



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ELETRÔNICA E SISTEMAS
CURSO DE GRADUAÇÃO EM ENGENHARIA ELETRÔNICA

GRAYCE DE FÁTIMA TRINDADE FIGUEIRA

**ANÁLISE E CLASSIFICAÇÃO DE IMAGENS DE MANCHAS DA PELE POR
EXTRAÇÃO DE ATRIBUTOS DE TEXTURA PARA IDENTIFICAÇÃO DE
MELANOMA**

RECIFE

2019

GRAYCE DE FÁTIMA TRINDADE FIGUEIRA

**ANÁLISE E CLASSIFICAÇÃO DE IMAGENS DE MANCHAS DA PELE POR
EXTRAÇÃO DE ATRIBUTOS DE TEXTURA PARA IDENTIFICAÇÃO DE
MELANOMA**

Trabalho de Conclusão de Curso apresentado à
Universidade Federal de Pernambuco, como
parte dos requisitos para obtenção do título de
Engenheiro Eletrônico.

Orientador: Prof. Dr. Marcos Antônio Martins de Almeida.

RECIFE

2019

Catálogo na fonte
Bibliotecária Valdicéa Alves, CRB-4 / 1260

F475a Figueira, Grayce de Fátima Trindade.
Análise e classificação de imagens de manchas da pele por
extração de atributos de textura para identificação de melanoma. /
Grayce de Fátima Trindade Figueira - 2019.
102folhas, Ils. e Tab.

Orientador: Prof. Dr. Marcos Antônio Martins de Almeida.

TCC (Graduação) – Universidade Federal de Pernambuco. CTG.
Departamento de Graduação em Engenharia Eletrônica, 2019.
Inclui Referências , Anexos e Apêndices.

1. Engenharia Eletrônica. 2. Análise de textura. 3. Extração
de características. 4. Diagnóstico. 5. Melanoma. 6. Classificação
de imagens(Orientador). II. Título.

UFPE

621.381 CDD (22. ed.) BCTG/2019-131

GRAYCE DE FÁTIMA TRINDADE FIGUEIRA

**ANÁLISE E CLASSIFICAÇÃO DE IMAGENS DE MANCHAS DA PELE POR
EXTRAÇÃO DE ATRIBUTOS DE TEXTURA PARA IDENTIFICAÇÃO DE
MELANOMA**

Trabalho de Conclusão de Curso apresentado à
Universidade Federal de Pernambuco, como
parte dos requisitos para obtenção do título de
Engenheiro Eletrônico.

Aprovada em: 28/03/2019.

BANCA EXAMINADORA

Prof. Dr. Marcos Antônio Martins de Almeida (Orientador)

Universidade Federal de Pernambuco

Prof^ª. Dr^ª. Patrícia Silva Lessa

Universidade Federal de Pernambuco

AGRADECIMENTOS

Agradeço, primeiramente, à Deus por permitir que eu conseguisse tudo que conquistei até hoje e por me dar forças para continuar sempre, pois o caminho até aqui não foi fácil.

Agradeço à minha família, namorado e amigos, que sempre me apoiaram e confiaram no meu potencial, e acima de tudo, estiveram ao meu lado em todos os momentos da minha jornada, inclusive nos mais difíceis que pareciam intermináveis, sem deixar que eu desistisse.

Aos nossos professores, mestres, que sempre estiveram disponíveis para nossas dúvidas, nos ensinando, além dos assuntos acadêmicos, os caminhos da Engenharia Eletrônica, e especialmente ao meu professor orientador, Prof. Marcos Martins, pelo apoio e estímulo durante a execução deste trabalho.

RESUMO

O câncer de pele está entre os tipos de câncer mais frequentes no mundo. O melanoma é o tipo de câncer de pele que mais preocupa por ser o mais grave, devido à sua alta possibilidade de metástase e por ser fatal caso não tratada. O método de diagnóstico do melanoma mais comum é feito de forma visual, baseando-se em características como: assimetria, borda, cor, diâmetro e evolução, chamado de método ABCDE. Entretanto, esta análise visual é limitada pela capacidade visual, bem como a percepção e sensibilidade humana, além do fato de que nem todos os melanomas possuem estas características. Uma técnica de análise de imagens que tem como foco a extração de características intrínsecas da imagem, mas que não são percebidas ao olho humano, como por exemplo: brilho e cor que fornecem a ideia de rugosidade, suavidade, entre outras características, é a análise de textura. A análise de textura de imagens digitais diz respeito às técnicas que utilizam o processamento de imagens com o intuito de extrair características representativas das imagens estudadas, que são de grande importância na diferenciação de imagens. Este trabalho propõe uma aplicação deste tipo de análise em imagens de câncer de pele do tipo melanoma visando o desenvolvimento de um modelo matemático classificador, através da análise estatística, baseado nas técnicas de análise de textura, com o objetivo de auxiliar o dermatologista na identificação de melanomas.

Palavras-chave: Análise de textura. Extração de características. Diagnóstico. Melanoma. Classificação de imagens.

ABSTRACT

The skin cancer is among the most common types of cancer in the world. The melanoma is the type of cancer that most worries for being the most serious due to its high possibility of metastasis and for being fatal if left untreated. The most common diagnostic method is done visually, based on characteristics such as asymmetry, border, color, diameter and evolution, called the ABCDE method. However, visual analysis is limited by human visual ability, as well as human perception and sensitivity, in addition to the fact that not all melanomas have such characteristics. A technique of analysis that focuses on the extraction of intrinsic characteristics of the image such as: brightness and color that provide the idea of roughness, smoothness among other characteristics is texture analysis. The digital image texture analysis refers to techniques, that use image processing in order to extract representative features of the images studied, which have a great importance in the differentiation of images. This work proposes an application of this type of analysis in melanoma skin cancer images of the melanoma type, aiming the development of a mathematical classifier model, through the statistical analysis, based on texture analysis techniques, with the objective of assisting the dermatologist in the identification of melanomas.

Keywords: Texture analysis. Characteristics extraction. Diagnosis. Melanoma. Image classification.

LISTA DE FIGURAS

Figura 1 -	Diagrama de blocos detalhando o método proposto do trabalho.....	14
Figura 2 -	Processo de formação do câncer.....	16
Figura 3 -	Imagem de um melanoma	17
Figura 4 -	Camadas da pele.....	18
Figura 5 -	Processo de aquisição de imagem digital.....	20
Figura 6 -	Amostragem e quantização de uma imagem analógica.....	22
Figura 7 -	Eixos (x, y) de uma imagem digital.....	22
Figura 8 -	Efeitos da redução da resolução espacial para seis resoluções diferentes.....	24
Figura 9 -	(a) Imagem de 452×374 com 256 níveis de cinza (intensidade). (b)-(d) Imagem exibida com $L = 128, 64$ e 32 níveis de cinza enquanto o tamanho da imagem é mantido constante. (e)-(h) Imagem exibida com $L = 16, 8, 4$ e 2 níveis de cinza.....	24
Figura 10 -	Exemplo de imagens de manchas e suas texturas. (a) Mancha do tipo melanoma, com borda, contorno e textura irregular, (b) mancha benigna que não apresenta ressaltos e pouca irregularidade na textura.....	25
Figura 11 -	Exemplo do histograma para imagens de (a) mancha do tipo melanoma e (b) mancha benigna.....	32
Figura 12 -	Distribuição normal.....	28
Figura 13 -	Comparação entre a distribuição normal e t de <i>Student</i>	34
Figura 14 -	Exemplo da seleção de 16 regiões de interesses diferentes selecionadas sobre a imagem de um melanoma do banco de dados.....	49
Figura 15 -	<i>BoxPlot</i> do parâmetro de textura Momento da Diferença Inversa para o grupo de imagens de manchas benignas versus melanomas.....	52

Figura 16 - *BoxPlot* do parâmetro de textura Entropia para o grupo de
imagens de manchas benignas versus melanomas..... 52

LISTA DE TABELAS

Tabela 1 -	Número de bytes de armazenamento de uma imagem digital quadrada de lado N. L representa o número de níveis de cinza.....	23
Tabela 2 -	Coeficientes de correlação entre as variáveis X, Y e Z.....	38
Tabela 3 -	Matriz de correlação simplificada entre as variáveis X, Y e Z.....	38
Tabela 4 -	Parâmetros de texturas obtidos através do <i>MaZda</i> , e seus respectivos valores de média, desvio padrão (DP) e valor-p, obtidos através dos testes t e Mann-Whitney.....	51
Tabela 5 -	Coeficiente de correlação (r) para o parâmetro de textura Correlação.....	53
Tabela 6 -	Coeficiente de correlação (r) para o parâmetro de textura Momento da Diferença Inversa.....	54
Tabela 7 -	Coeficiente de correlação (r) para o parâmetro de textura Variância.....	55
Tabela 8 -	Coeficiente de correlação (r) para o parâmetro de textura Assimetria.....	55
Tabela 9 -	Estatísticas das variáveis (parâmetros) do modelo.....	59
Tabela 10 -	Testes de qualidade e validação do modelo.....	60
Tabela 11 -	Teste de qualidade do ajuste dos coeficientes ao modelo.....	60
Tabela 12 -	Predição para o parâmetro Assimetria do grupo de imagens de manchas benignas.....	62
Tabela 13 -	Predição para o parâmetro Variância do grupo de imagens de manchas benignas.....	63
Tabela 14 -	Predição para o parâmetro Momento da Diferença Inversa do grupo de imagens de manchas de melanomas.....	63
Tabela 15 -	Predição para o parâmetro Correlação do grupo de imagens de manchas de melanomas.....	64
Tabela 16 -	Resultado dos testes de classificação.....	65

SUMÁRIO

1	INTRODUÇÃO.....	12
1.1	MOTIVAÇÃO.....	13
1.2	OBJETIVOS.....	13
1.2.1	Objetivo Geral.....	13
1.2.2	Objetivos Específicos.....	14
1.3	METODOLOGIA.....	14
1.4	ORGANIZAÇÃO DO TRABALHO.....	15
2	FUNDAMENTAÇÃO TEÓRICA.....	16
2.1	O CÂNCER.....	16
2.1.1	Câncer de Pele.....	17
2.1.2	Câncer de Pele Tipo Melanoma.....	17
2.2	FUNDAMENTOS DO PROCESSAMENTO DIGITAL DE IMAGEM.....	19
2.2.1	Aquisição de Imagens Utilizando Sensores.....	19
2.2.2	Imagem Digital.....	20
2.2.3	Representação de Imagens Digitais em Tons de Cinza.....	21
2.2.4	Amostragem e Quantização de Imagens Digitais.....	21
2.2.5	Resolução Espacial e de Intensidade de Cinza.....	23
2.3	TEXTURAS E CARACTERIZAÇÃO DE IMAGENS.....	25
2.3.1	Extração de Características e Análise de Textura.....	26
2.4	ANÁLISE ESTATÍSTICA.....	30
2.4.1	Teste de Significância.....	30
2.4.2	Testes de Hipóteses.....	30
2.4.3	Teste de Correlação.....	37
2.4.4	Regressão Linear.....	39
2.4.5	Regressão Logística Binária.....	40
2.4.5.1	Estimando os Coeficientes do Modelo.....	42
2.5	SOFTWARE MAZDA.....	45
3	EXPERIMENTO REALIZADO.....	46

3.1	IDENTIFICAÇÃO DO PROBLEMA.....	46
3.2	METODOLOGIA DE SOLUÇÃO DO PROBLEMA.....	47
3.2.1	Aquisição de Imagem.....	47
3.2.2	Extração de Características e Análise de Textura.....	47
3.2.3	Análise Estatística.....	50
3.2.3.1	Testes t e Mann-Whitney.....	50
3.2.3.2	Predição Linear.....	53
3.2.3.3	Modelagem pela Regressão Logística Binária.....	56
4	CONCLUSÕES.....	61
4.1	CONSIDERAÇÕES FINAIS.....	66
5	TRABALHOS FUTUROS.....	68
	REFERÊNCIAS.....	69
	APÊNDICE A – BOX-PLOT DOS PARÂMETROS DE TEXTURA UTILIZADOS NA REGRESSÃO LOGÍSTICA.....	76
	APÊNDICE B - TABELAS DE PARÂMETROS EXTRAÍDOS DA IMAGENS- TESTES.....	84
	APÊNDICE C - CORRELAÇÃO ENTRE AS VARIÁVEIS USADAS NA MODELAGEM DE REGRESSAO LOGÍSTICA.....	86
	ANEXO A - EQUAÇÕES DAS ESTATÍSTICAS DE TEXTURA.....	87
	ANEXO B - AMOSTRAS DE IMAGENS PARA CRIAÇÃO DO MODELO E IMAGENS USADAS NOS TESTES DE VALIDAÇÃO.....	100

1 INTRODUÇÃO

O presente Trabalho de Conclusão de Curso apresenta técnicas de análise e diferenciação de imagens médicas através de ferramentas de processamento de imagem, com o objetivo de auxiliar o diagnóstico clínico do câncer de pele do tipo melanoma.

O melanoma é um tipo de câncer de pele que se desenvolve rápido e se dissemina facilmente por outros órgãos, tornando-se fatal se não for identificado precocemente (INCA, 2018a).

Com quase 300 mil casos previstos, o melanoma foi considerado o 19º tumor mais incidente no mundo em 2018, segundo o estudo da Agência Internacional de Pesquisa do Câncer (IARC, na sigla em inglês) e da Organização Mundial da Saúde (OMS) (NAKANO, 2018).

Nos últimos anos com o avanço da tecnologia, a medicina diagnóstica evoluiu a ponto de permitir que médicos detectem a probabilidade de alguém desenvolver uma doença. Em outros casos, as doenças são identificadas tão cedo que as chances de sucesso se multiplicam no tratamento (ALBAN, 2019).

Muitas pesquisas estão sendo feitas para aproveitar o poder da computação no diagnóstico de doenças. Segundo Friedman (2008), os sistemas de análise de imagens por computador facilitam a detecção precoce dos melanomas e podem diminuir o número de biópsias realizadas, em lesões benignas, para descartar o melanoma.

Diante deste panorama, sistemas de processamento e análise de imagens digitais podem facilitar a detecção e diagnóstico precoce do melanoma, sendo este, um método não invasivo.

Este trabalho aborda a aplicação de técnicas estatísticas para investigação dos dados obtidos através de fundamentos do processamento da imagem digital, para extração e análise de características da textura de imagens de melanomas e imagens de manchas benignas objetivando criar um método para diferenciá-las.

1.1 MOTIVAÇÃO

Sempre que se trata de diagnosticar o câncer de pele, a lacuna entre a suspeita clínica e a verificação histomorfológica pode ser preenchida pela aplicação da tecnologia de imagem. Embora os sinais clínicos sugestivos de câncer estejam razoavelmente bem estabelecidos, como no caso do melanoma, que se dá pela regra ABCDE, a tecnologia de imagem e os procedimentos de avaliação de caráter morfológico microespacial têm constantemente sofrido refinamento e amplificação contínuos, refletindo também indicadores biológicos da doença (VOIGT, 2002)

Segundo Voigt (2002) a análise da imagem digital amplifica os métodos de visualização de imagens usando recursos de processamento de imagem e interfaces gráficas adequadamente rápidas, em combinação com os *softwares* deste tipo de processamento. Além do mais, a análise digital de imagens oferece o potencial de detectar, além de aplicar sub-estruturas nunca vistas antes, ainda que presentes e potencialmente indicativas de doença. Este tipo de análise, que é capaz de detectar alterações de estruturas sutis, principalmente invisíveis ao olho humano, complementa a investigação diagnóstica e auxilia o clínico no diagnóstico precoce do câncer.

Dado o grande potencial do processamento de imagens digitais na análise micro-estruturas bem como a grande limitação visual humana de características imperceptíveis, mas significantes do órgão estudado, além da grande taxa de cura no tratamento do melanoma em seu estado inicial, o método de classificação de imagens a partir de análise de texturas proposto neste trabalho é de grande valia no auxílio do diagnóstico precoce do melanoma.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Esse estudo tem como objetivo apresentar uma maneira alternativa para a análise e classificação do câncer de pele do tipo melanoma, baseando-se em técnicas de processamento digital de imagens com a extração de características de texturas bem como em métodos estatísticos, com o intuito de obter a identificação da doença

de maneira cada vez mais rápida e precoce, aumentando assim as chances de cura do paciente.

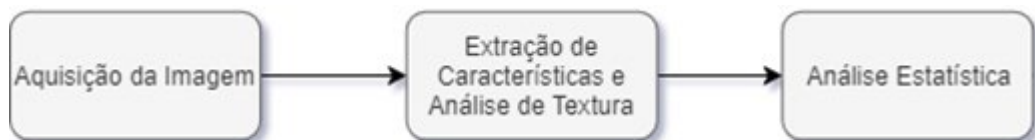
1.2.2 Objetivos Específicos

- Realizar o processamento digital de imagens, para extração de características de textura, para diagnóstico do melanoma.
- Analisar as texturas de manchas de pele benignas e de melanomas com o objetivo de encontrar diferenças texturais entre elas.
- Utilizar métodos estatísticos para aplicação de um modelo matemático classificador de imagens das manchas de pele, de acordo com suas malignidade: melanoma ou benigna.

1.3 METODOLOGIA

Na Figura 1 é apresentado o diagrama de blocos do método proposto para este estudo.

Figura 1 – Diagrama de blocos detalhando o método proposto do trabalho.



Fonte: A autora, 2019.

O método proposto pode ser dividido da maneira apresentada a seguir:

- No bloco de Aquisição de Imagem é mostrado como foram obtidas as imagens, baseando-se no tipo de lesão da pele estudadas, que estão disponíveis nos bancos de dados. (DERMNETNZ), (ATLAS OF CLINICAL DERMATOLOGY), e (ATLAS DERMATOLÓGICO).

- No bloco de Extração de Características e Análise de Textura é mostrado como foram obtidos os parâmetros de textura das imagens, tanto do câncer de pele

tipo melanoma quanto da mancha benigna, através do *software* de extração de características de textura, *MaZdA*. Para cada grupo de imagens (benignas ou melanomas), foram extraídos os parâmetros mais representativos para a textura de uma imagem.

- No bloco de Análise Estatística são mostrados os processos e testes estatísticos que foram usados para a diferenciação dos dois tipos de manchas estudadas.

1.4 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em 6 capítulos divididos da seguinte forma:

No capítulo 1, encontram-se as considerações iniciais do trabalho. Neste capítulo está contido a introdução, a motivação, objetivo e a estrutura do trabalho.

O capítulo 2, trata da fundamentação teórica onde são abordados os temas relevantes aplicados no desenvolvimento deste trabalho.

O capítulo 3, trata da análise do problema proposto bem como o desenvolvimento da metodologia para solução utilizada, onde são abordadas técnicas de extração e análise de textura utilizadas no presente estudo, bem como técnicas estatísticas usadas na classificação das imagens.

O capítulo 4, trata das conclusões encontradas analisando os resultados obtidos bem como das últimas considerações do projeto.

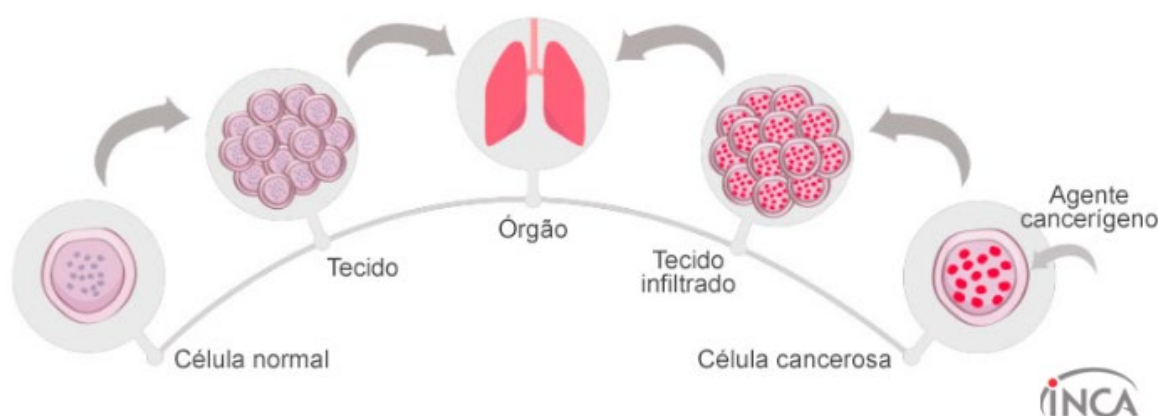
E por fim, o capítulo 5 complementa ideias sobre trabalhos futuros a serem estudados analisando as perspectivas de melhoria e extensão do método proposto.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 O CÂNCER

Segundo o INCA (2018), o câncer é o nome genérico para um grupo de mais de 100 doenças. Embora existam muitos tipos de câncer, todos começam devido ao crescimento e multiplicação anormal e descontrolado das células. A Figura 2 mostra o processo de formação do tecido normal e do câncer de pele.

Figura 2 – Processo de formação do câncer.



Fonte: INCA (2017).

Os diferentes tipos de câncer correspondem aos vários tipos de células do corpo e, quando começam em tecidos epiteliais, como pele ou mucosas, são denominados carcinomas. Se o ponto de partida do câncer são os tecidos conjuntivos, como osso, músculo ou cartilagem, são chamados sarcomas (INCA, 2018d).

Segundo a OPAS (ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2018) o câncer é a segunda principal causa de morte no mundo e é responsável por 9,6 milhões de mortes em 2018. A nível global, uma em cada seis mortes são relacionadas à doença.

A mortalidade por câncer pode ser reduzida se os casos forem detectados e tratados precocemente e nessa fase o câncer pode responder melhor a um tratamento eficaz. Isso resulta em uma maior probabilidade de sobrevivência, menor morbidade e um tratamento menos dispendioso. Melhorias significativas podem ser feitas na vida dos pacientes com câncer por meio da detecção precoce e cuidados em tempo oportuno (ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2018).

2.1.1 Câncer de Pele

O câncer de pele é o de maior incidência no Brasil e o mais comum entre os seres humanos. Estima-se que 1 entre cada 4 casos de câncer diagnosticados se origine na pele ou nas mucosas. Curável, em sua grande maioria, quando detectado precocemente (INCA, 2018c).

Existem diversos tipos de câncer de pele, divididos em dois grupos principais: os melanomas e os carcinomas (também conhecidos como “câncer de pele não-melanoma”), em que características dos carcinomas são bem distintas das do melanoma, desde o aspecto da lesão até o prognóstico.

Os carcinomas são o tipo mais comum entre os cânceres de pele, representam cerca de 95% dos tumores malignos de pele, sendo o carcinoma basocelular, o mais comum e menos agressivo, e o carcinoma espinocelular, o mais grave.

2.1.2 Câncer de Pele Tipo Melanoma

O melanoma é menos frequente que outros tumores de pele, porém costuma ter comportamento mais agressivo. Tem origem a partir dos melanócitos, que são as células responsáveis pela produção do pigmento (melanina) que dá cor à pele. Por isso, costumam se manifestar como pintas de cor escura (negro ou castanho) (OSWALDO CRUZ, c2017). A Figura 3 mostra um exemplo de uma imagem de mancha de pele com o diagnóstico de melanoma.

Figura 3 – Imagem de um melanoma.

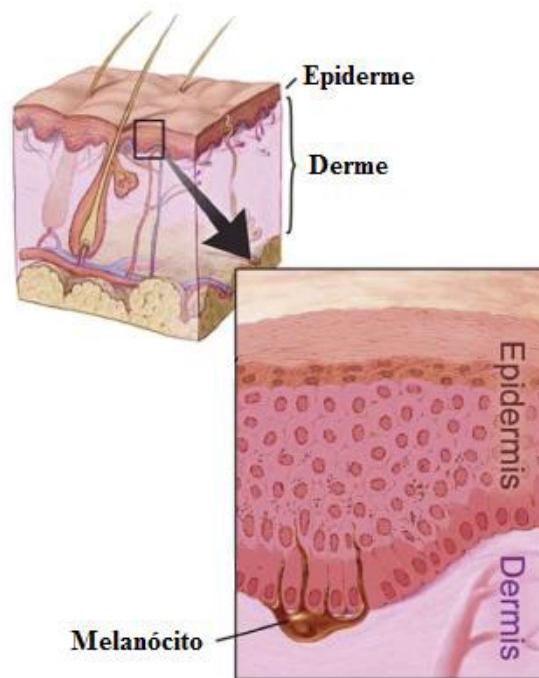


Fonte: Santos, (c2019).

O grande fator de risco do melanoma deve-se ao fato de que ele costuma começar nas camadas mais profundas da pele. Quando este tipo de câncer surge, pode se expandir a outras áreas ou órgãos do corpo (ONCOGUIA, 2017).

A Figura 4 mostra as camadas da pele e onde, entre elas, se encontram os melanócitos.

Figura 4 - Camadas da pele.



Fonte: Adaptado de Gordon (2013).

O prognóstico desse tipo de câncer pode ser considerado bom se detectado em sua fase inicial. Nos últimos anos, houve grande melhora na sobrevivência dos pacientes com melanoma, principalmente devido à detecção precoce do tumor e à introdução dos novos medicamentos imunoterápicos (INCA, 2018a).

O sinal de alerta mais importante para o melanoma, é o aparecimento de uma nova mancha na pele ou uma mancha pré-existente, que está mudando de tamanho, forma ou cor. Outro sinal importante é uma mancha que parece diferente de todas as outras manchas da pele (AMERICAN CANCER SOCIETY, 2016).

A regra **ABCDE** é outro guia, além dos exames clínicos, laboratoriais ou radiológicos, para os sinais habituais de melanoma. O paciente deve ficar atento e informar o médico sobre manchas com qualquer um dos seguintes características:

- Assimetria: uma metade do sinal é diferente da outra;
- Bordas irregulares: contorno mal definido;

- Cor variável: presença de várias cores em uma mesma lesão (preta, castanha, branca, avermelhada ou azul);
- Diâmetro: maior que 6 milímetros;
- Evolução: mudanças observadas em suas características (cor, tamanho, forma, textura). Quando o melanoma se desenvolve em uma mancha pré-existente, a textura da mancha pode mudar e tornar-se dura ou rugosa.

2.2 FUNDAMENTOS DO PROCESSAMENTO DIGITAL DE IMAGEM

Nesta seção serão apresentados alguns fundamentos e técnicas de processamento digital de imagens tais como: processo e obtenção de imagens, conceito de uma imagem digital, amostragem e quantização e por fim algumas relações básicas entre pixels.

2.2.1 Aquisição de Imagens Utilizando Sensores

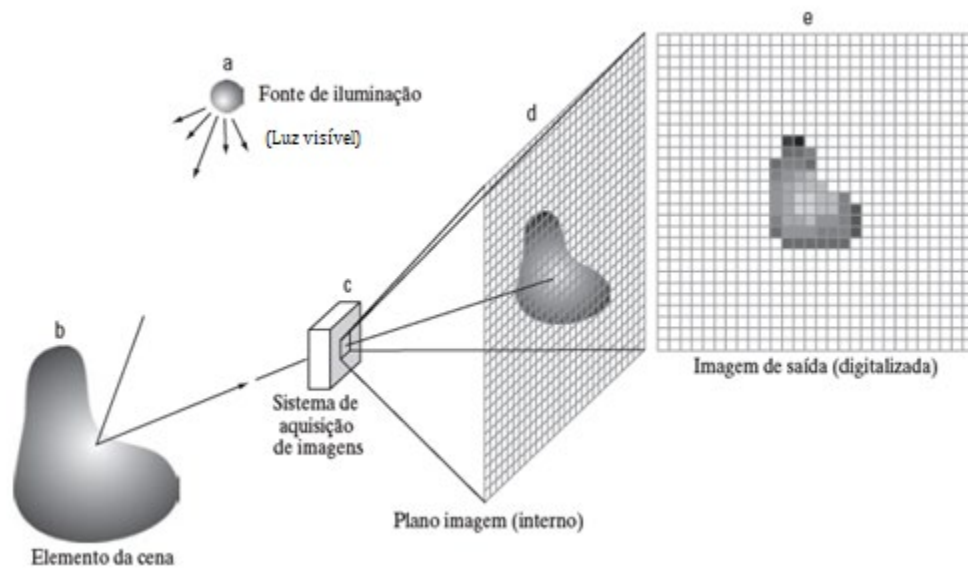
Segundo Gonzalez e Woods (2009), a associação de uma fonte de luz com elementos de uma cena produz o que chama-se imagem. Tais elementos são capazes de refletir ou absorver energia dessa fonte de luz.

Uma fonte de iluminação ou fonte de energia pode ter várias origens e a mais conhecida é a luz visível, mas também pode ser originada de outras fontes de energia como por exemplo a fonte eletromagnética e o infravermelho.

Supondo que a fonte de iluminação seja a luz visível, a Figura 5(a) exemplifica como esta luz é refletida a partir de um elemento de cena, Figura 5(b), sendo este elemento, qualquer objeto. O sistema de aquisição de imagens, Figura 5(c), realiza, em primeiro lugar, a coleta da luz absorvida da fonte, e projeta-lá em um plano imagem. Como o tipo de iluminação é a luz visível, a entrada frontal do sistema de aquisição de imagens é uma lente ótica que projeta a cena vista sobre o plano focal da lente, como mostra a Figura 5(d). O arranjo de sensores, que coincide com o plano focal, produz saídas proporcionais à integral da luz recebida em cada sensor (GONZALEZ e WOODS, 2009).

A saída destes sensores é uma imagem digital, como mostra o esquema da Figura 5(e).

Figura 5 - Processo de aquisição de imagem digital.



Fonte: Adaptado de Gonzalez e Woods (2009).

2.2.2 Imagem Digital

Uma imagem digital pode ser definida como uma função de duas dimensões, $f(x, y)$, onde x e y são coordenadas espaciais (planas), e a amplitude da função $f(x, y)$, em qualquer par de coordenadas, é chamado de intensidade ou nível de cinza (ou intensidade de brilho) da imagem naquele ponto (GONZALEZ e WOODS, 2009).

Na medicina, a imagem digital é muito utilizada para fins de diagnóstico e é de grande importância no tratamento correto de doenças sendo feito, algumas vezes, de maneira pouco ou não-invasiva e, por isso, existe um grande interesse neste método de diagnóstico.

A ressonância magnética (RM) e a Tomografia Computadorizada (TC), que são exames que têm como resultado uma imagem em alta resolução da estrutura de um órgão ou do corpo inteiro (GILTON, 2011), são exemplos do uso da imagem digital no diagnóstico de doenças.

2.2.3 Representação de Imagens Digitais em Tons de Cinza

De acordo com Gonzalez e Woods (2009), a luz sem cor é chamada de luz monocromática (ou acromática). O único atributo da luz monocromática é sua intensidade. Pelo fato da intensidade da luz monocromática ser percebida como variações de tons de cinza, do preto até chegar ao branco. O termo “nível de cinza” costuma ser utilizado para denotar a intensidade monocromática.

Os níveis de cinza de uma imagem monocromática em quaisquer coordenadas (x_0, y_0) são expressos por:

$$l = f(x_0, y_0) \quad (1)$$

Se l está na faixa $L_{\min} < l < L_{\max}$ tem-se que o intervalo $[L_{\min}, L_{\max}]$ é chamado de escala ou intensidade de cinza.

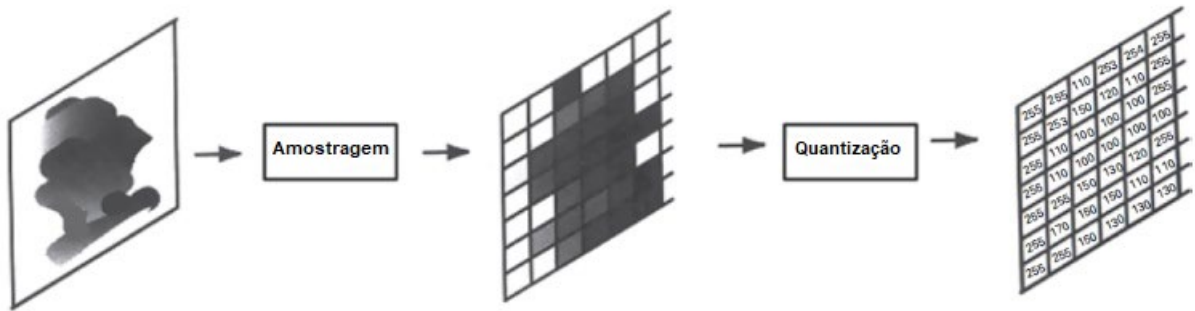
O l está localizado no intervalo $[0; L - 1]$, com $l = 0$ representando o preto e $l = L - 1$ o branco, na escala de cinza tendo assim, todos os valores intermediários como tons de cinza variando do preto ao branco.

2.2.4 Amostragem e Quantização de Imagens Digitais

Segundo Gonzalez e Woods (2009) para criar uma imagem digital, a imagem analógica precisa passar pelos processos de amostragem e quantização como ilustra a Figura 5.

A imagem analógica é representada por uma função contínua $f(x, y)$ e, para convertê-la para o formato digital, é preciso realizar o processo de amostragem da função em ambas as coordenadas gerando a matriz $M \times N$. Já a quantização é o processo no qual existe uma digitalização da amplitude da função, para cada ponto (x, y) , da coordenada digitalizada.

Figura 6 – Amostragem e quantização de uma imagem analógica.



Fonte: Adaptado de Adessowiki (2011).

Pode-se escrever a representação do resultado da digitalização (quantização), como uma matriz numérica $M \times N$, obtida através da amostragem, como a seguir:

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0, N-1) \\ f(1,0) & f(1,1) & \dots & f(1, N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1, N-1) \end{bmatrix} \quad (2)$$

Cada elemento dessa matriz é chamado de elemento de imagem, elemento pictórico, pixel ou pel.

Figura 7 – Eixos (x, y) de uma imagem digital.



Fonte: Glenda (2007).

O número L de níveis discretos de intensidade de cinza, devido à representação em *bits* dos dados computados, é representado como uma potência de 2, e é definido como:

$$L = 2^k. \quad (3)$$

Assumindo valores de k inteiros e que estejam no intervalo $[0; L - 1]$. O número de *bits* necessários para armazenar uma imagem digitalizada é

$$b = M \times N \times k, \quad (4)$$

em que b é chamado de profundidade da imagem (PEDRINI, 2008), mas, a unidade de armazenamento padrão da imagem é o byte que corresponde a 8 bits.

A Tabela 1 mostra o número de bytes que são utilizados para a representação de uma imagem monocromática para diferentes dimensões $N \times N$.

Tabela 1 - Número de bytes de armazenamento de uma imagem digital quadrada de lado N . L representa o número de níveis de cinza.

N	L=2	L=4	L=8	L=16	L=32	L=64	L=256
32	128	256	384	512	640	768	1024
64	512	1024	1536	2048	2560	3072	4096
128	2048	4096	6144	8192	10240	12288	16384
256	8204	16384	24576	32018	40960	49152	65536

Fonte: Adaptado de Gonzalez e Woods (2002).

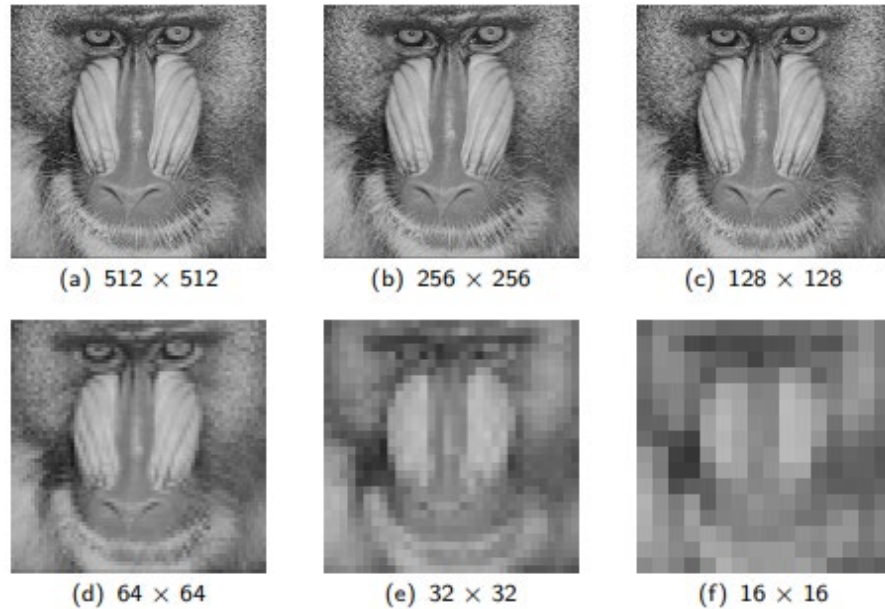
Para $L = 256$, cada pixel pode ter associado a si, um valor de cinza entre 0 e 255. A profundidade da imagem, neste caso, é de 8 bits por pixel. (PEDRINI, 2008)

2.2.5 Resolução Espacial e de Intensidade de Cinza

Segundo Gonzalez e Woods (2009), a resolução espacial pode ser definida como a medida do menor detalhe distinguível em uma imagem. A Figura 8 mostra como a resolução espacial interfere na visualização da imagem. Analisando a figura, percebe-se que quanto menor a resolução espacial, menos detalhes visuais serão percebidos ou, de acordo com Pedrini (2008), quanto menor o intervalo de

amostragem entre os pixels da imagem ou, quanto maior a densidade de pixels em uma imagem, maior será a resolução desta imagem.

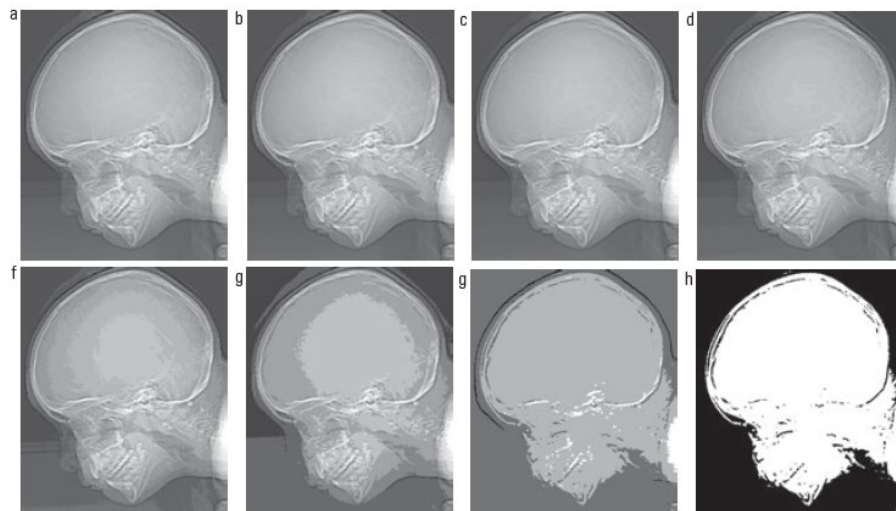
Figura 8 - Efeitos da redução da resolução espacial para seis resoluções diferentes.



Fonte: Pedrini (2008).

Na resolução de intensidade de cinza, observa-se, na Figura 9, que mesmo mantendo o tamanho da imagem em 452x374 pixels, a diminuição da quantidade de níveis de cinza causa uma perda de informações visuais.

Figura 9 - (a) Imagem de 452 x374 com 256 níveis de cinza (intensidade). (b)-(d) Imagem exibida com L= 128, 64 e 32 níveis de cinza enquanto o tamanho da imagem é mantido constante. (e)-(h) Imagem exibida com L= 16, 8, 4 e 2 níveis de cinza.



Fonte: Gonzalez e Woods (2002).

2.3 TEXTURAS E CARACTERIZAÇÃO DE IMAGENS

A análise de textura vem sendo muito estudada em vários ramos, como na computação gráfica e principalmente na área da medicina com o uso de imagens capturadas por raios X, tomografia computadorizada, ressonância magnética e ultrassonografia. De acordo com Kavitha, Suruliandi e Nagarajan (2007), as características da textura são consideradas fatores importantes na detecção do câncer de pele do tipo melanoma.

Segundo Rosenholtz (2014), as variações de superfície que levam à textura tátil, como em um exame clínico, também tendem a levar a variações na intensidade da luz que atinge nossos olhos, produzindo o que é conhecido como “textura visual” (ou simplesmente “textura”) como é mostrado na Figura 10, com o exemplo de uma imagem de melanoma e uma imagem de uma mancha benigna.

Figura 10 - Exemplo de imagens de manchas e suas texturas. (a) Mancha do tipo melanoma, com borda, contorno e textura irregular, (b) mancha benigna que não apresenta ressaltos e pouca irregularidade na textura.



(a)



(b)

Fonte: (DERMNETNZ)

Compreender o processamento de percepção da textura requer a capacidade de sintetizar texturas com as propriedades desejadas.

Segundo Ito, Kim e Salcedo (2009) as texturas são usadas naturalmente pela percepção humana no reconhecimento de objetos e de padrões e, portanto, são características importantes na análise de imagens. Além disso as texturas também podem ser representadas em função das relações tonais e estruturais entre as

primitivas¹. O tom é baseado principalmente nas propriedades de intensidade de pixel (níveis de cinza) das primitivas, enquanto a estrutura é a relação espacial (local) entre as primitivas. Cada pixel pode ser caracterizado por suas propriedades tonais e de localização (SRINIVASAN e SHOBHA,2008).

A classificação de imagens por textura consiste em categorizar uma amostra de imagem de classe desconhecida a uma das classes de texturas conhecidas. Um problema na análise de textura no mundo real é, frequentemente, a sua não-uniformidade devido a mudanças de orientação, escala e outros aspectos visuais (ITO, KIM e SALCEDO, 2009).

Segundo Haralick, Shanmugam e Dinsten (1973) as texturas definem características como: uniformidade, densidade, aspereza, regularidade, intensidade, entre outras características em uma imagem e para Sklansky (1978), uma região da imagem apresenta uma textura uniforme se o conjunto de características estatísticas das propriedades locais apresentar pouca variação de tons de cinza.

2.3.1 Extração de Características e Análise de Textura

A extração de características é geralmente o primeiro estágio da análise de textura da imagem. Os resultados obtidos, a partir desta etapa, são utilizados para discriminação, classificação ou segmentação de texturas (SZCZYPIŃSKI *et al.*, 2009).

De acordo com Zayed (2015), a extração de características é o processo de obtenção de informações de alto nível de uma imagem, como cor, forma e textura. Os métodos estatísticos de textura analisam a distribuição espacial dos níveis de cinza, calculando os recursos locais em cada ponto da imagem e obtendo como resultado, um conjunto de estatísticas das distribuições das características locais.

Três abordagens principais usadas em classificação de imagens para a descrição de texturas são a estatística, a estrutural e a espectral (ROCHA, 2002):

- Abordagens Estruturais: utilizam a ideia de que texturas são compostas de primitivas dispostas de forma aproximadamente regular e repetitiva, de acordo com

¹ Uma primitiva de textura é um conjunto contíguo de pixels com algum tom e/ou propriedade local que pode ser descrito por sua intensidade média, intensidade máxima e mínima, tamanho, forma e etc. (SRINIVASAN e SHOBHA,2008).

regras bem definidas. Como exemplo, pode-se citar a descrição da textura baseada em linhas paralelas regularmente espaçadas;

- Abordagens Espectrais: baseiam-se em propriedades do espectro de Fourier, sendo principalmente utilizadas na detecção de periodicidade.

- Abordagens Estatísticas: Segundo Haralick, Shanmugam e Dinsten (1973), este tipo de estatística utiliza propriedades das distribuições e relações entre os níveis de cinza de uma imagem. Medidas estatísticas comuns incluem entropia, correlação, contraste, variância, média e etc.

Para realização deste trabalho foram utilizadas as abordagens estatísticas.

Dentre as abordagens estatísticas estão as estatísticas de primeira ordem e de segunda ordem e são definidas como:

1. Estatísticas de primeira ordem:

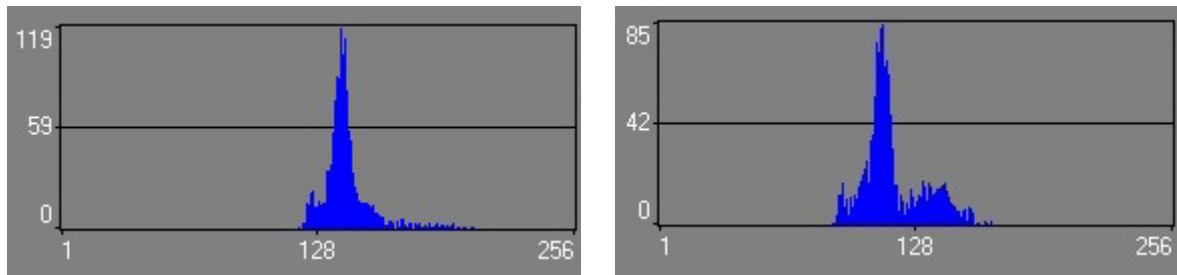
Segundo Medeiros (2013), a análise estatística de primeira ordem diz respeito à distribuição de níveis de cinza em uma imagem, onde se usa o histograma de primeira ordem como base para extração de suas características.

Medidas de textura de primeira ordem são estatísticas que não consideram relacionamentos de vizinhança de pixels da imagem digital (SRINIVASAN e SHOBHA, 2008). Tais características são: média, desvio-padrão, coeficiente de assimetria e curtose.

Segundo Srinivasan e Shobha (2008), a análise de textura baseada somente no histograma de níveis de cinza sofre com a limitação de não fornecer informações sobre a posição relativa dos pixels entre si. Por exemplo, duas imagens completamente diferentes, cada uma com 50% de preto e 50% de pixels brancos podem produzir o mesmo histograma de nível de cinza. Portanto, não se pode distinguir entre as duas imagens usando apenas a análise estatística de primeira ordem.

De acordo com a Figura 11 pode-se observar como se comporta o histograma de uma imagem de uma mancha de um melanoma (a) e uma imagem de uma mancha benigna (b). Nesta figura, percebe-se o quão semelhantes estes histogramas são, mostrando assim como esse tipo de análise isolada, não representa uma boa diferenciação entre imagens.

Figura 11 - Exemplo do histograma para imagens de (a) mancha do tipo melanoma e (b) mancha benigna.



Fonte: A autora, 2019.

A seguir será mostrado os parâmetros de textura baseados na estatística de primeira ordem. Suas análises detalhadas serão mostradas no Anexo A.

i. Histograma:

- a) Média;
- b) Variância;
- c) Assimetria;
- d) Curtose.

1. Estatísticas de segunda ordem:

A abordagem estatística de segunda ordem considera o posicionamento espacial relativo da ocorrência dos níveis de cinza, característica não atendida quando se faz uso do histograma de primeira ordem (MEDEIROS, 2013).

A seguir serão apresentados os parâmetros de textura que são baseados nas estatísticas de segunda ordem. A análise detalhada de cada parâmetro, será mostrada no Anexo A.

i. Gradiente Absoluto:

- a) Média do Gradiente;
- b) Variância do Gradiente;
- c) Assimetria do Gradiente;
- d) Curtose do Gradiente.

ii. Matriz de Co-ocorrência dos Níveis de Cinza:

- a) Segundo Momento Angular;
- b) Contraste;
- c) Correlação;
- d) Soma dos Quadrados;
- e) Momento da Diferença Inversa;
- f) Entropia;
- g) Média da Soma;
- h) Variância da Soma;
- i) Entropia da Soma;
- j) Variância da Diferença;
- k) Entropia da Diferença.

iii. Matriz de Run-Length:

- a) Momentos Inversos de Ênfase de Curto Prazo (MIECP);
- b) Momentos de Ênfase de Longo Prazo (MIELP);
- c) Não Uniformidade do Nível de Cinza (NUNC);
- d) Não uniformidade do Comprimento de Execução (*run length*) (NUCE);
- e) Fração da Imagem em Execuções (FIE).

iv. Modelo Auto-regressivo:

- a) Teta 1;
- b) Teta 2;
- c) Teta 3;
- d) Teta 4;
- e) Sigma.

2.4 ANÁLISE ESTATÍSTICA

O objetivo principal da análise estatística é determinar se os resultados obtidos são estatisticamente significantes de acordo com limites preestabelecidos. Em outras palavras, permite concluir se os resultados de uma dada pesquisa são reais ou fruto da casualidade (WITTE & WITTE, 2005).

2.4.1 Teste de Significância

Segundo Oliveira (2010), ao estudar uma característica de uma certa população podem ser observadas diferenças decorrentes de flutuações normais da amostra. Neste caso, os valores encontrados não são estatisticamente significantes e as diferenças na amostra são fruto do acaso. Por outro lado, as diferenças encontradas podem não ser decorrentes de flutuações inerentes à amostra, neste caso, essas diferenças representam um efeito real e não podem ser atribuídas ao acaso. Dessa forma, podemos dizer que os resultados obtidos são “estatisticamente significantes”. A definição sobre o que é ou não “estatisticamente significativo” é baseada em limites pré-estabelecidos, chamados de níveis de significância. O nível de significância é o limite no qual podemos afirmar que o desvio presente numa amostra é fruto do acaso ou não. Este limite, chamado valor p , é geralmente considerado estatisticamente significativo se $p \leq 0,05$ ou $p \leq 0,01$. Neste trabalho usou-se o nível de significância $p \leq 0,05$.

2.4.2 Testes de Hipóteses

Um teste de hipótese é um tipo de teste que especifica se deve-se aceitar ou rejeitar uma alegação sobre uma população, de acordo com as provas fornecidas por uma amostra de dados. Este teste examina duas hipóteses opostas sobre uma população: a hipótese nula e a hipótese alternativa. A hipótese nula, denotada por H_0 , é a declaração que está sendo testada. Normalmente, a hipótese nula é uma declaração de "nenhum efeito" ou "nenhuma diferença". A hipótese alternativa, denotada por H_1 , é a declaração que você quer ser capaz de concluir que é verdadeira com base em evidências fornecidas pelos dados da amostra e com base nos dados amostrais. O objetivo deste teste é determinar se devemos rejeitar a hipótese nula, e

para, isso usa-se um valor-p para fazer esta determinação. Se o valor de p for menor que o nível de significância, então você pode rejeitar a hipótese nula.

A formulação de hipóteses estatística é muito empregada em pesquisas nas mais diversas áreas do conhecimento. Contudo, para decidir se uma determinada hipótese é confirmada por um conjunto de dados, é necessário ter um procedimento objetivo segundo o qual não rejeita-se ou rejeita-se a hipótese (SIEGEL e CASTELLAN, 2006).

Os testes de hipóteses dividem-se em paramétricos e não paramétricos. Os paramétricos são aqueles que utilizam os parâmetros da distribuição amostral, ou uma estimativa destes, para o cálculo de sua estatística. Normalmente, estes testes são mais rigorosos e possuem mais pressuposições para sua validação como o fato de exigirem que as amostras possuam distribuição normal (gaussiana), se tiverem dimensões menores que 30. Já os não paramétricos utilizam, não exigem pressuposições de normalidade como os paramétricos (REIS e RIBEIRO, 2007).

Neste trabalho serão usados um teste paramétrico e um não paramétrico e eles serão detalhados a seguir.

1. Teste paramétrico: Os testes paramétricos assumem que sua distribuição, com relação a frequências dos erros amostrais, seja normalmente distribuída (Olivera, 2016). Se os dados não satisfazem as suposições assumidas pelas técnicas tradicionais, métodos não paramétricos devem ser usados.

- O modelo Normal

Segundo Magalhães e Lima (2005), diz-se que uma variável aleatória contínua x tem distribuição normal se sua função de densidade é dada por:

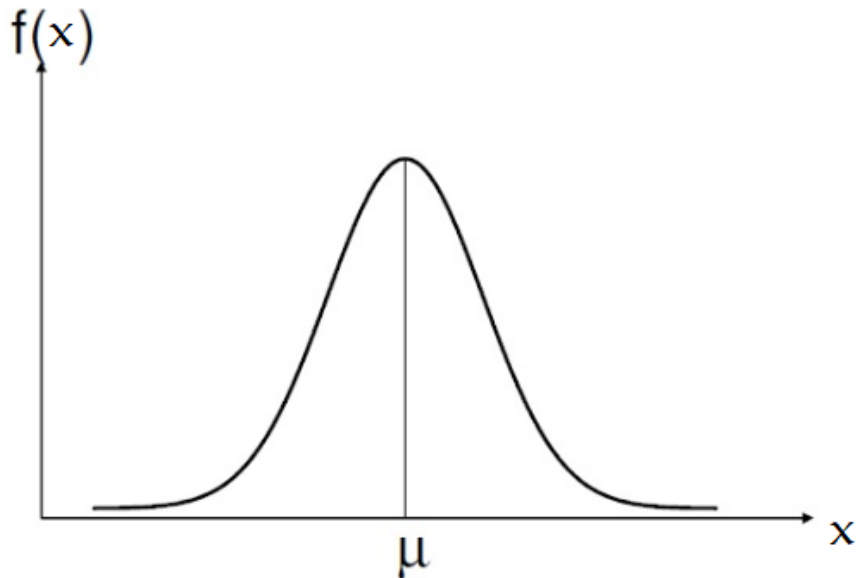
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad (5)$$

onde $x \in \mathbb{R}$, μ e σ , são a média e o desvio-padrão da distribuição gaussiana.

A distribuição normal possui algumas propriedades que podem ser observadas a partir do seu gráfico mostrado na Figura 12.

- i. $f(x)$ é simétrica em relação a μ ;
- ii. $f(x)$ tende à 0 quando x tende à $\pm\infty$;
- iii. o valor máximo de $f(x)$ se dá para $x = \mu$.

Figura 12 – Distribuição normal.



Fonte: Magalhães e Lima (2005).

- O teorema Central do Limite (TCL)

Seja $x_1, \dots, x_n$, onde existem n variáveis aleatórias independentes e identicamente distribuídas com média μ e desvio-padrão σ . A distribuição da soma das n variáveis tende a apresentar um comportamento Normal, como segue:

$$X_1 + X_2 + \dots + X_n = \sum X_i \sim \text{Normal}(n\mu; \sqrt{n\sigma^2}). \quad (6)$$

Este resultado pode ser utilizado para compreender o comportamento probabilístico da média amostral $\bar{X}$ utilizando-se propriedades da Esperança e da Variância de variáveis aleatórias:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \sim Normal\left(\mu; \frac{\sigma}{\sqrt{n}}\right). \quad (7)$$

Onde a Esperança (E) pode ser definida como:

$$E(X) = \sum_{i=1}^n P_i X_i, \quad (8)$$

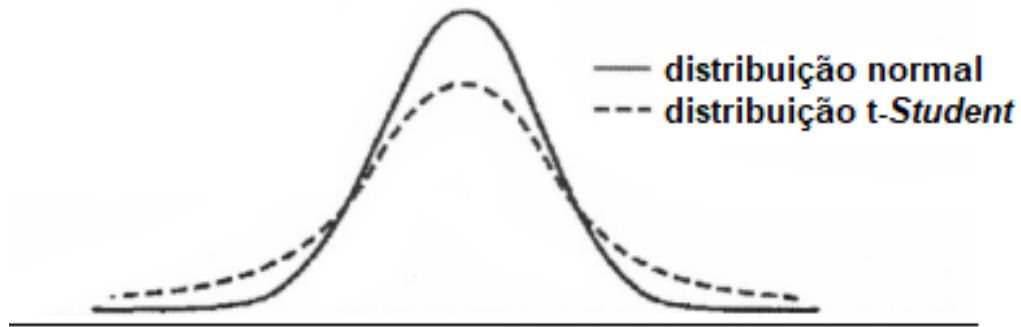
em que P_i é a probabilidade do evento.

De acordo com o TCL, para amostras grandes ($n \geq 30$), qualquer que seja a distribuição original da variável na população, a média amostral $\bar{X}$ tem distribuição aproximadamente normal. Ou seja, se cada observação tem média μ e desvio-padrão σ , então a média amostral tem distribuição aproximadamente normal com média μ e desvio-padrão $\sigma/\sqrt{n}$, sendo n o tamanho da amostra superior a 30 (VIEIRA, SILVEIRA & SIQUEIRA, 2017).

Teste t

O teste t de *Student* foi desenvolvido por Willian Sealy Gosset em 1908, que usou o pseudônimo *Student* em função da confidencialidade requerida por seu empregador (cervejaria Guinness) que considerava o uso de estatística na manutenção da qualidade como uma vantagem competitiva. O teste t de *Student* tem diversas variações de aplicação, mas sempre há a limitação do mesmo ser usado na comparação de duas (e somente duas) médias e as variações dizem respeito às hipóteses que são testadas (ALVES, 2016). O teste t deriva da distribuição normal conforme ilustrado na Figura 13.

Figura 13 – Comparação entre a distribuição normal e t de *Student*.



Fonte: Magalhães e Lima (2002).

De acordo com Magalhães e Lima (2002), se o desvio-padrão é desconhecido, ele precisa ser estimado, então supondo que a amostra aleatória seja representada pelo vetor de variáveis aleatórias $(X_1, \dots, X_n)$, todas elas com densidade normal de média μ e variância σ^2 que é a variância amostral. O desvio padrão das populações será estimado através do estimador denominado variância amostral S^2 , que é dado por:

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{(n-1)}, \quad (9)$$

resultando em:

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}. \quad (10)$$

Sendo T uma variável aleatória e, apesar de X ter uma distribuição normal, o denominador envolve a variável aleatória S^2 , que fará com que a função de densidade T seja diferente da normal.

O teste t *Student*, ou simplesmente teste t é o método mais utilizado para se avaliar as diferenças entre as médias de dois grupos. (VILELA, 2001)

2. Teste não-paramétricos: Estes tipos de testes pressupõem não possuem suposições sobre a normalidade da distribuição das amostras. Estes testes geralmente são chamados testes livres de distribuição (REIS e RIBEIRO, 2007).

Teste de Mann – Whitney

Inicialmente proposto por Wilcoxon e aprimorado por M H. B. Mann e D. R. Whitney, este teste é utilizado para verificar se duas amostras independentes originam-se de amostras com medias iguais.

Considere duas populações P_1 e P_2 das quais não tem-se informações a respeito de suas distribuições, mas as variáveis envolvidas tem uma escala de medida pelo menos ordinal.

O objetivo é testar se as distribuições são iguais em localização, isto é, tem-se interesse em saber se uma população tende a ter valores maiores do que a outra, ou se elas têm a mesma mediana.

O teste de Wilcoxon-Mann-Whitney é baseado nos postos dos valores obtidos combinando-se as duas amostras. O posto de uma observação é sua posição relativa às outras observações, quando os dados são colocados em ordem crescente (Oliveira, 2016).

Estas duas amostras podem ser representadas por:

- $X_1, X_2, \dots, X_m$ sendo amostras aleatórias da população P_1 e
- $Y_1, Y_2, \dots, Y_n$ uma amostra aleatória da população P_2 ,

de modo que os X_i 's são independentes e identicamente distribuídos e os Y_i 's são independentes e identicamente distribuídos. Além disso, suponha que os X_i 's e os Y_i 's são mutuamente independentes e tome a amostra Y aquela com o menor tamanho amostral, isto é, $n \leq m$.

Supondo que F e G sejam as funções de distribuição correspondentes as populações P_1 e P_2 , respectivamente e, neste caso, considera-se como hipótese nula, a hipótese:

$$H_0 : F(t) = G(t), \text{ para todo } t. \quad (11)$$

A hipótese alternativa consiste em considerar que Y tende a ser maior (ou menor) que X . Um modelo útil para descrever esta alternativa é um modelo de translação chamado modelo de mudança de posição. Neste modelo temos que,

$$G(t) = F(t - \Delta), \text{ para todo } t. \quad (12)$$

e, desta forma, a hipótese nula H_0 reduz-se a:

$$H_0: \Delta = 0. \quad (13)$$

Com isto, é estabelecida uma das seguintes hipóteses em um teste de Wilcoxon-Mann-Whitney:

$$\begin{cases} H_0 : \Delta = 0 \\ H_1 : \Delta \neq 0 \end{cases} \quad \begin{cases} H_0 : \Delta = 0 \\ H_1 : \Delta > 0 \end{cases} \quad \begin{cases} H_0 : \Delta = 0 \\ H_1 : \Delta < 0 \end{cases}, \quad (14)$$

Em seguida, ordena-se todos os valores (das duas amostras) em ordem crescente e coloca-se os postos associados.

Considere S_m e S_n as somas dos postos relacionados aos elementos das amostras X e Y respectivamente. A partir dos valores S_m e S_n , os valores são calculados como:

$$U_m = S_m - \frac{1}{2}m(m+1), \quad (15)$$

e

$$U_n = S_n - \frac{1}{2}n(n+1). \quad (16)$$

Como $S_m + S_n$ é igual à soma de todos os postos (das duas amostras), isto é,

$$S_m + S_n = \frac{1}{2}(m+n)(m+n+1). \quad (17)$$

Percebe-se que os valores de U_m e U_n estão relacionados da seguinte maneira,

$$U_m = mn - U_n. \quad (18)$$

Apenas um dos U_m , U_n precisa ser calculado e, através da equação acima encontra-se o valor do outro de maneira simples.

No teste de Wilcoxon-Mann-Whitney, a estatística W do teste será dada por U_n .

2.4.3 Teste de Correlação

A correlação procurar medir o grau de relacionamento entre as variáveis quantitativas distintas. Este teste informa se a relação é forte ou fraca e se é no mesmo sentido ou em sentido contrário. A medida usada é para isto é o coeficiente de correlação (r), que é um valor numérico, ou seja, uma medida, para o grau de associação entre as variáveis (SILVA, 2015).

Dadas duas variáveis X e Y definidas para uma amostra de tamanho n , o coeficiente de correlação linear $r(X, Y)$ é calculado com a seguinte fórmula:

$$r_{XY} = \frac{\sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i / n}{\sqrt{[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2] - [n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2]}} \quad (19)$$

As propriedades do coeficiente de correlação são:

- O coeficiente de correlação independe das unidades de medida das variáveis; é um número adimensional que varia entre -1 e $+1$, isto é, $-1 \leq r \leq +1$.
- O coeficiente de correlação de uma variável e ela mesma é igual a $+1$.

A interpretação do coeficiente de correlação (r) pode ser descrita como:

- $r = +1$: correlação positiva perfeita. Existe uma relação linear perfeita entre as variáveis cujos pares de valores se situam numa reta com inclinação positiva.
- r próximo de $+1$: correlação positiva forte. A maioria dos pares de valores das variáveis se situa próxima a uma reta com inclinação positiva.
- r positivo e próximo de zero: correlação positiva fraca.

- $r = 0$: correlação nula. Os pares de valores formam uma nuvem de pontos sem nenhuma tendência. As variáveis tendem a variar sem nenhuma relação uma com a outra.
- r negativo e próximo de zero: correlação negativa fraca.
- r próximo de -1 : correlação negativa forte.
- $r = -1$: correlação negativa perfeita. Existe uma relação linear perfeita entre as variáveis.

O coeficiente de correlação linear simples sempre se refere à duas variáveis. Num estudo com mais de duas variáveis pode-se calcular o coeficiente para diferentes pares de variáveis. Neste caso, os coeficientes de correlação são organizados em uma tabela, denominada matriz de correlação, de tamanho definido pelo número de variáveis. Para três variáveis, X, Y e Z, por exemplo, os possíveis coeficientes de correlação das três variáveis, tomadas duas a duas, podem ser registrados como na Tabela 2.

Tabela 2 – Coeficientes de correlação entre as variáveis X, Y e Z.

	X	Y	Z
X	r_{XX}	r_{XY}	r_{XZ}
Y	r_{YX}	r_{YY}	r_{YZ}
Z	r_{ZX}	r_{ZY}	r_{ZZ}

Fonte: Adaptado de Silva (2015).

Observando-se que o coeficiente de correlação de uma variável com ela mesma é igual a $+1$ e que a permutação das variáveis não altera o resultado do coeficiente, a Tabela 2 pode ser simplificada como se apresenta na Tabela 3.

Tabela 3 – Matriz de correlação simplificada entre as variáveis X, Y e Z.

	X	Y	Z
X	1		
Y	r_{YX}	1	
Z	r_{ZX}	r_{ZY}	1

Fonte: Adaptado de Silva (2015).

2.4.4 Regressão Linear

A regressão linear é chamada "linear" porque se considera que a relação da resposta às variáveis é uma função linear de alguns parâmetros. Os modelos de regressão que não são uma função linear dos parâmetros se chamam modelos de regressão não-linear. Sendo uma das primeiras formas de análise regressiva a ser estudada rigorosamente, e usada extensamente em aplicações práticas. Isso acontece porque modelos que dependem de forma linear dos seus parâmetros desconhecidos, são mais fáceis de ajustar que os modelos não-lineares aos seus parâmetros, e porque as propriedades estatísticas dos estimadores resultantes são fáceis de determinar (REIS, 1994).

A predição linear é uma equação para se estimar a condicional (valor esperado) de uma variável y , dados os valores de algumas outras variáveis x . A predição, advém do modelo de regressão linear.

A equação de regressão linear para predição pode ser definida como (BUSSAB E MORETTIN, 1988):

$$y = \alpha + \beta x + e, \quad (20)$$

onde

y - variável explicada ou dependente (aleatória).

x - variável explicativa ou independente

α - coeficiente de regressão linear, que representa o intercepto (parâmetro desconhecido do modelo -> a estimar)

β - coeficiente de regressão, que representa o declive (inclinação) (parâmetro desconhecido do modelo -> a estimar)

e - representa fatores residuais da medição.

Esse método será utilizado no processo de predição dos parâmetros de textura utilizados nesse trabalho.

2.4.5 Regressão Logística Binária

A regressão logística binária é um tipo de modelagem que é utilizada quando a variável resposta tem natureza dicotômica, ou seja, quando tal resposta assume apenas dois valores possíveis.

Este tipo de regressão será usado, neste trabalho, para classificação das imagens quanto a sua malignidade

Segundo Figueira (2006), a regressão logística é um modelo que relaciona a distribuição aleatória da variável dependente no experimento com a parte sistemática através de uma função chamada função de ligação *logit*, com componente aleatória binomial.

Inicialmente considera-se $\pi(X)$ uma função de caráter monotônico, com valores que variam entre zero e um, quando x varia na reta real, ou seja, $\pi(X)$ é uma função de distribuição de probabilidade.

Sendo Y , uma variável dicotômica, que assume os valores: 1 para ocorrência de sucesso da proposição de interesse e 0 para o fracasso da proposição de interesse.

De acordo com Souza (2006), considera-se uma série de eventos binários, em que $(Y_1, Y_2, \dots, Y_n)$ são variáveis aleatórias independentes com distribuição de Bernoulli, possuindo probabilidade de sucesso $\pi(x)$. Então, $\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x)$ e $0 < \pi(x) < 1$.

Seja z o conjunto de variáveis independentes denotadas por $X = (X_1, \dots, X_p)^T$, tal que, $j = 0, 1, \dots, z$, em que $x = (x_1, \dots, x_p)^T$ é um valor particular e uma v.a dependente binária Y . Denotando $B = (\beta_0, \beta_1, \dots, \beta_p)^T$ o vetor de parâmetros desconhecidos e β_j sendo o j -ésimo parâmetro associado à variável explicativa x_j .

Como $\pi(x)$ varia entre zero e um, uma representação linear simples para π sobre todos os valores de x não é adequada, caso contrario seria considerado um modelo de regressão linear, então considera-se a transformação logística de $\pi(x)$ sob a forma,

$$\ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = g(x), \quad (21)$$

onde,

$$g(x) = \text{logit}[P(x)] = (\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p). \quad (22)$$

Então, o modelo para prever $P(Y = 1|x) = \pi(x)$ é

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}, \quad (23)$$

que é dita probabilidade de sucesso do modelo logístico.

Já a probabilidade de fracasso é dada como:

$$\pi(x) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}. \quad (24)$$

A função exponencial transforma o valor esperado da variável dicotômica estudada, em uma função logística e nota-se então que $\pi(X)$ pode assumir qualquer valor real e continuar restrita ao intervalo $[0,1]$, como desejado. (BARRETO, 2011)

De acordo com Souza (2006), no problema de regressão a quantidade a ser modelada é a esperança da variável aleatória dependente, dado o valor da variável independente, ou seja, $E(Y | X = x)$.

Na regressão logística tem-se que $0 \leq E(Y | X = x) \leq 1$, enquanto a regressão linear varia conforme $-\infty \leq E(Y | X = x) \leq \infty$. Usando a definição de variáveis aleatórias discretas, tem-se que, na regressão logística:

$$E(Y | X = x) = 1P(Y = 1|X = x) + 0P(Y = 0|X = x) = \pi(x).$$

Quando Y é dicotômica, uma observação pode ser expressa como $y = E(Y_i|x_i) + \epsilon$, supondo que ϵ possa ter duas possibilidades que são:

1. $y = 1$ então $\epsilon = 1 - \pi(x)$ com probabilidade $\pi(x)$,
2. $y = 0$ então $\epsilon = -\pi(x)$ com probabilidade $1 - \pi(x)$.

Então, obtém-se uma distribuição de Bernoulli com média zero e variância $\pi(x)[1 - \pi(x)]$.

Segundo Figueira (2006), dada n observações independentes de cada Y , chamadas $(y_1, \dots, y_n)$ associadas aos valores de $x_i = (x_{i1}, \dots, x_{ip})$, para $i \in \{1, \dots, n\}$, tem-se que a função *logit* dada pela equação (58) é dada como:

$$y_i = \frac{\exp[g(x_i)]}{1 + \exp[g(x_i)]} + \epsilon_i, \quad (25)$$

onde e os erros, ϵ_i , seguem as seguintes suposições, para todo $i, l \in \{1, \dots, n\}$:

- i. $E((\epsilon_i|x_i)) = 0$.
- ii. $Var(\epsilon_i|x_i) = \pi(x_i)[1 - \pi(x_i)]$.
- iii. $Cov(\epsilon_i|\epsilon_l) = 0$, se $i \neq l$.

Desta forma, as v.a's $Y_1, \dots, Y_n$ satisfazem um modelo logístico múltiplo se uma amostra de tamanho um de cada Y_i pode ser expressa como

$$y_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})} + \epsilon_i, \quad (26)$$

para qual x_{ij} é constante conhecida, β_j é parâmetro desconhecido do modelo e os erros ϵ_i possuem as suposições dadas.

2.4.5.1 Estimando os Coeficientes do Modelo

Para estimar a equação que define o modelo de regressão logística é necessário estimar os parâmetros desconhecidos $\beta_0, \beta_1, \dots, \beta_p$ com probabilidade condicional $P(Y_i = 1|x_i) = p_i(x_i)$. Supõe-se que uma amostra independente de n pares de observações em que y_i apresenta o valor da variável resposta dicotômica e x_i é o valor da variável independente da i -ésima observação em que $i = 1, 2, \dots, n$.

O método da máxima verossimilhança é utilizado para estimar os parâmetros. A função de distribuição de probabilidade de Y_i para o modelo de regressão logística é dada por:

$$f(y_i, \pi(x_i)) = \pi_i^{y_i} (1 - \pi(x_i))^{1-y_i}. \quad (27)$$

Como as observações são independentes, a função de distribuição de probabilidade conjunta será:

$$\prod_{i=1}^n f(y_i, \pi(x_i)) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi(x_i))^{1-y_i}, \quad y_i \in [0,1]. \quad (28)$$

Então, a função de verossimilhança é dada por:

$$L(\mathbf{B}) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}, \quad \mathbf{B} \in \mathbb{R}^{(2)}. \quad (29)$$

O princípio da máxima verossimilhança é estimar o valor de $\mathbf{B}$ que maximiza $L(\mathbf{B})$.

Aplicando o logaritmo, a expressão é definida como:

$$l(\mathbf{B}) = \ln[L(\mathbf{B})] = \ln \left[\prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \right], \quad (30)$$

$$= \sum_{i=1}^n [y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))]. \quad (31)$$

Para encontrar o valor de $\mathbf{B}$ que maximiza $l(\mathbf{B})$, deriva-se $l(\mathbf{B})$ utilizou-se o processo iterativo de Newton-Raphson, sendo preciso derivar $l(\mathbf{B})$, em função de cada parâmetro.

$$\frac{\delta l(\mathbf{B})}{\delta \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \pi(x_i) = 0, \quad (32)$$

$$\frac{\delta l(\mathbf{B})}{\delta \beta_j} = \sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n x_{ij} \pi(x_i) = 0, \quad j \in \{1, \dots, p\}. \quad (33)$$

Então, como base na equação (31), o estimador pelo método da máxima verossimilhança $\hat{\mathbf{B}}$, é a solução das equações de verossimilhança

$$\sum_{i=1}^n (y_i - \pi(x_i)) = 0, \quad (34)$$

$$\sum_{i=1}^n (y_i - \pi(x_i)) = 0, \quad j \in \{1, \dots, p\}. \quad (35)$$

Segundo Silva (1992), o vetor $U(B)$, denominado vetor *score*, pode ser escrito da seguinte maneira:

$$U(B) = X^T y - X^T \pi = X^T (y - \pi). \quad (36)$$

A solução das equações de verossimilhança é dada pelo método de Newton-Raphson, também conhecido apenas como método de Newton, consiste em uma técnica de iterações muito utilizada para encontrar os zeros de uma função ou mesmo para a resolução de sistemas não lineares. Uma importante característica desse método em relação aos demais é que a sua convergência para a solução buscada é feita por uma taxa quadrática, tornando o tempo de cálculo normalmente menor quando comparado aos outros métodos (SOUZA, 2006).

Sendo o conjunto de equações iterativas dadas por

$$B^{(t+1)} = B^{(t)} + [I(B^{(t)})]^{-1} U(B^{(t)}); \quad t = 0, 1, 2 \dots \quad (37)$$

$$= B^{(t)} + [X^T Q^{(t)} X]^{-1} X^T (y - \pi^{(t)}), \quad (38)$$

onde $B^{(t)}$ e $B^{(t+1)}$ são os vetores dos parâmetros estimados nos passos t e $t+1$, $Q = \text{diag}[\pi(x)(1 - \pi(x))]$ e X é denominada matriz de dados, onde a sua inversa $[I(B)]^{-1}$ é a matriz de variâncias e covariâncias das estimativas de máxima verossimilhança, ou seja, estimativas sobre o sinal ou ruído dos parâmetros.

O chute inicial é dado com todos os coeficientes igualados a zero. Esses valores iniciais são substituídos no lado direito da equação (37), que dará o resultado para a primeira iteração B^1 . Em seguida, os valores são novamente substituídos no lado direito, $U(B)$ e $I(B)$ são recalculados, encontrando assim B^2 . Esse processo é repetido até a convergência, obtendo-se $B = B^{(t+1)}$.

Um dos critérios para verificar a convergência poderia ser:

$$\sum_{i=1}^n \left(\frac{B_j^{(t)} - B_j^{(t+1)}}{B_j^{(t)}} \right)^2 < \xi, \quad (39)$$

onde $\xi \ll 1$.

Geralmente não é possível encontrar distribuições exatas para os estimadores então trabalha-se com resultados assintóticos, considerando-se que o modelo escolhido irá satisfazer as condições de regularidade (SOUZA, 2006).

Nesses problemas regulares, a função escore $U(B) = \frac{\delta l(B)}{\delta B}$ tem valor esperado igual a zero, e a estrutura da covariância é igual a matriz de informação de Fisher $I(B) = E \left[\frac{\delta^2 l(B)}{\delta B \delta B^T} \right] = X^T Q X$.

Então B é dada por:

$$B \sim N_p(B, I(B)^{-1}). \quad (40)$$

2.5 SOFTWARE MAZDA

O MaZdA é um *software* usado para a extração de características de uma imagem que pode ser 2D ou 3D. Este *software* foi escrito para extrair uma variedade de características de textura, cor da imagem, incluindo reconhecimento de imagem, detecção e segmentação dentro de regiões de interesse arbitrariamente formadas. Também, inclui procedimentos para análise estatística de conjuntos de recursos computados, para auxiliar na classificação de imagens e reconhecimento de conteúdo de imagem (SZCZYPIŃSKI *et al.*, 2009).

O nome “MaZdA” é derivado de ‘*macierz zdarzen*’ que significa, em polonês “matriz de co-ocorrência” (TECHNICAL UNIVERSITY OF LODZ, 2010).

Neste trabalho foram usadas apenas os recursos de textura no qual, obteve-se 301 parâmetros, dentre eles os parâmetros das estatísticas de primeira e segunda ordem, usados neste trabalho e que são caracterizadores das imagens de melanomas e das imagens de manchas benignas.

3 EXPERIMENTO REALIZADO

Para o desenvolvimento desse trabalho foi escolhido, como foco de estudo, o câncer de pele do tipo melanoma e, neste capítulo, serão abordados a identificação do problema a ser estudado, bem como a metodologia com a qual foram realizadas as etapas de solução para o problema proposto no projeto.

3.1 IDENTIFICAÇÃO DO PROBLEMA

A escolha do melanoma, para o desenvolvimento deste trabalho se deu pelo fato de que, segundo a *American Cancer Society* (2016) alguns melanomas não se encaixam na regra **ABCDE** e podem conduzir a um falso diagnóstico mesmo para médicos com grande experiência e, por isso, é tão importante a pesquisa de novas formas de diagnóstico desse tipo de doença, que tem uma taxa de mortalidade tão alta, se não for descoberta no início.

A análise de textura de uma imagem, técnica que foi usada nesse trabalho, é bastante utilizada quando não há diferenças perceptíveis visualmente, ou quando essas diferenças são muito sutis.

Segundo Voigt (2002), o olho humano é capaz de discriminar 64 níveis de cinza e aproximadamente 200.000 cores. A identificação de características estruturais distintas pressupõe que o nível de cinza e os gradientes possam ser percebidos e alocados como uma matriz complexa de distribuição de níveis de cinza dentro das dimensões da área de imagem visualizada. Em imagens digitais, cada pixel de imagem recebe um valor de $2^8 = 256$ níveis de cinza distinguíveis, se forem usados sistemas de discriminação de 8 bits. Esse valor é significativamente maior do que o olho humano é capaz de discriminar e ilustra o grande potencial de diagnóstico se utilizadas imagens digitais para análise da doença. Somente se os níveis de cinza forem distintos um do outro, eles criarão estruturas que podem ser reconhecidas e diferenciadas. Já as imagens que contêm níveis cinzentos semelhantes, são deficientes em informações processáveis para distinção de estruturas.

Com a ideia de diminuir essa fragilidade na diferenciação visual dos tipos de manchas, bem como diminuir a quantidade de exames invasivos feitos para o diagnóstico desse tipo de câncer, foi feito este estudo que aborda a extração e análise das características de textura da imagem de melanomas e manchas benignas, afim

de encontrar parâmetros que sejam representativos para o processo de diferenciação e classificação do melanoma.

3.2 METODOLOGIA DE SOLUÇÃO DO PROBLEMA

3.2.1 Aquisição de Imagem

O primeiro passo do método desenvolvido é a aquisição da imagem, isto é, adquirir imagens de forma digitalizada para serem utilizadas no desenvolvimento do trabalho. Para tanto, foram utilizadas imagens digitais de melanomas e de imagens benignas de bancos de dados de imagens do (DERMNETNZ), (ATLAS OF CLINICAL DERMATOLOGY), e (ATLAS DERMATOLÓGICO).

Neste estudo, foram selecionadas 32 imagens de manchas de pele benignas e 32 imagens de câncer de pele do tipo melanoma.

3.2.2 Extração de Características e Análise de Textura

Inicialmente, foi feita a padronização do tamanho de algumas imagens com o objetivo de facilitar as seleções de regiões de interesse necessárias para análise pois algumas imagens, principalmente de manchas benignas, possuíam tamanhos menores que as demais, dificultando a seleção de regiões de interesse através do software utilizado para extração de características, o *MaZda*

Este software tem como padrão de entrada imagens em formato BMP e, antes de submeter as imagens ao programa, foi feita uma conversão em seus formatos, inicialmente em JPEG, para o formato BMP² através de um *software* de edição de fotos.

Todas as imagens usadas neste trabalho estão no formato BMP e possuem 24 bits por pixel. Imagens que possuem esse formato apresentam 256 níveis de cinza (que corresponde à profundidade de pixels $b=8$) então, as estatísticas de texturas deste trabalho foram calculadas para matrizes de dimensões 256x256.

² Um arquivo BMP é um arquivo bitmap, ou seja, um arquivo de imagem gráfica que é representado por uma matriz de bits que especifica a cor de cada pixel em uma matriz retangular de pixels. O número de bits dedicados a um pixel individual determina o número de cores que podem ser atribuídos a esse pixel (TEOTÔNIO, 2017).

Para a transformação da imagem colorida para tons de cinza, o *MaZda* utiliza o seguinte algoritmo de conversão, entre os existentes no *software* para conversão em outras escalas:

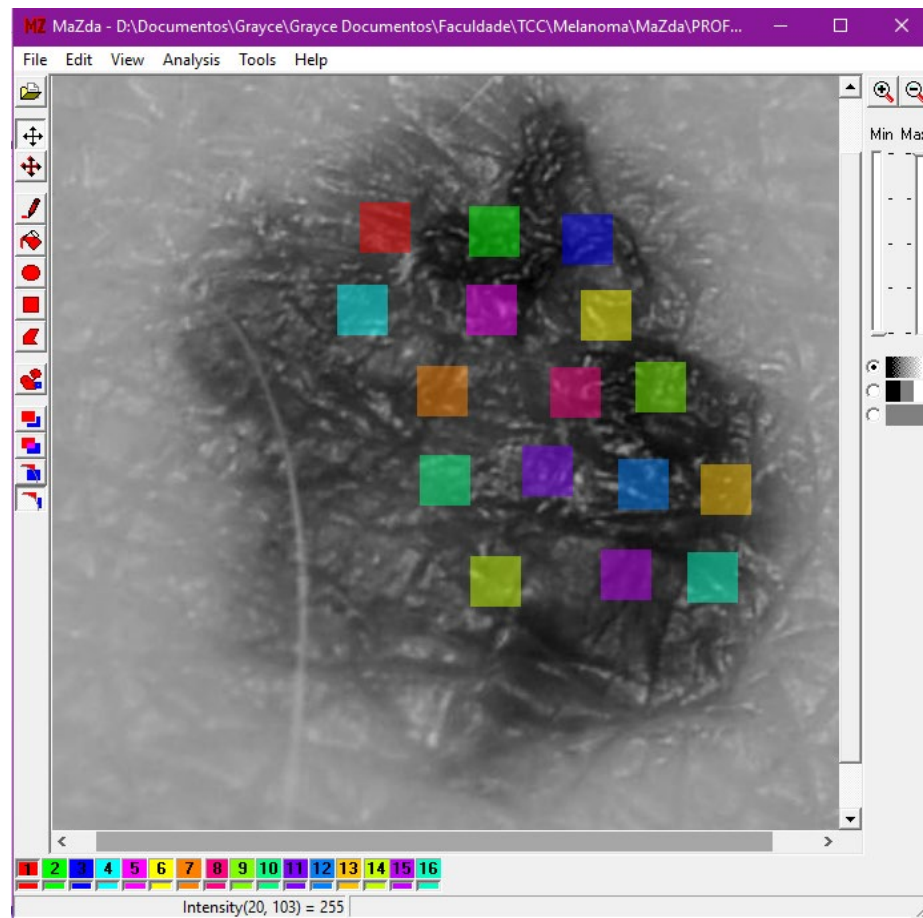
$$\text{Intensidade de Cinza} = (299R + 587G + 114B)/1000, \quad (41)$$

onde RGB (Red, Green e Blue), do português Vermelho, Verde e Azul respectivamente, é o sistema de cores, que formam a imagem colorida tricromática, composta pelas três cores primárias.

No *MaZda*, foram selecionadas 16 regiões de interesse (ROI) (número máximo de ROI's que o programa disponibiliza) de cada imagem do grupo das 32 imagens de manchas benignas e foi feito o mesmo procedimento para as 32 imagens de melanomas. Este processo foi feito separadamente para cada imagem.

Este programa calcula, para cada ROI, diversos parâmetros de textura, incluindo os parâmetros de primeira ordem e segunda ordem, que, quando selecionados em regiões distintas da mesma imagem, fornecerão um maior número de informações sobre de textura daquela imagem em estudo. Cada ROI que foi selecionada tem forma quadrada e possui tamanho de 40x40 pixels, tamanho esse escolhido pois, no estudo do tamanho da imagem, observou-se que conseguiríamos usar a quantidade máxima de regiões de interesse e que elas caberiam dentro da imagem analisada. Esse processo foi feito de maneira manual, forma como é disponibilizada pelo programa.

Figura 14 - Exemplo da seleção de 16 regiões de interesses diferentes selecionadas sobre a imagem de um melanoma do banco de dados.



Fonte: A autora, 2019.

No MaZda, cada ROI possui uma cor para diferencia-las entre si. Com as ROIs selecionadas, como mostra a Figura 14, foi escolhida a opção “*Analysis*” no Mazda que faz uma análise de cada região de interesse e entrega, como saída, uma tabela com os mais de 300 parâmetros caracterizadores da textura de cada região de interesse.

3.2.3 Análise Estatística

3.2.3.1 Testes t e Mann-Whitney

Nesta etapa serão aplicados testes estatísticos com a finalidade de comparar os dois grupos de imagens estudadas.

Para a análise estatística, foi feita a exportação da tabela de resultados da análise de textura do MaZda para o Excel, e em seguida foram extraídas as médias de cada parâmetro, para cada imagem estudada. Posteriormente foram feitas as análises e testes estatísticos.

A tabela criada no Excel, continha todos os parâmetros calculados para as 16 regiões de interesse de cada imagem. Esse processo foi feito tanto para o grupo das 32 imagens benignas quanto para o grupo das 32 imagens de melanomas.

As médias calculadas a partir dos parâmetros extraídos das imagens utilizadas, foram usadas nos cálculos dos testes estatísticos t e Mann-Whitney, que são testes que fazem a comparação estatística entre dois grupos, afim de encontrar quais parâmetros são significantemente diferenciadores de cada grupo de imagem.

Segundo Oliveira (2016), embora o teste-t deva ser aplicado para amostras com distribuição normal, o seu uso também é válido no caso de amostras grandes. Segundo Vieira e Siqueira (2017), em amostras grandes (mais que 30 observações) o TLC é válido e podemos aplicar testes paramétricos. Na área médica, em particular, é comum o uso de testes paramétrico em amostras com $n \geq 30$. No caso deste trabalho, o número de amostras foi $n = 32$, o que justifica uso do teste-t.

A utilização dos dois testes apresentados foi feita pois, o teste de Mann-Whitney diferenciou algumas estruturas que o teste t não diferenciou e que são importantes para a análise e de texturas e, por consequência, para a caracterização das imagens estudadas.

A Tabela 4 mostra os parâmetros que apresentaram diferenças estatisticamente significantes ($p < 0,05$) de acordo com os testes estatísticos que foram realizados.

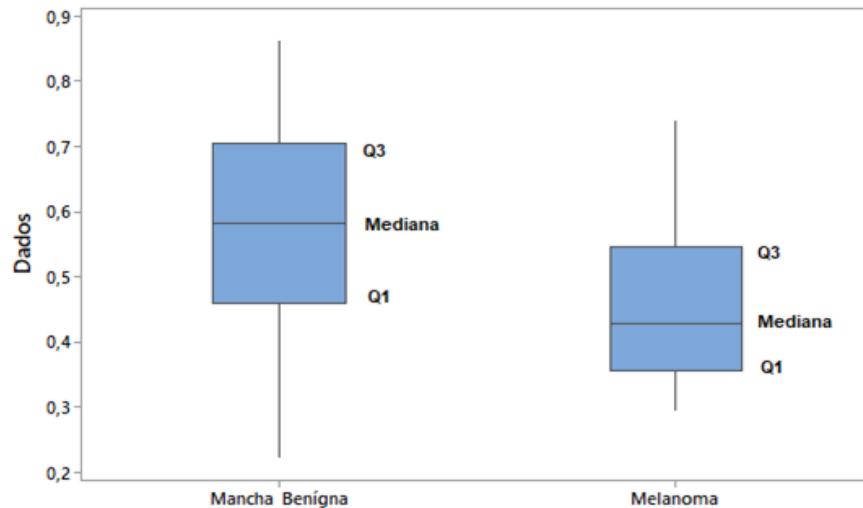
Tabela 4 – Parâmetros de texturas obtidos através do MaZda, e seus respectivos valores de média, desvio padrão (DP) e valor-p, obtidos através dos testes t e Mann-Whitney.

	Benignas		Melanomas		Teste t	Teste Mann-Whitney
	Média	±DP	Média	± DP	Valor-p	
Variância	166	203	226	183	0,212	0,006
Assimetria	0,613	0,514	0,812	0,472	0,113	0,023
Segundo Momento Angular (ASM)	0,685	0,614	0,0293	0,024	0,002	0,001
Contraste	8,3	12,9	11,35	9,39	0,289	0,006
Correlação	0,641	0,179	0,5856	0,0909	0,122	0,011
Soma dos Quadrados	10,2	12,6	14	11,3	0,215	0,005
Momento da Diferença Inversa	0,574	0,173	0,455	0,107	0,002	0,002
Entropia das Somas	1.137	0.213	1.288	0.15	0.002	0,001
Entropia	1,574	0,441	1,884	0,292	0,002	0,001
Variância da Diferença	3,87	6,15	5,3	4,54	0,295	0,007
Entropia da Diferença	0,587	0,24	0,742	0,16	0,004	0,002
NUCE Hor.	254	264	378	192	0,037	0,003
MIELP Hor.	71,1	16,8	81,8	14,1	0,011	0,009
MIECP Hor.	0,493	0,22	0,653	0,131	0,001	0,002
FIE Hor.	0,417	0,214	0,557	0,136	0,003	0,003
NUCE Vert.	254	273	409	184	0,01	0,001
MIELP Vert.	71,7	16,8	81,8	14,1	0,011	0,009
MIECP Vert.	0,492	0,216	0,673	0,123	0,001	0,001
FIE Vert.	0,418	0,212	0,579	0,128	0,001	0,001
NUCE 45°	345	298	529	203	0,005	0,002
MIECP 45°	0,576	0,203	0,659	0,12	0,001	0,001
FIE 45°	0,505	0,207	0,659	0,12	0,001	0,001
NUCE 135°	351	303	528	196	0,008	0,003
MIECP 135°	0,578	0,206	0,738	0,101	0,001	0,003
FIE 135°	0,509	0,209	0,659	0,115	0,001	0,002
Média do Gradiente Absoluto	0,589	0,489	0,357	0,212	0,001	0,002
Assimetria do Gradiente Absoluto	1,679	0,950	0,929	0,469	0,001	0,001
Curtose do Gradiente Absoluto	3,66	6,37	0,62	1,88	0,014	0,007

Nota: ASM = *Angular Second Moment*, Hor. = Horizontal e Ver. = Vertical. Fonte: A autora, 2019.

Na análise do *box-plot* para os parâmetros mais significantes de textura das imagens do estudo, apresentados na tabela 4, percebe-se que realmente há uma diferença entre os parâmetros das manchas benignas e dos melanomas, mas eles apresentam uma certa sobreposição.

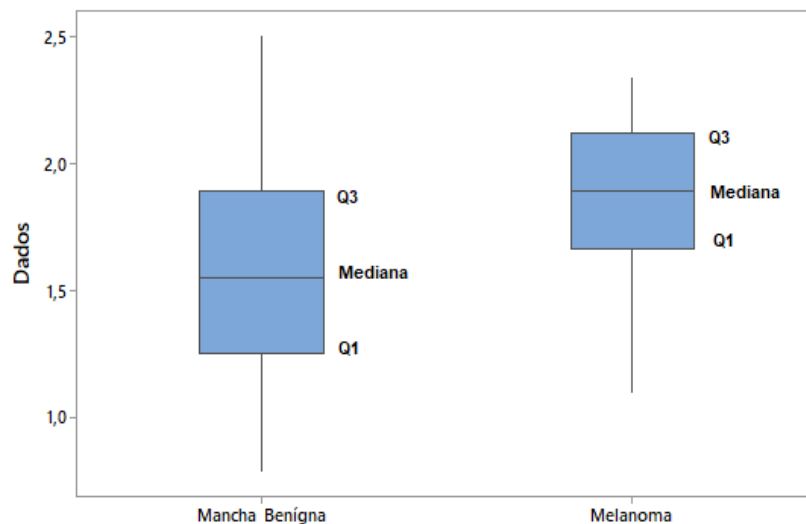
Figura 15 – *BoxPlot* do parâmetro de textura Momento da Diferença Inversa para o grupo de imagens de manchas benignas *versus* melanomas.



Nota: Q1= 1º Quartil e Q3 = 3º Quartil. Fonte: A autora, 2019.

As Figuras 15 e 16 mostram os *box-plots* do Momento Inverso da Diferença e Entropia. Os gráficos dos outros parâmetros representativos estão mostrados no Apêndice A.

Figura 16 – *BoxPlot* do parâmetro de textura Entropia para o grupo de imagens de manchas benignas *versus* melanomas.



Nota: Q1= 1º Quartil e Q3 = 3º Quartil. Fonte: A autora, 2019.

3.2.3.2 Predição Linear

Para avaliar a capacidade de predição entre os parâmetros de textura, foi realizado o processo estatístico de regressão linear para predição de alguns parâmetros, para cada grupo de imagens. Para isso, foi analisada a tabela de correlação e, os parâmetros escolhidos, foram os que apresentaram maiores números de parâmetros com correlações baixas entre si, ou seja, que possuíam coeficientes de correlação (r) entre -0,2 e +0,2. As Tabelas 5, 6, 7 e 8 mostram os coeficientes de correlação para cada parâmetro em relação ao parâmetro analisado.

Para o grupo de imagens de melanomas foram selecionados os parâmetros Correlação e o Momento da Diferença Inversa.

Para o parâmetro Correlação foi obtida a seguinte equação de regressão linear:

$$\begin{aligned}\hat{Y}_{Correlação} = & 0,560 - 0,0451(Média) - 0,00788(Contraste) + 0,0145(Perc. 10\%) \\ & + 0,0169(Perc. 50\%) + 0,0123(Perc. 90\%) \\ & + 0,700(Momento da Diferença Inversa) + 0,003(Média da Soma) \\ & + 0,00239(Variância da Soma) - 0,02001(NUNC 135^\circ) \\ & - 0,0488(Assimetria do Gradiente) - 0,224(Sigma).\end{aligned}\quad (42)$$

Tabela 5 - Coeficiente de correlação (r) para o parâmetro de textura Correlação.

Parâmetros	$r < 0.2$
Média	0,133328
Contraste	0,012551
Perc. 10%	0,187979
Perc. 50%	0,171032
Perc. 90%	0,00134
Momento da Diferença Inversa	0,019556
Média da Soma	0,131405
Variância da Soma	-0,1982
NUNC 135°	-0,16793
Assimetria do Gradiente	0,135279
Sigma	-0,14486

Fonte: A autora, 2019.

Para o parâmetro Momento da Diferença Inversa foi obtida a seguinte equação de regressão linear:

$\hat{Y}_{\text{Momento da Diferença Inversa}}$

$$\begin{aligned}
 &= 1,0262 + 0,000061 (\text{Média}) - 0,0070 (\text{Curtose}) + 0,0144 (\text{Contraste}) \\
 &- 0,018 (\text{Correlação}) + 0,0283 (\text{Soma dos Quadrados}) \\
 &+ 0,535 (\text{Entropia da Soma}) - 0,647 (\text{Entropia}) \\
 &- (0,000243 \text{ NUNC HOR}) + 0,00035 (\text{Curtose do Gradiente}) \\
 &- 0,0260 (\text{Teta1}).
 \end{aligned} \tag{43}$$

Tabela 6 - Coeficientes de correlação (r) para o parâmetro de textura Momento da Diferença Inversa.

Parâmetros (x)	r <0,2
Média	0,133328
Curtose	0,187979
Contraste	0,107681
Correlação	0,171032
Soma dos Quadrados	0,00134
Entropia da Soma	0,019556
Entropia	0,131405
NUNCE HOR	-0,1982
Curtose do Gradiente	-0,16793
Teta1	0,135279

Fonte: A autora, 2019.

Já para o grupo de imagens de manchas benignas, foram selecionados os parâmetros de textura Variância e Assimetria cujas equações de regressão, bem como tabela de correlação, são mostrados a seguir.

Para o parâmetro Variância foi obtida a seguinte equação de regressão linear:

$$\begin{aligned}
 \hat{Y}_{\text{Variância}} &= 326,2 + 20,85 (\text{Perc. 01\%}) - 60,3 (\text{Perc. 10\%}) - 31,8 (\text{Perc. 50\%}) \\
 &- 34,5 (\text{Perc. 90\%}) + 763 (\text{Segundo Momento Angular}) \\
 &+ 0,85 (\text{Contraste}) - 8,9 (\text{Média da Soma}) + 1,59 (\text{NUNC HOR.}) \\
 &- 0,06 (\text{NUNC VERT.}) - 160 (\text{Sigma}).
 \end{aligned} \tag{44}$$

Tabela 7 - Coeficientes de correlação (r) para o parâmetro de textura Variância.

Parâmetros (x)	r <0,2
Perc.01%	-0,08535
Perc.10%	-0,02102
Perc.50%	0,095049
Perc.90%	0,061337
Segundo Momento Angular	-0,1074
Contraste	0,010461
Média da Soma	0,150473
NUNC HOR	-0,05458
NUNC VERT	-0,0724
Sigma	0,165037

Fonte: A autora, 2019.

Para o parâmetro Assimetria foi obtida a seguinte equação de regressão linear:

$$\begin{aligned}
 \hat{Y}_{Assimetria} = & + 0,1562 (Perc.01\%) - 0,132 (Perc.10\%) - 0,3082 (Perc.50\%) \\
 & - 1,92 (Segundo Momento Angular) + 0,015 (Média da Soma) \\
 & + 0,276 (Entropia da Soma) + 0,001585 (NUCE HOR.) \\
 & - 0,0051 (NUNC 45^\circ) + 0,0080 (NUNC 135^\circ) + 1,374 (Teta1).
 \end{aligned}
 \tag{45}$$

Tabela 8 - Coeficientes de correlação (r) para o parâmetro de textura Assimetria.

Parâmetros (X_i)	r <0,2
Perc.01%	0,007416
Perc.10%	0,043363
Perc.50%	0,098973
Segundo Momento Angular	-0,1603
Média da Soma	0,143854
Entropia da Soma	0,139009
NUCE HOR	0,118152
NUNC 45°	0,101333
NUNC 135°	0,116572
Teta1	0,051916

Fonte: A autora, 2019.

Para realização dos testes de acerto e precisão da regressão linear, foram utilizadas imagem de manchas com diagnósticos previamente conhecidos. Em

seguida foram extraídas as características de textura da imagem, através do software *MaZda*, para 16 regiões de interesse. Como resposta, foram obtidos os parâmetros texturais da imagem analisada e de posse desses valores, foram extraída as médias das 16 ROI's, mesmo procedimento que foi realizado inicialmente, com todas as 64 imagens, para obtenção do modelo final.

Em seguida à predição, foram realizados os testes de comparação entre os valores destes parâmetros e os valores obtidos para as imagens utilizadas nos testes de acerto e validação, com o objetivo de obter a taxa de erro das modelagens encontradas na regressão linear. Estes resultados, estão mostradas no capítulo 4 nas Tabelas 12, 13, 14 e 15.

3.2.3.3 Modelagem pela Regressão Logística Binária

Para perceber a influência dos parâmetros de textura na probabilidade da mancha ser benigna ou ser um melanoma, foi utilizada uma técnica estatística de probabilidade chamada regressão logística, em que a variável dependente (Y) desta regressão foi construída a partir da identificação do tipo da imagem, e estabelecida a partir das variáveis independentes (X_i), que no nosso estudo, foram todos os parâmetros de textura caracterizadores da imagem, que foram estatisticamente significativos para a diferenciação ($p < 0,05$) de acordo com os testes t e Mann-Whitney e que foram mostrados na Tabela 4.

A modelagem do estudo proposto, obtida através da regressão logística, foi construída para fornecer uma relação entre os parâmetros de textura usados e a variável dependente na forma de uma equação.

A seleção das variáveis independentes que serão usadas no modelo pode ser feita, segundo Hair *et al.* (1998), através da exclusão de todas as variáveis que apresentaram um Fator de Inflação da Variância (FIV) alto ($VIF > 5$). Valores ideais para o FIV estão entre 1 e 5. Este processo é repetido até se encontrar o modelo que contenha apenas os parâmetros que apresentaram um FIV baixo e como consequência disso, obtém-se também, valores $p < 0,05$, para o teste de Wald, para os parâmetros restantes na modelagem. Este processo foi realizado neste trabalho para a obtenção dos parâmetros do modelo.

O FIV mede a quantidade de multicolinearidade existente em uma análise de regressão. Segundo Mohamad (2016), a multicolinearidade na regressão logística ocorre quando há maior correlação entre as variáveis independentes do modelo.

Para a verificação sobre a não correlação entre as variáveis utilizadas na modelagem, foi montada a matriz de correlação, que está mostrada na tabela do Apêndice C, que exhibe os coeficientes de correlação (r). Nesta análise, foi definido como correlação baixa, valores de $r < 0,2$.

A variável dependente do modelo tem característica dicotômica e é representada, no nosso estudo, pelo questionamento sobre a malignidade da mancha de pele (se a mancha é um melanoma, ou se é uma mancha benigna).

A modelagem, obtida através da regressão logística, pode ser representada na forma matricial para cada grupo de imagens estudado (manchas de imagens benignas e melanomas), como:

- Imagens de melanomas:

$$\begin{bmatrix} Y_{m1} \\ Y_{m2} \\ \vdots \\ Y_{m32} \end{bmatrix} = \begin{bmatrix} x_{m1,1} & \dots & x_{m1,18} \\ x_{m2,1} & \dots & x_{m2,18} \\ \vdots & \ddots & \vdots \\ x_{m32,1} & \dots & x_{m32,18} \end{bmatrix} * \begin{bmatrix} w_{m1} \\ w_{m2} \\ \vdots \\ w_{m32} \end{bmatrix} + \begin{bmatrix} b_{m1} \\ b_{m2} \\ \vdots \\ b_{m32} \end{bmatrix}, \quad (46)$$

onde,

Y_m - variável dependente e dicotômica.

X_m - variáveis independentes (parâmetros obtidos no MaZda) da análise.

w_m - coeficientes das equações.

b_m - constantes.

- Imagens de manchas benignas

$$\begin{bmatrix} Y_{b1} \\ Y_{b2} \\ \vdots \\ Y_{b32} \end{bmatrix} = \begin{bmatrix} x_{b1,1} & \dots & x_{b1,18} \\ x_{b2,1} & \dots & x_{b2,18} \\ \vdots & \ddots & \vdots \\ x_{b32,1} & \dots & x_{b32,18} \end{bmatrix} * \begin{bmatrix} w_{b1} \\ w_{b2} \\ \vdots \\ w_{b32} \end{bmatrix} + \begin{bmatrix} b_{b1} \\ b_{b2} \\ \vdots \\ b_{b32} \end{bmatrix}, \quad (47)$$

onde,

Y_b - variável dependente e dicotômica.

X_b - variáveis independentes (parâmetros obtidos no MaZda) da análise.

w_b - coeficientes das equações.

b_b - constantes.

Usando o *software* estatístico Minitab 18.1 (MINITAB, 2017), aplicou-se o modelo de regressão logística aos parâmetros selecionados e à variável dependente e assim, foi obtida como saída, a seguinte equação de modelagem geral:

$$\begin{aligned}
 Y = & -290 - 0,452(\text{Variância}) + 2,52(\text{Assimetria}) + 983(\text{SegundoMomento Angular}) \\
 & - 4,24(\text{Contraste}) + 24,4(\text{Correlação}) + 6,80(\text{Soma dos Quadrados}) \\
 & + 80(\text{Momento da Diferença Inversa}) - 92(\text{Entropia da Soma}) \\
 & + 252(\text{Entropia}) + 8,65(\text{Variância da Diferença}) \\
 & - 192(\text{Entropia da Diferença}) - 0,0314(\text{NUCE Hor}) \\
 & - 0,929(\text{MELP Hor}) + 0,0077(\text{NUCE } 45^\circ) - 55(\text{MIECP } 45^\circ) \\
 & + 98(\text{FIE } 45^\circ) - 0,906(\text{Assimetria do Gradiente}) \\
 & - 1,310(\text{Curtose do Gradiente}).
 \end{aligned} \tag{48}$$

onde, a probabilidade do evento escolhido acontecer é determinada por $\pi(x) = P(Y)$, e é definida por $P(Y) = \frac{\exp(Y)}{(1+\exp(Y))}$.

Nesta modelagem, P significa a probabilidade de a variável dependente Y , ser classificada como uma mancha do tipo melanoma (evento escolhido) ou como uma mancha benigna, através da equação de exponencial de probabilidade $P(Y)$.

Todas as variáveis independentes do modelo são contínuas, ou seja, os valores possíveis para cada uma delas é um valor dentro do conjunto de números reais, e representam um conjunto de variáveis qualitativas e quantitativas que impactam no valor de P podendo assim, influenciar na probabilidade da imagem da macha estudada ser de um melanoma.

Como, para a criação do modelo foram usadas 32 (50%) imagens de manchas benignas e 32 (50%) de imagens de melanomas então, foi criado um valor de corte para $P(Y)$, onde: resultados da probabilidade $P(Y)$ maiores que 0,5 são classificados como melanomas, e resultados da probabilidade $P(Y)$ abaixo de 0,5 são considerados de manchas benignas.

Além, da modelagem, foi obtida a tabela das varáveis da equação e seus respectivos coeficientes como mostra a Tabela 9.

A Tabela 9 mostra o teste de Wald e sua significância. Segundo Hair *et al.* (1998) o teste de Wald mede a importância de cada variável da regressão logística, então, variáveis que apresentam níveis de significância menores que 0,05, são consideradas variáveis importantes para o modelo.

Tabela 9 – Estatísticas das variáveis (parâmetros) do modelo.

Termo	Coeficiente	Estatística de Wald	Valor-p	FIV
Constante	-290	11,82	0,028	*
Variância	-0,452	6,01	0,042	1,25
Assimetria	2,52	9,64	0,034	3,01
Segundo Momento Angular	983	2,28	0,023	2,54
Contraste	-4,24	2,27	0,023	3,77
Correlação	24,4	11,38	0,023	3,82
Soma dos Quadrados	6,80	3,93	0,041	1,03
Momento da Diferença Inversa	80	3,52	0,031	2,65
Entropia da Soma	-92	4,45	0,042	1,80
Entropia	252	4,38	0,016	1,29
Variância da Diferença	8,65	8,32	0,021	3,98
Entropia da Diferença	-192	9,41	0,036	3,12
NUCE HOR	-0,0314	15,95	0,044	3,74
MELP HOR	-0,929	21,07	0,038	2,69
NUCE 45°	0,0077	9,14	0,029	1,71
MIECP 45°	-55	6,44	0,038	3,00
FIE 45°	98	7,67	0,005	1,39
Assimetria do Gradiente	-0,906	4,23	0,037	2,61
Curtose do Gradiente	-1,310	7,73	0,024	1,28

Fonte: A autora, 2019.

Estes testes de qualidade do ajuste são usados para determinar se as probabilidades preditas fogem das probabilidades observadas. Se o valor-p dos testes de qualidade do ajuste for menor do que seu nível de significância (0,05), as probabilidades preditas desviam-se das probabilidades observadas de uma maneira que a distribuição binomial não prediz. Em outras palavras, se valor-p do teste for maior que o nível de significância, o modelo encontrado prediz.

As Tabelas 10 e 11 mostram os testes de qualidade do ajuste e validação do modelo, que foram obtidos após ser realizado o processo de regressão logística.

Os testes de qualidade obtidos na modelagem foram o teste *Deviance* e o Hosmer-Lemeshow e são definidos como:

- *Deviance*: Usado para detectar erros no ajuste. Tal resultado é uma estatística que representa a bondade do ajuste (SOUZA, 2006).
- Hosmer-Lemeshow: A finalidade desse teste é verificar se existem diferenças significativas entre as classificações realizadas pelo modelo e a realidade observada (HAIR *et al.*, 1998).

Tabela 10 – Testes de qualidade e validação do modelo.

Testes	Qui- Quadrado	Valor-p
Hosmer-Lemeshow	3,22	0,913
<i>Deviance</i>	31,05	0,944

Fonte: A autora, 2019.

Outro teste de qualidade obtido da regressão logística foi o R^2 mostrado na Tabela 11. Segundo Hair *et al.* (1998), este teste é considerado como proporção do desvio na resposta das variáveis que explicam o modelo e mostra o quão bom o modelo se ajusta aos dados. Valores maiores que 75% são considerados bons resultados para a modelagem.

Tabela 11 – Teste de qualidade do ajuste dos coeficientes ao modelo.

Teste	Resultado
R^2	75,01%

Fonte: A autora, 2019.

Para realização dos testes de acerto e precisão do modelo, foi realizado o mesmo procedimento de análise das imagens-testes que foi utilizado para regressão linear. Em seguida, foram analisadas quantas imagens passaram nos testes de validação da modelagem afim de encontrar a porcentagem de acerto do modelo. Estes resultados estão mostrados na Tabela 16 no capítulo 4, que trata das conclusões.

4 CONCLUSÕES

Percebe-se então que a análise de textura realmente apresenta uma boa capacidade de diferenciação das imagens de cada grupo e, analisando o resultado dos testes diferenciadores de grupos para alguns parâmetros nota-se que, o Momento da Diferença Inversa, que mede a homogeneidade (suavidade) da distribuição de níveis de cinza, diminuem nas imagens de melanomas. Esse é um resultado esperado e identificado no estudo, pois a mancha benigna possui uma textura mais homogênea em relação ao melanoma. Também, foi possível perceber que o Contraste aumenta nas imagens de melanoma, o que também é um resultado esperado e identificado, já que este parâmetro representa a quantidade de variação local dos níveis de cinza, onde um valor mais elevado desse parâmetro, indica a presença bordas, ruídos ou rugas na imagem, características tais que são pouco observadas em imagens de manchas benignas.

Além dos resultados já mencionados, um outro resultado importante encontrado neste estudo foi o parâmetro Entropia, que mede o grau de desordem da imagem, que foi maior nas imagens de melanomas. Este dado é esperado e também verificado, pois imagens de melanomas possuem uma maior desordem na distribuição dos níveis de cinza em comparação com a imagem de uma mancha benigna. Já o Segundo Momento Angular, que mede a uniformidade (ou ordem) da imagem diminuiu para as imagens de melanomas. Este resultado pode ser interpretado como o melanoma degenerando a pele de pacientes que possuem a doença.

Analisando os dados obtidos através dos testes diferenciadores de grupos, percebe-se que o teste t indicou que os valores da Variância e Assimetria, que são baseadas no histograma em níveis de cinza de primeira ordem, a Variância da Soma, Entropia da Soma, Entropia e Entropia da Diferença obtidos através da matriz de co-ocorrência e os parâmetros derivados da Matriz de *Run Length* foram significativamente maiores para as imagens de melanoma do que nas imagens de manchas benignas. Já o Segundo Momento Angular, da matriz de co-ocorrência, e os parâmetros baseados no Gradiente Absoluto foram significativamente maiores para a imagem de manchas benignas do que nas imagens de melanomas, resultado obtido também pelo teste-t.

Já de acordo com o teste de Mann-Whitney, foi possível encontrar parâmetros que são estatisticamente significantes na diferenciação dos grupos e que não foram

encontrados pelo teste t. Tais parâmetros foram: Contraste, Soma dos Quadrados e Variância da Diferença, que foram maiores para imagens de melanomas, e o Correlação, que foi menor para imagens de manchas benignas.

Observando os dados, percebe-se que os parâmetros de texturas indicam que existe realmente uma quebra na uniformidade e homogeneidade na textura das imagens de manchas de melanomas em relação às imagens de manchas benignas, mesmo que essas não-uniformidades não sejam perceptíveis visualmente.

Analisando os valores das médias e desvios padrões de alguns parâmetros trabalhados que estão na Tabela 4, percebe-se que há uma certa superposição entre as distribuições dos grupos. Apesar disso, os testes estatísticos usados no estudo mostram que existe uma diferença entre tais grupos.

Para a verificação da capacidade de predição dos parâmetros bem como a validação e teste de acerto do modelo de regressão logística, foram feitos testes, aplicando às modelagens, 10 imagens com diagnósticos conhecidos, das quais, 5 foram imagens de manchas benignas e 5 imagens de melanomas e que não participaram da criação do modelo.

Para a predição linear, realizada em relação aos parâmetros do modelo foram obtidos os seguintes resultados mostrados nas Tabelas 12, 13, 14 e 15.

A Tabela 12 mostra a predição para o parâmetro Assimetria do grupo de imagens de manchas benignas.

Tabela 12 - Predição para o parâmetro Assimetria.

	Valor Obtido na Predição	Valor Real	 Erro %
Imagem 1	1,268354	1,727366	26,57
Imagem 2	0,363114	0,333996	8,72
Imagem 3	0,762753	0,625369	21,96
Imagem 4	0,433405	0,567349	23,61
Imagem 5	0,428947	0,482366	11,07
Erro Médio $\pm$ DP			18,39 $\pm$ 7,13

Nota: DP=Desvio Padrão. Fonte: A autora, 2019.

A Tabela 13 mostra a predição para o parâmetro Variância do grupo de imagens de manchas benignas.

Tabela 13 - Predição para o parâmetro Variância.

	Valor Obtido na Predição	Valor Real	 Erro %
Imagem 1	130,001	138,0881	5,86
Imagem 2	140,493	113,8188	23,34
Imagem 3	227,906	186,1641	22,42
Imagem 4	400,854	336,5818	19,09
Imagem 5	519,781	412,1065	26,13
Erro Médio $\pm$ DP			19,37 $\pm$ 7,12

Nota: DP=Desvio Padrão. Fonte: A autora, 2019.

A Tabela 14 mostra a predição para o parâmetro Momento da Diferença Inversa do grupo de imagens de manchas de melanomas.

Tabela 14 - Predição para o parâmetro Momento da Diferença Inversa.

	Valor Obtido na Predição	Valor Real	 Erro %
Imagem 1	0,30583	0,309264	1,11
Imagem 2	0,343191	0,340468	0,79
Imagem 3	0,397401	0,394865	0,64
Imagem 4	0,403742	0,403718	0,0059
Imagem 5	0,296319	0,298151	0,61
Erro Médio $\pm$ DP			0,63 $\pm$ 0,36

Nota: DP=Desvio Padrão. Fonte: A autora, 2019.

A Tabela 15 mostra a predição, para o último parâmetro analisado, que foi a Correlação do grupo de imagens de manchas de melanomas.

Tabela 15 - Predição para o parâmetro Correlação do grupo de imagens de manchas de melanomas.

	Valor Obtido na Predição	Valor Real	Erro %
Imagem 1	0,497456	0,475358	4,65
Imagem 2	0,681032	0,622801	9,35
Imagem 3	0,74213	0,682856	8,68
Imagem 4	0,710314	0,703234	1,00
Imagem 5	0,402522	0,490874	17,98
Erro Médio $\pm$ DP			8,33 $\pm$ 5,68

Nota: DP=Desvio Padrão. Fonte: A autora, 2019.

Analizando os resultados que foram obtidos utilizando a predição linear, e que foram mostrados nas Tabelas 12, 13, 14 e 15, foi possível perceber que a porcentagem de erro, obtida entre as médias de alguns parâmetros estudados, das 32 imagens de melanomas e manchas benignas e as imagens usadas para a realização dos testes, foi significativamente pequena (abaixo de 50%), o que mostra que realmente existe uma boa capacidade de classificação através dos parâmetros de textura das imagens.

Observa-se também que o parâmetro Momento da Diferença Inversa, para imagens de manchas de melanomas, foi o que apresentou um percentual de erro menor entre todos os parâmetros testados, tornando-o um parâmetro com boa capacidade de classificação.

Para a modelagem da regressão logística com os parâmetros selecionados, foram obtidas as estatísticas de Wald, que como já mencionado, mede a importância de cada variável para o modelo ($p < 0,05$), e de acordo com a Tabela 9, percebe-se que os parâmetros selecionados possuíram valores de significância (valor-p) menores que 0,05, sendo então ditos importantes para a modelagem.

Além da estatística de Wald, foram obtidos testes de qualidade do modelo. Tais testes foram o de Hosmer e Lemeshow, *Deviance* e o R^2 .

O teste Hosmer e Lemeshow mede a correspondência dos valores efetivos previstos da variável dependente. Segundo Hair *et al.* (1998), um bom ajuste do modelo é caracterizado por um valor, *chi-quadrado* não significativo. De acordo com a Tabela 10, o teste de Hosmer e Lemeshow mostrou um valor de $p=0,913$, o que representa e comprova que o modelo apresentou um bom ajuste, informando que o modelo pode ser utilizado para estimar a probabilidade de uma mancha ser um

melanoma ou não, a partir das variáveis independentes utilizadas. O teste *Deviance* também apresentou um valor p maior que nível de significância e, segundo Souza (2006), este resultado mostra o quão bom o modelo se mostra, concluindo assim que o modelo representa a realidade de forma satisfatória.

A análise da Tabela 11 mostra o fator R^2 , o chamado fator de qualidade do modelo, é uma medida de quão bem o modelo se ajusta aos dados.

Segundo Hair *et al.* (1998) valores de R^2 acima de 75% provam que o modelo estimado possui um bom ajuste em relação aos dados usados. Para a regressão logística realizada neste trabalho, o teste R^2 apresentou um resultado de 75,01%, concluindo-se assim, que o modelo apresentou um resultado satisfatório, no que diz respeito ao ajuste do modelo.

Em seguida, foi feita a análise dos resultados obtidos através dos testes de acerto do modelo, para regressão logística. Estes resultados encontram-se na Tabela 16.

Tabela 16 – Resultado dos testes de classificação.

Imagem testada	Classificação da mancha	Porcentagem de chances de ser um melanoma	Resultado
Imagem 1	Melanoma	99,91%	Acerto
Imagem 2	Melanoma	0,3%	Erro
Imagem 3	Melanoma	99,98%	Acerto
Imagem 4	Melanoma	100%	Acerto
Imagem 5	Melanoma	98,76%	Acerto
Imagem 6	Mancha Benigna	29,2%	Acerto
Imagem 7	Mancha Benigna	9,34%	Acerto
Imagem 8	Mancha Benigna	97,71%	Erro
Imagem 9	Mancha Benigna	99,62%	Erro
Imagem 10	Mancha Benigna	0%	Acerto

Nota: Porcentagens acima de 50% indicam que a mancha é um melanoma e porcentagens abaixo de 50% indicam que a mancha é benigna. Fonte: A autora, 2019.

Ao analisar a Tabela 16, percebe-se que das 10 imagens testadas, a modelagem conseguiu classificar corretamente 7 imagens, obtendo então 70% de acertos no teste de classificação.

Este resultado evidencia que o modelo proposto é potencialmente significativo para a classificação de imagens quanto ao seu tipo (benigna ou melanoma).

4.1 CONSIDERAÇÕES FINAIS

A classificação do câncer de pele ainda é uma questão desafiadora na medicina. A imagenologia, que é o estudo dos órgãos e sistemas do corpo humano através de exames de imagem, depende do conhecimento prévio anatômico do órgão a ser analisado para a interpretação dos resultados do exame. Este fator depende da qualidade e desempenho do meio que produz tais imagens, estimulando assim a busca das técnicas de processamento de imagem para o auxílio na detecção de doenças.

O objetivo desse trabalho foi a extração de características de textura de imagens de manchas de pele, e benignas de melanomas, afim de encontrar parâmetros que fossem representativos para diferenciar estes tipos de imagens. Para isso, foram estudadas algumas técnicas de processamento de imagem para extração de texturas, bem como a análise estatística de cada característica.

Cada parâmetro de textura foi analisado e comparado estatisticamente através de testes paramétricos e não-paramétricos, com os quais foram retirados resultados bastante significativos. A partir destes parâmetros, foi criado um modelo matemático classificador que forneceu resultados bastante importantes na identificação do tipo da lesão da pele, podendo ser de grande importância no auxílio ao diagnóstico do melanoma.

Juntamente à modelagem classificadora, as predições lineares de parâmetros das imagens, realizadas entre as médias dos parâmetros das 32 imagens e as imagens usadas nos testes, obtiveram bons resultados mostrando assim que os parâmetros de textura são bons identificadores e classificadores de imagens.

A técnica de análise de textura das imagens visa encontrar parâmetros que façam a diferenciação entre duas imagens. Todos esses parâmetros, que passaram nos testes estatísticos diferenciadores de grupos (teste t e Mann-Whitney), são ditos diferenciadores entre a imagem de um tecido com mancha benigna e um com melanoma e, por isso, representam a base da caracterização das imagens de cada

grupo. Estes parâmetros são adequados para a caracterização das distribuições de cinza das ROI's e podem ser considerados bons biomarcadores³.

Diante dos resultados obtidos, foi possível perceber que o objetivo geral do trabalho, que é a classificação de imagens a partir de descritores de textura, foi atendido, mostrando assim que este tipo de técnica é promissora para análise e caracterização de imagens em geral. Essa metodologia pode ser aplicada em outras áreas, não apenas na medicina, como foi proposto por este trabalho.

³ Biomarcadores ou marcadores biológicos são entidades que podem ser medidas experimentalmente e indicam a ocorrência de uma determinada função normal ou patológica de um organismo ou uma resposta a um agente farmacológico (MONTEIRO, 2019).

5 TRABALHOS FUTUROS

Espera-se que esse projeto venha a agregar conceitos e possibilidades à trabalhos futuros que seguirem este mesmo modelo de pesquisa.

Os resultados positivos obtidos de acordo com as imagens estudadas mostram o potencial deste trabalho como auxílio de diagnóstico e melhor compreensão de patologias que podem ser averiguadas através de imagens.

As perspectivas para a continuação e melhoria do trabalho desenvolvido são:

- Obtenção de imagens padronizadas e completamente aptas para realização do estudo, ou seja, que exista um maior controle no processo de aquisição das imagens para que estas possuam uma boa resolução e não apresentem interferência externas como reflexos ou outros fatores, que impeçam que a análise e extração de características seja feita a partir uma imagem que esteja mais próxima da realidade quanto possível.
- Uso de um maior número de imagens (amostras) para realização do estudo.
- Exploração de outras abordagens de análise de textura, além da que foi utilizada neste trabalho (abordagem estatística), como:
 - Abordagens Estruturais e
 - Abordagens Espectrais.

REFERÊNCIAS

A. MATERKA; M. STRZELECKI, “**Texture Analysis Methods – A Review**,” Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels 1998. Disponível em: <http://www.eletel.p.lodz.pl/cost/pdf_1.pdf>. Acesso em: 17 jan. 2019.

About Kids Health. **Moles (acquired nevi)**. 2015. Disponível em: <<https://www.aboutkidshealth.ca/Article?contentid=2305&language=English>>. Acesso em: 13 jan. 2019.

ADESSOWIKI. Collaborative Scientific Writing and Programming. Cap1. **Processamento e análise de imagens digitais**. 2011. Disponível em: <<http://adessowiki.fee.unicamp.br/adesso/wiki/courseIA369O1S2011/Cap1/view/>>. Acesso em: 08 fev. 2019.

AGGARWAL, NAMITA; AGRAWAL, R. K. **First and second order statistics features for classification of magnetic resonance brain images**. Journal of Signal and Information Processing, v. 3, n. 02, p. 146, 2012. Disponível em: <https://file.scirp.org/pdf/JSIP20120200017_13782492.pdf>. Acesso em: 15 jan. 2019.

ALBAN, E. **Como a tecnologia pode salvar vidas**. Exame. 2015. Disponível em: <<https://exame.abril.com.br/tecnologia/como-a-tecnologia-pode-salvar-vidas/>>. Acesso em: 29 jan. 2019.

ALMEIDA, ANTÔNIO B. Informática Médica, **Usando o Computador para Processamento de Imagens Médicas**, 1998. Disponível em: <<http://www.informaticamedica.org.br/informaticamedica/n0106/imagens.htm>>. Acesso em: 30 jan. 2019.

ALVES, M. C. **Proc ttest: Comparando duas médias**, 2016. Disponível em: <http://cmq.esalq.usp.br/wiki/lib/exe/fetch.php?media=publico:syllabvs:lcf5759:proc_ttest-_teste_de_hipoteses.pdf>. Acesso em: 25 Janeiro 2019.

AMERICAN CANCER SOCIETY. **Signs and Symptoms of Melanoma Skin Cancer**, 2016. Disponível em: <<https://www.cancer.org/cancer/melanoma-skin-cancer/detection-diagnosis-staging/signs-and-symptoms.html>>. Acesso em: 28 jan. 2019.

APTE, A. P. **GLCM_global_features**, 2018. Disponível em: <https://github.com/cerr/CERR/wiki/GLCM_global_features>. Acesso em: 24 jan. 2019.

Banco de Imagens **Atlas Dermatológico**. Disponível em: <<http://www.atlasdermatologico.com.br/index.jsf>>. Acesso em: 29 jan. 2019.

Banco de Imagens **Atlas of Clinical Dermatology**. Disponível em: <<http://www.danderm-pdv.is.kkh.dk/atlas/index.html>>. Acesso em 15 jan. 2019.

Banco de Imagens **Dermnet NZ**. Disponível em: <<http://www.dermnetnz.org/sitemap.html>>. Acesso em: 10 jan. 2019.

BARRETO, A. S. **Modelos de Regressão: Teoria e Aplicação com o Programa Estatístico R**, Edição do Autor, 1 Edição, p. 109-114, Brasília 2011.

BOTELHO, G. **Processamento de Imagens no Auxílio a Detecção de Câncer de Pele utilizando redes Neurais Artificiais**. Universidade Federal de Goiás, 2007.

BUSSAB, W. O.; MORETTIN, P. A. – **Estatística Básica**. Atual Editora, São Paulo, 1988.

CONTI, F. **Teste de hipóteses e significância**, 2011. Disponível em: <<http://www.cultura.ufpa.br/dicas/biome/biotestes.htm>>. Acesso em: 23 jan. 2019.

FERREIRA, J. C.; PATINO, C. M. **What does the p value really mean?**. Jornal Brasileiro de Pneumologia, v. 41, n. 5, p. 485, 2015.

FIGUEIRA, C. V. **Modelos de regressão logística**. 2006. Dissertação (Mestrado em Matemática) - Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

FRIEDMAN, R. J. *et al.* **Performance de Dermatoscopistas Versus Visão Computadorizada**. Boletim dermatológico UNIFESP, São Paulo, n. 18, p. 1, 2008.

GALLOWAY, M. M. **Texture classification using gray level run length**. Computer graphics and image processing, v. 4, n. 2, p. 172-179, 1975.

GILTON, Z. **Entenda a diferença entre tomografia computadorizada e ressonância magnética**, 2011. Disponível em: <<https://gauchazh.clicrbs.com.br/comportamento/noticia/2011/12/entenda-a-diferenca-entre-tomografia-computadorizada-e-ressonancia-magnetica-3597315.html>>. Acesso em: 06 jan. 2019.

GONZALEZ, R. C.; WOODS, R. E. **Processamento Digital de Imagens**. 3ª. ed. São Paulo: Pearson, 2009.

GORDON, R. **Skin Cancer: an Overview of Epidemiology and Risk Factors**. 2013. *Seminars in Oncology Nursing*, v. 29, n. 3.

HAIR, Joseph F. *et al.* **Multivariate data analysis**. 7ª Edição. Pearson, 2014.

HARALICK, R. M. **Statistical and Structural Approaches to texture**. [S.l.]: *Proceedings of the IEEE*, v. 67, 1979. p. 786-804.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. **Textural features for images classification**. 6ª. ed. [S.l.]: *IEEE transactions on systems, man, and cybernetics*, v. 3, 1973. 610-621 p.

INCA. **Câncer de pele melanoma**, 2018a. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-melanoma>>. Acesso em: 14 jan. 2019.

INCA. **Câncer de Pele Não Melanoma**, 2018b. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pele_ao_melanoma>. Acesso em: 15 jan. 2019.

INCA. **Como surge o câncer**. 2018e. Disponível em: <<https://www.inca.gov.br/como-surge-o-cancer>>. Acesso em: 28 jan. 2019.

INCA. **Estimativa 2018: Incidência de Câncer no Brasil**, 2018c. Disponível em: <<http://www1.inca.gov.br/estimativa/2018/casos-taxas-brasil.asp>>. Acesso em: 15 jan. 2019.

INCA. **O que é câncer**. 2018d. Disponível em: <<https://www.inca.gov.br/o-que-e-cancer>>. Acesso em: 30 jan. 2019.

ITO, R. H.; KIM, H. Y.; SALCEDO, W. J. **Classificação de texturas invariante a rotação usando matriz de co-ocorrência**. In: 8th International Information and Telecommunication Technologies Symposium. 2009. p. 1-6.

KAVITHA, J. C.; SURULIANDI A.; D. NAGARAJAN. **Melanoma detection in dermoscopic images using global and local feature extraction**. *Int. J. Multimedia Ubiquit. Eng*, v. 12, n. 5, p. 19-28, 2017.

KAZMIER, LEONARD J. **Estatística Aplicada à Administração e Economia-4**. Bookman Editora, 2006.

MAGALHÃES, M. N.; LIMA, A. C. P. **Noções de probabilidade e estatística**. São Paulo: Editora da Universidade de São Paulo, 2002.

MEDEIROS, G. C. F.. **Uso de texturas para o acompanhamento da evolução do tratamento de úlceras dermatológicas**. 2001. Dissertação (Mestrado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2001. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18133/tde-26042001-125441/pt-br.php>>. Acesso em: 28 jan. 2019.

MINITAB. **O que é um teste de hipótese?**, c2017a. Disponível em: <<https://support.minitab.com/pt-br/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/what-is-a-hypothesis-test/>>. Acesso em: 13 jan. 2019.

MINITAB. **Sobre as hipóteses nula e alternativa**, c2017b. Disponível em: <<https://support.minitab.com/pt-br/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/null-and-alternative-hypotheses/>>. Acesso em: 13 jan. 2019.

MINITAB Inc. Statistical Software Data Analysis Software. Version 18.1, 2017.

MOHAMAD, Nor Amira et al. **Multinomial logistic regression modelling of stress level among secondary school teachers in Kubang Pasu District, Kedah**. In: AIP Conference Proceedings. AIP Publishing, 2016. p. 060018. Disponível em: <<https://aip.scitation.org/doi/pdf/10.1063/1.4954623>>. Acesso em: 13 fev. 2019.

MONTEIRO, J. **O QUE SÃO BIOMARCADORES?**, c2019. Disponível em: <<https://www.roche.pt/corporate/index.cfm/farmaceutica/ensaios-clinicos-profissionais-de-saude/biomarcadores/o-que-sao-biomarcadores/>>. Acesso em: 28 jan. 2019.

Moreira, L. M. C. **Tamanho e resolução da imagem**. Universidade de Porto. Disponível em: <https://paginas.fe.up.pt/~ee03037/tmp/api11/imagem_resol.pdf>. Acesso em: 08 jan. 2019.

NAKANO, S. Melanoma Brasil. **Melanoma será 19º tumor mais incidente no mundo, segundo a OMS**, 2018. Disponível em: <<http://www.melanomabrasil.org/melanoma-sera-19o-tumor-mais-incidente-no-mundo-segundo-a-oms/>>. Acesso em 30 jan. 2019.

OLIVEIRA, L. F. *et al.* **Extração de Características de Forma utilizando matriz de co-ocorrência e Atributos de Haralick**. In: Workshop de Visão Computacional (WVC). Programa de Pós-graduação em Ciência e Tecnologia de Sementes, Universidade Federal de Pelotas, Pelotas. 2012.

OLIVEIRA, M. S. **Análise de Textura em Imagens Cerebrais: aplicações em Acidente Vascular Cerebral Isquêmico, Epilepsia Mioclônica Juvenil, Doença de Machado-Joseph, Déficit Cognitivo Leve e Doença de Alzheimer.** Universidade Estadual de Campinas, 2010.

ONCOGUIA. **Quais os fatores de risco para o câncer de pele melanoma?**, 2017. Disponível em: <<http://www.oncoguia.org.br/conteudo/quais-os-fatores-de-risco-para-o-cancer-de-pele-melanoma/4555/699/>>. Acesso em: 26 jan. 2019.

OPAS. Organização Pan-Americana de Saúde, **Folha infotmativa – Câncer**, 2018. Disponível em: <https://www.paho.org/bra/index.php?option=com_content&view=article&id=5588:folha-informativa-cancer&Itemid=839>. Acesso em: 30 jan. 2019.

OSWALDO CRUZ - Centro Especializado em Oncologia. **Câncer de Pele.** c2017. Disponível em: <<http://centrodeoncologia.org.br/tudo-sobre-cancer/cancer-de-pele/>>. Acesso em: 30 jan. 2019.

PATHAK, Biswajit; BAROOAH, Debajyoti. Texture analysis based on the gray-level co-occurrence matrix considering possible orientations. **International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering**, v. 2, n. 9, p. 4206-4212, 2013.

PEDRINI, Hélio. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações.** Editora Thomson, 2008.

PORTAL ACTION. **Técnicas não-paramétricas.** 2017?a. Disponível em: <<http://www.portalaction.com.br/tecnicas-nao-parametricas>>. Acesso em: 29 jan. 2019.

PORTAL ACTION. **Teste de Wilcoxon-Mann-Whitney.** 2017?b. Disponível em: <<http://www.portalaction.com.br/tecnicas-nao-parametricas/teste-de-wilcoxon-mann-whitney-amstras-independentes>>. Acesso em: 29 jan. 2019.

REIS, E. **Estatística Descritiva** (2ª ed.). Lisboa: Edições Sílabo, 1994.

REIS, G. M.; RIBEIRO JUNIOR, J. I. **Comparação de testes paramétricos e não paramétricos aplicados em delineamentos experimentais.** Simpósio Acadêmico de Engenharia de Produção, v. 3, p. 2007-3, 2007.

ROCHA, A. R.; LEITE, N. J. Classificação de texturas a partir de vetores de atributos e função de distribuição de probabilidades. **Campinas, SP. Universidade Estadual de Campinas**, 2002.

ROSENHOLTZ, Ruth. **Texture perception**. Oxford handbook of perceptual organization, p. 167-186, 2014.

SANTOS, R. **Prevenção do Câncer**, c2019. Disponível em: <http://www.prevencaodecancer.com.br/principais_tipos/melanoma.php>. Acesso em: 21 jan. 2019.

SILVA, G. L. **Modelos logísticos para dados binários**. 1992. 118f. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1992.

SILVA, J. **Análise de Correlação – Métodos Quantitativos e Econometria**. 2015. Economia e Finanças. Disponível em: <https://pt.slideshare.net/sahsilva988/anlise-de-correlao?from_action=save>. Acesso em: 22 fev. 2019.

SKLANSKY, J. **Image Segmentation and Feature Extracrion**. IEE Transction on Systems, Man and Cybernetics, vol. 13, 1978.

SOUZA, E. C. **Análise de influência local no modelo de regressão logística**. 2006. 101f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2006.

SPADARI, A. CCM. **O formato BMP**, 2017. Disponível em: <<https://br.ccm.net/contents/720-o-formato-bmp>>. Acesso em: 25 jan. 2019.

SRINIVASAN, G. N.; SHOBHA, G. **Statistical texture analysis**. In: Proceedings of world academy of science, engineering and technology. p. 1264-1269, 2008.

SZCZYPIŃSKI, P. M. *et al.* **MaZda – The Software Package for Textural Analysis of Biomedical Images**. **Computers in Medical Activity**, Berlin, v. 65, 2009. Disponível em: <http://www.eletel.p.lodz.pl/pms/Publikacje/2007_ICCMA_Lodz.pdf>. Acesso em: 21 dez. 2018.

Technical University of Lodz. **Brief history of MaZda**, 2010. Disponível em: <<http://www.eletel.p.lodz.pl/programy/mazda/index.php?action=history>>. Acesso em: 20 dez. 2018.

TEOTÔNIO, P. **TIPOS DE BITMAPS**. 2017. Disponível em: <<https://docs.microsoft.com/pt-br/dotnet/framework/winforms/advanced/types-of-bitmaps>>. Acesso em: 14 jan. 2019.

VIANA, A. **3 principais tipos de Câncer de pele**. c2017. Disponível em: <<https://www.tuasaude.com/tipos-de-cancer-de-pele/>>. Acesso em: 29 jan. 2019.

VIEIRA, J.; SILVEIRA, P. S.; SIQUEIRA, J.P. **Os Testes Estatísticos (em Medicina)**. Atheneu, 2017. Disponível em: <https://www.researchgate.net/publication/326522403_OS_TESTES_ESTATISTICO_S_EM_MEDICINA>. Acesso em: 12 jan. 2019.

VILELA JUNIOR, G. D. B. V. **ESTATÍSTICA: Teste t *student***, 2001. Disponível em: <http://www.cpaqv.org/estatistica/teste_t.pdf>. Acesso em: 25 jan. 2019.

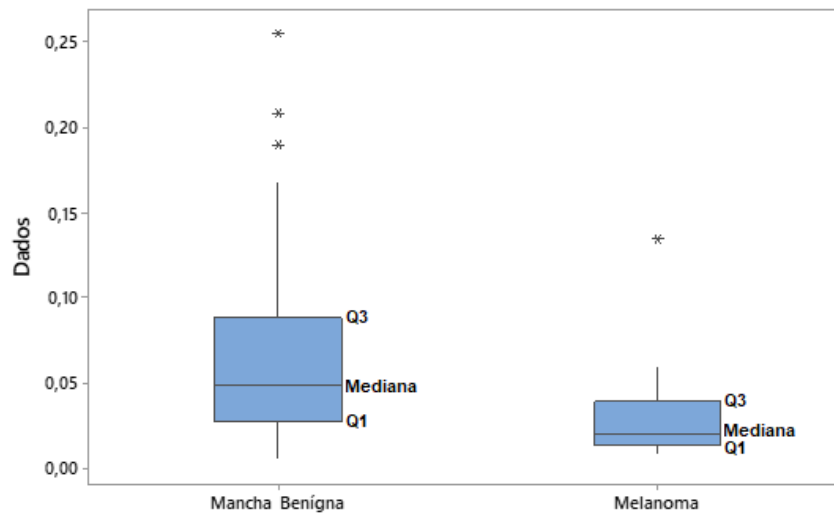
VOIGT, Holger. **The ‘digital eye’ at the threshold of cancer diagnosis, Expert Review of Anticancer Therapy**, 2:5, 479-480, 2002.

WITTE, Robert S.; WITTE, John S. **Estatística**. Livros Técnicos e Científicos, 2005.

ZAYED, Nourhan; ELNEMR, Heba A. **Statistical analysis of haralick texture features to discriminate lung abnormalities**. Journal of Biomedical Imaging, v. 2015, p. 12, 2015.

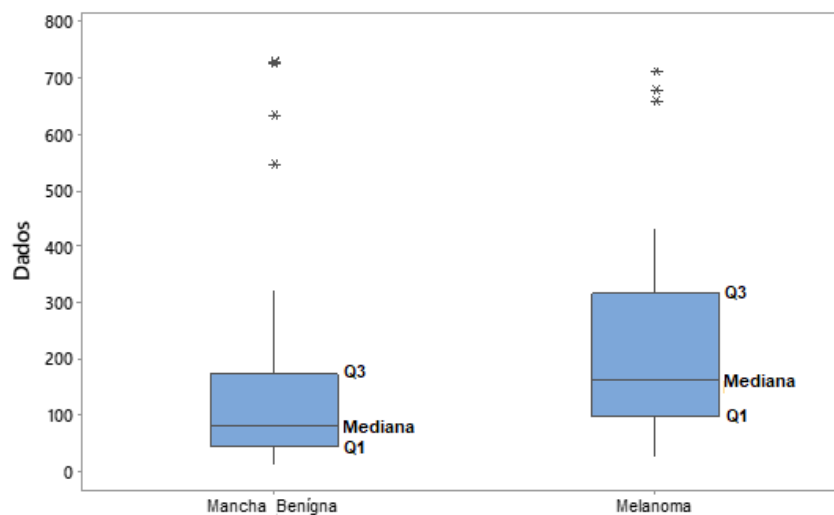
APÊNDICE A – *BOX-PLOT* DOS PARÂMETROS DE TEXTURA UTILIZADOS NA REGRESSÃO LOGÍSTICA

Figura A1 - *BoxPlot* do parâmetro de textura Variância para o grupo de imagens de manchas benignas *versus* melanomas.



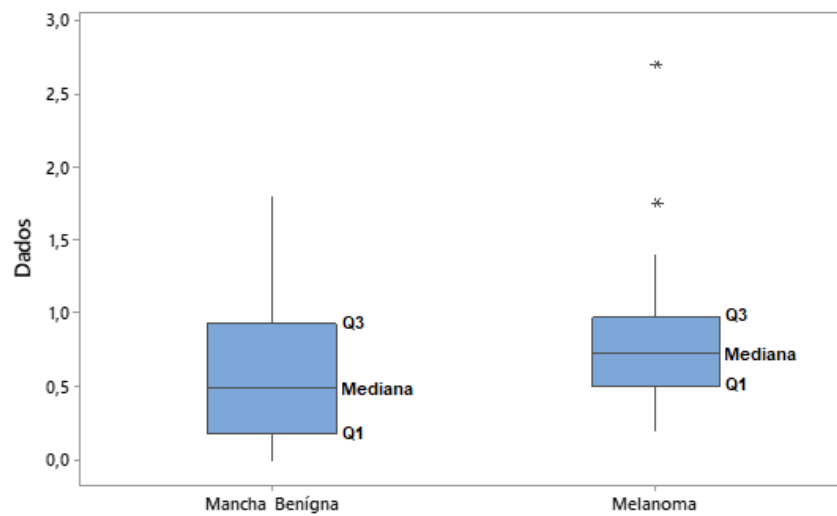
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A2 - *BoxPlot* do parâmetro de textura Segundo Momento Angular para o grupo de imagens de manchas benignas *versus* melanomas.



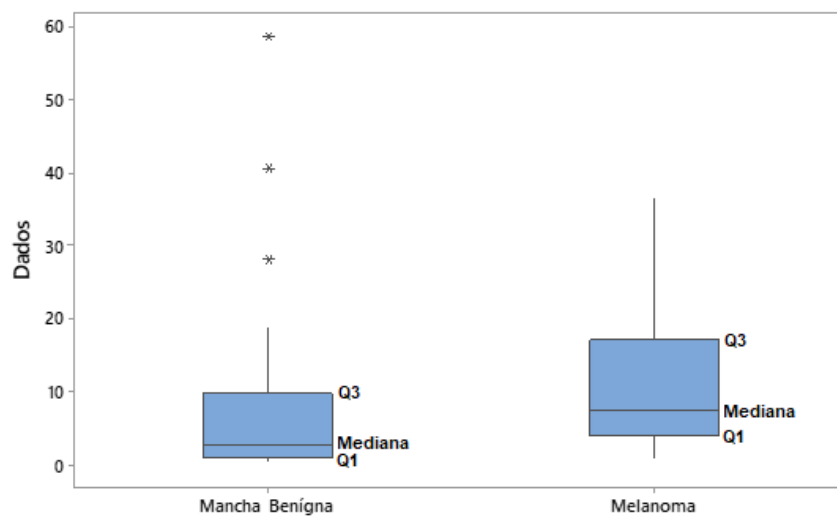
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A3 - *BoxPlot* do parâmetro de textura Segundo Momento Angular para o grupo de imagens de manchas benignas *versus* melanomas.



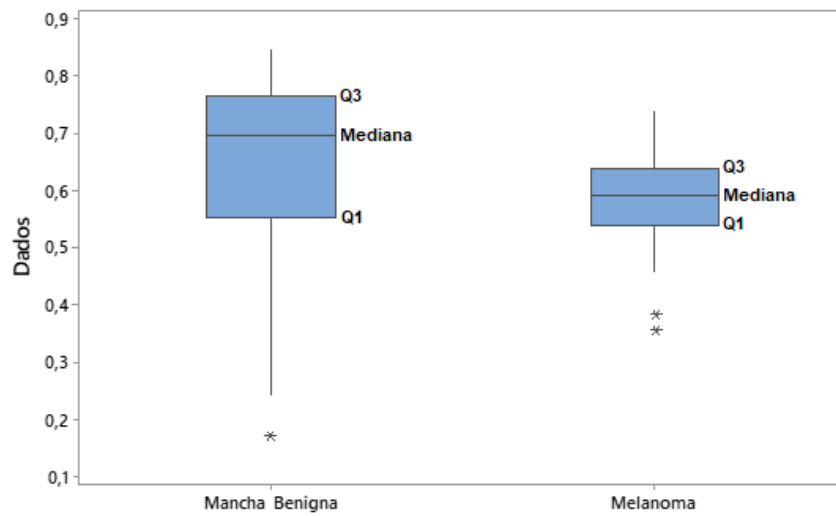
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A4 - *BoxPlot* do parâmetro de textura Contraste para o grupo de imagens de manchas benignas *versus* melanomas.



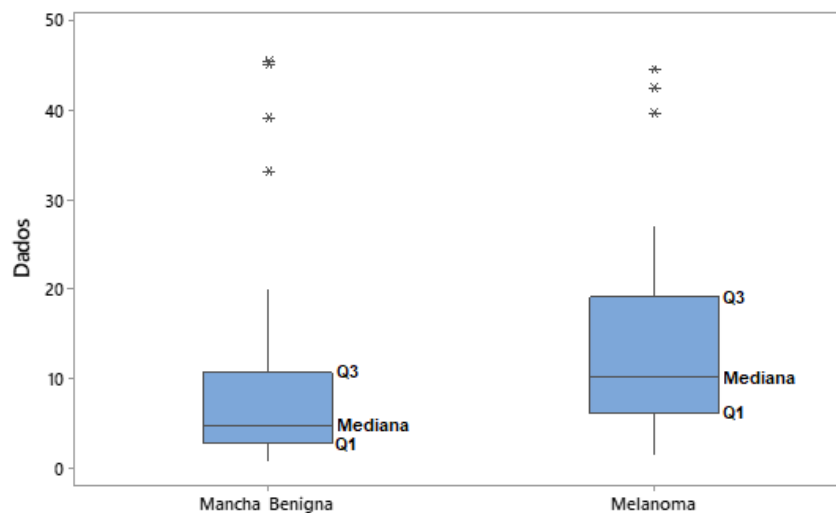
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A5 - *BoxPlot* do parâmetro de textura Correlação para o grupo de imagens de manchas benignas *versus* melanomas.



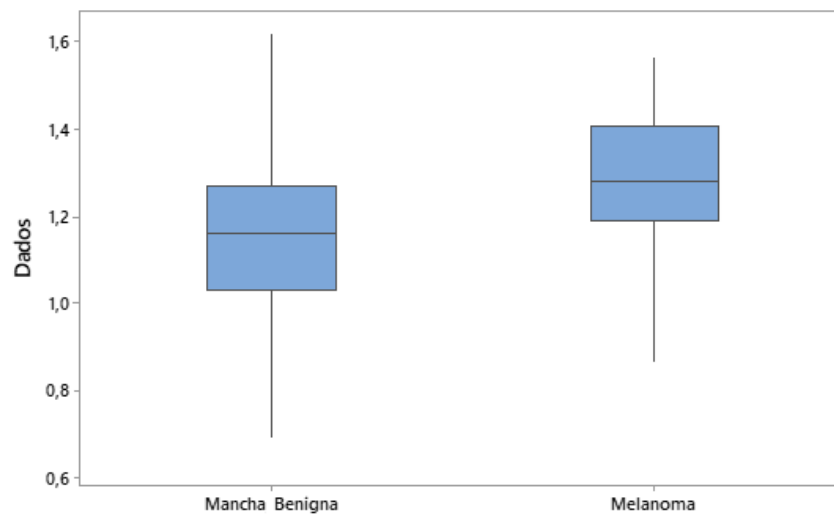
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A6 - *BoxPlot* do parâmetro de textura Soma dos Quadrados para o grupo de imagens de manchas benignas *versus* melanomas.



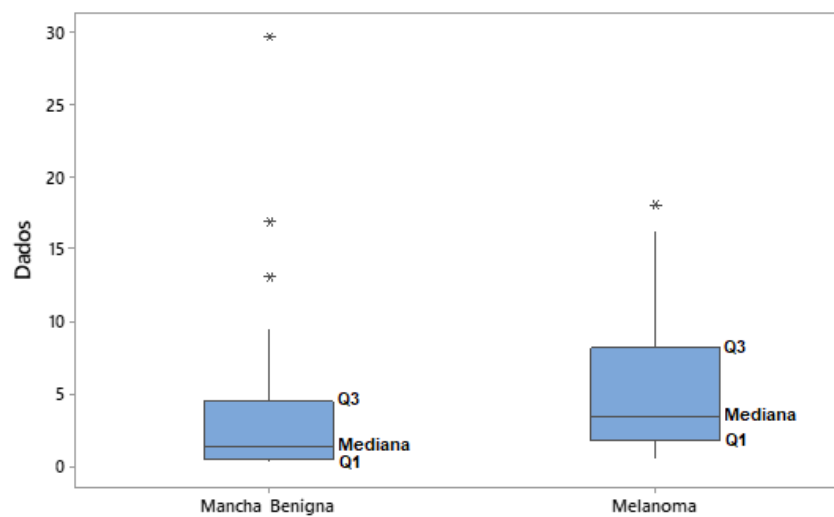
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A7 - *BoxPlot* do parâmetro de textura Entropia da Soma para o grupo de imagens de manchas benignas *versus* melanomas.



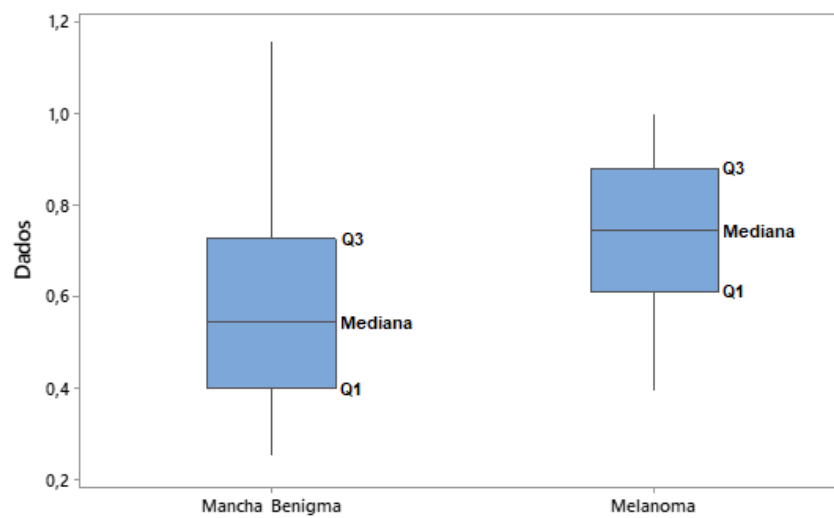
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A8 - *BoxPlot* do parâmetro de textura Entropia da Soma para o grupo de imagens de manchas benignas *versus* melanomas.



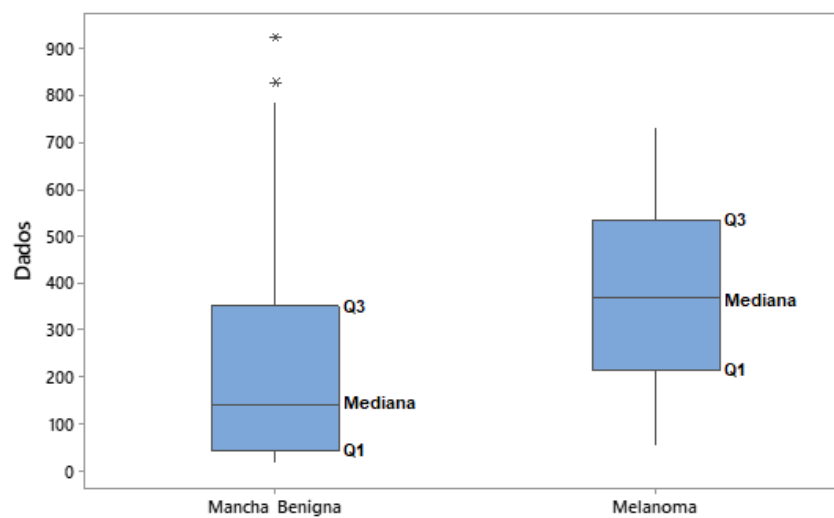
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A9 - *BoxPlot* do parâmetro de textura Entropia da Diferença para o grupo de imagens de manchas benignas *versus* melanomas.



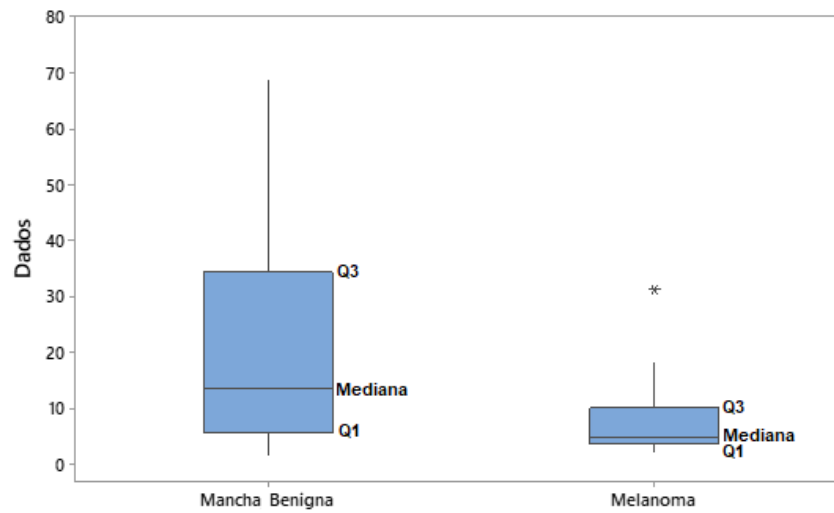
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A10 - *BoxPlot* do parâmetro de textura NUCE Hor. para o grupo de imagens de manchas benignas *versus* melanomas.



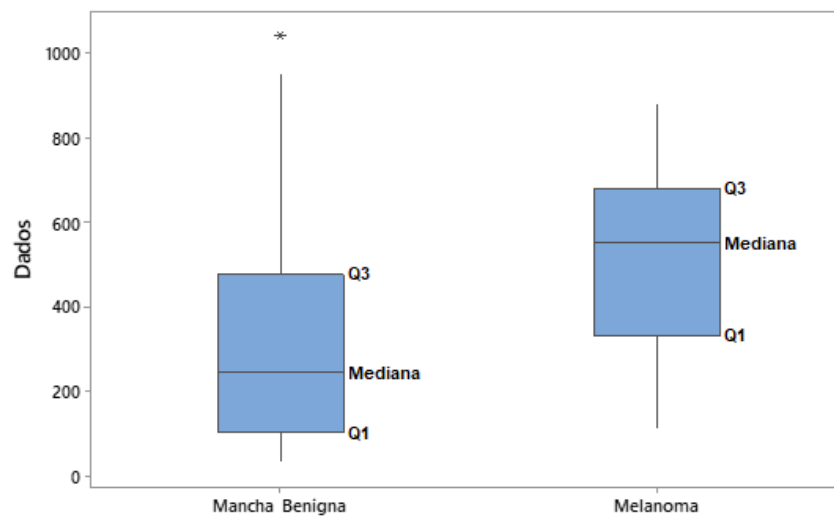
Nota: Q1= 1º Quartil, Q3 = 3º Quartil e Hor.= Horizontal.

Figura A11 - *BoxPlot* do parâmetro de textura MIELP Hor. para o grupo de imagens de manchas benignas *versus* melanomas.



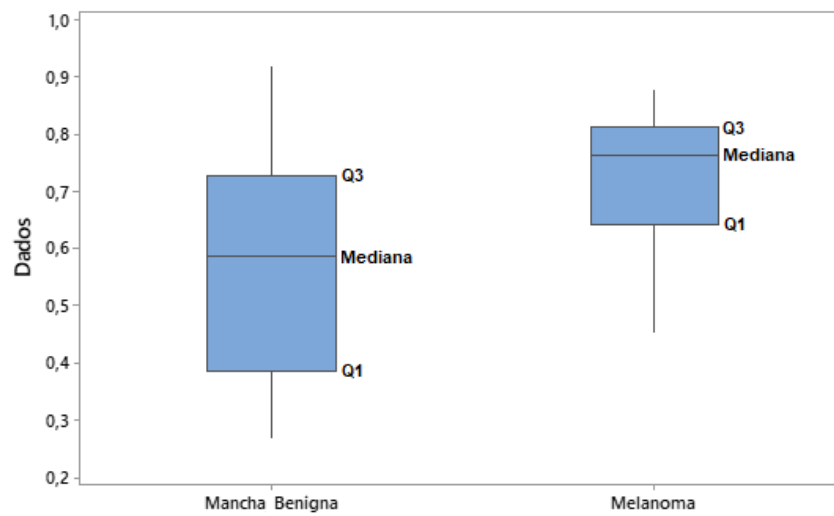
Nota: Q1= 1º Quartil, Q3 = 3º Quartil e Hor.= Horizontal.

Figura A12 - *BoxPlot* do parâmetro de textura NUCE 45° para o grupo de imagens de manchas benignas *versus* melanomas.



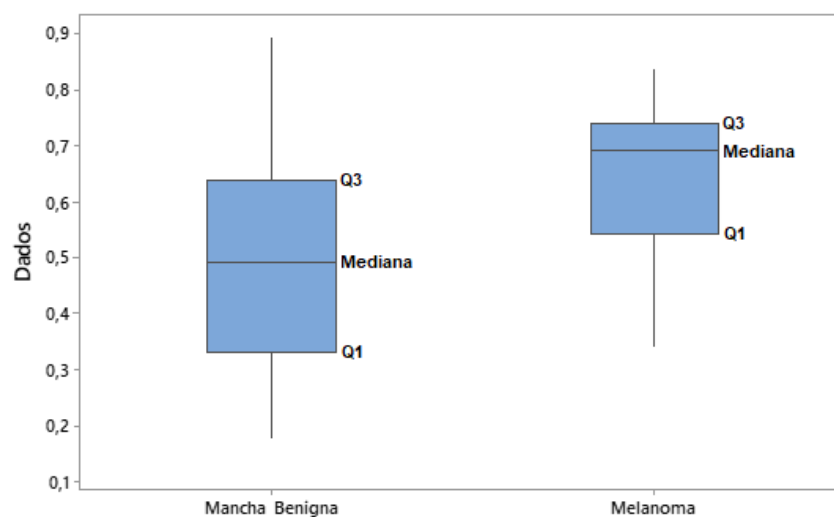
Nota: Q1= 1º Quartil, Q3 = 3º Quartil.

Figura A13 - *BoxPlot* do parâmetro de textura MIECP 45° para o grupo de imagens de manchas benignas *versus* melanomas.



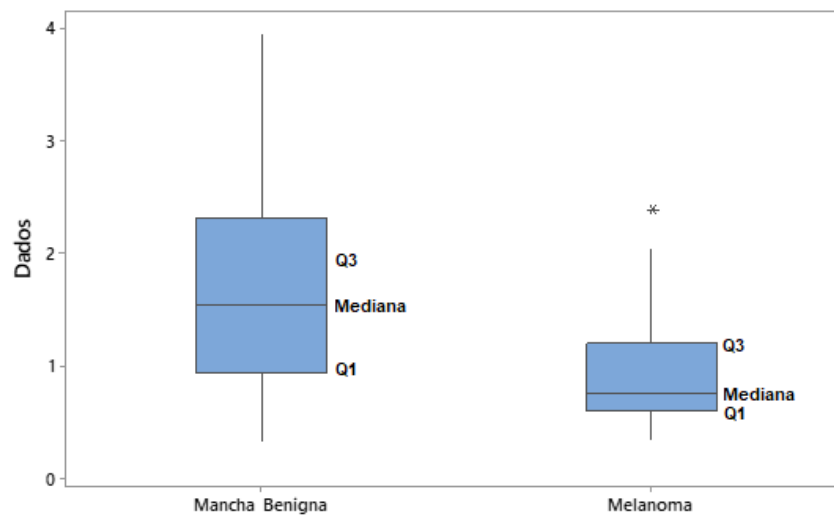
Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A14 - *BoxPlot* do parâmetro de textura FIE 45° para o grupo de imagens de manchas benignas *versus* melanomas.



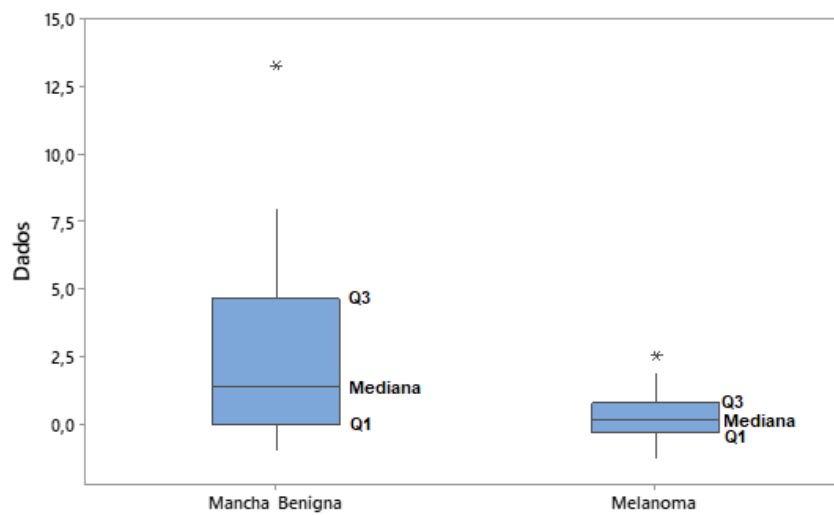
Nota: Q1= 1º Quartil, Q3 = 3º Quartil.

Figura A15 - *BoxPlot* do parâmetro de textura Assimetria do Gradiente para o grupo de imagens de manchas benignas *versus* melanomas.



Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

Figura A16 - *BoxPlot* do parâmetro de textura Curtose do Gradiente para o grupo de imagens de manchas benignas *versus* melanomas.



Nota: Q1= 1º Quartil e Q3 = 3º Quartil.

APÊNDICE B – TABELAS DE PARÂMETROS EXTRAÍDOS DA IMAGENS-TESTES

Tabela B1 – Média das dezesseis regiões de interesse das imagens-teste de melanomas.

Parâmetro	Imagem 1	Imagem 2	Imagem 3	Imagem 4	Imagem 5
	Médias				
Variância	270,646	878,290	559,597	370,909	268,524
Assimetria	0,505178	1,08068	1,30826	0,710892	0,231025
Segundo Momento Angular (ASM)	0,0089321	0,0147555	0,0184845	0,0154749	0,0082450
Contraste	17,1529	41,1803	20,9439	13,4949	15,1300
Correlação	0,475358	0,622801	0,682856	0,703234	0,490874
Soma dos Quadrados	16,9284	54,1299	34,5826	22,8515	16,1564
Momento da Diferença Inversa	0,309264	0,340468	0,394865	0,403718	0,298151
Entropia das Somas	1,41947	1,54476	1,46692	1,45648	1,42088
Entropia	2,22754	2,29640	2,13044	2,10965	2,24889
Variância da Diferença	6,76949	17,4617	10,4889	5,57051	6,19449
Entropia da Diferença	0,910217	0,941333	0,869854	0,789996	0,916546
NUCE Hor.	637,847	679,122	446,803	329,482	812,948
MIELP Hor.	2,74903	3,76507	5,45139	6,60019	1,82585
NUCE 45°	834,748	785,276	609,382	597,769	904,650
MIECP 45°	0,864407	0,839192	0,790771	0,781115	0,883799
FIE 45°	0,817434	0,770300	0,698321	24,0296	0,847386
Assimetria do Gradiente Absoluto	0,271506	0,665459	0,737885	0,451148	0,583870
Curtose do Gradiente Absoluto	-0,202022	-0,134235	-0,0520409	-0,952344	1,02193

Nota: ASM = *Angular Second Moment* e Hor. = Horizontal.

Tabela B2– Média das dezesseis regiões de interesse das imagens-teste de manchas benignas.

Parâmetro	Imagem 6	Imagem 7	Imagem 8	Imagem 9	Imagem 10
	Médias				
Variância	138,088	113,819	186,164	336,582	412,106
Assimetria	1,72737	0,333996	0,625369	0,567349	0,482366
Segundo Momento Angular (ASM)	0,0340212	0,0392611	0,0143022	0,0142858	0,0076979
Contraste	9,73475	2,29490	11,9094	15,4403	21,8672
Correlação	0,439334	0,822490	0,485245	0,629267	21,8672
Soma dos Quadrados	8,49585	6,89177	11,8748	20,7891	0,566253
Momento da Diferença Inversa	0,476609	0,601482	0,364901	0,378792	25,3872
Entropia das Somas	1,20581	1,24334	1,33067	1,43659	0,308858
Entropia	1,78202	1,60825	2,04803	2,12993	1,51361
Variância da Diferença	5,43538	0,937042	4,71174	5,86707	2,32633
Entropia da Diferença	0,738464	0,502527	0,815218	0,811063	8,99203
NUCE Hor.	400,025	100,264	492,842	597,658	0,931733
MIELP Hor.	5,26649	16,9654	3,93961	3,61949	657,530
NUCE 45°	555,260	127,558	686,390	681,066	2,81559
MIECP 45°	0,760720	0,440718	0,819507	0,810864	743,885
FIE 45°	0,684600	0,368650	0,753073	0,735111	0,837703
Assimetria do Gradiente Absoluto	1,57796	1,45461	0,202009	0,377221	0,779432
Curtose do Gradiente Absoluto	4,06719	0,366611	1,12373	-0,877205	-0,391050

Nota: ASM = *Angular Second Moment* e Hor. = Horizontal.

APÊNDICE C - CORRELAÇÃO ENTRE AS VARIÁVEIS USADAS NA MODELAGEM DE REGRESSÃO LOGÍSTICA

	Variação	Assimetria	ASM	Contraste	Correlação	Soma dos Quadrados	Momento da Diferença Inversa	Entropia da Soma	Entropia	Variação da Diferença	Entropia da Diferença	NUCE Hor.	MIELP Hor.	NUCE 45°	MIELP 45°	FIE 45°	Assimetria do Gradiente	Curtose do Gradiente
Variação	1																	
Assimetria	0.0285826	1																
ASM	-0.028093	-0.1131363	1															
Contraste	0.1581511	0.04591538	-0.11005	1														
Correlação	-0.139736	-0.1386147	0.101169	-0.15697	1													
Soma dos Quadrados	0.0977873	0.0289089	-0.09679	0.0612204	-0.043357	1												
Momento da Diferença Inversa	-0.104514	-0.1956655	0.163039	-0.192269	0.164236	-0.0049788	1											
Entropia da Soma	0.1075887	0.12828912	-0.18799	0.1021381	-0.1103003	0.19641737	-0.13736	1										
Entropia	0.0673617	0.07994637	-0.10675	0.1001341	-0.0314796	0.06722512	-0.09842	0.1781635	1									
Variação da Diferença	0.0347185	0.1876063	-0.08943	0.1244935	-0.1589709	0.03777904	-0.167	0.0018732	0.180014	1								
Entropia da Diferença	0.1629287	0.07456827	-0.12197	0.0523654	-0.1394583	0.16386791	-0.18746	0.1982017	0.198071	0.123855	1							
NUCE Hor.	0.1851019	0.11689313	-0.16782	0.0050361	-0.1628916	0.0881631	-0.19597	0.1094754	0.002316	0.0349845	0.0525379	1						
MIELP Hor.	-0.007337	-0.0964396	0.134583	-0.121797	0.0055241	-0.0057599	0.082785	-0.081306	-0.12196	-0.004073	-0.12247	-0.197825	1					
NUCE 45°	0.077308	0.02086404	-0.04582	0.1193793	-0.1449475	0.17876184	-0.03998	0.1559761	0.039903	0.1959418	0.073276	0.0010681	-0.164414	1				
MIELP 45°	0.1214859	0.09910075	-0.15856	0.0038526	-0.08251	0.12095671	-0.19779	0.189169	0.053041	0.1826788	0.1207141	0.0990138	-0.011455	0.0424882	1			
FIE 45°	0.1290462	0.19184961	-0.104	0.1358217	-0.1036332	0.02932778	-0.0867	0.1920643	0.160482	0.0136126	0.1977342	0.1292495	-0.192467	0.1745015	0.093825	1		
Assimetria do Gradiente	-0.041114	-0.1772901	0.120998	-0.074209	0.0819194	-0.0086833	0.124219	-0.005039	-0.11863	-0.032561	-0.004007	-0.13902	0.1531	-0.02462	-0.187515	-0.03661	1	
Curtose do Gradiente	-0.136115	-0.1285768	0.015759	-0.174055	0.1820436	-0.1333351	0.007988	-0.128096	-0.01571	-0.147371	-0.076065	-0.048018	0.078232	-0.123921	-0.074973	-0.15751	0.0518122	1

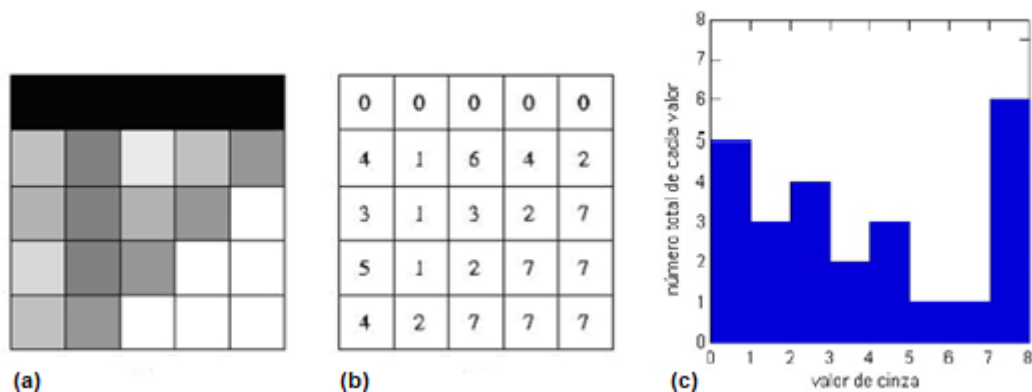
ANEXO A – EQUAÇÕES DAS ESTATÍSTICAS DE TEXTURA

1. Estatística de primeira ordem:

i. Histograma:

Segundo Oliveira (2010), o histograma de uma imagem fornece a quantidade de pixels que possuem um certo valor de cinza como pode ser visto de acordo com a Figura A1 (a)-(c).

Figura A1 - (a) Imagem sobre a qual será calculado o histograma, (b) representação numérica da imagem digital e (c) o histograma.



Fonte: Adaptada Oliveira (2010).

Seja a variável i representando o nível de cinza da região da imagem como exemplo da Figura 12(a). O histograma de primeira ordem $P(i)$ é definido por (AGGARWAL, 2012):

$$P(i) = \frac{\text{Número de pixels com intensidade de cinza } i}{\text{Total de pixels na região}}, \quad (\text{A.1})$$

onde

$$i = 1, 2, \dots, L,$$

em que L é o número de possíveis níveis de cinza.

De acordo com $P(i)$, as características do histograma podem ser definidas como a seguir (AGGARWAL,2012) (MAZDA USER'S MANUAL, 2006):

a) Média é medida da intensidade média do histograma e é dada por:

$$Média = \mu = \sum_{i=1}^L iP(i). \quad (A.2)$$

b) Variância:

$$Variância = \sigma^2 = \sum_{i=1}^L (i - \mu)^2 P(i). \quad (A.3)$$

c) A assimetria é uma medida do grau de assimetria do histograma em torno da média e é dada por:

$$Assimetria = \sigma^{-3} \sum_{i=1}^L (i - \mu)^3 P(i). \quad (A.4)$$

d) A Curtose é uma medida da nitidez do histograma e é dada por:

$$Curtose = \sigma^{-4} \sum_{i=1}^L (i - \mu)^4 P(i) - 3. \quad (A.5)$$

2. Estatística de segunda ordem:

i. Gradiente Absoluto

Segundo Oliveira (2010), o gradiente de uma imagem é uma medida das variações espaciais dos níveis de cinza através da imagem. Se em um ponto da imagem o nível de cinza varia abruptamente de preto para branco, o valor do gradiente neste ponto é alto. Porém, se varia suavemente do cinza escuro para o cinza claro, por exemplo, o valor do gradiente será baixo. O gradiente pode ser negativo ou positivo, dependendo se o nível de cinza varia do tom mais escuro para o mais claro ou vice-versa.

Como, em geral, o interesse é em saber simplesmente se a transição abrupta ou suave, usa-se o valor absoluto do gradiente. Os parâmetros derivados do gradiente absoluto são: Média do Gradiente, Variância do Gradiente, Assimetria do Gradiente e Curtose do Gradiente.

Para o cálculo dos parâmetros do gradiente absoluto da imagem, o valor do pixel vizinho à $x(i, j)$ pode ser definido como:

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>
<i>K</i>	<i>L</i>	$x(i, j)$	<i>N</i>	<i>O</i>
<i>P</i>	<i>Q</i>	<i>R</i>	<i>S</i>	<i>T</i>
<i>U</i>	<i>V</i>	<i>W</i>	<i>Y</i>	<i>Z</i>

Baseado no seu vizinho, o valor do gradiente absoluto ($VABS(i, j)$) é calculado para cada pixel:

$$VABS(i, j) = \sqrt{(R - H)^2 + (N - L)^2} \quad (A.6)$$

Sendo $ABSV$ uma matriz de M elementos, as características do gradiente são definidas como a seguir:

a) Média do Gradiente:

$$GrMed = \frac{1}{M} \sum_{i,j} VABS(i,j). \quad (A.7)$$

b) Variância do Gradiente:

$$GrVar = \frac{1}{M} \sum_{i,j} (VABS(i,j) - GrMed)^2. \quad (A.8)$$

c) Assimetria do Gradiente:

$$GrAss = \frac{1}{(\sqrt{GrVar})^3} \frac{1}{M} \sum_{i,j} (VABS(i,j) - GrMed)^3. \quad (A.9)$$

d) Curtose do Gradiente:

$$GrCurt = \frac{1}{(\sqrt{GrVar})^4} \frac{1}{M} \sum_{i,j} (VABS(i,j) - GrMed)^4 - 3. \quad (A.10)$$

ii. Matriz de Co-ocorrência dos níveis de cinza

De acordo com MaZda User's Manual (2006), Materka e Strzelecki (1998) e Haralick, Shanmugam e Dinsten (1973), o histograma de segunda ordem é definido como a matriz de co-ocorrência $h_{d\theta}(i,j)$. Quando dividido pelo número total de pixels vizinhos $R(d,q)$ na ROI (Região de Interesse), essa matriz se torna a estimativa da probabilidade conjunta, $P_{d\theta}(i,j)$, de dois pixels, a uma distância d separada ao longo de uma determinada direção q tendo particulares (co-ocorrentes) valores i e j . Formalmente, dada a imagem $f(x,y)$ com um conjunto de L níveis discretos de intensidade, a matriz $h_{d\theta}(i,j)$ é definida de tal forma que sua entrada (i,j) é igual ao número de vezes que $f(x_1, y_1) = i$ e $f(x_2, y_2) = j$, onde

$$(x_2, y_2) = (x_1, y_1) + (d\cos\theta, d\sin\theta). \quad (A.11)$$

Isto produz uma matriz quadrada de dimensão igual ao número de níveis de intensidade na imagem, para cada distância d e orientação q ;

As distâncias $d = 1, 2, 3, 4$ e 5 pixels com ângulos $q; = 0^\circ, 45^\circ, 90^\circ$ e 135° são considerados. A redução do número de níveis de intensidade (por quantificação para menos níveis de intensidade) ajuda a aumentar a velocidade de computação, com alguma perda de informação textural.

Os parâmetros de Haralick derivados da matriz de co-ocorrência são:

- Segundo Momento Angular,
- Contraste,
- Correlação,
- Soma dos Quadrados,
- Momento da Diferença Inversa,
- Entropia,
- Média da Soma,
- Variância da Soma,
- Entropia da Soma,
- Variância da Diferença e
- Entropia da Diferença.

Estes parâmetros são definidos pelas equações que seguem, onde μ_x, μ_y e σ_x, σ_y denotam a média e os desvios padrões das somas de linha e coluna da matriz de co-ocorrência, respectivamente (relacionadas a as distribuições marginais $P_x(i)$ e $P_y(j)$).

$$\mu_x = \sum_i \sum_j iP_d(i, j), \quad (\text{A.12})$$

$$\mu_y = \sum_i \sum_j jP_d(i, j), \quad (\text{A.13})$$

$$\sigma_x = \sqrt{\sum_i \sum_j (i - \mu_x)^2 P_d(i, j)}, \quad (\text{A.14})$$

$$\sigma_y = \sqrt{\sum_i \sum_j (i - \mu_y)^2 P_d(i, j)}. \quad (\text{A.15})$$

a) **Energia:** derivada do Segundo Momento Angular (ASM) também chamada, mede a uniformidade local dos níveis de cinza. Quando os pixels forem muito similares, o ASM ou Energia será muito alta.

$$Energia = \sqrt{ASM}, \quad (\text{A.16})$$

$$ASM = \sum_i \sum_j P_d^2(i, j). \quad (\text{A.17})$$

b) **Contraste:** medida de intensidade ou variação do nível de cinza entre o pixel de referência e o pixel vizinho. Este parâmetro relaciona o brilho ou escurecimento da imagem, entre uma área de interesse e sua vizinhança. Um alto contraste reflete em uma alta diferença de intensidade na Matriz de Co-ocorrência.

$$Contraste = \sum_i \sum_j (i - j)^2 P_d(i, j). \quad (\text{A.18})$$

c) **Correlação:** representa a dependência linear dos valores de níveis de cinza na matriz de co-ocorrência. Imagens mais ruidosas são menos correlacionadas.

$$Correlação = \sum_i \sum_j P_d(i, j) \frac{(i - \mu_x)(j - \mu_y)}{\sigma_x \sigma_y}. \quad (\text{A.19})$$

d) **Soma dos Quadrados:** representa a dispersão da distribuição dos níveis de cinza, em relação à média.

$$Soma\ dos\ Quadrados = \sum_{i=1}^L \sum_{j=1}^L (i - \mu_x)^2 P(i, j). \quad (\text{A.20})$$

e) Homogeneidade, também chamado de Momento da Diferença Inversa: representa a suavidade ou homogeneidade da distribuição dos níveis de cinza da imagem.

$$\text{Momento da Diferença Inversa} = \sum_i \sum_j \frac{1}{1 + (i - j)^2} P_d(i, j). \quad (\text{A.21})$$

f) Média da Soma: representa a média da distribuição da soma dos níveis de cinza.

$$\text{Média da Soma} = \sum_{i=2}^{2L} i P_{x+y}(i) = \mu_{x+y}, \quad (\text{A.22})$$

onde,

$$P_{x+x}(k) = \sum_{i=1}^L \sum_{\substack{j=1 \\ i+j=k}}^L P(i, j), \quad (\text{A.23})$$

$$k = 2, 3, \dots, 2L.$$

g) Variância da Soma: medida da dispersão da soma dos níveis de cinza, em relação à média.

$$\text{Variância da Soma} = \sum_{i=1}^{2L} [i - (\text{Média da Soma})]^2 P_{x+y}(i). \quad (\text{A.24})$$

h) Entropia: mede o grau de desordem entre os pixels na imagem.

$$\text{Entropia} = - \sum_i \sum_j P_d(i, j) \ln P_d(i, j). \quad (\text{A.25})$$

i) Entropia da Soma: medida da desordem referente à distribuição da soma dos níveis de cinza da soma da imagem.

$$\text{Entropia da Soma} = - \sum_{i=1}^{2L} P_{x+y}(i) \log(P_{x+y}(i)). \quad (\text{A.26})$$

Existem também as estatísticas de diferenças que são um subconjunto da matriz de co-ocorrência. Estas características são baseadas na distribuição da probabilidade $P_{x-y}(k)$ que pode ser definida como:

$$P_{x-y}(k) = \sum_{i=1}^{2L} \sum_{\substack{j=1 \\ |i-j|=k}}^{2L} P(i, j), \quad (\text{A.27})$$

onde,

$$k = 0, 1, \dots, L - 1.$$

j) **Variância da Diferença:** medida de heterogeneidade que coloca pesos mais altos em pares de níveis de intensidade diferentes que se desviam mais da média (APTE, 2018).

$$\text{Variância da Diferença} = \sum_{i=0}^{L-1} (i - \mu_{x-y})^2 P_{x-y}(i), \quad (\text{A.28})$$

onde,

$$\mu_{x-y} = \sum_0^{L-1} i P_{x-y}(i). \quad (\text{A.29})$$

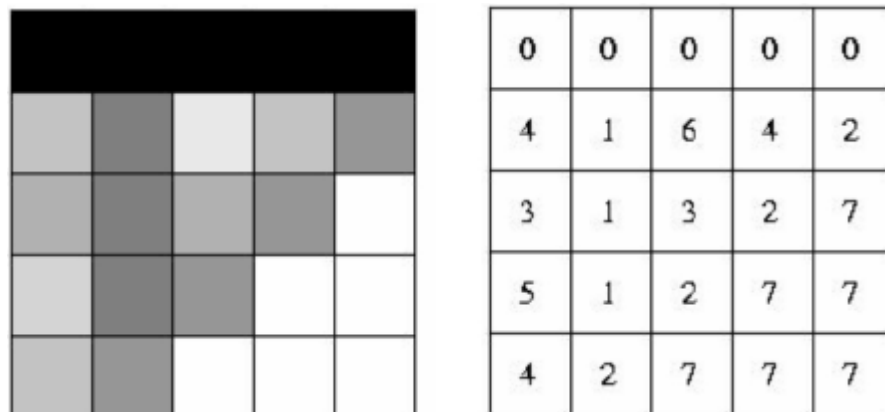
k) **Entropia da Diferença:** medida da desordem relacionada com distribuição da diferença de níveis de cinza da imagem.

$$\text{Entropia da Diferença} = - \sum_{i=0}^{L-1} P_{x-y}(i) \log\{P_{x-y}(i)\}. \quad (\text{A.30})$$

iii. Matriz de *Run-Length*

A matriz de *run-length* armazena as probabilidades de encontrar repetições de tamanho definido de pixels com mesmo valor de cinza, estas repetições são chamadas execuções. Dada uma direção, a matriz de *run-length* calcula, para cada valor de cinza permitido, quantas vezes ele aparece em grupos de 2, 3, 4, etc. pixels. Normalmente calcula-se 4 matrizes, uma horizontal, uma vertical e duas diagonais. A Figura A2 mostra a imagem com a qual será calculada a matriz de *run-length* bem como sua representação numérica.

Figura A2 - Imagem sobre a qual será calculada a matriz de *run-length* e sua representação numérica.



Fonte: Oliveira (2010).

A Figura A3 mostra uma matriz horizontal e uma diagonal calculada com os valores da imagem mostrada na Figura A2.

Como a imagem é pequena, existem vários zeros distribuídos ao longo da matriz. Os únicos elementos diferentes de zero correspondem a valores de cinza 0, 2 e 7, que são os únicos valores que se repetiam de forma consecutiva nas direções selecionadas.

Figura A3 - Matriz de *run-length* calculada na direção horizontal e diagonal da Figura A2.

	2	3	4	5	comprimento de execução		2	3	4	5	comprimento de execução
0	4	3	2	1		0	0	0	0	0	
1	0	0	0	0		1	0	0	0	0	
2	0	0	0	0		2	3	2	1	0	
3	0	0	0	0		3	0	0	0	0	
4	0	0	0	0		4	0	0	0	0	
5	0	0	0	0		5	0	0	0	0	
6	0	0	0	0		6	0	0	0	0	
7	3	1	0	0		7	3	1	0	0	
nível de cinza	execuções na direção horizontal					nível de cinza	execuções na direção 45°				

Fonte: Oliveira (2010).

Seja $p(i, j)$ o número de vezes que há um percurso de comprimento (*length*) j com nível de cinza i . Seja L o número de níveis de cinza e Nr o número de execuções (*run*). As características da matriz run-length são definidas como a (GALLOWAY, 1975):

- a) Momentos Inversos de Ênfase de Curto Prazo (MIECP): mede a distribuição de pequenos espaços onde um maior valor indica texturas finas e é definido como:

$$MIECP = \left(\sum_{i=1}^L \sum_{j=1}^L \frac{P(i, j)}{j^2} \right) / C. \quad (A.31)$$

- b) Momentos de Ênfase de Longo Prazo (MIELP): mede a distribuição de longos espaços onde um maior valor indica texturas grossas e é definido como:

$$MIELP = \left(\sum_{i=1}^L \sum_{j=1}^L j^2 P(i, j) \right) / C. \quad (A.32)$$

- c) Não Uniformidade do Nível de Cinza (NUNC): mede a distribuição de trechos acima dos valores de cinza. O valor do parâmetro é baixo quando as execuções são distribuídas igualmente ao longo dos níveis de cinza. Assim, um valor menor indica maior similaridade nos valores de intensidade.

$$NUNC = \left(\sum_{i=1}^L \left(\sum_{j=1}^L P(i, j) \right)^2 \right) / C. \quad (A.33)$$

- d) Não uniformidade do Comprimento de Execução (*run length*) (NUCE): mede a distribuição das repetições nos comprimentos de execução (Galloway, 1975). O valor do recurso é baixo quando as execuções são distribuídas igualmente ao longo dos comprimentos de execução.

$$NUCE = \left(\sum_{j=1}^L \left(\sum_{i=1}^L P(i, j) \right)^2 \right) / C. \quad (A.34)$$

- e) Fração da Imagem em Execuções (FIE): mede a fração do número de execuções realizadas e o número máximo de execuções potenciais.

$$FIE = \left(\sum_{i=1}^L \sum_{j=1}^L P(i, j) \right) / \left(\sum_{i=1}^L \sum_{j=1}^L j P(i, j) \right). \quad (A.35)$$

Em que o coeficiente C é definido por:

$$C = \left(\sum_{i=1}^L \sum_{j=1}^L P(i, j) \right). \quad (\text{A.36})$$

As 5 características acima são calculadas para 4 direções: horizontal, vertical, inclinada à 45° e inclinada à 135°.

iv. Modelo Auto-regressivo

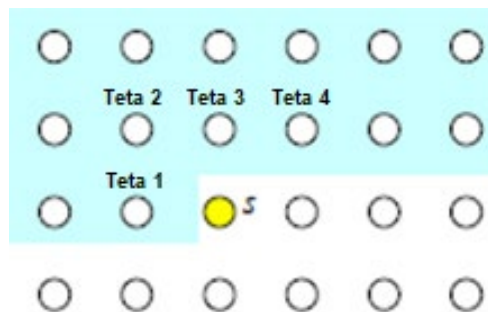
Segundo Materka & Strzelecki (1998), o modelo auto-regressivo é uma característica pressupõe interações locais entre os pixels da imagem na qual o valor de cinza do pixel é uma soma ponderada dos valores de cinza dos pixels vizinhos. Em outras palavras, é uma maneira de descrever as formas dentro da imagem através de relações entre grupos de pixels vizinhos.

Assumindo que a imagem f é um campo aleatório possuindo média zero, um modelo causal auto-regressivo pode ser definido como

$$f_s = \sum_{r \in N_s} Teta_r f_r + e_s, \quad (\text{A.37})$$

onde f_s é a intensidade da imagem no local s , e_s denota um ruído independente e identicamente distribuído, N_s é uma vizinhança de s e $Teta$ é um vetor de parâmetros do modelo. Os vizinhos locais calculados no modelo auto-regressivo, representados por quatro parâmetros, são mostrados na Figura A4. A área sombreada indica a região onde a vizinhança do modelo auto-regressivo de semi-plano válido, pode estar localizada em geral.

Figura A4 - Vizinho local do elemento de imagem f_s .



Fonte: Adaptada de Materka e Strzelecki (1998).

A utilização do modelo auto-regressivo para segmentação de imagens consiste em identificar os parâmetros do modelo para uma dada região de imagem e, em seguida, utilizar os valores de parâmetros obtidos para a discriminação de texturas. No caso da vizinhança simples de pixels mostrada na Figura A4, que compreende 4 vizinhos imediatos de pixels, existem 5 parâmetros de modelo desconhecidos: o desvio padrão σ da direção dos ruídos e_s e o vetor de parâmetros do modelo $Teta = (Teta1, Teta2, Teta3, Teta4)$. Os parâmetros podem ser estimados minimizando a soma do erro quadrático.

$$\sum_s e_s^2 = \sum_s (f_s - \widehat{Teta} w_s)^2, \quad (A.38)$$

o que fornece as seguintes equações:

$$\widehat{Teta} = \left(\sum_s w_s w_s^T \right)^{-1} \left(\sum_s w_s f_s \right), \quad (A.39)$$

$$\sigma^2 = N^{-2} \sum_s (f_s - \widehat{Teta} w_s)^2, \quad (A.40)$$

onde $w_s = col[f_i, i \in N_s]$ e é assumida uma imagem quadrada $N \times N$.

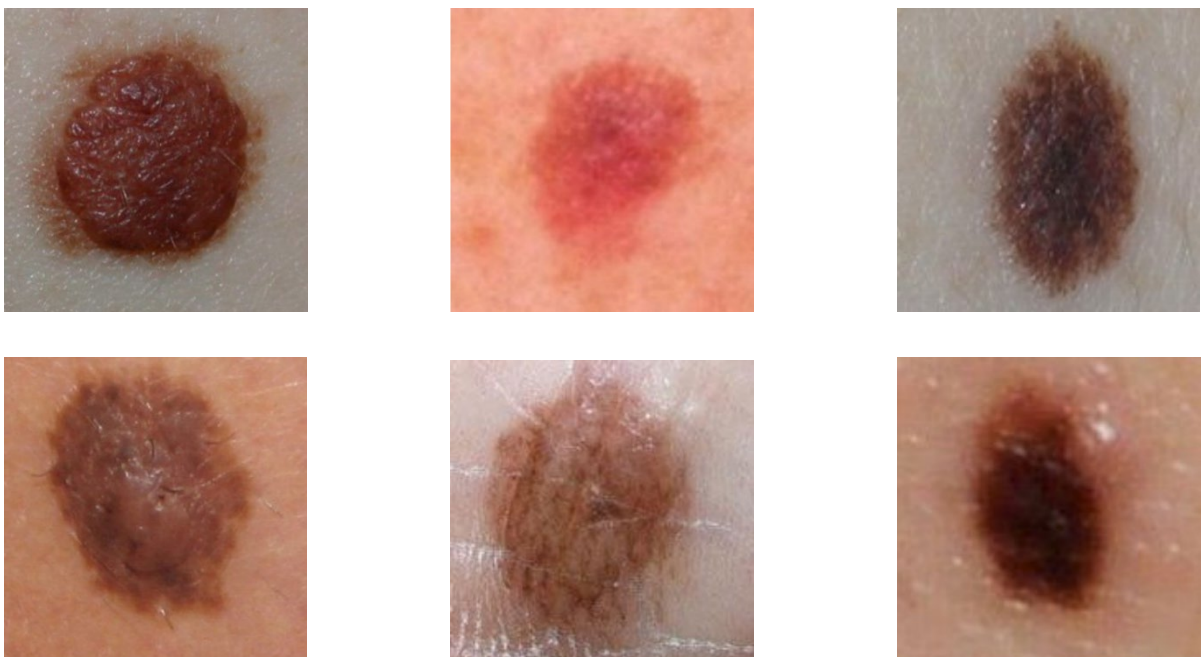
ANEXO B– AMOSTRAS DE IMAGENS PARA CRIAÇÃO DO MODELO E IMAGENS USADAS NOS TESTES DE VALIDAÇÃO

Figura B1 – Imagens de Melanomas.



Fonte: (DERMNETNZ).

Figura B2 – Imagens de manchas Benignas



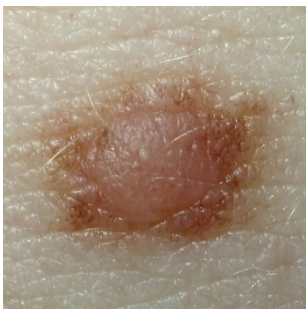
Fonte: (DERMNETNZ).

Figura B3 – Imagens de Melanomas usadas nos testes.



Fonte: (ATLAS OF CLINICAL DERMATOLOGY).

Figura B4 – Imagens de manchas benignas usadas nos testes.



Fonte: (DERMNETNZ) e (ABOUT KIDS HEALTH, 2016).