram essenciais!

As meninas da república, pelo apoio!

À uma pessoa querida que tive a oportunidade de conhecer, Valdirene. Uma bênção de Deus pra mim.

À Capes pela concessão da bolsa.

À todos aqueles que de forma direta e indireta contribuíram para a realização deste trabalho.

Meus sinceros agradecimentos!

SUMÁRIO

LISTA DE FIGURAS.....	xi
LISTA DE TABELAS.....	xii
RESUMO.....	xiii
SUMMARY.....	xiv
1. INTRODUÇÃO GERAL	15
2. REVISÃO BIBLIOGRÁFICA	17
2.1 Cacau	17
2.2 Identificação geográfica	18
2.3 Fatores que influenciam a qualidade do cacau e seus derivados.....	20
2.4 Quimiometria.....	23
2.4.1 Análise de Componentes Principais (ACP).....	24
2.4.2 Análises Discriminante (AD)	25
2.4.3 Redes Neurais Artificiais (RNA).....	26
2.5 Espectroscopia de Infravermelho com Transformada de Fourier (FTIR)	28
3. Objetivo geral	29
3.1 Objetivos específicos.....	29
4. MATERIAL E MÉTODOS.....	30
4.1 Coleta de dados.....	30
4.2 Caracterização física das amêndoas	30
4.2.1 Prova de corte	30
4.3 Caracterização química, físico-química e espectroscópicas das amêndoas	31
4.3.1 Determinação do pH.....	31
4.3.2 Determinação da acidez titulável.....	31
4.3.3 Determinação de umidade	31
4.3.4 Determinação de cinzas	32

4.3.5 Determinação de proteína.....	32
4.3.6 Determinação de lipídios.....	32
4.3.7 Determinação de açúcares redutores e não-redutores.....	33
4.3.8 Determinação de atividade de água (aW).....	33
4.3.9 Análise de cor.....	33
4.3.10 Quantificação de Teobromina, Cafeína, Catequina e Epicatequina por Cromatografia Líquida de Alta Eficiência – CLAE.....	33
4.3.11 Compostos fenólicos totais.....	35
4.3.12 Espectroscopia de Infravermelho com Transformada da Fourier – FTIR.....	35
4.4 Análises quimiométricas.....	36
4.4.1 Análise de Componentes Principais (ACP).....	37
4.4.1.1 Quantidade de CP e descarte de variáveis.....	38
4.4.2 Análise Discriminante.....	40
4.4.3 Redes Neurais Artificiais.....	40
4.4.3.1 Treinamento e validação.....	43
5. RESULTADOS E DISCUSSÃO.....	44
5.1 Análise Multivariada de dados e Rede Neural Artificial para avaliação da qualidade de amêndoas de cacau utilizando dados de composição.....	44
5.1.1 Análise de Componentes Principais.....	44
5.1.2 Análise Discriminante.....	46
5.1.3 Rede Neural Artificial.....	48
5.2 Análise Multivariada de dados e Rede Neural Artificial para classificação de amêndoas de cacau de diferentes regiões da Bahia utilizando dados de composição.....	49
5.2.1 Análise de Componentes Principais.....	49
5.2.2 Análise Discriminante.....	51
5.2.3 Rede Neural Artificial.....	53
5.3 Análise Multivariada de dados e Rede Neural Artificial para avaliação da qualidade de amêndoas de cacau utilizando dados espectroscópicos do FTIR.....	54
5.3.1 Análise de Componentes Principais.....	55

5.3.2 Análise Discriminante	57
5.3.3 Rede Neural Artificial	58
5.4 Análise Multivariada de dados e Rede Neural Artificial para classificação de amêndoas de cacau de diferentes regiões da Bahia utilizando dados espectroscópicos do FTIR.....	58
5.4.1 Análise de Componentes Principais	58
5.4.2 Análise Discriminante	59
5.4.3 Rede Neural Artificial	61
6. CONCLUSÃO.....	62
7. REFERÊNCIAS BIBLIOGRÁFICAS	64
8. ANEXOS	73

LISTA DE FIGURAS

FIGURA 1. Rede com dois neurônios na camada de entrada, quatro neurônios em cada camada intermediária e um neurônio na saída.....	26
FIGURA 2. Configurações das redes escolhidas para dados de composição e espectrais na classificação de amêndoas de cacau (a) composição/qualidade (b) composição/região (c) FTIR/qualidade (d) FTIR/região.....	42
FIGURA 3. <i>Scree Plot</i> dos autovalores associados aos componentes principais.....	44
FIGURA 4. Gráfico de dispersão das amêndoas de cacau produzidas na Bahia em relação aos componentes principais CP1 e CP2 com dados de composição.....	46
FIGURA 5. Gráfico do erro quadrático médio (RMSE) em função do número de iterações nas etapas de treinamento e validação das redes neurais de melhor arquitetura (a) composição/qualidade (b) composição/região (c) FTIR/qualidade (d) FTIR/região.....	49
FIGURA 6. Gráfico de dispersão das amêndoas dos territórios regiões Litoral Sul (L), Médio Rio das Contas (M), Vale do Jequiriçá (J) e Baixo Sul (B) em relação aos componentes principais CP1 e CP2 com dados de composição.....	51
FIGURA 7. Espectro de amêndoa de cacau por FTIR.....	55
FIGURA 8. Gráfico de dispersão das amêndoas de cacau produzidas na Bahia em relação aos componentes principais CP1 e CP2 utilizando comprimentos de onda.....	56
FIGURA 9. Gráfico de dispersão das amêndoas dos territórios regiões Litoral Sul (L), Médio Rio das Contas (M), Vale do Jequiriçá (J) e Baixo Sul (B) em relação aos componentes principais CP1 e CP2 utilizando comprimentos de onda.....	59

LISTA DE TABELAS

TABELA 1. Tolerância de defeitos para amêndoas de cacau, expressa em porcentagem e respectiva classificação do produto em tipos.....	31
TABELA 2. Parâmetros da CLAE para análise dos padrões e das amostras.....	34
TABELA 3. Regiões espectrais em FTIR.....	36
TABELA 4. Arquitetura de redes.....	41

RESUMO

BATISTA, A. S. **Identificação da qualidade e origem de amêndoas de cacau produzidas no estado da Bahia usando quimiometria.** Itapetinga-BA: UESB, 2018, 75 p. (Dissertação - Mestrado em Engenharia e Ciência de Alimentos; Área de Concentração: Engenharia de Alimentos). *

O cacauzeiro é uma cultura que encontrou condições ambientais apropriadas para seu desenvolvimento na Bahia, sendo um importante negócio para a economia das regiões produtoras. Em geral, a qualidade do cacau é afetada pela falta de infraestrutura e por práticas inadequadas de pós-colheita, sendo que a região de origem é utilizada como uma maneira de diferenciar e classificar suas características e, conseqüentemente, sua qualidade. Nesse sentido, objetivou-se avaliar dados químicos, físicos, físico-químicos e espectroscópicos no infravermelho aliado a análise estatística multivariada para fornecer informações sobre as principais variáveis envolvidas na qualidade da amêndoa do cacau, e proporcionar a identificação da origem e qualidade das amêndoas. Realizou-se a determinação de pH, acidez titulável, atividade de água (aW), análise de cor, umidade, cinzas, proteína, lipídios, carboidratos, compostos fenólicos totais, teores de cafeína, catequina, epicatequina e teobromina e foram obtidos dados espectroscópicos no FTIR para realização da Análise de Componente Principal (ACP), Análise Discriminante (AD) e Rede Neural Artificial (RNA) de amostras de quatro diferentes regiões produtoras de cacau, sendo elas: Litoral Sul, Baixo Sul, Médio Rio das Contas e Vale do Jequiçá. Os resultados revelaram que, pela ACP, foi possível separar as amostras de acordo com a sua região utilizando os dados espectroscópicos, enquanto que com os dados de composição, houve um agrupamento menos disperso de amêndoas das regiões Médio Rio das Contas, Baixo Sul e Vale do Jequiçá. Para o método de AD as melhores taxas de discriminação das amostras foram obtidas com os dados de FTIR, tanto para separação por qualidade quanto por região. As RNA's apresentaram os melhores resultados quando comparada às demais análises multivariadas, onde foram obtidas taxas de classificação total acima de 65% para qualidade e região utilizando os dois dados (composição e espectroscópicos).

Palavras-chave: Cacau, Quimiometria, Qualidade, Origem.

* Orientador(a): DSc. UESB. Leandro Soares Santos; Co-orientadores: DSc. UESB. Ivan Oliveira Pereira; DSc. UESB, Silmara Carvalho

SUMMARY

BATISTA, A. S. **Identification of the quality and origin of cocoa beans produced in the state of Bahia using chemometrics**. Itapetinga-BA: UESB, 2018, 75 p. (Dissertation – MSc. in Food Science and Engineering, Area of concentration: Food Engineering). *

Cacao is a crop that has found suitable environmental conditions for its development in Bahia, being an important business for the economy of the producing regions. In general, the quality of cocoa is affected by the lack of infrastructure and inadequate postharvest practices, and the region of origin is used as a way to differentiate and classify its characteristics and, consequently, its quality. The aim of this study was to evaluate chemical, physical, chemical and spectroscopic data in the infrared, combined with multivariate statistical analysis to provide information on the main variables involved in the quality of cocoa beans and to provide identification of the origin and quality of the almonds. PH, titratable acidity, water activity (aW), color analysis, moisture, ashes, protein, lipids, carbohydrates, total phenolic compounds, caffeine, catechin, epicatechin and theobromine were measured and spectroscopic data in the FTIR to perform Principal Component Analysis (ACP), Discriminant Analysis (AD) and Artificial Neural Network (RNA) of samples from four different cocoa producing regions, being: South Coast, South Low, Rio das Contas and Vale Middle of Jequiçá. The results showed that, through the ACP, it was possible to separate the samples according to their region using the spectroscopic data, whereas with the composition data, there was a less dispersed clustering of almonds from the Middle Rio das Contas, Southern Low and Vale regions of Jequiçá. For the AD method the best discrimination rates of the samples were obtained with the FTIR data for both quality and region separation. The RNAs presented the best results when compared to the other multivariate analyzes, where total classification rates above 65% were obtained for quality and region using the two data (composition and spectroscopy).

Keywords: Cocoa, Chemometrics, Quality, Origin.

* Advisor(a): **DSc.** UESB. Leandro Soares Santos; Co-advisors: **DSc.** UESB. Ivan Oliveira Pereira; **DSc.** UESB, Silmara Carvalho

1. INTRODUÇÃO GERAL

O cacauzeiro é uma planta da família *Sterculiaceae*, gênero *Theobroma*, da espécie *Theobroma cacao L.* que cresce em regiões tropicais em todo o mundo. Embora, esta espécie seja nativa do centro e norte da América do Sul, quase 70% da safra mundial é atualmente produzida na África Ocidental (DIOMANDE et al., 2015). A renda de cerca de 40 a 50 milhões de pessoas depende da produção de cacau, que tem uma produção mundial anual de 4,2 milhões de toneladas, com valor de US\$ 11,8 bilhões e crescimento a uma taxa de 3% ao ano, comparada a última década. A demanda de cacau deverá aumentar em 30% até 2020, mas sem capacitação e investimento em pequenos agricultores, a indústria terá dificuldade em fornecer oferta suficiente. No Brasil, a Bahia ostentou desde o século passado a posição de maior produtor de cacau, tendo sido responsável, até os anos 70, por quase 60% da arrecadação do estado (BEG et al., 2017). Entretanto, segundo dados do IBGE (2017), o estado do Pará se consolidou como o maior produtor de cacau em grão do país, desbancando a Bahia pela segunda vez consecutiva, respondendo por quase 50% da produção nacional. Do estado, sairão 116 mil toneladas neste ano, acima das 105 mil toneladas previstas para a Bahia.

O processo de obtenção do cacau até a produção de seus derivados é complexa. Em razão disto, para além da produção, é necessário promover discussões acerca da qualidade das amêndoas de cacau produzidas no Brasil como forma de fomentar alternativas que permitam aumentar a rentabilidade das fazendas de cacau e valorizar a matéria-prima, por meio de certificações de indicação de procedência. Segundo o Instituto Nacional da Propriedade Industrial - INPI (2018), a Associação dos Produtores de Cacau do Sul da Bahia certificou a Indicação de Procedência (IP) que é uma das modalidades da Indicação Geográfica (IG) para as amêndoas de cacau (*Theobroma Cacao L.*) da região.

Antes que as amêndoas de cacau possam ser comercializadas e processadas nos seus produtos industriais finais, principalmente chocolate em pó, manteiga de cacau e liquor de cacau, elas são submetidas ao processamento pós-colheita que compreende os passos de abertura da vagem e remoção das bagas, fermentação por 6 a 7 dias e secagem até um teor de umidade de 7%. Nesta sequência, a fermentação constitui um passo crítico e essencial para o desenvolvimento dos atributos de qualidade da amêndoa, tendo em vista que as amêndoas não fermentadas apresentam sabor adstringente e demasiadamente amargo, sendo características indesejáveis para produção do chocolate. Essas mudanças no perfil de sabor acompanham uma mudança de cor de roxo pálido (não fermentado) para marrom (totalmente fermentado)

(BIEHL e VOIGT, 1996; WOLLGAST e ANKLAM, 2000; NIELSEN et al., 2008; PAYNE et al., 2010; KADOW et al., 2013).

No Brasil, o grau de fermentação, e conseqüentemente a qualidade das amêndoas, é realizada pelo teste de corte. Neste teste, as amêndoas são classificadas como Tipo 1, Tipo 2, Tipo 3 ou fora do tipo a depender de algumas características, como presença de mofo, fumaça, amêndoas danificadas por insetos, ardósia, germinadas e achatadas. O teste consiste em cortar longitudinalmente as amêndoas e contar a proporção de roxa e marrom em uma amostra seca representativa de 300 amêndoas (WOOD e LASS, 1985; BRASIL, 2008).

Há uma série de indicadores químicos, como cafeína, teobromina, catequina, epicatequina, pirimidina B, conteúdo de nitrogênio amoniacal, peptídeos, aminoácidos livres, açúcares redutores, etc. que podem ser utilizados como indicadores de qualidade das amêndoas de cacau, resultando em grande número de análises a serem realizadas para avaliação da composição das amêndoas. Geralmente essas variáveis são avaliadas individualmente, o que muitas vezes tem dificultado a compreensão sistêmica da qualidade. Portanto, alternativas de análise e/ou técnicas analíticas rápidas e práticas que expressem toda a composição metabólica do cacau em uma única análise são de grande interesse.

A espectrometria de infravermelho associada à análise multivariada pode ser uma estratégia adequada por ser considerada rápida, sensível, não invasiva, não destrutiva, de custo relativamente baixo, fornecendo uma grande quantidade de informações com apenas um teste, pois demanda pouco ou nenhum pré-tratamento da amostra. É limpa, uma vez que dispensa o uso de reagentes químicos e evita a geração de resíduos danosos ao ambiente, além dos equipamentos serem relativamente fáceis de manusear (LOHUMI et al., 2013; LOHUMI et al., 2015).

A análise quimiométrica ou multivariada consiste no tratamento matemático de dados resultantes de análises que caracterizam uma amostra, com intuito de realizar uma análise exploratória, calibração e classificação dos dados, bem como a otimização de um experimento. Algumas técnicas quimiométricas podem ser citadas, como por exemplo a análise das componentes principais (ACP), análise fatorial (AF), análise de correlações canônicas (ACC), análise discriminante (AD) e análise de agrupamentos (*Cluster*).

O presente estudo tem como objetivo avaliar dados físicos, químicos, físico-químicos e espectroscópicos no infravermelho aliado a análise estatística multivariada para fornecer informações sobre as principais variáveis envolvidas na qualidade da amêndoa do cacau, proporcionar a identificação da amêndoa de qualidade e identificação de sua origem.

2. REVISÃO BIBLIOGRÁFICA

2.1 Cacau

O cacauzeiro é uma planta da família *Sterculiaceae*, gênero *Theobroma*, da espécie *Theobroma cacao L.* É uma planta encontrada no continente Sul Americano, que em condições naturais pode ultrapassar os cem anos. Possui uma altura em média de 6 - 8 m, e para um bom crescimento necessita de sombreamento adequado (BRASIL, 2014).

Todas as partes desse fruto são utilizadas, desde a casca até as sementes. Este fruto possui sementes com tamanhos variando de 2 a 3 cm de comprimento, são as responsáveis pelo valor agregado desse fruto, possuindo uma coloração no seu interior que pode variar do branco ao violeta, sendo que, quanto mais violáceo maior adstringência e amargor da semente, também são indispensáveis para comercialização, pois, são usadas para fabricação de chocolate e para extração de manteiga, que é um produto fundamental para a indústria farmacológica e de cosméticos. Essas sementes são envolvidas por uma polpa branca com sabor entre mais adocicado ao mais ácido, que é rica em açúcares, sendo utilizada na fabricação de geleias, vinho, suco, entres outros (SALES e CÂNDIDA, 2016; BRASIL, 2014).

A composição centesimal da amêndoa do cacau consiste em: 48-57% de lipídios, 2-5% de umidade, 11-16% de proteínas, 6-9% de amido, 2,6-4,2 de cinzas, 2,21- 3,2% de fibras, 0,8-1,4% de teobromina e 0,1-0,7% de cafeína (BECKETT, 2009). De acordo com Hashimoto (2015), os valores reportados anteriormente são apenas referências, pois, podem variar de acordo com a origem de procedência, variedade do fruto, grau de fermentação e secagem, condições de armazenamento, métodos analíticos utilizados, dentre outros. Esses fatores podem afetar de diferentes formas a composição química das amêndoas de cacau, bem como o aroma e a coloração das amêndoas que estão diretamente ligados à qualidade deste produto.

Os cultivos mais antigos de cacau no Brasil, foram estabelecidos no estado do Pará, depois essas sementes foram transportadas e plantadas na Bahia, em 1746. A cultura encontrou condições ambientais apropriadas para seu desenvolvimento no sul da Bahia, onde a expansão no cultivo de cacau ocorreu entre o final do século XIX e o início do XX (1896-1930), alterando a monocultura da cana-de-açúcar e progredindo sobre as áreas de floresta. Devido as vantagens adquiridas pela combinação de condições favoráveis para o cultivo, com a disponibilidade de terra e mão de obra, a cultura do cacau se transformou na principal atividade econômica da Bahia no século XX. No Estado da Bahia, quatro das quinze Regiões

Econômicas se destacam no cultivo de cacau: Litoral Sul, Extremo Sul, Sudoeste e Recôncavo Sul. As três primeiras Regiões Econômicas compõem o Sudeste da Bahia (PIASENTIN e SAITO, 2012; PIASENTIN et al., 2014).

Os cacauzeiros são classificados basicamente em três grupos que se diferenciam entre si pelas suas características. O Criollo produz um cacau de excelente qualidade, considerado como a mais nobre variedade de cacau, apresenta uma casca fina e rugosa, com amêndoas grandes e coloração clara ou rosácea. É de fácil fermentação, e possui baixa acidez e pouco sabor amargo, porém é muito sensível a doenças. O Forasteiro apresenta amêndoas achatadas de coloração violeta, exige uma fermentação prolongada e possui um acentuado sabor adstringente. É uma variedade do cacau de maior resistência a doenças e assim, é considerado produtivo com algumas subespécies que produzem cacau de excelente qualidade. Já os Trinitários são frutos híbridos, resultantes do cruzamento natural do Criollo e Forasteiro, que possui uma excelente qualidade devido a junção das características das variedades do cruzamento (KREIBICH, 2016).

2.2 Identificação geográfica

A região de origem dos produtos agrícolas e agroalimentares tradicionais são utilizadas como uma maneira de diferenciar e classificar suas características e conseqüentemente das suas qualidades. Os produtos agropecuários conhecidos como de qualidade superior, representam em média 30% em valores, das vendas no comércio mundial de bebidas e alimentos (CABO et al., 2015). Estabelecer a identidade de origem de um produto, como o cacau, não está associada somente ao fato da região ficar conhecida como a melhor produtora, mas também é uma garantia ou uma alternativa do aumento da rentabilidade das fazendas de cacau. Conseqüentemente o comércio do cacau será afetado de forma positiva, onde a região produzirá com qualidade e será reconhecida pelo que produz.

Segundo Brasil (2008) a identificação geográfica é uma forma de agregar valor e credibilidade a um produto ou serviço, concedendo-lhe um diferencial de mercado em função das características de seu local de origem. As IG's funcionam como uma protetora dos produtos e regiões contra falsificações e usurpações indevidas, além de convirem como garantia para o consumidor, indicando que se trata de um produto especial e diferenciado.

A identificação geográfica possui duas modalidades, a indicação de procedência (IP) e a denominação de origem (DO). Na indicação de procedência, o produto é indicado apenas pelo renome da localidade em que é produzido, independentemente de outros fatores, já a denominação de origem traz uma complexidade maior às IG's, pois, envolve não só aspectos

de comércio, mas também fatores geográficos e influência da mão-de-obra. A associação dos produtores de cacau do Sul da Bahia conseguiu certificar a IP das amêndoas de cacau (*Theobroma Cacao L.*), sendo uma conquista inestimável para região (INPI, 2018).

Segundo Valente et al. (2012), é necessário obter informações sobre o produto, se existe somente naquele local com tais características, produto típico, de sabor particular, um processo específico, uma ligação histórica forte, se são produtos diferenciados e intimamente ligados a um local. O diferencial na qualidade do produto pode estar principalmente relacionado com o meio em que é obtido, como solo, clima e vegetação. Essa interpretação é válida no caso de produtos vegetais, pois são influenciados diretamente pelo solo e clima cultivados. Os autores citam, como exemplo, que uma mesma variedade de uva plantada em locais distintos resultará em vinhos com características que se diferem, como estrutura e aroma.

De acordo com Diomande et al. (2015), as características das origens geográficas das amêndoas de cacau podem estar ligadas a qualidade do produto, como também o processo de fermentação, secagem e variedade do fruto. Assim, há uma necessidade de desenvolver métodos adequados para que haja rastreabilidade dessa matéria prima, para poder diferenciá-la de acordo com sua qualidade e também sua origem. A identificação de origem do cacau vem sendo estudado em muitos trabalhos que buscam relacioná-la a diferentes características dessa matéria-prima (HUE et al., 2014; CAMBRAI et al., 2016).

Vázquez-Ovando et al. (2015) utilizou a Análise de Componentes Principais, uma ferramenta quimiométrica para classificar amostras de cacau de diferentes regiões do México de acordo com características químicas, físicas e físico-químicas como tamanho da amêndoa, peso, conteúdo de lipídios, proteína, polifenóis e capacidade antioxidante. De acordo com os dados obtidos, eles conseguiram visualizar a formação de seis agrupamentos distintos.

Segundo Caligiani et al. (2016), as transformações ocorridas com os peptídeos e aminoácidos durante o beneficiamento do cacau precisam ser estudadas de forma mais profunda, pois essas mudanças que ocorrem são as que determinam o bom aroma do cacau. Eles estudaram as alterações bioquímicas que ocorrem durante a fermentação de amostras de cacau fermentadas, não fermentadas e sob fermentadas dos principais países produtores, a fim de obter uma melhor definição de qualidade do cacau e identificação de origem das amostras. Concluíram que há diferença na composição peptídica entre amostras fermentadas e não fermentadas, porém, é possível também diferenciá-las por região de origem.

2.3 Fatores que influenciam a qualidade do cacau e seus derivados

Considerando que o cacau seja uma das *commodities* agrícolas de maior dificuldade nas negociações de mercado, o produtor deve ser capaz de produzir cacau com qualidade, menor custo e excelente rendimento. Muitos pontos foram desenvolvidos e aperfeiçoados para suprir essa demanda, facilitar distinções de produtos, além de agregar valor como, o sistema de rastreabilidade, certificações ambientais, denominação de origem e clareza nos processos produtivos, promovendo um importante diferencial competitivo para as agroindústrias de cacau/chocolate (CEPLAC, 2013; FONTES, 2013).

O cacau de alta qualidade, tipo fino e/ou especial possui propriedades sensoriais de sabor, cor e aroma bem definidas. Esta qualidade está relacionada com as boas práticas agrícolas e o correto beneficiamento das amêndoas, desde a colheita até o armazenamento. Hoje, este é um dos grandes gargalos do Brasil, pois, em geral a qualidade do cacau é afetada pela falta de infraestrutura e por práticas inadequadas de pós-colheita (PWC, 2012; NOGUEIRA, 2015). Em 2012, o cacau de alta qualidade representava apenas 0,5% da produção nacional e 5% da produção mundial, com preço 100% maior quando comparado a amêndoas convencionais. Considerando que esse produto é importante para a economia mundial, a avaliação do seu aspecto nutricional, reações bioquímicas e químicas que desenvolvem as características desejáveis para seus derivados devem ser bem compreendidas para que se possa otimizar as etapas de produção (PWC, 2012; NOGUEIRA, 2015).

O índice denominado como *Cocoa Quality Index* foi proposto por Araujo et al. (2014), com a finalidade de avaliar de uma forma abrangente quais fatores mais influenciam na qualidade do cacau. A cadeia de produção do cacau é complexa desde a etapa de colheita até a confecção de seus derivados e, como já visto, o desenvolvimento de estudos a respeito da qualidade dessa matéria-prima vem sendo realizados com a finalidade de contribuir para uma melhor compreensão do processo de obtenção das amêndoas de cacau e posteriores transformações em novos produtos (EFRAIM et al., 2010; FERNANDES et al., 2013; VIEIRA et al., 2015).

A qualidade do cacau está associada a vários fatores envolvidos nas etapas do beneficiamento das amêndoas. Quando há um controle nessas etapas, principalmente a fermentação e secagem, tem-se a garantia de obtenção de um produto de qualidade, pois, ao realizar essas etapas de forma adequada, ocorrerão algumas alterações como a acidificação e aumento da temperatura que são necessárias para que as reações se processem de forma satisfatória e promovam o desenvolvimento de características pertinentes ao cacau e seus derivados. Ou seja, na etapa de fermentação ocorre a redução da adstringência e do amargor

nas amêndoas do cacau, que proporcionarão um sabor mais agradável para produção de seus derivados (KRÄHMER et al., 2015; OLIVEIRA, 2016).

A dedicação dos produtores aos procedimentos realizados desde o cultivo até a colheita de cacau, são essenciais para obtenção de amêndoas de qualidade. Alguns fatores que merecem atenção e controle são: homogeneização dos lotes da área de cultivo, identificação da variedade do cacau que será cultivado, utilização de produtos para proteção contra insetos e doenças prejudiciais ao fruto e o processo de adubação e sombreamento do cacauzeiro. Sendo assim, após tomadas as decisões de como será o manejo do fruto, é necessário também que haja conhecimento por parte dos produtores nas posteriores etapas envolvidas na produção de uma amêndoa de cacau de qualidade, que são compreendidas em: colheita, quebra, transporte, fermentação, secagem e armazenamento (FERREIRA et al., 2013; KRÄHMER et al., 2015).

Saltini et al. (2013) ressaltaram a influência das práticas agrícolas sobre as amêndoas de cacau, e que é possível discernir as eventuais causas que alteram os aspectos de qualidade e sabor do cacau e, conseqüentemente, do seu principal derivado, o chocolate. De acordo com Crafacck et al. (2014), a origem do cacau e o processo de fermentação são fatores importantes na formação dos compostos voláteis. Eles observaram que o efeito da fermentação é maior do que o do cultivo do cacau, e assim indicam que a condução da fermentação é de grande relevância na formação dos voláteis no chocolate.

A fermentação é uma etapa decisiva do processo de obtenção de amêndoas de cacau de qualidade, afetando significativamente a composição dessa matéria-prima. Comparativamente, a semente de cacau não fermentada possui elevados teores de polifenóis e amido, sabor mais amargo e adstringente, não desenvolvendo as características e/o sabor desejável para uma posterior utilização na produção do seu derivado, o chocolate (ACULEY et al., 2010; KIM e KEENEY, 2006; WOLLGAST e ANKLAM, 2000; BIEHL e VOIGT, 1996).

O processo de fermentação natural das sementes do cacau é, o mais empregado no Brasil. Em razão da composição da polpa, contendo mais de 80% de água, 11-13% de carboidratos e 0,3% de ácido cítrico, as fermentações são tipicamente alcoólicas, lácticas e acéticas promovendo mudanças bioquímicas e enzimáticas nas amêndoas (LOPEZ e DIMICK, 1995).

O pH inicial da polpa, juntamente com baixos níveis de oxigênio, favorece a colonização de leveduras, que utilizam carboidratos para produzir etanol. À medida que a aeração da massa de fermentação aumenta, as bactérias acéticas tornam-se dominantes. O etanol é oxidado em ácido acético por bactérias em uma reação altamente exotérmica,

aumentando a temperatura da massa (BAREL, 1998; SCHWAN e WHEALS, 2004). A produção de ácido acético, juntamente com o aumento da temperatura, gera mudanças drásticas dentro das amêndoas de cacau, como a morte do embrião e a quebra do compartimento celular (QUESNEL, 1965). Como resultado, o contato entre componentes bioquímicos é facilitado e a cor da amêndoa muda de roxo para marrom (ACULEY et al., 2010). Os precursores de aroma são formados, as proteínas são degradadas (BUYUKPAMUKCU et al., 2001; DE BRITO et al., 2001; LERCETEAU et al., 1999) e os polifenóis são oxidados e polimerizados (FORSYTH, 1952; HANSEN, 1998). Uma correlação forte foi demonstrada entre a formação de precursores de aroma de cacau e degradação de proteína durante a fermentação (KIRCHHOFF, BIEHL e CRONE, 1989).

Muitas substâncias contribuem para o sabor do cacau e são precursores de sabor de chocolate, como por exemplo, peptídeos, aminoácidos livres e açúcares redutores (KADOW et al., 2013). Os componentes de sabor amargo são a teobromina e a cafeína. Entretanto, demonstrou-se também que as dicetopiperazinas contribuem para o típico sabor amargo do cacau (STARK et al., 2006). A nota ácida é um outro atributo de sabor típico e deve-se principalmente ao ácido láctico e o acético formado durante o processo de fermentação. A adstringência é atribuída aos compostos fenólicos, que no cacau contribuem com, aproximadamente, 15% do peso seco (KADOW et al., 2013). De acordo com o Cocoa Atlas 2010, as substâncias mencionadas podem ser utilizadas como parâmetros de qualidade bioquímica para a descrição detalhada da qualidade do cacau (ROHSIUS et al., 2010).

Vários métodos de cromatografia como CLAE (cromatografia líquida de alta performance) ou CG (cromatografia gasosa) foram descritos para a quantificação das substâncias relevantes para o sabor (ACULEY et al., 2010; ELWERS et al., 2009; HUMSTON et al., 2010; MACHONIS et al., 2014; PEREIRA-CARO et al., 2013; ROHSIUS et al., 2006, SYNOVEC et al., 2012). Além disso, vários métodos tradicionais para determinação de minerais, sólidos totais, açúcar redutor, pH e acidez também foram utilizados (AFOAKWA et al., 2013a; SUAZO et al., 2014).

Hue et al. (2014) utilizaram o Infravermelho com Transformada de Fourier (FTIR) para comparar níveis de fermentação (obtidas por uma técnica demorada que quantifica a produção de voláteis) de amêndoas de cacau de seis diferentes origens geográficas. Eles obtiveram correlação entre os resultados das técnicas utilizadas e validaram uma calibração precisa e confiável para determinação de produtos desenvolvidos na fermentação e que indicam a qualidade da amostra avaliada. Essa correlação foi apontada como um método de

rotina que poderia ser aplicada na classificação de amêndoas de cacau de acordo com seu grau de fermentação pelos fabricantes de chocolate.

Um estudo de caracterização foi desenvolvido através de análises físicas, químicas, físico-químicas e por espectros obtidos no infravermelho próximo de amostras de lotes comerciais de amêndoas de cacau produzidas em diferentes estados brasileiros (Bahia, Pará e Rondônia) e na Costa do Marfim. Os dados espectrais foram utilizados na análise quimiométrica para determinação dos parâmetros de composição química e de qualidade anteriormente analisados pelos métodos analíticos convencionais (prova de corte, pH, acidez, teores de umidade, lipídios, proteína, compostos fenólicos totais, teobromina e cafeína). Chegou-se a conclusão que a aplicação da Espectroscopia no Infravermelho próximo associada à quimiometria é uma alternativa promissora, devido a possibilidade do desenvolvimento de métodos não destrutivos, rápidos e que proporcionam a interpretação de muitas variáveis obtidas a fim de caracterizar uma amostra (HASHIMOTO, 2015).

2.4 Quimiometria

A quimiometria pode ser compreendida como uma ciência que busca correlacionar variáveis químicas baseados na álgebra linear, enfocando de maneira geral em dois temas fundamentais: o planejamento experimental e a análise de dados multivariados. A evolução da quimiometria ocorreu no século 21 devido a sua importância nas análises de dados obtidos através da química analítica, ou seja, as ferramentas quimiométricas tentam encontrar as possíveis relações entre as amostras analisadas e as variáveis estudadas, com objetivo de diminuir a complexidade dos dados de um experimento (BRERETON, 2007; KUMAR, et al. 2017).

Análise multivariada está relacionada as técnicas estatísticas que realizam análises de várias medidas simultaneamente dos indivíduos ou objetos sob investigação. Dessa forma, é considerada como multivariada qualquer análise que, de forma simultânea, trata mais do que duas variáveis de um planejamento experimental (MINGOTI, 2007; HAIR et al., 2010).

As ferramentas quimiométricas se baseiam em conceitos importantes, como o espaço químico, que pode ser identificado como uma área multidimensional, em um sistema de coordenadas com múltiplos eixos que caracteriza uma quantidade de propriedades físicas, químicas e físico-químicas. Essas ferramentas consideram que as posições das estruturas químicas no sistema de coordenadas estão baseadas nos diferentes valores de seus respectivos eixos. Sendo possível navegar de forma eficiente e inteligente através do espaço químico e

extrair informações significativas a partir da grande quantidade de dados obtidos (LINUSSON, et al. 2000).

Segundo Mingoti (2005), a análise quimiométrica ou multivariada consiste no tratamento matemático de dados resultantes de análises químicas com intuito de realizar uma análise exploratória, calibração e classificação dos dados, bem como uma otimização de um experimento. Algumas técnicas quimiométricas podem ser citadas, como por exemplo a análise das componentes principais (ACP), análise fatorial (AF), análise de correlações canônicas (ACC), análise discriminante (AD) e análise de agrupamentos (*Cluster*) (FERREIRA et al., 1999; MINGOTI 2005).

2.4.1 Análise de Componentes Principais (ACP)

A análise de componente principal tem sido utilizada como um ponto de partida para a análise multivariada de dados, pois, de maneira rápida, fornece uma visão geral das informações mais importantes, as quais se tornam mais visíveis e fáceis de serem analisadas e compreendidas. A ACP permite uma visualização geral das amostras ou tabelas de dados observadas, assim os padrões como tendências, desvios, agrupamentos, valores divergentes, podem ser identificados (TEÓFILO, et al. 2009).

A ACP é uma ferramenta aplicada para reduzir o conjunto de dados mantendo o máximo de informações originais possível, que tem como princípio buscar similaridade entre os dados, de tal forma que estes possam ser representados pelas componentes mais representativas do sistema, ditas componentes principais. Assim, a ACP permite a realização de uma análise rigorosa dos dados, a fim de encontrar as relações entre as variáveis medidas, agrupando-as e detectando variáveis que pouco contribuem para o entendimento sistêmico dos dados. A maioria das pesquisas avaliam muitas variáveis que na maioria das vezes dificultam a interpretação do comportamento dos dados, a utilização da ACP possibilita a utilização de um menor número de variáveis que podem indicar a mesma quantidade de informação contida no conjunto total (FERREIRA, 2011; PACHECO et al., 2013).

Sendo assim, a ACP, além proporcionar o resumo das informações contidas no complexo de variáveis, permite a retirada de variáveis que pouco contribuem em termos de variação, no grupo avaliado. Ou seja, além de reduzir e simplificar os dados originais, a ACP proporciona detecção de amostras (*outliers*) responsáveis por maiores variações entre os dados, seleção de variáveis mais importantes e previsões do comportamento dos dados (SHARAF et al., 1986; PEREIRA e PEREIRA, 2004).

Atualmente os estudos acadêmicos estão sendo direcionados para a rastreabilidade e identificação de origem dos alimentos. Cambrai et al. (2016) identificaram polifenóis como epicatequina, catequina e polímeros de procianidinas com intuito de investigar o país de origem do chocolate, utilizando análises cromatográficas e quimiométricas de 47 amostras provenientes de 12 países de dois continentes. Constataram que através do perfil polifenólico, é possível classificar amostras de chocolates de diferentes regiões geográficas utilizando a ACP, para isso verificaram que 2 componentes principais conseguiram explicar 84% da variabilidade total dos dados.

2.4.2 Análises Discriminante (AD)

A estatística multivariada possui uma técnica utilizada para discriminar e classificar objetos de uma população em duas ou mais classes, esta técnica é conhecida como Análise Discriminante ou critério de Fisher. A AD se caracteriza por formar uma combinação linear das variáveis originais, separando o máximo possível as populações formadas, ou seja, é capaz de maximizar a razão entre as somas dos quadrados entre os grupos formados e a soma de quadrados dentro desses grupos (CHEN et al., 2011; KHATTREE e NAIK, 2000; MINGOTI, 2005; DAS e NENADIC, 2008).

A AD foi estudada por Fisher (1936) e por isso é conhecida como critério de Fisher. A discriminação e classificação de uma população em duas ou mais classes é realizada a partir de funções matemáticas que proporcionam uma baixa probabilidade de má classificação e que buscam relacionar uma variável dependente com várias variáveis independentes através da criação de índices que identificam similaridade ou dissimilaridade entre os grupos formados. A função discriminante é uma combinação linear responsável por separar o conjunto de dados em regiões distintas e seus coeficientes são utilizados para mostrar quais variáveis originais mais influenciam ou contribuem para sua formação (REGAZZI, 2000).

De acordo com Chen et al (2011) a análise discriminante linear é uma das técnicas supervisionadas mais utilizadas entre os métodos de análise multivariada, sendo usada para encontrar a combinação linear de características de determinado experimento, a função gerada resulta na classificação linear de grupos existentes no conjunto de dados. A ACP é utilizada antes de aplicar a análise discriminante com intuito de reduzir o conjunto de dados e facilitar uma possível geração de uma função capaz de separar o conjunto de dados nos respectivos grupos existentes (BERRUETA et al., 2007).

Teye e colaboradores (2013) desenvolveram um estudo com amostras de cacau de sete diferentes regiões de Gana, aplicando a técnica de Espectroscopia de Infravermelho próximo

associado a análises estatísticas multivariada com intuito de discriminar essas amostras de acordo com sua origem geográfica, pois, é essencial para um controle de qualidade e rastreabilidade. Eles concluíram que a AD é uma técnica rápida e confiável para classificar amostras de cacau, obtendo uma taxa de discriminação de 96,15% e 90,63% para os conjuntos de treinamento e predição, respectivamente.

2.4.3 Redes Neurais Artificiais (RNA)

A motivação para o desenvolvimento de tecnologia de redes neurais resultou do desejo de desenvolver um sistema artificial que pudesse executar tarefas "inteligentes" semelhantes às exercidas pelo cérebro humano. A RNA é um método computacional com capacidade de aprendizagem e manutenção desses conhecimentos. É uma das grandes vantagens das redes neurais é a capacidade de aprender as relações diretamente a partir dos dados experimentais que representam tanto o comportamento linear como o não linear (RAI, et al., 2005; SILVA et al., 2010).

A Rede Neural Artificial (RNA) pode ser compreendida como um grupo de neurônios artificiais ou de unidades de processamento interconectadas entre si formando vários subgrupos estruturais de camadas (entrada, intermediárias e saída) e conexões entre as camadas. A RNA é apta para acumular os conhecimentos baseando-se em aprendizagem (experiência) e disponibilizar essas experiências para posterior utilização (HAYKIN, 1999). Ou seja, é um método computacional que se destaca por sua capacidade de conseguir, através de um só modelo, avaliar, prever ou classificar quaisquer dados fornecido a ela.

Na figura 1 está esquematizado um exemplo de uma rede de múltiplas camadas.

Figura 1. Rede com dois neurônios na camada de entrada, quatro neurônios em cada camada intermediária e um neurônio na saída.



Fonte: GARCIA, 2004.

A partir de uma base de dados, a RNA tem a capacidade de capturar as relações funcionais existentes entre suas variáveis de entrada e saída. Essas redes neurais artificiais são valiosas e aplicadas em diversos sistemas de processamento de alimentos, que envolvem estimativa, predição, otimização e classificação. Mesmo sem a descrição prévia da relação funcional que possa existir entre as variáveis, a RNA pode lidar com múltiplas variáveis dependentes ou não, simultaneamente (SINGH et al., 2009; ZHOU e THERDTHAI, 2010; EIM et al., 2013).

Os sinais captados através dos neurônios artificiais são transferidos por conexões matemáticas para as camadas intermediárias. Este processamento de informação ocorre em paralelo, em todos os neurônios, e assim, é distribuída automaticamente para todas as unidades que compõem a rede. A função de ativação é responsável por transmitir as somas ponderadas de informações. Geralmente utiliza-se a função sigmoideal, devido à grande quantidade de sinais informativos, que são convertidos a valores definidos durante a etapa de treinamento e teste (EERIKAINEN et al., 1993; HORIMOTO et al., 1995).

As redes do tipo Perceptron de Múltiplas Camadas (MLP), são muito utilizadas para a solução de vários problemas que envolvem altos graus de não-linearidade. Através do tipo de treinamento elas são classificadas como supervisionadas e utilizam o algoritmo Error Backpropagation (retropropagação do erro) cuja regra de aprendizagem se baseia em corrigir o erro durante o processo de treinamento. O *Resilient Propagation*, uma versão aperfeiçoada do algoritmo *Backpropagation*, atualmente também é bem difundido e tem por objetivo tornar o processo de convergência mais eficiente (HAYKIN, 1994; SILVA et al., 2010).

As redes neurais podem ser compostas por uma série de algoritmos de aprendizagem, onde cada tipo de algoritmo opera de forma diferente. O algoritmo ajusta os pesos das interconexões de maneira a reduzir o erro entre o valor obtido e o valor desejado. Assim, através do método de tentativa e erro, o programa continua a alterar as conexões internas até alcançar um menor erro possível (RAI et al., 2005).

As RNA's tem sido aplicadas com sucesso em várias áreas de pesquisa, tais como: química, matemática, engenharia, medicina e psicologia. A boa adaptação em várias áreas de pesquisa, bem como bons resultados obtidos justificam essa tendência (RAMADHAS et al. 2006).

Durante a etapa de treinamento da rede, os valores previstos são consecutivamente comparados com os valores verdadeiros e, assim as conexões entre os neurônios de processamento, com seus devidos pesos, são alterados com a finalidade de minimizar o desvio entre os valores. A rede é capaz de medir o erro entre o valor obtido e o valor desejado

ajustando os pesos das interconexões através do algoritmo, reduzindo assim, o erro nas determinações subsequentes. Dessa forma, o programa continua a fazer modificações nas conexões entre as unidades, por meio do processo de tentativa e erro, até alcançar um determinado nível de certeza (WYTHOFF, 1990 e KATZ, 1992).

O treinamento continua até que o menor erro possível (abaixo de um limiar pré-definido) é atingido. Assim, é possível saber a quantidade de interações que serão necessárias para chegar a esse erro mínimo (JAIN, MAO e MOHIUDDIN, 1996). As iterações são alteradas (aumentadas) no decorrer do treinamento e validação dos dados e, ao observar a minimização do RMSE, pode-se definir a quantidade necessária de iterações.

As redes neurais artificiais tem sido utilizadas com sucesso por pesquisadores no controle de qualidade de amêndoas de cacau. León-Roque et al. (2016) aplicaram a RNA para validar um procedimento quantitativo baseado na medida de cor RGB para prever o índice de fermentação de amêndoas de cacau de três regiões do Peru. Concluíram que é possível desenvolver um modelo quantitativo e simples baseado na medição de cores e prever o índice de fermentação de amêndoas de cacau, podendo ser desenvolvidos aplicativos que facilitem o controle de qualidade e níveis de fermentação.

2.5 Espectroscopia de Infravermelho com Transformada de Fourier (FTIR)

A espectroscopia de infravermelho com transformada de Fourier se baseia nos movimentos vibracionais dos grupos funcionais encontrados na amostra avaliada quando submetida à radiação infravermelha, ou seja, quando a radiação infravermelha for absorvida por um grupo funcional, ocorrerá uma vibração molecular. A depender da quantidade de absorção e do tipo de ligação, serão formados picos variando de largos a estreitos, altos ou baixos, no amplo comprimento de onda da região infravermelha. Esta técnica produz gráficos com os dados espectrais obtidos e se tornou o método preferido da espectroscopia de infravermelhos para avaliar adulterações em diversos produtos alimentícios, como a manteiga de cacau e chocolate e para determinar a qualidade das amêndoas de cacau (BATISTA et al., 2016; THERMO, 2016).

O FTIR é muito utilizado por ser uma técnica rápida, simples e precisa, sendo de interesse para indústria que tenha alta produção e que procura análises rápidas para controlar a qualidade de seus produtos. Juntamente com a análise multivariada, a espectroscopia de infravermelho é utilizada para desenvolvimento de modelos matemáticos que são capazes de dar resultados rápidos de presença ou não de componentes nutricionais ou prejudiciais em diversos tipos de alimentos (HU et al., 2016).

Batista et al. (2016) utilizaram o FTIR e determinaram o conteúdo fenólico total, capacidade antioxidante e os teores de cafeína e teobromina em amostras de chocolate. Nessa pesquisa, concluíram que a técnica espectroscópica é indicada para avaliar antioxidantes e quantificar parâmetros fenólicos presentes em derivados do cacau de uma maneira rápida e confiável.

O FTIR foi utilizado em mais de um trabalho com amêndoa de cacau, com o objetivo de determinar o conteúdo total de compostos fenólicos, atividade antioxidante e grupo funcional das amostras em seis tempos diferentes de fermentação (24, 48, 60, 72, 96 e 120 horas). Com essa pesquisa, foi observado que as amostras que foram parcialmente fermentadas (24 horas) apresentaram maior conteúdo fenólico e as amostras mais fermentadas (120 horas), continham menor quantidade de compostos fenólicos. A atividade antioxidante teve alta correlação com o perfil fenólico, ou seja, quanto maior a atividade antioxidante, maior a composição de fenólicos nas amostras avaliadas (UTAMI et al., 2016).

Hu et al. (2016) concluíram que a espectroscopia no infravermelho pode ser utilizada para quantificar e prever a capacidade antioxidante, compostos fenólicos e concentração de catequina em chocolate. Sendo uma técnica rápida, que obtém os mesmos dados que análises químicas demoradas, sendo não destrutivo e reduzindo custos com reagentes e tempo para obtenção de resultados precisos.

3. Objetivo geral

O presente estudo tem como objetivo avaliar características químicas, físicas, físico-químicas e espectroscópicas no infravermelho aliado a análise estatística multivariada para fornecer informações sobre as principais variáveis envolvidas na qualidade da amêndoa do cacau e proporcionar a identificação de qualidade e de origem das amêndoas de cacau.

3.1 Objetivos específicos

- a) Avaliar a qualidade das amêndoas de cacau por meio de atributos externos e internos do teste de corte;
- b) Obter dados de composição químicas, físicas e físico-química de amêndoas de cacau de diferentes regiões do estado da Bahia;
- c) Obter dados de espectros de amostras de amêndoas de cacau de diferentes regiões do estado da Bahia utilizando FTIR;
- d) Aplicar a técnica de ACP, para redução da dimensionalidade e definição das variáveis mais importantes na identificação de origem e qualidade das amostras;

- e) Utilizar a AD, para desenvolver funções discriminantes que possam separar em grupos distintos de acordo com as regiões e qualidade da amostra;
- f) Desenvolver uma Rede Neural Artificial para identificar as amêndoas de acordo com a origem e com a sua qualidade;

4. MATERIAL E MÉTODOS

4.1 Coleta de dados

Foram coletadas 87 amostras das regiões (Litoral Sul, Baixo Sul, Vale do Jiquiriçá e Médio Rio das Contas) em “consulta pública” por meio de um edital do Instituto Federal Baiano - *campus* Uruçuca e ampla divulgação em mídias digitais. Os escritórios regionais da CEPLAC também contribuíram na divulgação, auxílio em visitas nas fazendas e como local de recebimento de amostras. Para a coleta das amostras, foram entregues um kit formado por um formulário (Anexo 1) que contém informações sobre beneficiamento das amêndoas, uma embalagem plástica e um lacre.

Após a coleta das amostras, as amêndoas foram submetidas às análises física e físico-químicas no laboratório de pesquisa do Instituto Federal Baiano, *campus* Uruçuca e nos laboratórios da Universidade Estadual do Sudoeste da Bahia.

4.2 Caracterização física das amêndoas

4.2.1 Prova de corte

As amêndoas foram avaliadas por meio de um corte longitudinal para a classificação comercial como recomenda o Regulamento Técnico da Amêndoa de Cacau, quando foram observados atributos internos relacionados ao grau de fermentação, como colorações marrom, parcialmente marrom ou violácea, e defeitos como: mofo, fumaça, danos causados por insetos, aspecto de ardósia e germinação (BRASIL, 2008). O Anexo 2 apresenta o laudo de avaliação das amostras de cacau.

Após avaliar as características das amêndoas de acordo com os atributos e defeitos citados acima, as amostras foram classificadas de acordo com sua qualidade em dois grupos: Qualidade Superior (QS) e Qualidade Inferior (QI). Os critérios utilizados para essa classificação foram o índice de fermentação (porcentagem de amêndoas marrons e parcialmente marrons) e os defeitos listados na Tabela 1. Para uma amostra ser considerada de Qualidade Superior ela deveria apresentar um índice de fermentação igual ou superior a 60% e limites nos defeitos de 4% e 5% para mofo e ardósia, respectivamente, além da ausência de fumaça.

Tabela 1. Tolerância de defeitos para amêndoas de cacau, expressa em porcentagem e respectiva classificação do produto em tipos.

Classificação	Defeitos					
	Mofadas	Fumaça	Danificadas por Insetos	Ardósia	Germinadas	Achatadas
Tipo I	0 a 4%	0 a 1%	0 a 4%	0 a 5%	0 a 5%	0 a 5%
Tipo II	4% a 6%	1% a 4%	4% a 6%	5% a 10%	5% a 6%	5% a 6%
Tipo III	6% a 12%	4% a 6%	6% a 8%	10% a 15%	6% a 7%	6% a 7%
Fora de Tipo	12% a 25%	> 6%	> 8%	> 15%	> 7%	> 7%

Fonte: BRASIL, 2008.

4.3 Caracterização química, físico-química e espectroscópicas das amêndoas

Para realização das análises de pH, acidez titulável, umidade, cinzas, lipídeo, atividade de água (aW), cor, determinação de proteína, carboidratos, compostos fenólicos totais, análises cromatográficas e espectroscópicas utilizou-se amostras secas e trituradas em um moedor de grãos portátil de modelo DCG 20 (para redução do tamanho das partículas e homogeneização).

4.3.1 Determinação do pH

O pH foi medido em suspensão aquosa da amostra (5 g de amostra em 50 mL de água), utilizando-se um medidor de pH digital (modelo HI 8314) com as instruções do (IAL, 2008).

4.3.2 Determinação da acidez titulável

A acidez titulável das amostras foi medida através da titulação potenciométrica com um medidor de pH digital (modelo HI 8314) e o titulante utilizado foi o NaOH 0,1M, de acordo com o (IAL, 2008).

4.3.3 Determinação de umidade

A determinação de umidade foi realizada pelo método de secagem direta em estufa de 105° C, para isso, foram pesadas cerca de 3 g da amostra em cápsula de porcelana, previamente tarada e desengordurada. Em seguida foram aquecidas, durante 3 horas, resfriadas em dessecador até a temperatura ambiente e então pesadas. O processo de aquecimento e resfriamento foi repetido até obtenção de peso constante, de acordo com o (IAL, 2008).

4.3.4 Determinação de cinzas

A determinação de cinzas foi realizada por método gravimétrico usando forno mufla a 550° C. Os cadinhos foram desengordurados em estufa a 105° C durante três horas e foram registrados seus pesos. Três gramas da amostra foram adicionados a este recipiente e colocados por três horas, após este tempo, foram retiradas e resfriadas em dessecador. O peso final foi registrado e a quantidade de cinzas foi calculada pela diferença entre a massa inicial e a final. O percentual se deu pela razão desta diferença com o peso inicial da amostra, multiplicado por 100, de acordo com o (IAL, 2008).

4.3.5 Determinação de proteína

A determinação de proteínas foi realizada pelo método de Kjeldahl. Pesou-se 0,3 g da amostra que foi transferida para o balão de Kjeldahl. Adicionou-se 5 mL de ácido sulfúrico e cerca de 2 g da mistura catalítica. Em seguida, os tubos foram levados ao aquecimento de forma gradativa em chapa elétrica, na capela, até a solução se tornar azul-esverdeada e livre de material não digerido (pontos pretos). Ao esfriar, adicionou-se aproximadamente 10 mL de água destilada nos tubos e foram acoplados ao conjunto de destilação. Na sequência, foram adicionados 25 ml de solução de NaOH a 50%, aquecidos à ebulição e destilados até obter cerca de 100 mL do destilado em um erlenmeyer contendo 20 mL de solução de ácido bórico. Por fim, titulou-se o destilado com ácido sulfúrico 0,05 M. A quantificação da proteína foi realizada utilizando-se 6,25 como fator de conversão do nitrogênio, de acordo com o (IAL, 2008).

4.3.6 Determinação de lipídios

A determinação de lipídios totais foi realizada em um extrator de Soxhlet com refluxo de quatro horas. Inicialmente foi pesado 500 mg de amostra dentro de cartucho feito de papel filtro. O balão do Soxhlet foi encaixado no intermediário e foi adicionado o solvente éter de petróleo em quantidade suficiente para permitir o sifonamento do mesmo. Os balões foram aquecidos em manta de aquecimento e a parte superior do intermediário foi conectada a um destilador. O processo de gotejamento foi uniforme e ocorreu durante 4 horas. Logo após, os cartuchos foram colocados em estufa de ventilação forçada a 55 ° C durante 16 horas. Em seguida, essas amostras foram postas para secar em estufa a 105 ° C por três horas, e então pesadas (AOAC, 2005).

4.3.7 Determinação de açúcares redutores e não-redutores

A determinação de açúcar redutor foi realizada utilizando o método DNS (Ácido 3,5-dinitrosalicílico), onde a curva padrão foi preparada a partir de uma solução de glicose a 1,0 g/L. As amostras foram pesadas e adicionadas de 10 mL de água destilada, em seguida foram agitadas em um vórtex por dois minutos. Na sequência, os extratos foram centrifugados a 1600 RCF durante 15 minutos em um equipamento Daiki de modelo DT4500. Este procedimento foi realizado por três vezes. Após coletar o sobrenadante, foi realizado o teste DNS, em que foi pipetado 1 mL e adicionado 1 mL do reagente DNS, e então, aquecidos em banho maria a 100 ° C (em ebulição) por 5 minutos. Em seguida, as amostras foram resfriadas em banho de gelo por 5 minutos. Na sequência, realizou-se a leitura da absorbância em espectrofotômetro a 540 nm, após zerar o aparelho com o branco.

Para quantificação de açúcar não-redutor pipetou-se 2 mL do sobrenadante obtido anteriormente, e foram adicionados 2,0 mL de HCl 2N em tubos, em seguida, foram aquecidos em banho maria em ebulição por 10 minutos. Após o aquecimento, a amostra foi resfriada em banho de gelo e acrescentou-se 2,0 mL de NaOH 2N. Em seguida, foi realizado o teste de DNS (CECCHI, 2003).

4.3.8 Determinação de atividade de água (aW)

A atividade de água foi determinada através da quantificação da fugacidade de água, por meio da constante dielétrica em equipamento AQUALAB Lite, com exatidão $\pm 0,015$ e resolução 0,001.

4.3.9 Análise de cor

A cor das amostras foi determinada no sistema L*, a*, b* e dE em colorímetro modelo Color Quest XE (Hunter Lab). Nesse sistema de cores L* representa a luminosidade (L*=0 – preto e L*=100 – branco), a* e b* são as coordenadas de cores, responsáveis pela cromaticidade (+a* = vermelho e - a* é o verde, +b* é o amarelo e -b* é o azul) e dE corresponde à diferença total de cor (CIE, 1996).

4.3.10 Quantificação de Teobromina, Cafeína, Catequina e Epicatequina por Cromatografia Líquida de Alta Eficiência – CLAE

Para o preparo das amostras foi retirada a casca das amêndoas (testa), obtendo-se somente os cotilédones (nibs). Os cotilédones foram triturados em moinho analítico e pesados aproximadamente 0,02g. Adicionou-se 5 mL de solução extratora (85% H₂O, acidificado com

0,3% de ácido acético, e 15% Metanol) em tubos de ensaio, sob agitação por 15 minutos em um banho ultrassônico. Na sequência, os tubos foram colocados em banho maria a 60°C por 10 minutos e, logo após, na centrifuga (3000g por 15 minutos) para uma melhor separação. O extrato foi filtrado com filtros estéreis de 0,22µm.

A tabela 2 apresenta os parâmetros utilizados para a análise cromatográfica dos padrões e das amostras do presente estudo.

Tabela 2. Parâmetros da CLAE para análise dos padrões e das amostras.

Parâmetros da fase móvel	Condições		
	Tempo (min)	% Hac 0,3	%MeqOH
Gradiente (frações/volume)	Inicial	95	5
	9	95	5
	9,1	75	25
	12,5	75	25
	12,6	5	95
	13,5	5	95
	13,6	95	5
	28	95	5
Fluxo	0,5 mL/min		
Volume de injeção	20 µm		
Coluna	Shim-Pack Prep-ODS (H) (octadecilsilano) de 25 cm e 4,6 mm		
Temperatura	30°C ± 2		
Comprimento de onda	Ultravioleta à 280 nm		
Tempo de execução	28 minutos		

Fonte: Do Autor, 2018.

Os alcaloides xantínicos (cafeína e teobromina) e compostos fenólicos majoritários do cacau ((+)-catequina e (-)-epicatequina) foram quantificados por meio do sistema HPLC-SHIMADZU DGU-20A5R composto por uma coluna Shim-pack PREP-ODS(H)KIT, 250x4,6 mm, comprimento de onda 280 nm a um fluxo de 0,5 mL/min e com volume de injeção de 20 µm. O gradiente composto de fase móvel foi estabelecido como: A (Metanol) e B (H₂O ultrapura com 0,3% de ácido acético) com um tempo total de corrida de 28 minutos (RISNER, 2008; GONÇALVES, 2016).

4.3.11 Compostos fenólicos totais

A extração dos compostos fenólicos foi feita em solvente aquoso, segundo método descrito por Lee et al. (2003). Inicialmente foi pesado 500 mg de material triturado, e adicionados 10 mL de água destilada a 100 ° C. Na sequência, os extratos foram centrifugados a 2010 RCF durante 10 minutos (em equipamento com marca Daiki de modelo DT4500). O sobrenadante foi coletado e o processo de extração foi repetido.

As amostras foram diluídas adequadamente e uma alíquota de 0,5 mL foi transferida para tubos de ensaio, adicionando-se 2,5 mL do reagente Folin-Ciocalteau 10%. A mistura permaneceu em repouso por 5 minutos. Em seguida, foram adicionados 2 mL de carbonato de sódio 4%. Após a incubação por 2 horas à temperatura ambiente e ao abrigo da luz, a absorbância foi medida em espectrofotômetro a 740 nm. Uma amostra em branco foi conduzida nas mesmas condições e os resultados dos compostos fenólicos totais foram expressos como equivalentes em epicatequina (mg AG.g^{-1} amostra em base úmida), calculados por meio do ajuste da curva de calibração de epicatequina com concentrações que variaram de 5 a 60 $\mu\text{g.mL}^{-1}$.

4.3.12 Espectroscopia de Infravermelho com Transformada da Fourier – FTIR

Os espectros foram obtidos por meio do FTIR-ATR – Espectroscopia de Infravermelho com Transformada de Fourier com Reflexão Total Atenuada, no equipamento Agilent Cary® 630, na região do infravermelho médio, utilizando-se a faixa espectral de 4000 a 500 cm^{-1} , com resolução de 4 cm^{-1} , 64 scans e leitura por meio do cristal de diamante. As amostras foram colocadas sobre a superfície do diamante para a realização da leitura. Antes de cada digitalização foi coletado o branco, para isto, foi realizada uma leitura sem amostra na superfície. O tempo de aquisição dos espectros foi de aproximadamente 30 segundos. Foram utilizados como variáveis os valores de números de onda onde ocorriam as máximas absorbâncias das bandas formadas nos espectros adquiridos na faixa de 4000 a 600 cm^{-1} e, para a análise estatística, foram empregados os valores de máxima absorbância citados, que variavam entre -0,5 e 0,5 unidades de absorbância (AU).

A Tabela 3 apresenta a codificação realizada para cada faixa de número de onda (variáveis) dos principais picos formados e seus respectivos grupos funcionais.

Tabela 3. Regiões espectrais em FTIR.

Códigos	Comprimento de onda (cm ⁻¹)	Grupos funcionais
C1	3283,90	O-H / N-H (Água associada/ proteína)
C2	2919,69	C=O (Lipídios)
C3	2849,12	CH ₂ / CH ₃ (Lipídios)
C4	2346,24	CH ₂ (Lipídios)
C5	2109,34	C-C (hidrocarbonetos- ácidos)
C6	1745,13	C=O (ésteres)
C7	1651,03	C=O de amidas (Proteína)
C8	1515,57	N-H / C=C (Proteína/ ácido carboxílico/ aromáticos)
C9	1439,32	CH ₂
C10	1374,43	CH ₃
C11	1234,10	N-H / C-O (Proteína/ ácido carboxílico)
C12	1145,68	C=O de ésteres, álcool e fenóis
C13	1016,70	(C - H) (C - O) (C = O) (carboidratos / lipídios)
C14	757,80	(C - H) / (C - O - H) / (C - O) / (C - C) / NH ₂ (carboidratos / lipídios/ proteína)

Fonte: Do Autor, 2018.

Após a obtenção dos espectros, as amostras foram organizadas em conjunto de dados numéricos, os quais foram manipulados por meio do Microsoft Office Excel 2010®. Foram realizadas as determinações de desvio padrão e coeficiente de variação. Os dados foram padronizados, criando uma variável (Z) com média zero e desvio padrão igual a 1, eliminando as diferenças entre as unidades de medidas das variáveis estudadas.

4.4 Análises quimiométricas

Os dados obtidos foram analisados utilizando o programa estatístico *Statistical Analysis System (SAS)*® *Studenty*, versão 9.0, para as análises de PCA e AD. Para o desenvolvimento das redes neurais artificiais, foi utilizado o *software Java Simulator Neural Network (Java NNS)* – *Java Neural Network Simulator 1.1* (Fischer et al., 2001), o qual que é baseado no *SNNS - Stuttgart Neural Network Simulator*.

4.4.1 Análise de Componentes Principais (ACP)

Para o desenvolvimento da ACP foi organizada uma matriz $m \times n$ (linhas x colunas), onde m representa a quantidade de amostras de amêndoas de cacau e n representa as variáveis analisadas em cada amostra. A matriz A1 foi composta por 82 amostras ($m=82$) e 18 variáveis ($n=18$), sendo as variáveis: proteína, umidade, cinzas, lipídios, aW, pH, acidez, açúcares redutores e não redutores, os parâmetros de cor L, a*, b* e dE, compostos fenólicos e teores de teobromina, cafeína, catequina e epicatequina. A matriz A2, composta pelas mesmas 82 amostras ($m=82$) e 14 variáveis ($n=14$) correspondentes aos números de onda em que ocorreram as maiores absorbâncias representadas nas bandas dos espectros obtidos no FTIR.

Após a formação das matrizes, os dados foram padronizados utilizando a Equação 1, para média igual a zero e desvio padrão igual a 1, a fim de eliminar a diferença das unidades de medidas das variáveis, os dados foram manipulados usando o programa Microsoft Office Excel 2010®.

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Em que: Z: variável padronizada; X: variável; μ : média; σ : desvio padrão

A partir dos dados padronizados foi obtida a matriz de variância e covariância (S), que foi fatorada numa matriz diagonal, cujos elementos são iguais a $a_{ii} < \lambda_i = 1, 2, \dots, 8$. A matriz foi obtida resolvendo a equação característica da matriz S, Equação 2, contendo autovalores com $\lambda_1 > \lambda_2 > \dots > \lambda_6$.

$$|S - \lambda I| = 0 \quad (2)$$

Em que: S é a matriz de variância e covariância; λ_i é o Autovalor e I é a Matriz Identidade.

Para cada autovalor (λ_i) foi determinado um autovetor α_i a partir da solução da Equação 3:

$$|S - \lambda I| \alpha_i = \emptyset \quad (3)$$

Em que: S é a matriz de variância e covariância; λ_i o Autovalor; I a Matriz Identidade; α_i o Autovetor e \emptyset : conjunto vazio.

A normalização dos autovetores (α_i) foram obtidos através da Equação 4.

$$\alpha_i = \frac{1}{\sqrt{\alpha_{i1}^2 + \alpha_{i2}^2 + \dots + \alpha_{i8}^2}} \begin{bmatrix} \alpha_{i1}^* \\ \alpha_{i2}^* \\ \vdots \\ \alpha_{i8}^* \end{bmatrix} = \frac{\alpha_{i1}^*}{\alpha_i} \quad (4)$$

Em que:

α_i é Autovetor normalizado;

$\alpha_{i1}, \alpha_{i2} \dots \alpha_{i8}$, são autovetores normalizados correspondentes aos grupos estudados;

$\alpha_{i1}^*, \alpha_{i2}^* \dots, \alpha_{i8}^*$, são autovetores não normalizados correspondentes aos grupos estudados.

Tomando-se os elementos do vetor α_i determinados como os coeficientes CP_i , temos que o i -ésimo CP é obtido pela Equação 5:

$$CP_i = a_{i1Z1} + a_{i2Z2} + \dots + a_{i8Z8} \quad (5)$$

Em que: $i = 1, 2, \dots, n$, e CP_i : i -ésimo componente principal; $\alpha_{i1}, \alpha_{i2}, \alpha_{in}$, elementos do i -ésimo autovetor (α_i) do i -ésimo componente principal (CP_i).

Para uma melhor visualização da dispersão dos coeficientes de correlação entre as variáveis e as CP's, foi utilizado o programa Sigma Plot, versão 11.0, para confecção dos gráficos.

4.4.1.1 Quantidade de CP e descarte de variáveis

Há três critérios mais difundidos na literatura acerca da escolha adequada do número de componentes principais necessários para explicar a variância dos dados. O primeiro, o critério de Kaiser, define que a quantidade necessária de componentes principais é igual ao número de autovalores > 1 para a análise dos dados padronizados. O segundo, o critério dos fatores interpretáveis, determina o número de componentes principais como sendo a soma do número de autovalores cuja soma represente mais que 70% da variação. Outro critério é a observação do *Scree Plot*, onde ocorrer a estabilização da curva é escolhido a quantidade de autovalores que representam a variância dos dados (OGILVIE et al., 2008).

Quanto ao descarte de variáveis, utilizou-se o método descrito por Pereira e Pereira (2004) cujo o critério é descartar as variáveis com menor valor da soma dos coeficientes de determinação nos primeiros componentes, geralmente utiliza-se os CPs selecionados pelos critérios estabelecidos no parágrafo acima. Assim, as variáveis mais importantes para os

componentes serão as de maior soma $R_{i1}^2 + R_{i2}^2 + R_{i3}^2$, e a que tiver menor soma deve ser descartada. Faz-se o descarte de uma variável por vez e reanalisa-se os dados por ACP. Onde, R_{i1}^2 corresponde a correlação da primeira variável com CP1, R_{i2}^2 corresponde a correlação da primeira variável com CP2 e assim sucessivamente.

4.4.2 Análise Discriminante

A análise discriminante consistiu em obter funções matemáticas capazes de classificar uma amostra de amêndoa de cacau em uma das populações π_i ($i=1$ – Litoral Sul, 2 – Baixo Sul, 3 – Médio Rio das Contas, 4 – Vale do Jequiriçá), buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente uma amostra em uma população π_1 , quando realmente pertence à população π_2 , π_3 ou π_4 . Também foram obtidas funções matemáticas capazes de classificar amêndoas de cacau em duas populações, uma que contém características de qualidade superior e outra de qualidade inferior.

Inicialmente, realizou-se a parte exploratória da análise, na qual foi estabelecido o número de grupos (4 para regiões e 2 para qualidade) e identificou-se as características que foram utilizadas com o objetivo de traçar o perfil geral de cada grupo visando realizar a discriminação ou separação das amostras e possibilitando a diferenciação das mesmas. Esta etapa consistiu no treinamento dos dados, onde é gerada a Função Discriminante de Anderson.

Os dados foram aleatorizados e divididos em dois grupos, um de teste e o outro de validação utilizando o algoritmo de Kenstone. Optou-se por usar a proporção 70% para treinamento e 30% para validação. Os dados de treinamento foram utilizados para obter a função discriminante linear e os dados de validação para testar a capacidade de generalização da função gerada.

A partir dos dados originais foram estimadas médias e matrizes de covariâncias das populações estudadas e obtidas as funções discriminantes de Anderson para o processo de classificação das amostras. A função discriminante de Anderson consistiu em uma combinação linear das variáveis originais utilizadas no estudo, a fim de maximizar a razão entre as classes e minimizar a dispersão dentro de cada grupo obtido, conforme se observa na Equação 6.

$$y = \beta + \beta_1 X_1 + \beta_i X_i + \varepsilon \quad (6)$$

Em que: Y é a Função linear discriminante; β é a Constante da função linear discriminante; $\beta_1 \dots \beta_n$, são os coeficientes das variáveis utilizadas no estudo; X são os escores das variáveis estudadas.

Após as fases anteriormente descritas, realizou-se a etapa de validação. A partir das equações lineares obtidas na fase de treinamento, os novos dados de validação foram aplicados a essas equações a fim de gerar as taxas de classificações e verificar a capacidade de generalização desses modelos.

4.4.3 Redes Neurais Artificiais

Foram testados dois conjuntos de configuração de rede neural artificial (RNA), o primeiro, a fim de possibilitar a classificação de amêndoas de qualidade superior (QS) e qualidade inferior (QI) e o segundo, com o objetivo de classificar as amêndoas em suas respectivas regiões (B, J, L e M) utilizando dados de composição e absorbâncias do espectro no infravermelho médio.

As redes neurais artificiais consistiram em uma estrutura complexa, composta por quatro camadas, a primeira de entrada, representada pelas variáveis medidas nas amêndoas, a segunda e terceira camadas ocultas, onde ocorre o processamento, aprendizado e estabelecimento da relação entre as variáveis de entrada e de saída, e a quarta camada, representada pelas saídas (qualidade ou região). Cada neurônio na camada de entrada recebe um sinal e propaga esse sinal para as camadas adjacentes, cuja força das ligações está diretamente relacionada aos pesos aplicados em cada neurônio durante a etapa de aprendizagem.

A metodologia supervisionada denominada perceptrons de multicamadas ou *multilayers perceptrons* (MLPs) foi a escolhida e utilizada para classificação das amostras, por ser amplamente utilizada dentre as classes de redes neurais supervisionadas. Foram utilizadas conexões “feed-forward” com parâmetros variando de 0 a 1 e os algoritmos testados foram o *Backpropagation* e o *Resilient Propagation*, a fim de encontrar qual deles melhor desempenharia o processo de convergência da rede.

A função logística foi utilizada devido ao fato da mesma ser uma das funções de ativações totalmente diferenciáveis, cujas derivadas de primeira ordem são conhecidas em todos os pontos de seu domínio de definição, assim, os dados foram padronizados entre 0 e 1, que representam os limites da imagem dessa função.

Com a finalidade de obter a melhor configuração da rede neural para classificação das amostras de cacau, foram feitos vários testes de configurações de rede que continham de 5 a

100 neurônios na primeira camada oculta e de 0 a 100 na segunda. A escolha do número de neurônios em cada camada escondida foi baseada na minimização da raiz do erro quadrático médio (Equação 7) e na maior porcentagem de acerto de classificação das amostras (RAI et al., 2005).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_d - x_p)^2} \quad (7)$$

Em que: n é o número de pontos de dados, e X_d e X_p são os valores experimentais e previstos em relação a qualidade ou região de cada amostra testada na rede.

A Tabela 4 apresenta os parâmetros utilizados na arquitetura das redes que foram testadas, em que se diferenciaram quanto a quantidade de neurônios das camadas de entrada, intermediárias e de saída e do algoritmo de aprendizado.

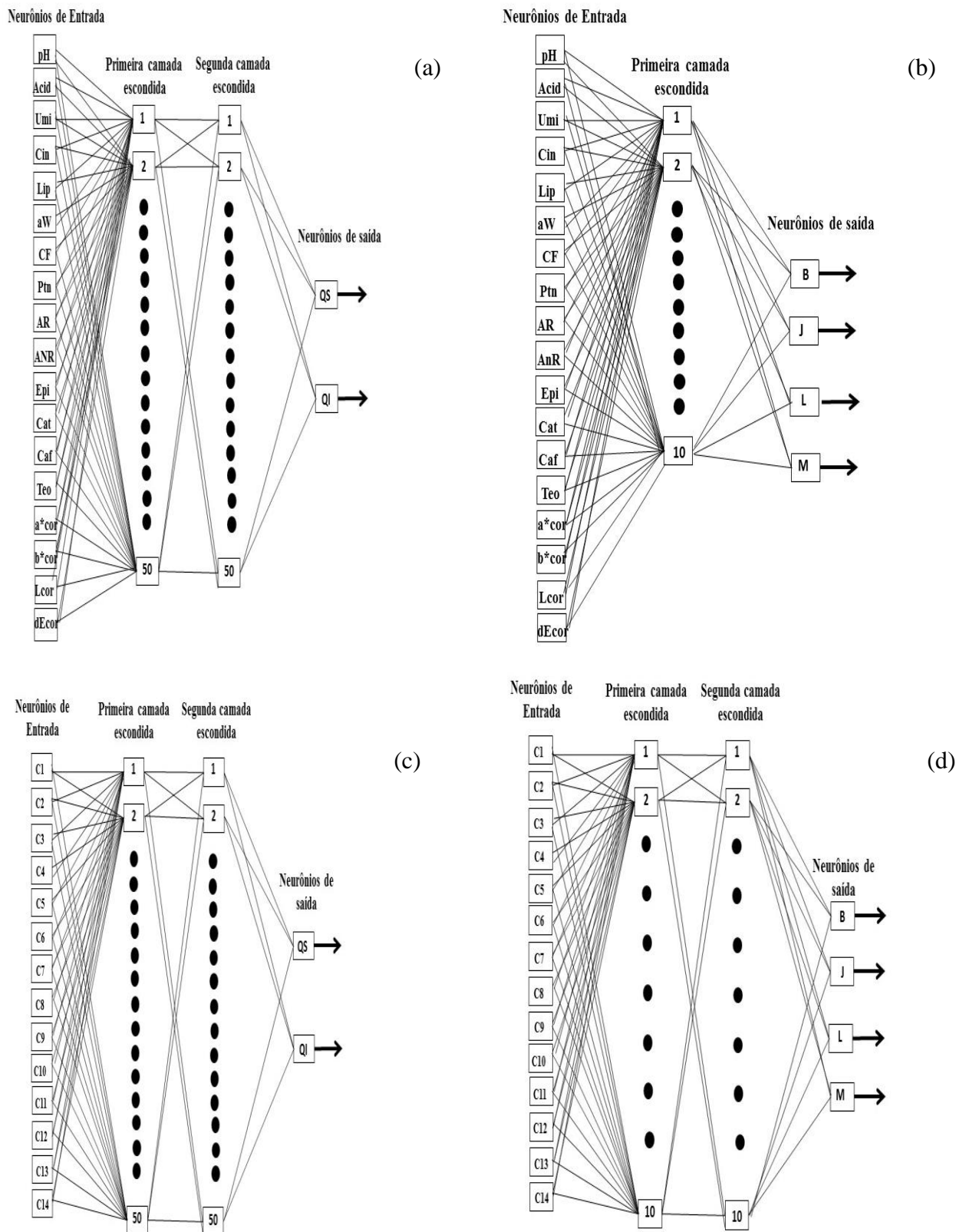
Tabela 4. Arquitetura de redes.

REDE USADA	REDE NEURAL DE RETROPROPAGAÇÃO
Aprendizagem	Supervisionado
Função de Ativação	Logística
Nº variáveis de entrada	14 e 18
Nº de camadas intermediárias	1 a 2
Nº de neurônios na primeira camada intermediária	5 a 100
Nº de neurônios na segunda camada intermediária	0 a 100
Nº variáveis de saída	2 e 4
Iteração	300

Fonte: Do Autor, 2018.

As redes escolhidas no presente estudo estão apresentadas na 2.

Figura 2. Configurações das redes escolhidas para dados de composição e espectrais na classificação de amêndoas de cacau (a) composição/qualidade (b) composição/região (c) FTIR/qualidade (d) FTIR/região.



As redes que utilizaram os dados de composição contêm 18 neurônios na camada de entrada que são representados pelas variáveis: pH, acidez, umidade, cinzas, lipídios, atividade de água, parâmetros de cor (L, a, b e dE), proteína, açúcar redutor e não redutor, compostos fenólicos totais, teor de cafeína, teor de teobromina, teor de catequina e teor de epicatequina. Quando utilizados os dados espectrais, as redes possuíam 14 neurônios na camada de entrada, referente aos picos (variáveis) formados pelas vibrações dos grupos funcionais presentes nas amostras. Em ambos os casos, a saída das redes variou de acordo com o que estava sendo avaliado, caso por região, continha 4 neurônios na camada de saída, caso por qualidade, continha 2 neurônios.

O software Java Neural Network Simulator, JavaNNS, versão 1.1 (Wilhelm-Schickard-Institute (WSI), Tübingen, Alemanha) foi utilizado para treinamento das Redes Neurais Artificiais.

4.4.3.1 Treinamento e validação

O conjunto de dados foi dividido em dois, o primeiro, com aproximadamente 70% dos dados, foi utilizado no treinamento da RNA, enquanto os 30% foram utilizados como conjunto de dados de validação, essa separação do conjunto de dados foi realizada por meio do algoritmo Kennard Stone.

A fim de classificar as amostras de acordo com a qualidade das amêndoas utilizando variáveis de composição e espectrais, foram utilizadas 83 amostras, das quais, 58 foram utilizadas no processo de treinamento e 25 na validação. A saída da RNA foi representada por vetores multidimensionais contendo 2 neurônios, correspondente aos grupos de QS e QI, dados criados em formato numérico, atribuindo 1 quando o grupo que foi corretamente classificado e 0, caso contrário. Sendo representados da seguinte maneira:

(1,0) = correspondente às amostras QS- Qualidade Superior;

(0,1) = correspondente às amostras QI- Qualidade Inferior;

Para classificação das amostras de acordo com a região de origem, utilizando variáveis de composição e espectrais, foram utilizadas 78 amostras, das quais, 55 foram utilizadas na etapa de treinamento e 23 na validação. Uma menor quantidade de amostras nas análises por região foi utilizada devido à falta de informações para identificar as suas origens. A saída da RNA também foi representada por vetores multidimensionais contendo 4 neurônios, correspondentes aos grupos B, J, L e M. Sendo representados da seguinte maneira:

(1, 0, 0, 0) = correspondente às amostras B – Baixo Sul;

(0, 1, 0, 0) = correspondente às amostras J – Vale do Jequiriçá;

(0, 0, 1, 0) = correspondente às amostras L – Litoral Sul;

(0, 0, 0, 1) = correspondente às amostras M – Médio Rio das Contas;

Os dados de entrada e saída foram normalizados pela equação 8:

$$x_{i, \text{norm}} = \frac{x_i - x_{\text{mín}}}{x_{\text{máx}} - x_{\text{mín}}} \quad (8)$$

Em que: $x_{i, \text{norm}}$ são os dados normalizados, x_i são os dados experimentais, $x_{\text{mín}}$ é o valor mínimo dos dados e $x_{\text{máx}}$ é o valor máximo dos dados.

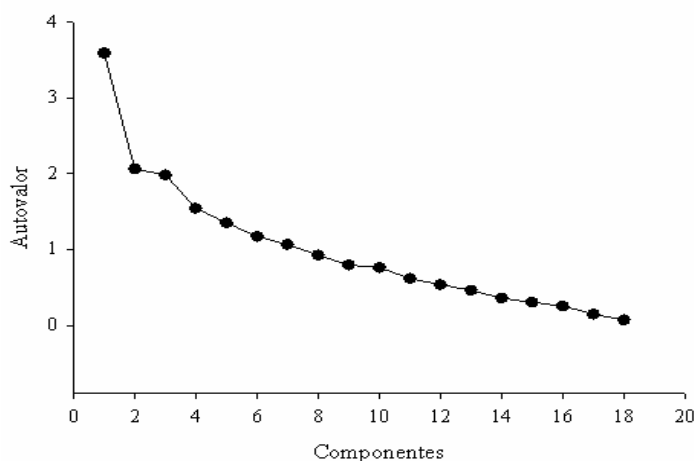
5. RESULTADOS E DISCUSSÃO

5.1 Análise Multivariada de dados e Rede Neural Artificial para avaliação da qualidade de amêndoas de cacau utilizando dados de composição

5.1.1 Análise de Componentes Principais

De acordo com o gráfico *Scree Plot* (Figura 3), do critério de Kaiser e dos fatores interpretáveis, pôde-se perceber a necessidade de 7 componentes principais para a explicação da variância do conjunto de dados obtidos. A partir desses resultados, observou-se uma redução de dimensionalidade 18 para 7, contudo essa redução não foi tão satisfatória.

Figura 3. *Scree Plot* dos autovalores associados aos componentes principais.



O excesso de componentes principais pode ser causado pela presença de variáveis redundantes, isto é, que forneçam informações já contidas nas outras variáveis. Por tanto, a identificação e descarte dessas variáveis podem levar à economia na obtenção de novos dados e simplificação na interpretação dos resultados. Foram excluídas as variáveis lipídios, proteína, compostos fenólicos totais, cinzas, açúcar não-reduzidor, umidade e teor de cafeína por

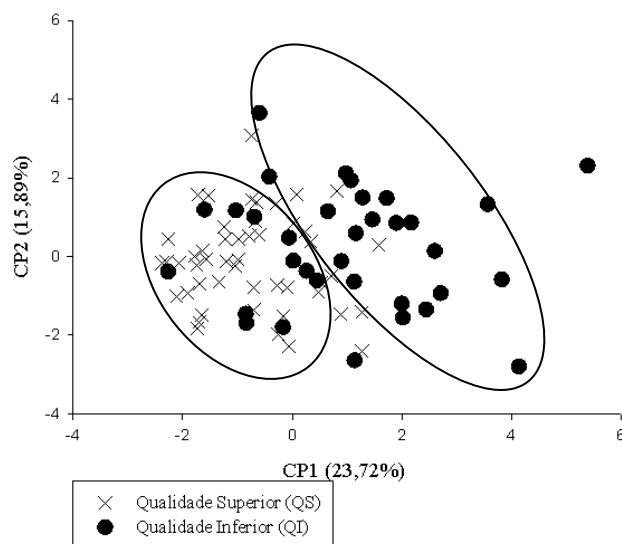
apresentarem redundância e não contribuírem para uma melhor explicação dos dados. A maioria das variáveis excluídas são trabalhosas e requer gasto com reagentes, isso leva a observar um ponto positivo para análise de ACP, que mostra que não há necessidade de realizar essas análises demoradas para obter informações necessárias para classificação das amostras de acordo com a qualidade.

A CP1 explica uma porcentagem de 23,7% da variação dos dados e as variáveis (pH, acidez, epicatequina e catequina) à ele relacionadas são importantes e representam o sabor ácido e adstringente das amêndoas, sendo parâmetros que possibilitam a diferenciação de qualidade, ou seja, quando uma amostra possui maior acidez e menor quantidade de compostos fenólicos majoritários das amêndoas significa que o processo de fermentação ocorreu de maneira adequada, e assim, essa amostra pode ser utilizada para uma posterior produção de chocolate.

A CP2 explica uma porcentagem de 15,89% da variação dos dados e as variáveis que apresentaram maior correlação foram os parâmetros de cor L e a* que são responsáveis pela luminosidade e variação da coloração do vermelho ao verde, indicando o quanto a amostra é capaz de refletir a luz. O índice de fermentação (IF) pode ser utilizado para diferenciar amostras de acordo com a sua qualidade, pois indica se o processo de fermentação ocorreu de maneira adequada. Sendo assim, uma amostra de QS deve apresentar maiores valores de L devido a coloração marrom ser mais clara, enquanto que a amostra de QI deve apresentar menores valores de L, devido a sua coloração ser mais escura e representada pela cor violeta, indicando maior teor de acidez e adstringência e que o processo de fermentação não ocorreu de maneira adequada.

A Figura 4 apresenta a disposição gráfica dos escores com relação à área que compreende CP1 e CP2, pois essa área proporciona uma melhor visualização dos grupos formados.

Figura 4. Gráfico de dispersão das amêndoas de cacau produzidas na Bahia em relação aos componentes principais CP1 e CP2 com dados de composição.



Pode-se observar que as amêndoas de qualidade superior apresentaram-se de forma mais agrupada quando comparada as amêndoas de qualidade inferior. A dispersibilidade das amêndoas de QI inviabilizou a verificação de padrões bem definidos das amêndoas de QS. As variáveis pH, acidez, epicatequina e catequina que estão correlacionadas com CP1 foram as responsáveis pela separação dos grupos formados.

5.1.2 Análise Discriminante

A Análise Discriminante é uma das técnicas multivariada que tem mais objetividade, pois, ou haverá uma boa classificação ou uma má classificação. Para que haja um resultado satisfatório é necessário fornecer as informações adequadamente e quanto maior a quantidade de amostras melhor será a estimativa ou separação da população.

Esta técnica teve a finalidade de separar as amostras da população em dois grupos, a saber, um com amêndoas de qualidade superior e outro de amêndoas com qualidade inferior. A primeira etapa, consistiu no treinamento ou parte exploratória da análise, que consiste em verificar características capazes de serem utilizadas para alocar as amostras nos dois grupos previamente definidos por meio do uso da função linear discriminante.

Em função do interesse em separar as amostras avaliadas em dois grupos, amostras de qualidade superior e qualidade inferior, duas funções discriminantes (Equação 9 e 10), foram geradas como combinações lineares das variáveis utilizadas no estudo, com o intuito de

aumentar a razão entre os grupos (QS e QI) e minimizar a dispersão dentro de cada grupo (FISHER, 1936).

$$Y_{QS} = - 0,3851 - 0,3316 * pH + 0,2300 * Acid - 0,2249 * aW + 5,3956 * Lcor + 0,3732 * acor + 0,2605 * bcor - 5,7830 * dEcor + 0,2737 * AR - 1,0020 * Epi - 0,1055 * Cat + 0,2984 * Teo \quad (9)$$

$$Y_{QI} = - 0,7730 + 0,4698 * pH - 0,3259 * Acid + 0,3187 * aW * - 7,6438 * Lcor - 0,5287 * acor - 0,3691 * bcor + 8,1926 * dEcor - 0,3878 * AR + 1,4195 * Epi + 0,1494 * Cat - 0,4227 * Teo \quad (10)$$

Em que: Y_{QS} : Função linear discriminante do grupo de amêndoas de qualidade superior; Y_{QI} : Função linear discriminante do grupo de amêndoas com qualidade inferior.

A segunda etapa, de validação, foi realizada para avaliar a capacidade de generalização das funções obtidas na etapa de treinamento utilizando os dados de validação. Ao analisar os coeficientes das equações, pode-se observar que L, dE e epicatequina são as variáveis que mais influenciaram para discriminação dos grupos. Sendo parâmetros que diferenciam as amêndoas pela sua coloração e adstringência, e que estão associados ao grau de fermentação. Em relação ao grupo de amêndoas com qualidade superior, o parâmetro de cor L influencia de forma positiva e a variável relacionada a adstringência influencia de forma negativa, sendo que influenciam de forma contrária quando em relação ao grupo de qualidade inferior. Essa observação é compatível com a caracterização das amêndoas, pois, amêndoas que possuem maiores valores de L e menor teor de epicatequina indicam que houve boa fermentação, e assim, são realmente amêndoas pertencentes ao grupo QS.

De acordo com as funções obtidas, obteve-se uma taxa de classificação de 88,2% e 75% na etapa de treinamento para o grupo de amêndoas com qualidade superior e qualidade inferior, respectivamente, e na etapa de validação de 26,7% de acerto para QS e 40% para QI. Evidenciando que o método foi capaz de diferenciar melhor as amostras na etapa de treinamento quando comparado com a etapa de validação, pois na etapa de validação houve baixa capacidade de discriminação entre as amostras indicando que a técnica não foi eficiente para indicar a qualidade de amêndoas de cacau.

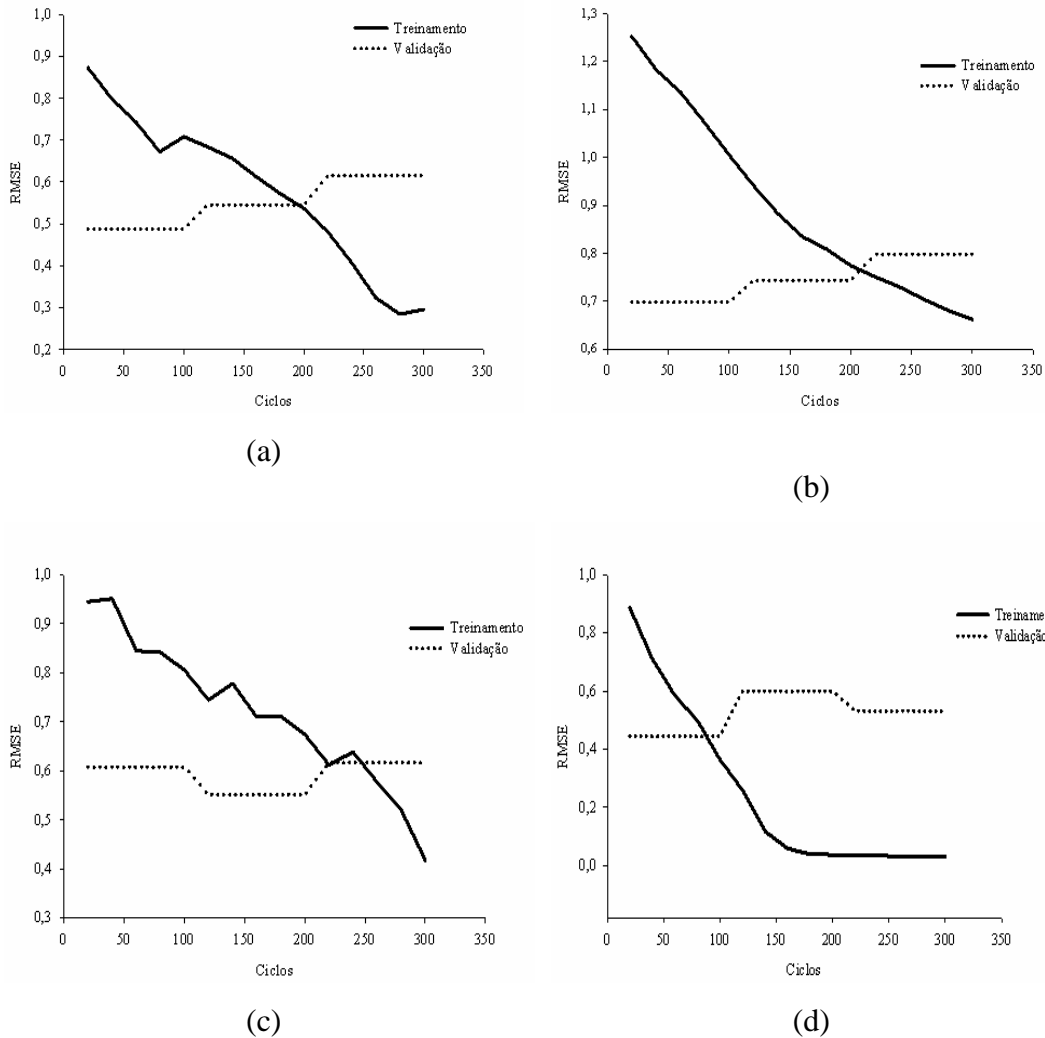
5.1.3 Rede Neural Artificial

Obteve-se a melhor configuração das RNA's testadas utilizando o algoritmo *backpropagation* com duas camadas intermediárias contendo cinquenta neurônios em cada uma (Figura 2a), apresentando um RMSE de 0,297 e 80% de taxa de classificação total correta, sendo que 86,6% e 70% foram as taxas de classificação para as amostras dos grupos de amêndoas de qualidade superior e inferior, respectivamente. Outras configurações testadas também mostraram baixo RMSE, porém foi escolhida a configuração de maior taxa de classificação das amostras.

No presente estudo, a RNA conseguiu generalizar os dados, pois foi capaz de discriminar corretamente as amostras em seus respectivos grupos de acordo com suas características e ainda com boas taxas de classificação ($\geq 70\%$), fato não observado ACP que mostrou a grande dispersão das amostras avaliadas como sendo de qualidade inferior, interferindo na formação de grupos homogêneos das amostras de acordo com sua qualidade e a AD que apresentou baixas taxas de discriminação para ambos os grupos existentes.

Geralmente o erro do treinamento apresenta-se maior que o de validação na fase inicial e ambos diminuem ligeiramente durante as primeiras iterações. Observou-se no gráfico da Figura 5a uma estabilização do erro nas primeiras 100 iterações e um leve aumento a cada 100 iterações. Esse aumento do RMSE da validação ocorre devido a rede já ter decorado o procedimento e perdendo a capacidade de generalização por meio da absorção dos ruídos.

Figura 5. Gráfico do erro quadrático médio (RMSE) em função do número de iterações nas etapas de treinamento e validação das redes neurais de melhor arquitetura (a) composição/qualidade (b) composição/região (c) FTIR/qualidade (d) FTIR/região.



5.2 Análise Multivariada de dados e Rede Neural Artificial para classificação de amêndoas de cacau de diferentes regiões da Bahia utilizando dados de composição

5.2.1 Análise de Componentes Principais

A partir das 18 variáveis de composição foi realizado a Análise de Componentes Principais, sendo a Matriz2 composta por 79 amostras e 18 variáveis que foram padronizados de acordo com Equação 1, e em seguida, realizado todo processo feito com as variáveis de composição das amostras em estudo.

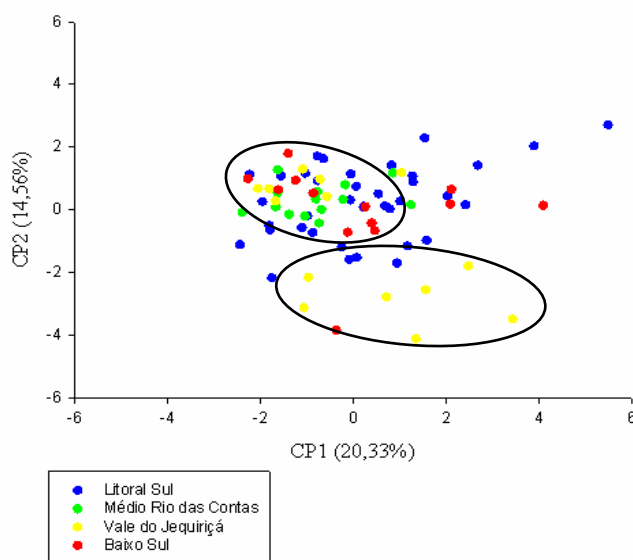
Foram necessários 7 componentes principais para a explicação da variância do conjunto de dados observados. A partir desses resultados observou-se uma redução de

dimensionalidade 18 para 7 indicando que a redução da dimensionalidade foi alcançada com sucesso. As variáveis lipídios, proteína, cinzas, compostos fenólicos totais e açúcar não-reduzidor foram consideradas redundantes e por isso foram descartadas, confirmando o resultado obtido na ACP para classificação de amêndoas de cacau de acordo com sua qualidade (seção 4.1.1), que descartou as mesmas variáveis, exceto o açúcar não-reduzidor.

A CP1 explica uma porcentagem de 20,33% da variação dos dados e as variáveis pH, acidez, açúcar redutor, epicatequina e catequina obtiveram maior correlação com esse CP. Essas variáveis são importantes por representar a acidez e adstringência das amêndoas, possibilitando diferenciá-las por qualidade. Já a CP2 explica uma porcentagem de 14,56% da variação dos dados e as variáveis que apresentaram maior correlação foram os teores de cafeína e teobromina que representam o gosto amargo nas amêndoas de cacau. Esses alcaloides tendem a diminuir conforme ocorre o processo de fermentação e secagem adequado, sendo assim, podem ser utilizados para verificação da qualidade do processo de beneficiamento das amostras.

A Figura 6 apresenta a disposição gráfica dos escores com relação à área que compreende CP1 e CP2, pois essa área proporciona uma melhor visualização do comportamento das amostras. Nesse gráfico é possível observar que as amêndoas do Médio Rio das Contas, Baixo Sul e um dos grupos formados com as amêndoas do Vale do Jequiriçá apresentaram comportamento similar e de forma menos dispersa. Já as amêndoas do Litoral Sul mostraram uma dispersão bem maior que as demais, ocupando todos os quadrantes do gráfico dificultando a verificação de padrões de separação das amêndoas por região. Essa grande dispersão das amostras do Litoral Sul pode ser devido ao fato de que essa região possui maior quantidade de produtores e da grande diferença no beneficiamento do cacau nessas diferentes fazendas. Assim, é necessário que haja momentos de capacitação para os produtores, mostrando a importância de se estabelecer padrões de qualidade para essa valiosa matéria prima.

Figura 6. Gráfico de dispersão das amêndoas dos territórios regiões Litoral Sul (L), Médio Rio das Contas (M), Vale do Jequiriçá (J) e Baixo Sul (B) em relação aos componentes principais CP1 e CP2 com dados de composição.



Resultado diferente ao do presente estudo foi o encontrado por Vázquez-Ovando et al (2015) que realizaram a Análise de Componentes Principais utilizando variáveis químicas (umidade, cinzas, gordura, proteínas, ácidos graxos), teores de polifenóis e capacidade antioxidante de 45 amostras de amêndoas de cacau da variedade Criollo coletadas em sete cidades do sul do México com objetivo de agrupá-las em suas respectivas regiões. As variáveis que mais contribuíram para a explicação da variância total dos dados foram os ácidos graxos, umidade e teor fenólico, onde foi possível classificar as amostras em sete grupos homogêneos. O teor de compostos fenólicos totais foi uma variável descartada na ACP do presente estudo e a sua não contribuição pode ser devido ao método de obtenção do extrato, no qual foi realizado uma extração aquosa, sendo que no estudo acima foi realizado uma extração hidroalcolica permitindo uma maior extração desses compostos.

5.2.2 Análise Discriminante

Esta técnica teve a finalidade de separar as amostras da população em quatro grupos, a saber, Litoral Sul, Médio Rio das Contas, Vale do Jequiriçá e Baixo Sul a partir das características mais pertinentes de cada região. A primeira etapa, consistiu no treinamento ou parte exploratória da análise, que consiste em verificar características capazes de serem

utilizadas para alocar as amostras nos quatro grupos previamente definidos por meio do uso da função linear discriminante.

Sendo assim, quatro funções discriminantes (Equação 11, 12, 13 e 14) foram geradas como combinações lineares das variáveis utilizadas no estudo, com o intuito de aumentar a razão entre os grupos (B, J, L e M) e minimizar a dispersão dentro de cada grupo (FISHER, 1936).

$$Y_B = - 1,1632 - 1,6094 * pH - 1,2365 * Acid + 0,2965 * aW - 0,5956 * Umid + 1,6152 * Lcor - 1,1282 * acor - 1,1704 * bcor - 2,2636 * dE + 0,4992 * AR - 0,7545 * Epi - 0,0368 * Cat - 0,1950 * Caf + 0,8495 * Teo \quad (11)$$

$$Y_J = - 1,3094 + 1,6508 * pH + 1,4967 * Acid + 0,2159 * aW + 0,2536 * Umid + 0,4983 * Lcor + 0,8687 * acor - 0,4370 * bcor + 0,9923 * dE - 0,1451 * AR - 0,9819 * Epi + 0,5025 * Cat - 0,4419 * Caf - 0,8411 * Teo \quad (12)$$

$$Y_L = - 0,1480 - 0,1172 * pH + 0,0498 * Acid - 0,0376 * aW + 0,1601 * Umid + 0,0060 * Lcor - 1,2111 * acor + 0,6524 * bcor - 1,0521 * dE + 0,0275 * AR + 0,5566 * Epi + 0,00037 * Cat + 0,1036 * Caf + 0,0580 * Teo \quad (13)$$

$$Y_M = - 0,3963 + 0,1143 * pH - 0,5184 * Acid - 0,3811 * aW - 0,1498 * Umid - 1,9684 * Lcor + 3,4167 * acor - 0,2712 * bcor + 3,8858 * dE - 0,3786 * AR + 0,1582 * Epi - 0,4704 * Cat + 0,3376 * Caf - 0,0801 * Teo \quad (14)$$

Em que: Y_B : Função linear discriminante do grupo de amêndoas da região do Baixo Sul; Y_J : Função linear discriminante do grupo de amêndoas da região do Vale do Jequiriçá; Y_L : Função linear discriminante do grupo de amêndoas da região do Litoral Sul; Y_M : Função linear discriminante do grupo de amêndoas da região do Médio Rio das Contas.

A segunda etapa, de validação, foi realizada para avaliar a capacidade de generalização das funções obtidas na etapa de treinamento utilizando os dados de validação. Ao analisar os coeficientes das equações, pode-se observar que os parâmetros de cor mais influenciaram na discriminação dos grupos B, L e M, ao passo que a função gerada para região J foi mais influenciada pelas variáveis pH e acidez.

De acordo com as funções obtidas, obteve-se uma taxa de classificação de 55,6%, 70%, 55,6% e 70% na etapa de treinamento para as regiões B, J, L e M, respectivamente, e na

etapa de validação de 25% para as regiões B e J, 27,3% para região L e 50% para região M. Nota-se que houve uma maior taxa de classificação das amostras em suas respectivas regiões na etapa de treinamento, e na etapa de validação uma baixa capacidade de discriminação entre as amostras, indicando que a técnica não foi eficiente para indicar a que região pertence as amostras qualidade de amêndoas de cacau.

As amêndoas de cacau de diferentes regiões possuem características similares e por isso houve uma considerável má discriminação entre elas. A região que mostrou características mais definidas foi a do Médio Rio das Contas, apresentando 50% de acerto na etapa de validação.

5.2.3 Rede Neural Artificial

Observando os dados obtidos das RNA's testadas a melhor configuração foi a RNA constituída pelo algoritmo *backpropagation* com apenas uma camada intermediária contendo dez neurônios (Figura 2b), apresentando um RMSE de 0,4879 e 65,2 de taxa de classificação total correta, sendo que 25%, 100%, 54,5% e 75% foram as taxas de classificação para as amostras dos grupos de amêndoas das regiões Baixo Sul, Vale do Jequiriçá, Litoral Sul e Médio Rio das Contas. As regiões Vale do Jequiriçá e Médio Rio das Contas obtiveram maior taxa de classificação correta e isso mostra que essas regiões possuem maior homogeneidade em suas características mais pertinentes. As amostras da região M utilizadas neste estudo apresentaram 100% de qualidade superior de acordo com suas características, ao passo que a região J apresentou 71% de qualidade inferior. Sendo assim, o resultado de classificação através da RNA pode ser atribuído a essa definição de características dessas regiões.

As demais configurações testadas apresentaram baixo RMSE, porém foi escolhida a configuração de menor RMSE e mais simples de neurônios devido a seu processo requerer menor esforço computacional. De acordo com Hagan (1996) não é recomendável utilizar uma rede de maior complexidade quando uma rede mais simples consegue reconhecer os dados de maneira satisfatória.

A RNA do presente estudo demonstrou sua capacidade em classificar amêndoas de cacau em suas respectivas regiões, e assim, é necessário o desenvolvimento de outros estudos a respeito da identificação de origem do cacau, pois esta matéria prima deve ser valorizada, como são os vinhos e os azeites de oliva com certificado de procedência. Permitindo o desenvolvimento do processo de rastreabilidade e contribuindo com a agregação de valor quando comprovado que esse produto apresenta características de qualidade superior ou está

associado a processos considerados ambientalmente corretos, como é o caso da produção de cacau cabruca na região do Litoral Sul da Bahia.

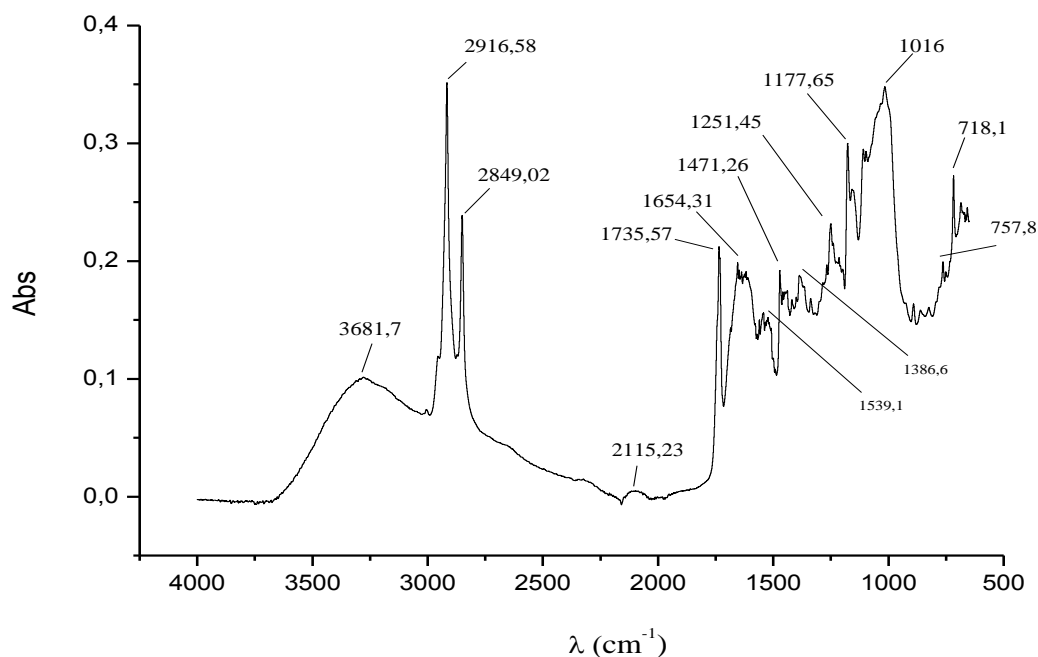
Essa indicação de procedência pode contribuir com a adoção de práticas que promovam uma melhoria na qualidade das amêndoas de cacau produzidas na Bahia. O momento é de pensar na qualidade das amêndoas produzidas nessas regiões e associar a imagem de produto sustentável que ajuda a preservar a mata atlântica como estratégia de viabilizar um aumento de receita perdido com a redução da produtividade devido a vassoura de bruxa.

5.3 Análise Multivariada de dados e Rede Neural Artificial para avaliação da qualidade de amêndoas de cacau utilizando dados espectroscópicos do FTIR.

A espectroscopia no Infravermelho FTIR foi empregada no intuito de realizar comparações entre os perfis espectroscópicos de amêndoas com características de qualidade superior e inferior.

O espectro FTIR da amêndoa de cacau é mostrado na Figura 7, e permite avaliar estruturalmente os grupos funcionais que contribuem para formação desses picos. Pode-se observar a formação de 14 picos e que a maioria deles são estreitos e pontiagudas, demonstrando que os modos vibracionais na amêndoa de cacau são bem definidos. As bandas formadas na faixa espectral entre $1800 - 800 \text{ cm}^{-1}$ são fundamentais para identificar grupos moleculares (LU e RASCO, 2012). Picos de absorção em 1016 , 1145 , 1150 e 1174 cm^{-1} são associados aos modos de alongamento dos grupos C-N, C-O e C-C. Enquanto que a banda $1368-1386 \text{ cm}^{-1}$ está associada á modos de flexão O-C-H, C-C-H e C-O-H (WILKERSON et al., 2013). Picos na região 1145 cm^{-1} se referem a estruturas aromáticas (C-OH) (MOVASAGHI, REHMAN e UR REHMAN, 2008). A banda larga encontrada $3563-3054 \text{ cm}^{-1}$ está associada ao alongamento do grupo (O-H) hidroxila (LU e RASCO, 2012).

A maioria das bandas observadas nos picos (909 , 1239 , 1472 , 1734 , 2917 e 2850 cm^{-1}), (989 , 1051 e 1067 cm^{-1}) e (1178 e 1650 cm^{-1}) foram associadas a lipídios, hidrocarbonetos (alcenos, aromáticos) e proteínas, respectivamente (CHE MAN et al., 2005; MOVASAGHI et al., 2008; SCHULZ e BARANSKA, 2007).

Figura 7. Espectro de amêndoa de cacau por FTIR.

5.3.1 Análise de Componentes Principais

A partir dos 14 comprimentos de onda obtidos, foi realizado a Análise de Componentes Principais, sendo a Matriz3 composta por 82 amostras e 14 variáveis codificadas como C1 a C14, que foram padronizados de acordo com Equação 1, e em seguida, realizado todo processo feito com as variáveis de composição das amostras em estudo.

Foram necessários 2 componentes principais para a explicação da variância do conjunto de dados observados. A partir desses resultados observou-se uma redução de dimensionalidade 14 para 2, indicando que a redução da dimensionalidade foi alcançada com sucesso sendo possível localizar cada indivíduo em um plano cartesiano. As variáveis C5, C6, C13 e C14 foram consideradas redundantes e por isso foram descartadas, sendo comprimentos de onda que representam os grupos funcionais presentes nas variáveis proteína, lipídios e carboidratos.

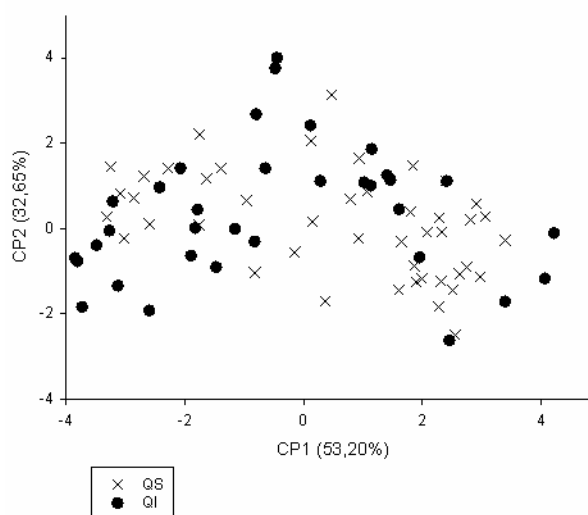
As variáveis codificada como C5 (2109 cm⁻¹) e C6 (1745 cm⁻¹) estão associadas ao grupo carboxílico (C=O) presente nos lipídios, enquanto que as variáveis C13 (1016 cm⁻¹) e C14 (757 cm⁻¹) estão associadas ao grupo amina (C-N) presente na proteína e ao grupo hidrocarbonetos aromáticos presentes nos compostos fenólicos, respectivamente. Essas variáveis (lipídios, proteínas e compostos fenólicos totais) também foram descartadas na ACP realizada com a composição das amêndoas de cacau. Sendo assim, há uma correlação entre dados espectrais e dados de composição para separação de amêndoas de acordo com a sua

qualidade. Para determinação da variável proteína e lipídios é necessário preparação, requer muito tempo e gasto com reagentes, e assim, esse descarte proporciona uma redução nas análises necessárias para classificar e diferenciar novas amostras.

Para a representação gráfica da ACP foram utilizados os CPs 1 e 2, que representam os seguintes percentuais de variação dos dados 53,20%, e 32,65%, respectivamente, totalizando 85,85% da variância total dos dados. As variáveis que obtiveram maior correlação com CP 1 foram as variáveis C2, C3, C4, C7, C8 e C9 (essas variáveis correspondem a grupos funcionais presentes em proteínas, lipídios e hidrocarbonetos), e com CP 2 foram as variáveis C1, C10, C11 e C12 que podem estar associadas a grupos funcionais presentes nas proteínas, lipídeos, fenóis e hidrocarbonetos alcenos ou aromáticos.

De acordo com a Figura 8, pode se observar que as variáveis que não foram descartadas promoveram um alto valor de explicação dos dados (CP1 e CP2), porém, não contribuíram para uma formação homogênea dos grupos de amêndoas de qualidade superior e inferior.

Figura 8. Gráfico de dispersão das amêndoas de cacau produzidas na Bahia em relação aos componentes principais CP1 e CP2 utilizando comprimentos de onda.



Houve uma grande dispersão das amêndoas cultivadas no estado da Bahia com relação a sua qualidade utilizando os dados espectrais, onde amostras com qualidade inferior se comportaram como sendo de qualidade superior, sendo que o contrário também ocorreu. Provavelmente as amostras possuem intensidades similares dos seus componentes, fato que dificultou a separação das mesmas. Vale lembrar que não foi possível visualizar uma separação bem definida na ACP com dados de composição, fortificando a idéia de que apesar

de terem diferença nos parâmetros avaliados para classificar as amêndoas como de qualidade superior e inferior possuem composição similar. E essa similaridade inviabilizou a formação dos grupos.

5.3.2 Análise Discriminante

De acordo com os dados espectrais foram geradas duas funções discriminantes (Equação 15 e 16) na etapa de treinamento, como combinações lineares das variáveis utilizadas no estudo, com o intuito de aumentar a razão entre os grupos (QS e QI) e minimizar a dispersão dentro de cada grupo (FISHER, 1936).

$$Y_{QS} = -0,6201 - 0,8702 * Y_{3283} - 2,8883 * Y_{2919} + 1,9043 * Y_{2849} + 0,8264 * Y_{2346} - 1,3239 * Y_{1651} - 0,3541 * Y_{1515} + 0,1714 * Y_{1439} - 1,2855 * Y_{1374} + 2,6249 * Y_{1234} - 1,1238 * Y_{1145} \quad (15)$$

$$Y_{QI} = -1,2444 + 1,325 * Y_{3283} + 4,0908 * Y_{2919} - 2,6969 * Y_{2849} - 1,1706 * Y_{2346} + 1,8755 * Y_{1651} + 0,5017 * Y_{1515} - 0,2430 * Y_{1439} + 1,8207 * Y_{1374} - 3,7182 * Y_{1234} + 1,5921 * Y_{1145} \quad (16)$$

Em que: Y_Q : Função linear discriminante do grupo de amêndoas de qualidade superior; Y_s : Função linear discriminante do grupo de amêndoas com qualidade inferior.

Na etapa de validação foi realizada a avaliação da capacidade de generalização das funções obtidas na etapa de treinamento utilizando os dados de validação. Ao analisar os coeficientes das equações, pode-se observar que as variáveis C2 (Y_{2919}) e C11 (Y_{1234}) mais influenciaram na discriminação dos grupos de qualidade superior e qualidade inferior.

Ao observar o percentual de acerto, houve altas taxas de classificação no treinamento para os grupos analisados, sendo 91,27% e 87,5% para QS e QI, respectivamente. Na etapa de validação as taxas de classificação correta foi de 50% e 60% para os grupos QS e QI, respectivamente, evidenciando que o método foi capaz de diferenciar as amostras em ambas as etapas com taxas de classificações corretas igual ou superior a 50%. Essa AD apresentou melhores resultados quando comparado com a AD que foi utilizado dados de composição para qualidade das amêndoas, mostrando que é possível discriminar as amostras de acordo com sua qualidade utilizando a espectroscopia no infravermelho associado a multivariada.

5.3.3 Rede Neural Artificial

A melhor configuração de RNA para classificação das amêndoas de cacau em relação a qualidade utilizando os dados do FTIR foi a constituída pelo algoritmo *backpropagation* com cinquenta neurônios nas duas camadas intermediárias (Figura 2c), apresentando um RMSE de 0,3039 e 79,2% de taxa de classificação total correta, sendo que 71,4% e 90% foram as taxas de classificação para as amostras dos grupos de amêndoas de qualidade superior e inferior, respectivamente. Outras configurações testadas também mostraram baixo RMSE, porém foi escolhida a configuração de maior taxa de classificação das amostras.

Pode-se observar que tanto utilizando dados de composição quanto dados espectrais foi possível obter taxas satisfatórias de classificação das amostras em relação a sua qualidade, confirmando a correlação existente entre os dados obtidos. Sendo assim, é oportuno dizer que é confiável a utilização dos dados espectrais fornecidos pelo equipamento Agilent Cary® 630 para poder identificar qualidade em amêndoas de cacau por meio da RNA. Dessa maneira haverá redução de tempo e de custos nos laboratórios, uma vez que o FTIR é uma técnica rápida e simples, que não precisa de reagentes nem de preparo laborioso com as amostras para reprodutibilidade dos espectros.

5.4 Análise Multivariada de dados e Rede Neural Artificial para classificação de amêndoas de cacau de diferentes regiões da Bahia utilizando dados espectroscópicos do FTIR.

5.4.1 Análise de Componentes Principais

A partir dos 14 comprimentos de onda obtidos, foi realizado a Análise de Componentes Principais, sendo a Matriz⁴ composta por 78 amostras e 14 variáveis codificadas como C1 a C14, que foram padronizados de acordo com Equação 1, e em seguida, realizado todo processo feito com as variáveis de composição das amostras em estudo.

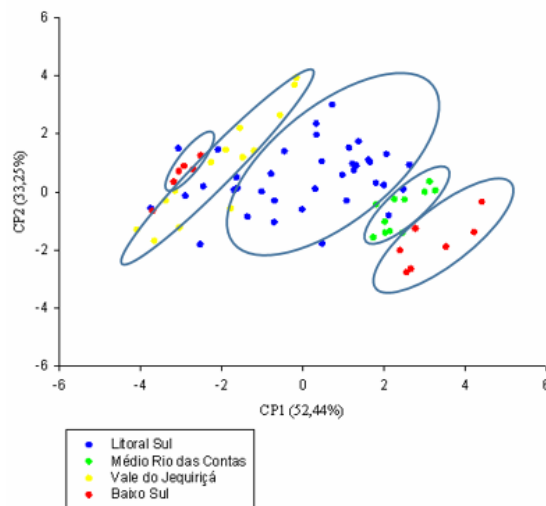
Foram necessários 2 componentes principais para a explicação da variância do conjunto de dados observados. A partir desses resultados observou-se uma redução de dimensionalidade 14 para 2, indicando que a redução da dimensionalidade foi alcançada com sucesso. As variáveis C5, C6, C13 e C14 foram consideradas redundantes e por isso foram descartadas, assim como ocorreu na ACP com dados espectrais na separação das amêndoas por qualidade. Isso já era o esperado, uma vez que são os mesmo dados utilizados com duas finalidades distintas, uma separar por qualidade e a outra separar por região.

Para a representação gráfica da ACP foram utilizados os CPs 1 e 2, que representam os seguintes percentuais de variação dos dados 52,44%, e 33,25%, respectivamente, totalizando

85,69% da variância total dos dados. As variáveis que obtiveram maior correlação com CP 1 e CP 2 foram as mesmas da ACP da seção 5.3.1, uma vez que os dados são os mesmos, mudando apenas o objetivo de classificação.

Para plotar a Figura 9 e visualizar a dispersão das amêndoas cultivadas nos diferentes territórios do estado da Bahia utilizou-se os escores dos CPs principais. Observou-se que houve formação de grupos distintos para cada região e apesar do grupo do Baixo Sul ter formado dois grupos, pode-se observar a forma compactada de cada um deles, podendo afirmar que em uma região pode-se encontrar distintas características nas amêndoas produzidas. Essa separação das amêndoas em suas respectivas regiões ocorre de acordo aos picos associados a CP1, enquanto que em relação a CP2 não diferem entre si.

Figura 9. Gráfico de dispersão das amêndoas dos territórios regiões Litoral Sul (L), Médio Rio das Contas (M), Vale do Jequiriçá (J) e Baixo Sul (B) em relação aos componentes principais CP1 e CP2 utilizando comprimentos de onda.



5.4.2 Análise Discriminante

Foram geradas quatro funções discriminantes (Equação 17, 18, 19 e 20) como combinações lineares das variáveis utilizadas no estudo, com o intuito de aumentar a razão entre os grupos (B, J, L e M) e minimizar a dispersão dentro de cada grupo (FISHER, 1936).

$$Y_B = -1,4413 + 0,1660 * Y_{3283} + 7,9823 * Y_{2919} - 7,4362 * Y_{2849} + 1,1923 * Y_{2346} - 1,8047 * Y_{1651} + 3,8702 * Y_{1515} - 1,1799 * Y_{1439} - 2,8334 * Y_{1374} + 2,7350 * Y_{1234} - 0,3238 * Y_{1145}$$

(17)

$$Y_J = - 2,3545 + 2,7536 * Y_{3283} + 6,8087 * Y_{2919} - 5,0844 * Y_{2849} - 1,4043 * Y_{2346} - 1,6755 * Y_{1651} + 3,2349 * Y_{1515} - 1,8931 * Y_{1439} + 4,3122 * Y_{1374} - 2,8513 * Y_{1234} - 0,8953 * Y_{1145}$$

(18)

$$Y_L = - 0,5720 - 0,0067 * Y_{3283} - 6,7287 * Y_{2919} + 6,4922 * Y_{2849} - 1,4901 * Y_{2346} + 0,3751 * Y_{1651} - 0,8294 * Y_{1515} + 0,7423 * Y_{1439} + 1,2700 * Y_{1374} - 1,8653 * Y_{1234} + 0,4574 * Y_{1145}$$

(19)

$$Y_M = - 4,8880 - 3,2055 * Y_{3283} + 4,6384 * Y_{2919} - 6,3909 * Y_{2849} + 4,8384 * Y_{2346} + 2,5409 * Y_{1651} - 4,9761 * Y_{1515} + 1,0565 * Y_{1439} - 5,7680 * Y_{1374} + 6,0292 * Y_{1234} - 0,0537 * Y_{1145}$$

(20)

Em que: Y_B : Função linear discriminante do grupo Baixo Sul; Y_J : Função linear discriminante do grupo Vale do Jequiriçá; Y_L : Função linear discriminante do grupo Litoral Sul; Y_M : Função linear discriminante do grupo Médio Rio das Contas.

A segunda etapa, de validação, foi realizada para avaliar a capacidade de generalização das funções obtidas na etapa de treinamento, utilizando os dados de validação. Ao analisar os coeficientes das equações, pode-se observar que todas as variáveis contribuíram para geração das funções discriminantes, porém as maiores contribuições foram dos comprimentos de onda C2, C3, C10 e C11 que estão associadas as vibrações de grupos funcionais encontrados em proteína, lipídios e compostos fenólicos.

Ao observar o percentual de acerto, houve uma taxa de classificação de 66,7%, 90%, 77,8% e 100% no treinamento e na validação de 50%, 25%, 63,6% e 75% para as regiões Baixo Sul, Vale do Jequiriçá, Litoral Sul e Médio Rio das Contas, respectivamente, evidenciando que o método foi capaz de diferenciar as amostras em ambas as etapas com taxas de classificações corretas acima de 50% exceto na validação da região Vale do Jequiriçá, onde as amostras foram classificadas 75% como sendo da região Litoral Sul, mostrando através desta análise que essa região possui uma baixa definição em suas características mais pertinentes. Esse resultado foi o melhor entre as análises discriminantes realizadas no presente estudo, evidenciando que é possível discriminar as amostras em suas respectivas regiões utilizando a espectroscopia no infravermelho associado a multivariada.

5.4.3 Rede Neural Artificial

Ao utilizar os dados espectrais do FTIR foram testadas várias configurações de RNA e a melhor configuração para classificar as amêndoas de cacau de acordo com sua região de origem foi a constituída pelo algoritmo *resilient propagation* com dez neurônios na primeira e segunda camada intermediária, visto na (Figura 2d), apresentando um RMSE de 0,1980 e 82,6% de taxa de classificação total correta, sendo que 50% foi a taxa de classificação para as amostras dos grupos de amêndoas das regiões Baixo Sul e Vale do Jequiriçá, e 100% para as regiões do Litoral Sul e Médio Rio das Contas.

Segundo Cybenko (1988), uma camada intermediária é essencial para aproximar qualquer função contínua, enquanto duas camadas intermediárias são capazes de aproximar qualquer função matemática. Neste caso, foi necessárias duas camadas intermediárias, porém com uma menor quantidade de neurônios. As demais configurações testadas apresentaram baixo RMSE, porém foi escolhida a configuração de menor RMSE e mais simples de neurônios devido a seu processo requerer menor esforço computacional.

A RNA apresenta um desempenho superior as demais análises multivariadas e isso pode ser atribuído ao fato de que são consideradas aproximadores universais de função. Esse método computacional não depende de nenhum conhecimento prévio ou premissa do relacionamento funcional entre as entradas e as saídas, a importante recomendação é que a amostra selecionada contenha informação suficiente para descrever a variável de resposta, bem como a sua distribuição, para selecionar apropriadamente a função não linear de ativação (MATIGNON, 2005).

A classificação de amostras de cacau em sua respectiva região é importante para contribuição de identificação de origem deste produto. E a utilização de dados obtidos pelo FTIR traz resultados que contribuem para essa designação de origem. Dados esses que são obtidos de maneira rápida e sem uso de materiais e reagentes que tornaria essa técnica inviável do ponto de visto do custo, tempo e realização de análises laboriosas.

6. CONCLUSÃO

A Espectroscopia de Infravermelho com Transformada de Fourier é capaz de fornecer informações sobre as principais variáveis envolvidas na qualidade das amêndoas de cacau e proporcionam a possibilidade de classifica-las de acordo sua origem quando aliada a análise multivariada.

A Análise de Componentes Principais (ACP) apresentou uma melhor visualização na separação das amostras quando foi utilizado os dados espectrais na diferenciação por região. Na ACP com dados de composição pôde-se observar a formação de dois possíveis grupos, um de forma menos dispersa formado pelas amêndoas das regiões Médio Rio das Contas, Baixo Sul e Vale do Jequiriçá, e o outro com amêndoas do Vale do Jequiriçá com comportamento mais disperso. Em relação a qualidade não houve uma boa separação dos grupos em nenhum dos dados utilizados (composição e espectrais). Um ponto interessante que pôde ser observado, foi a etapa de descarte das variáveis, pois, as variáveis representadas pelos comprimentos de onda que foram descartadas estão associadas as variáveis também descartadas na ACP com dados de composição, mostrando a possibilidade de utilizar dados do FTIR para classificar amostras de diferentes origens.

A Análise Discriminante apresentou baixa taxa de classificação ($\leq 50\%$) na fase de validação quando utilizados dados de composição, tanto em relação a separação por qualidade quanto por região. A AD com dados espectrais demonstrou maior capacidade de discriminação das amostras com taxas igual ou superior a 50% para qualidade e para origem, apresentando baixa classificação apenas para o grupo de amêndoas da região Vale do Jequiriçá. Demonstrando ser uma análise que possibilita a separação de amêndoas de cacau tanto por qualidade quanto por origem.

As RNA se apresentaram mais eficazes na classificação das amostras em relação às demais análises multivariadas, proporcionando erros satisfatórios. Foram obtidas taxas de classificação total de 80% e 79,2% para a classificação de acordo com a qualidade das amostras utilizando dados de composição e espectrais, respectivamente. E taxas de classificação total de 82,6 e 65,2% para a classificação de acordo com a região para dados de composição e espectrais, respectivamente. Assim, a Rede Neural Artificial pode ser utilizada para classificar amêndoas de cacau tanto em relação a qualidade quanto em relação a região de origem.

Este trabalho fornece uma alternativa estratégica para a análise e classificação rápida e simples das amêndoas produzidos nas regiões do Litoral Sul, Baixo Sul, Vale do Jequiriçá e

Médio Rio das Contas, além de contribuir para a autenticidade proteção da Designação de Origem.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ACULEY, P. C., SNITKJAER, P., OWUSU, M., BASSOMPIERE, M., TAKRAMA, J., NØRGAARD, L., et al. Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. **Journal of Food Science**, v. 75(6), p. S300–S307, 2010.

AFOAKWA, E., KONGOR, J., TAKRAMA, J., BUDU, A. Changes in acidification, sugars and mineral composition of cocoa pulp during fermentation of pulp preconditioned cocoa (*Theobroma cacao*) beans. **International Food Research Journal**, v. 20, p. 1215–1222, 2013a.

ARAUJO, Q. R., FERNANDES, C. A. F., RIBEIRO, D.O., EFRAIM, P., STEINMACHER, D., LIEBEREI, R., BASTIDE, P., ARAUJO, T. G. Cocoa Quality Index – A proposal, **Food Control**. 46, 49–54, 2014.

BAREL, M. **Première transformation du cacao**. In: J. Pontillon (Ed.), *Cacao et chocolat, production, utilisation, caractéristiques*. Paris: Lavoisier TEC e DOC, p. 96–116, 1998.

BATISTA, N. N. ANDRADE, D. P. RAMOS, C. L. DIAS, D. R. SCHWAN, R. F. Antioxidant capacity of cocoa beans and chocolate assessed by FTIR. **Food Research International**, 2016.

BECKETT, S.T. **Industrial Chocolate Manufacture and use, Fourth Edi, Wiley-Blackwell**, York, 2009.

BEG, M. S., AHMAD, S., JAN, K., BASHIR, K. Status, supply chain and processing of cocoa - A review. **Food Science e Technology**, v. 66, p. 108 e 116, 2017.

BERRUETA, L. A., ALONSO-SALCES, R. M. HEBERGER, K. Supervised pattern recognition in food analysis. **Journal of Chromatography A**, v. 1158, p. 196–214, 2007.

BIEHL, B., e VOIGT, J. **Biochemistry of chocolate flavour precursors**. International Cocoa Conference, Salvador de Bahia, 1996.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento-Instrução Normativa n° 57, de 12 de nov. de 2008. Regulamento Técnico da Amêndoa de Cacao. **Diário Oficial da União**, Brasília, 13 nov. 2008. Seção I, pg.2.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento-Instrução Normativa n° 38, de 23 de jun. de 2008. Regulamento Técnico da Amêndoa de Cacao. **Diário Oficial da União**, Brasília, 30 jun. 2008. Seção I, pg.2.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Implantação do cacauero em sistemas agroflorestais / Ministério da Agricultura, Pecuária e Abastecimento. **Comissão Executiva do Plano da Lavoura Cacaueira**. – Brasília: Mapa/ACS, 2014. 76 p.

BRERETON, R. G. **Applied chemometrics for scientists**.1ª. ed. Cichester, John Wiley, p.379, 2007.

BUYUKPAMUKCU, E., GOODALL, D. M., HANSEN, C. E., KEELY, B. J., KOCHHAR, S., WILLE, H. Characterization of peptides formed during fermentation of cocoa bean. **Journal of Agricultural and Food Chemistry**, v. 49(12), p. 5822–5827, 2001.

- CABO, P. S. A.; RIBEIRO, M. I. B.; FERNANDES, A. J.G.; MATOS, A. M. V. Hábitos e preferências dos consumidores de produtos tradicionais regionais certificados: o caso de trás-os-montes. **XXV Jornadas Hispanolusas de Gestión Científica**, 2015.
- CALIGIANI, A., MARSEGLIA, A., PRANDI, B., PALLA, G., SFORZA, S., Influence of Fermentation Level and Geographical Origin on Cocoa Bean Oligopeptide Pattern, **Food Chemistry**, 2016.
- CAMBRAI, A.; MARCHIONI, E.; JULIEN-DAVID, D.; MARCIC, C. Discrimination of Cocoa Bean Origin by Chocolate Polyphenol Chromatographic Analysis and Chemometrics. **Food Anal. Methods**, 2016.
- CECCHI, H. M. **Fundamentos teóricos e práticos em análise de alimentos**. 2.ed. Campinas: Unicamp, 2003. p. 73.
- CHE MAN, Y., SYAHARIZA, Z., MIRGHANI, M. E. S., JINAP, S., BAKAR, J. Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy. **Food Chemistry**, v. 90(4), p. 815–819, 2005.
- CHEN, Q.; CAI, J.; WAN, X.; ZHAO, J. Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. **Food Science and Technology**, v. 44, p. 2053-2058, 2011.
- CIE. Commission Internationale de L'Éclairage. Colorimetry. Vienna: CIE publication, 2^o ed., 1996.
- CONSELHO NACIONAL DO COMÉRCIO EXTERIOR (Brasil). Resolução nº 42. Rio de Janeiro, 1968. 9p.
- CRAFACK, M.; KEUL, H.; ESKILDSEN, C. E.; PETERSEN, M. A.; SAERENS, S.; BLENNOW, A.; SKOVMAND-LARSEN, M.; SWIEGERS, J. H.; PETERSEN, G. B.; HEIMDAL, H.; NIELSEN, D. S. Impact of starter cultures and fermentation techniques on the volatile aroma and sensory profile of chocolate. **Food Research International**, v. 63, p. 306–316, 2014.
- DAS, K. NENADIC, Z. Approximate information discriminant analysis: A computationally simple heteroscedastic feature extraction technique. **Pattern Recognition**, v. 41, n. 5, p. 1548-1557, 2008.
- DE BRITO, E. S., PEZOA, G. N. H., GALLAO, M. I., CORTELAZZO, A. L., FEVEREIRO, P. S., BRAGA, M. R. Structural and chemical changes in cocoa (*Theobroma cacao* L) during fermentation, drying and roasting. **Journal of the Science of Food and Agriculture**, v. 81(2), p. 281–288, 2001.
- DECOSTER, J. **Overview of Factor Analysis**, 1998. Disponível em: <http://www.stathelp.com/notes.html>.
- DIOMANDE, D., ANTHEAUME, I., LEROUX, M., LALANDE, J., BALAYSSAC, S., REMAUD, G. S., TEA, I. Multi-element, multi-compound isotope profiling as a means to distinguish the geographical and varietal origin of fermented cocoa (*Theobroma cacao* L.) beans. **Food Chemistry**, v. 188, p. 576–582, 2015.
- DUNCAN, R. C. **Factor Analysis, A Farm Management Research Tool**. Review of Marketing and Agricultural Economics, v.36, n.1, p.37-46, 1968.

- EERIKAINEN, T.; LINKO, P.; LINKO, S.; SIIMES, T.; ZHU, y-h. Fuzzy Logic and Neural Network Applications in Food Science and Technology. **Trends Food Science and Technology**, n.4, p. 237-242, 1993.
- EIM, V. S. SIMAL, S. ROSSELLÓ, C. José, B. A. F. Optimisation of the addition of carrot dietary fibre to a dry fermented sausage (sobrassada) using artificial neural networks. **Meat Science**, 94(3), pp. 341–348, 2013.
- EFRAIM, P. et al. Teores de compostos fenólicos de sementes de cacau de diferentes genótipos. **Brazilian Journal of Food Technology**, v. 9, n. 4, p. 229-236, 2006
- EFRAIM, P., PEZOA-GARCÍA, N. H., JARDIM, D.C.P., NISHIKAWA, A., HADDAD, R., EBERLIN, M. N. Influência da fermentação e secagem de amêndoas de cacau no teor de compostos fenólicos e na aceitação sensorial, **Ciência E Tecnol. Aliment.** 30, 142–150, 2010.
- ELWERS, S., ZAMBRANO, A., ROHSIUS, C. e LIEBEREI, R. Differences between the content of phenolic compounds in Criollo, Forastero and Trinitario cocoa seed (*Theobroma cacao* L.). **European Food Research and Technology**, v. 229, p. 937–948, 2009.
- FERNANDES, V. A., MÜLLER, A. J., SANDOVAL, A. J. Thermal, structural and rheological characteristics of dark chocolate with different compositions, **J. Food Eng.** 116, 97–108, 2013.
- FERREIRA, M. M. C; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: Calibração multivariada, um tutorial. **Química Nova**, 22(5): 724-731, 1999.
- FERREIRA, D. F. **Estatística multivariada**. Lavras: Editora UFLA, 2011. 675p.
- FERREIRA, A. C. R.; AHNERT, D.; NETO, B. A. M.; MELLO, D. L. N. **Guia de Beneficiamento de Cacau de Qualidade Instituto Cabruca**. Ilhéus, Bahia: 2013.
- FISCHER, I.; HENECKE, F.; BANNES, C.; ZELL, A. **Java Neural Network Simulator – JavaNNS, Version 1.1**. Wilhelm-Schickarad-Institute for Computer Science – Department of Computer Architecture, University of Tübingen, Alemanha, 2001.
- FISHER, R. A. Has Mendel’s work been rediscovered? **Annals of Science**, v. 1, p. 115–137, 1936.
- FONTES, M. J. V. Do Cacau ao Chocolate: trajetória, inovações e perspectivas das pequenas agroindústrias de cacau/chocolate. Tese (Doutorado), Universidade Federal Rural do Rio de Janeiro, Instituto de Ciências Humanas e Sociais, RJ, 200 p., 2013.
- FORSYTH, W. G. C. Cacao polyphenolic substances. II. Changes during fermentation. **Biochemistry Journal**, v. 51(4), p. 516–520, 1952.
- GONÇALVES, C.G. Avaliação do impacto da fermentação, secagem e torração sobre o perfil dos compostos fenólicos e da capacidade antioxidante de sementes de cacau (*Theobroma cacao* var. Forasteiro) produzidos no Estado do Pará. Dissertação (mestrado), Universidade Federal do Pará, Belém, 2016.
- GUTTMAN, L. **A new approach to factor analysis: The radex**. In P.F. Lazarsfeld (Ed.) **Mathematical Thinking in the Social Sciences**. New York: Free Press, 1954.
- HAGAN, M. T.; DEMUTH, H. B.; BEALE, M. H **Neural network design**. Boston: Pws Pub., 1996. <http://dx.doi.org/10.1109/72.329697>

HAIR JR.; J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. *Multivariate Data Analysis*. 7ed. Pearson Prentice Hall, 2010. 816 p.

HANSEN, C. E., DEL OLMO, M., BURRI, C. Enzyme activities in cocoa beans during fermentation. **Journal of the Science of Food and Agriculture**, v. 77(2), p. 273–281, 1998.

HASHIMOTO, J. C. **Caracterização de amêndoas de cacau produzidas em diferentes estados brasileiros e aplicação de espectroscopia no infravermelho próximo e quimiometria como alternativa para o controle de qualidade**. Tese (Doutorado), Universidade Estadual de Campinas, São Paulo, 2015.

HAYKIN, S. **Neural networks: A comprehensive foundation**. 2nded. Upper Saddle River: PrenticeHall, 1999. 842 p.

HAYKIN, S. **Neural Networks a Comprehensive Foundation**. New York: Macmillan Collge Publishing, 1994.

HORIMOTO, Y.; Durance, T.; Nakai, S.; Lukow, O.M. Neural Networks vs Principal Component Regression for Prediction of Wheat Flour Loaf Volume in Baking Tests. **Journal of Food Science**, n.60, p.429-433, 1995.

HU, Y., PAN, Z. J., LIAO, W., Li, J., GRUGET, P., KITTS, D. D., e LU, X. Determination of antioxidant capacity and phenolic content of chocolate by attenuated total reflectance-Fourier transformed-infrared spectroscopy. **Food Chemistry**, 202, 254–261, 2016.

HUE, C., GUNATA, Z., BERGOUNHOU, A., ASSEMAT, S., BOULANGER, R., SAUVAGE, F. X., et al. Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification. **Food Chemistry**, 148, 240–245, 2014.

HUE, C., GUNATA, Z., BREYSSE, A., DAVRIEUX, F., BOULANGER, R., SAUVAGE, F. X. Impact of fermentation on nitrogenous compounds of cocoa beans (*Theobroma cacao* L.) from various origins. **Food Chemistry**. 192, 958–964, 2016.

HUMSTON, E. M., KNOWLES, J. D., MCSHEA, A., e SYNOVEC, R. E. Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics. **Journal of Chromatography A**, v. 1217, p. 1963–1970, 2010.

IAL (INSTITUTO ADOLFO LUTZ). **Normas Analíticas do Instituto Adolfo Lutz. Métodos químicos e físicos para análise de alimentos**, v.1, 3. ed. São Paulo: IMESP, 2008.

INPI. **Instituto Nacional da Propriedade Industrial**. Pedidos de indicação geográfica concedidos e em andamento. Disponível em: < <http://www.inpi.gov.br/menu-servicos/indicacao-geografica/pedidos-de-indicacao-geografica-no-brasil>>. Acesso em 22 fev. 2018.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. **Artificial neural networks: a tutorial**. **Computer**, v. 29, n. 3, p. 31- 44, 1996.

JAIN, P. M.; SHANDLIYA, V. K. A survey paper on comparative study between Principal Component Analysis (ACP) and Exploratory Factor Analysis (EFA). **International Journal Of Computer Science And Applications**, v. 6, n. 2, p.373-375, 2013.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 816p, 1999.

- KADOW, D., BOHLMANN, J., PHILLIPS, W., LIEBEREI, R. Identification of main fine or flavour components in two genotypes of the cocoa tree (*Theobroma cacao* L.). **Journal of Applied Botany and Food Quality**, v. 86, p. 90–98, 2013.
- KATZ, W.T.; SNELL, J.W.; MERICKEL, M.B. **Artificial Neural Networks. Methods in Enzymology**, n. 210, p. 610-636, 1992.
- KHATTREE, R. e NAIK, D.N. **Multivariate data reduction and discrimination with SAS software**. Cary, NC, USA: SAS Institute Inc., 2000. 558 p.
- KIM, H., e KEENEY, P. Epicatechin content in fermented and unfermented cocoa beans. **Journal of Food Science**, v. 49(4), p. 1090–1092, 2006.
- KIRCHHOFF, P. M., BIEHL, B., CRONE, G. Peculiarity of the accumulation of free amino acids during cocoa fermentation. **Food Chemistry**, v. 31(4), p. 295–311, 1989.
- KRÄHMER, A., ENGEL, A., KADOW, D., Ali, N., UMAHARAN, P., KROH, L. W. e SCHULZ, H. Fast and neat – Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. **Food Chemistry**, 181, 152–159, 2015.
- KREIBICH, H. H. **Qualidade e segurança das amêndoas de cacau (*Theobroma cacao* L.) e seus produtos com relação aos contaminantes biológicos e a descontaminação de fungos toxigênicos com ozônio gasoso**. Dissertação (mestrado) – Universidade Federal de Santa Catarina, Centro Ciências Agrárias, Florianópolis- SC, 2016.
- KUMAR, N.; BANSAL, A.; SARMA, G. S.; RAWAL, R. K. Chemometrics tools used in analytical chemistry: An overview. **Talanta**, v. 123, p. 186–199, 2014.
- LEE, K. W; KIM, Y. K.; LEE, H. J.; LEE, C. Y. Cocoa has more phenolic phytochemicals and a higher antioxidant capacity than teas and red wine. **Journal of Agricultural and Food Chemistry**, Washington, v. 51, n. 25, p. 7292-7295, 2003.
- LEÓN-ROQUE, N., ABDERRAHIM, M., NUÑEZ-ALEJOS, L., ARRIBAS, S. M., CONDEZO-HOYOS, L. Prediction of fermentation index of cocoa beans (*Theobroma cacao* L.) based on color measurement and artificial neural networks. **Talanta**, v. 161, p. 31–39, 2016.
- LERCETEAU, E., ROGERS, J., PETIARD, V., CROUZILLAT, D. Evolution of cacao bean proteins during fermentation: A study by two-dimensional electrophoresis. **Journal of the Science of Food and Agriculture**, v. 79(4), p. 619–625, 1999.
- LIMA, L. J. R., ALMEIDA, M. H., NOUT, M. J. R., ZWIETERING, M. H. *Theobroma cacao* L., “The Food of the Gods”: Equality determinants of commercial cocoa beans, with particular reference to the impact of fermentation. **Critical Reviews in Food Science and Nutrition**, v. 51(8), p. 731–761, 2011.
- LINUSSON, A. et al. **Statistical molecular design of building blocks for combinatorial chemistry**. *J. Med. Chem.* v. 43, p.1320-1328, 2000.
- LOHUMI, S. H.; LEE, S.; LEE, H.; CHO, B-K. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. **Trends in Food Science e Technology**, v. 46, n.1, p. 85-98, 2015.
- LOHUMI, S.; MO, C.; KANG, J-S.; HONG, S-J.; CHO, B-K. Nondestructive evaluation for the viability of watermelon (*Citrullus lanatus*) seeds using fourier transform near infrared spectroscopy. **Journal of Biosystems Engineering**, v. 38, n. 4, 2013.

LOPEZ, A. S., e DIMICK, P. S. **Cocoa fermentation. biotechnology second, completely revised edition**. Wiley-VCH Verlag, p. 561–577, 1995.

LU, X., & RASCO, B. A. Determination of antioxidant content and antioxidant activity in foods using infrared spectroscopy and chemometrics: A review. **Critical Reviews in Food Science and Nutrition**, v. 52(10), p. 853–875, 2012.

MACHONIS, P. R., JONES, M. A., & KWIK-URIBE, C. Analysis of cocoa flavanols and procyanidins (DP 1–10) in cocoa-containing ingredients and products by rapid resolution liquid chromatography: Single-laboratory validation. **Journal of AOAC International**, v. 97, p. 166–172, 2014.

MATIGNON, R. **Neural network modeling using SAS enterprise miner**. AuthorHouse, 2005.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005. 297p.

MOVASAGHI, Z., REHMAN, S., UR REHMAN, D. I. Fourier transform infrared (FTIR) spectroscopy of biological tissues. **Applied Spectroscopy Reviews**, v. 43(2), p. 134–179, 2008.

NIELSEN, D. S., SNITKJAER, P., VAN DEN BERG, F. Investigating the fermentation of cocoa by correlating Denaturing Gradient Gel Electrophoresis profiles and Near Infrared spectra. **International Journal of Food Microbiology**, v. 125, p. 133–140, 2008.

NOGUEIRA, B. L. **Processamento do cacau: avaliação do teor nutricional do chocolate e dos outros derivados do cacau**. Monografia, Escola de Engenharia de Lorena da Universidade de São Paulo, 2015.

NØRSKOV-LAURITSENA, L. e H. B. Bürgi, “Cluster Analysis of periodic distributions; application to conformational analysis”, **J. Comput. Chem.**, v. 6, p. 216-228, 1985

OGILVIE, D.; MITCHELL, R.; MUTRIE, N.; PETTICREW, M.; PLATT, S. Perceived characteristics of the environment associated with active travel: development and testing of a new scale. **International Journal of Behavioral Nutrition and Physical Activity**, v. 5, n. 32, p. 1-10, 2008.

OLIVEIRA, L. F.; BRAGA, S. C. G. N.; AUGUSTO, F.; HASHIMOTO, J. C.; EFRAIM, P.; POPPI, R. J. Differentiation of cocoa nibs from distinct origins using comprehensive two-dimensional gas chromatography and multivariate analysis. **Food Research International**, n. 90, p. 133-138, 2016.

PACHECO, J.; CASADO, S.; PORRAS, S. Exact methods for variable selection in principal component analysis: Guide functions and pre-selection. **Computational Statistics and Data Analysis**, v.57, p. 95–111, 2013.

PAYNE, M. J., HURST, W. J., MILLER, K. B., RANK, C., STUART, D. A. Impact of fermentation, drying, roasting, and Dutch processing on epicatechin and catechin content of cocoa beans and cocoa ingredients. **Journal of Agricultural and Food Chemistry**, v. 58, p. 10518–10527, 2010.

PEREIRA, E. B. B.; PEREIRA, M. B. Um critério para o descarte de variáveis na análise de componentes principais. **Revista Universidade Rural**, Série Ciências Exatas e da Terra. v.23, n. 1-2, p. 1-7, 2004.

- PEREIRA-CARO, G., BORGES, G., NAGAI, C., JACKSON, M. C., YOKOTA, T., CROZIER, A., ASHIHARA, H. Profiles of phenolic compounds and purine alkaloids during the development of seeds of *Theobroma cacao* cv. Trinitario. **Journal of Agricultural and Food Chemistry**, v. 61, p. 427–434, 2013.
- PIASENTIN, F. B.; SAITO, C. H. Caracterização do cultivo de cacau na região econômica Litoral Sul, Sudeste da Bahia. **ESTUDO e DEBATE**, Lajeado, v. 19, n. 2, p. 63-80, 2012.
- PIASENTIN, F. B.; SAITO, C. H. The different methods of cocoa farming in southeastern Bahia, Brazil: historical aspects and perceptions. **Cienc. Hum.**, Belém, v. 9, n. 1, p. 61-78, 2014.
- PWC Brasil. **A Cadeia Produtiva do Cacau no Brasil- Pesquisa de Mercado**. Centro de Serviços em Agribusiness e Agribusiness Research e Knowledge Center Pwc Brasil - Ribeirão Preto, SP, 40p. 2012.
- QUESNEL, V. C. Agents inducing the death of cacao seeds during fermentation. **Journal of the Science of Food and Agriculture**, v. 16, p. 441–447, 1965.
- RAI, P., MAJUMDAR, G.C., DASGUPTA, S., DE, S. Prediction of the viscosity of clarified fruit juice using artificial neural network: a combined effect of concentration and temperature. **Journal of Food Engineering**, v. 68, n.4, p. 527-533, 2005.
- RAMADHAS, A. S. et al. Artificial neural networks used for the prediction of the cetane number of biodiesel. **Renew. Energ.**, v. 31, n. 15, p. 2524-2533, dez. 2006.
- REGAZZI, A. J. **Análise multivariada, notas de aula INF 766**, Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.
- RISNER, C. H. Simultaneous Determination of Theobromine, (+)-Catechin, Caffeine, and (–)-Epicatechin in Standard Reference Material Baking Chocolate 2384, Cocoa, Cocoa Beans, and Cocoa Butter. **Journal of Chromatographic Science**, Vol. 46, p. 892-899, 2008.
- ROHSIUS, C., MATISSEK, R., LIEBEREI, R. Free amino acid amounts in raw cocoas from different origins. **European Food Research and Technology**, v. 222, p. 432–438, 2006.
- ROHSIUS, C., ELWERS, S., LIEBEREI, R. **Cocoa Atlas**. German Cocoa and Chocolate Foundation, 2010.
- SALES, J. H.; CÂNDIDA. T. Efeito da temperatura sobre a amêndoa de cacau: secador vertical. **Revista GEINTEC**, v. 6, n. 3, p. 3437-3446, 2016.
- SALTINI, R, AKKERMAN, R., FROSCH, S. Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. **Food Control**, n. 29, p. 167-187, 2013.
- SCHWAN, R. F., WHEALS, A. E. The microbiology of cocoa fermentation and its role in chocolate quality. **Critical Reviews in Food Science and Nutrition**, v. 44(4), p. 205–221, 2004.
- SCHULZ, H., & BARANSKA, M. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. **Vibrational Spectroscopy**, v. 43(1), p. 13–25, 2007.
- SHARAF, M. A.; ILMAN, D. L.; KOWALSKI, B. R. **Chemometrics**. 1a. ed. New York, Wiley, p. 332, 1986.

- SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais: para engenharia e ciências aplicadas**. São Paulo: Editora Artliber Ltda, 2010, 399p.
- SINGH, K. P. et al. Artificial neural network modeling of the river water quality: A case study. **Ecological Modelling**, n. 220, v. 6, p. 888-895, 2009.
- STARK, T., BAREUTHER, S., HOFMANN, T. Molecular definition of the taste of roasted cocoa nibs (*Theobroma cacao*) by means of quantitative studies and sensory experiments. **Journal of Agricultural and Food Chemistry**, v. 54, p. 5530–5539, 2006.
- SUAZO, Y., DAVIDOV-PARDO, G., & AROZARENA, I. Effect of fermentation and roasting on the phenolic concentration and antioxidant activity of cocoa from Nicaragua. **Journal of Food Quality**, v. 37, p. 50–56, 2014.
- SYNOVEC, R. E., HUMSTON, E. M., SKOGERBOE, K. J., HOGGARD, J. C. Cacao bean quality assessment using 2D gas chromatography combined with time-of-flight mass spectrometry and chemometrics. **Abstracts of Papers of the American Chemical Society**, 243, 2012.
- TEÓFILO, R. F. et al. Sorting Variables by Using Informative Vectors as a Strategy for Feature Selection in Multivariate Regression. **J. Chemom.** v. 23, p. 32-48, 2009.
- TEYE, E., HUANG, X., DAI, H., CHEN, Q. Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 114, p. 183–189, 2013.
- THERMO. **Introduction to Fourier Transform Infrared Spectrometry**. Thermo Nicolet, Madison, USA. Disponível em: http://www.niu.edu/analyticallab/_pdf/ftir/FTIRintro.pdf. Acesso em: 10 de Agosto de 2017.
- UTAMI, R. R., ARMUNANTO, R., RAHARDJO, S., SUPRIYANTO. Effects of Cocoa Bean (*Theobroma cacao* L.) Fermentation on Phenolic Content, Antioxidant Activity and Functional Group of Cocoa Bean Shell. **Pakistan Journal of Nutrition**, 15: 948-953, 2016.
- VALENTE, M. E. R.; PEREZ, R.; RAMOS, A. M.; CHAVES, J. B. P. Indicação geográfica de alimentos e bebidas no Brasil e na União Europeia. **Ciência Rural**, v. 42, n. 3, p. 551- 558, 2012
- VÁZQUEZ-OVANDO, A., MOLINA-FREANER, F., NUÑEZ-FARFÁN, J., BETANCUR-ANCONA, D., SALVADOR-FIGUEROA, M. Classification of cacao beans (*Theobroma cacao* L.) of southern Mexico based on chemometric analysis with multivariate approach, **Eur. Food Res. Technol.** 240,1117–1128, 2015.
- VIEIRA, L. R., EFRAIM, P., VAN DE WALLE, D., DE CLERCQ, N., DEWETTINCK, K. Influence of Brazilian Geographic Region and Organic Agriculture on the Composition and Crystallization Properties of Cocoa Butter, **JAOCS, J. Am. Oil Chem. Soc.** 92, 1579–1592, 2015.
- WILLIAMS, B.; BROWN, T.; ONSMAN, A. Exploratory factor analysis: A five-step guide for novices. **Australasian Journal of Paramedicine**, v.8, n.3, p.1-13, 2010.
- WILKERSON, E. D., ANTHON, G. E., BARRETT, D. M., SAYAJON, G. F. G., SANTOS, A. M., RODRIGUEZ- SAONA, L. E. Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. **Journal of Agricultural and Food Chemistry**, v. 61(9), p. 2088–2095, 2013.

WYTHOFF, B.J. et al. Spectral Peak Verification and Recognition Using a Multi Layered Neural Network. **Analytical Chemistry**, n. 62, p. 702-709, 1990

WOLLGAST, J., e ANKLAM, E. Review on polyphenols in Theobroma cacao: changes in composition during the manufacture of chocolate and methodology for identification and quantification. **Food Research International**, v. 33, p. 423 e 447, 2000.

WOOD, G. A. R., e LASS, A. R. **Cocoa**. John Wiley e Sons, 1985.

ZHOU, W. THERDTHAI, N. **Artificial neural networks in food processing**. M.M. Farid (Ed.), Mathematical modeling of food processing, CRC Press, Boca Raton, pp. 901–920. 2010.

8. ANEXOS

ANEXO 1



MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA BAIANO
IF BAIANO – CAMPUS URUÇUCA

FORMULÁRIO A SER PREENCHIDO PARA CADA AMOSTRA

17 de Abril (data limite para envio das amostras)		
Tipo de dados	Descrição	Respostas
A. Pessoa que preencheu o formulário		
Nome completo do remetente:		
Organização:		
Endereço de e-mail/telefone		
Data informações enviadas:	___/___/____	
B. Tipo de amostra		
Nome da região/território		
Número de amostra	Identificação da amostra de cacau	
Tipo da amostra?	Comercial: (1) SIM (2) NÃO	
Especifique		
Produção anual		
Área plantada		
C. Informações sobre as origens da amostra		
Nome completo do produtor da amostra: Nome do agricultor, associação cooperativa ou outros		
Local da exploração ou	Nome da Fazenda:	

plantação		
	Endereço:	
	Cidade/Distrito:	
	Território:	
Clima	Início do período seco (mês do ano):	
	Final do período seco (mês do ano):	
As práticas agrícolas	Tipo de agricultura: tradicional / orgânicos / outros, especifique.	
	Certificado? (1) SIM (2) NÃO	
	Uso de fertilizante? (1) SIM (2) NÃO	
	Se usar fertilizante, por favor, especifique qual tipo? (1) Químico (2) Orgânico	
	Uso de pesticidas? (1) SIM (2) NÃO	
	Se usar pesticidas, por favor, especifique qual tipo?	
	Uso de sombra permanente? (1) SIM (2) NÃO	
	Sistema (1) Cabruca (2) Agroflorestal (3) Pleno Sol	
Variedade	Nome local da variedade:	
	Genética da variedade: (1) Comum (2) Híbrido (3) Clonal (4) Misto	
Arvores	As árvores são plantadas a partir de (1) sementes (2) enxertia (3) clonal (4) outros	
D. Informações sobre quando e como foram preparadas as amostras de cacau		
Características amostra e práticas pós-colheita	Fermentadas (1) SIM (2) NÃO	
	Quando a amostra foi fermentada e seca (mês/ano):	
	Tempo entre a colheita e quebra (em dias):	
	Fermentação realizada em: (1) caixas de madeira (2) pilhas (3) em sacos (4) outros. Se outro, especifique	
	Massa total de fermentação em montes ou em caixas (kg):	
	Duração da fermentação (dias):	
	Número de revolvimento durante a fermentação:	
Secagem	(1)Barcaça (2) Estufa plástica (3) Secador (4) Outros	
	Duração do processo de secagem total (dias):	

ANEXO 2

Planilha1

INSTITUTO FEDERAL BAIANO – IF BAIANO

LAUDO DE AVALIAÇÃO FÍSICO-SENSORIAL PARA AMENDOAS DE CACAU

Amostra: _____ Data: _____

Avaliador(a): _____

Atributos externos – básico

<input type="text"/> % umidade <input type="text"/> Tamanho <input type="text"/> % ME <input type="text"/> %IMP <input type="text"/> Cor	<input type="text"/> % Mofo <input type="text"/> % Germinado <input checked="" type="checkbox"/> Insetos <input type="text"/> % Quebradas <input type="text"/> % Aglutinadas
--	--

Escala de intensidade do atributo

NOTAS	1	2	3	4	5
-------	---	---	---	---	---

Cheiro – amêndoa crua

ATRIBUTOS CACAU	DEFEITOS
<input type="checkbox"/> vinagre	<input type="checkbox"/> defumado
<input type="checkbox"/> cacau	<input type="checkbox"/> fumaça
<input type="checkbox"/> nozes	<input type="checkbox"/> mofado/bolor
<input type="checkbox"/> frutado	<input type="checkbox"/> podre/lixo
<input type="checkbox"/> especiarias	<input type="checkbox"/> rançoso
<input type="checkbox"/> floral	<input type="checkbox"/> carne
<input type="checkbox"/>	<input type="checkbox"/> animal

Gosto – amêndoas descascadas e moídas

acidez amargor adstringência

Aroma – amêndoas descascadas e moídas

ATRIBUTOS CACAU	DEFEITOS
<input type="checkbox"/> cacau	<input type="checkbox"/> defumado
<input type="checkbox"/> frutas frescas	<input type="checkbox"/> mofado/bolor
<input type="checkbox"/> frutas secas	<input type="checkbox"/> amônia
<input type="checkbox"/> vegetais	<input type="checkbox"/> podre/lixo
<input type="checkbox"/> ervas	<input type="checkbox"/> medicinal/fenólico
<input type="checkbox"/> floral	<input type="checkbox"/> rançoso/queijo
<input type="checkbox"/> nozes	<input type="checkbox"/> sulfuroso/borracha
<input type="checkbox"/> especiarias	<input type="checkbox"/> Pungente
<input type="checkbox"/> Gordura	<input type="checkbox"/> Metálico
<input type="checkbox"/> Maltado/caramelo	<input type="checkbox"/> Animal
<input type="checkbox"/> Amanteigado	<input type="checkbox"/>

Teste de corte %

<input type="checkbox"/> Marrom <input type="checkbox"/> Parcialmente marrom <input type="checkbox"/> Violeta <input type="checkbox"/> Violeta sem fissuras	<input type="checkbox"/> ardósia <input type="checkbox"/> mofo <input type="checkbox"/> insetos <input type="checkbox"/> sobre fermentada
--	--

IMPRESSÃO GLOBAL

ass. técnico(a) _____