

Wellington João da Silva Sousa Júnior  
Orientador: Prof. Dr. Antonio Oseas de Carvalho Filho

**Análise filogenética aplicada a exploração de  
textura de imagens de pele para auxílio no  
diagnóstico do melanoma**

Picos - PI  
Abril de 2019

Wellington João da Silva Sousa Júnior  
Orientador: Prof. Dr. Antonio Oseas de Carvalho Filho

## **Análise filogenética aplicada a exploração de textura de imagens de pele para auxílio no diagnóstico do melanoma**

Trabalho de Conclusão de Curso de Bacharelado em Sistemas de Informação do Campus Senador Helvídio Nunes de Barros da Universidade Federal do Piauí como parte dos requisitos para obtenção do Grau de Bacharel em Sistemas de Informação.  
Orientador: Prof. Dr. Antonio Oseas de Carvalho Filho

Universidade Federal do Piauí  
Campus Senador Helvídio Nunes de Barros  
Bacharelado em Sistemas de Informação

Picos - PI  
Abril de 2019

**FICHA CATALOGRÁFICA**  
**Serviço de Processamento Técnico da Universidade Federal do Piauí**  
**Biblioteca José Albano de Macêdo**

**S275a** Sousa Júnior, Wellington João da Silva.  
Análise filogenética aplicada a exploração de textura de imagens de pele para auxílio no diagnóstico do melanoma. / Wellington João da Silva Sousa Júnior. – Picos, PI, 2019.  
35 f.  
CD-ROM: 4 ¾ pol.

Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Universidade Federal do Piauí, Picos, 2019.

Orientador(A): Prof. Dr. Antonio Oseas de Carvalho Filho.

1. Imagens Digitais – Câncer de Pele. 2. Diagnóstico – Processamento Digital. 3. Sistemas de Informação - Saúde.  
Título.

**CDD 005.1**

TITULO

WELLINGTON JOÃO DA SILVA SOUSA JUNIOR

Monografia aprovado como exigência parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Data de Aprovação:

Picos - PI, 12 de junho de 2019

  
\_\_\_\_\_  
Prof. Dr. Antônio Oseas de Carvalho Filho

  
\_\_\_\_\_  
Prof. Me. Leonardo Pereira de Sousa

  
\_\_\_\_\_  
Téc. Esp. Norato Rodrigues de Sales Carvalho

# Agradecimentos

Primeiramente agradecer a Deus, por conceder a realização deste trabalho.

Ao meu pai Wellington João da Silva Sousa e a minha mãe Yara de Souza Cabral por todo o apoio e incentivo no decorrer do curso.

Para minhas irmãs Ana Leticia Cabral da Silva Sousa e Ana Livia Cabral da Silva Sousa por sempre estarem presentes quando precisei.

Para minhas Avós Camila Maria da Silva, Eronilda de Souza Cabral e ao meu avô João Arcênio de Sousa por todo o apoio.

Aos meus tios Washington João da Silva Sousa, Wesley João da Silva Sousa, Valkiria de Souza Cabral e Maria de Lourdes de Souza Cabral por toda a assistência familiar.

Para Cristiane Cardoso Dias, José Arilson Pires e José Arilson Pires Filho por todo o incentivo.

Ao meu Prof. Dr. Antonio Oseas de Carvalho Filho por sempre estar disponível para tirar minhas dúvidas e me auxiliar na realização deste trabalho, proporcionando o crescimento acadêmico no decorrer da pesquisa.

Aos meus professores Leonardo Pereira de Sousa, Ivenilton Alexandre de Souza Moura, Francisca Pamela Carvalho Nunes, Patricia Vieira da Silva Barros e aos demais professores, meus agradecimentos.

Aos meus amigos do curso Matheus Lima, Tomaz Bisneto, Davi Luis, Milton Junior, Naara Macedo, Douglas Siqueira, Walef Alves, Matheus Garcia, Brena Maia, Estevão Viana, Renesio Joaquim, aos meninos do pavic e aos demais colegas no decorrer do curso.

E por fim, a todos que, de forma direta ou indireta, contribuíram para à realização deste trabalho.

*Faça o teu melhor, na condição que você tem, enquanto você não tem condições melhores, para fazer melhor ainda.*

*Mario Sergio Cortella*

# Resumo

O câncer de pele é definido pelo crescimento anormal e descontrolado das células que compõem a pele. Entre os cânceres de pele, o menos frequente é o melanoma, porém possui o pior prognóstico, sendo o mais letal e correspondendo a 3% dos 33% de cânceres de pele no Brasil. Apesar de ser o mais nocivo, se houver uma detecção precoce as chances de cura são de 95%. O método aplicado para o diagnóstico precoce do melanoma é o exame dermatológico. Este trabalho mostra uma metodologia de classificação em melanoma e não melanoma, com o uso de descritores de textura baseados em índices de diversidade filogenética para extração de características, no reconhecimento de padrões é realizado a classificação usando Random Forest, SMO, MLP e o IBK. Como resultado, foi obtido uma acurácia de 99,8%, sensibilidade 100% e especificidade de 99,9%. O uso de índices de diversidade filogenéticos para detectar padrões em imagens de pele, mostrou-se relevante para categorizar o melanoma e auxiliar os especialistas no diagnóstico da doença.

**Palavras-chaves:** Câncer de Pele, Precoce, Diagnóstico.

# Abstract

*Skin cancer is determined by the abnormal and uncontrolled growth of the cells that make up the skin. Among skin cancers, the least common is melanoma, but it is the worst prognosis, being the most lethal and accounting for 3% of the 33% of skin cancers in Brazil. Although it is the most harmful, the chances of cure are 95%. The method used for the early diagnosis of melanoma is the dermatological examination. The work presents a methodology of classification in melanoma and non-melanoma, with the use of texture descriptors based on phylogenetic diversity indices for extraction of characteristics, no mark pattern is recognized using Random Forest, SMO, MLP and IBK. As a result, an accuracy of 99,8%, 100% sensitivity and 99,9% was obtained. The use of long-distance indices to identify patterns of skin images has proved to be relevant for categorizing melanoma and assisting specialists in the diagnosis of melanoma.*

**Keywords:** *Skin cancer, precocious, diagnosis.*

# Lista de ilustrações

Figura 1 – Imagem de pele do tipo melanoma . . . . .	13
Figura 2 – Etapas do processamento digital de imagens Knob (2011). . . . .	17
Figura 3 – (a) Representação da região analisada (b)Árvore filogenética extraída (c) Matriz com distâncias levando em consideração a quantidade de arestas da árvore filogenética que as separa as espécies analisadas. Dis- ponível em Carvalho et al. (2018). . . . .	19
Figura 4 – Metodologia Proposta. . . . .	26

# Lista de tabelas

Tabela 1 – Relação entre a metodologia e a biologia. . . . .	19
Tabela 2 – Matriz de Confusão. . . . .	21
Tabela 3 – Tabela de Concordância do Coeficiente Kappa. . . . .	22
Tabela 4 – Trabalhos Relacionados. . . . .	25
Tabela 5 – Testes em níveis de Cinza. . . . .	28
Tabela 6 – Testes com o canal vermelho (Red). . . . .	28
Tabela 7 – Testes com o canal verde (Green). . . . .	29
Tabela 8 – Testes com o canal azul (Blue). . . . .	29
Tabela 9 – Teste com o descritor GLCM. . . . .	30
Tabela 10 – Comparativo com os Trabalhos Relacionados. . . . .	31

# Lista de abreviaturas e siglas

A	Acurácia
AUC	Área sob curva ROC
E	Especificidade
FN	Falso Negativo
FP	Falso Positivo
IBK	<i>Instance Based Learning</i>
ISDIS	<i>International Society for Digital Imaging of the Skin</i>
ISIC	<i>Skin Imaging Collaboration</i>
K	Coeficiente Kappa
MLP	<i>Multi Layer Perceptron</i>
PD	<i>Phylogenetic Diversity</i>
PDI	<i>Digital Image Processing</i>
ROC	<i>Receiver Operating Characteristic</i>
ROI	<i>Region of Interest</i>
S	Sensibilidade
SMO	<i>Sequential Minimal Optimization</i>
SPD	<i>Sum of Phylogenetic Distances</i>
SVM	<i>Support Vector Machine</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

# Lista de símbolos

$A_i$	Abundância Média de Espécies
$A_m$	Abundância de Espécie $m$
$D_m$	Distância Entre as Espécies $m$ e $n$
$L_i$	Comprimento do Ramo
$P_e$	Taxa Hipotética da Aceitação
$P_o$	Taxa de Aceitação Relativa
$\Sigma$	Somatório

# Sumário

<b>1</b>	<b>Introdução</b>	<b>13</b>
1.1	Objetivos	14
1.2	Organização do Trabalho	14
<b>2</b>	<b>Referencial Teórico</b>	<b>16</b>
2.1	Melanoma	16
2.2	Processamento Digital de Imagens	17
2.3	Índices de Diversidade	18
2.3.1	Índice de Diversidade Filogenética	19
2.4	Reconhecimento de Padrões	20
2.5	Métricas de Validação	21
<b>3</b>	<b>Trabalhos Relacionados</b>	<b>24</b>
<b>4</b>	<b>Método Proposto</b>	<b>26</b>
4.1	Aquisição das Imagens	26
4.2	Extração de Características	27
4.3	Classificação e Validação dos Resultados	27
<b>5</b>	<b>Resultados e Discussões</b>	<b>28</b>
<b>6</b>	<b>Conclusão</b>	<b>32</b>
	<b>Referências</b>	<b>33</b>

# 1 Introdução

A interação entre a pele e a luz solar é inevitável. O potencial de dano depende do tipo e da duração da exposição. A luz solar tem efeitos profundos sobre a pele, e está associada a uma variedade de doenças (SAMIS, 2006). Dentre essas doenças, o câncer de pele.

O câncer de pele responde por 33% de todos os diagnósticos desta doença no Brasil, sendo que o Instituto Nacional do Câncer (INCA) registra, a cada ano, cerca de 180 mil novos casos (INCA, 2018). Entre os cânceres da pele, o melanoma é o menos frequente, porém o mais letal. Segundo o INCA (2018), no Brasil, embora o câncer de pele seja o mais frequente tendo 30% de todos os tumores malignos registrados, apenas 3% são do tipo melanoma, com 6.260 novos casos diagnosticados com essa doença, as mulheres são mais frequentes com um total de 3.340 casos e os homens 2.940, o número de mortes são de 1.794, sendo 1.012 homens e 782 mulheres.

A detecção precoce do câncer de pele aumenta a chance de cura dos pacientes, sendo realizado através de exames dermatológicos. Com o uso deste, a incidência de câncer de pele pode ser tratada de forma precoce, porém, alguns fatores podem comprometer os prognósticos, como a inexperiência do especialista, a qualidade do exame, diagnósticos de doenças parecidas na pele, acarretando em possíveis erros na detecção da doença, como afirma Bakos (2006) e INCA (2018).

No diagnóstico dermatológico, as lesões são examinadas clinicamente, utilizando primeiramente a técnica de análise das características e textura, para então diagnosticá-las e tratá-las. Para facilitar este processo, os dermatologistas podem dispor de sistemas computacionais, que analisam as características das lesões de forma mais precisa, utilizando imagens digitais, para auxiliar no seu diagnóstico (OLIVEIRA, 2012). A Figura 1 apresenta uma imagem do tipo melanoma.



Figura 1: Imagem de pele do tipo melanoma

O processamento digital de imagens tem um grande potencial nas diversas modalidades de imagens médicas. A análise e interpretação computadorizada de imagens médicas são fundamentais no auxílio a um diagnóstico preciso e no estudo da patologia da doença. Essa área ficou conhecida como diagnóstico por imagem. Com advento da inteligência artificial e sistemas inteligentes, o diagnóstico pode ser incorporado na análise informatizada da imagem (SOARES, 2008).

Como visto em Azevedo-Marques (2001), os sistemas de diagnóstico auxiliados por computador ( *computer-aided diagnosis, CADx*) são responsáveis por quantificar as características das imagens e reconhecer os padrões e classificar as imagens médicas em normais ou anormais. Geralmente são utilizados acoplados em equipamentos médicos com o intuito de auxiliar o especialista no diagnóstico de doenças.

Neste trabalho foi desenvolvido um sistema *CADx* para o diagnóstico automático do melanoma, categorizando em melanoma ou não. Para isso, são extraídos as características das imagens baseados em textura.

Por fim, é feito o reconhecimento de padrões com múltiplos classificadores para avaliar a relevância da pesquisa proposta, com base nos resultados obtidos em cada classificador e determinar se o trabalho proposto é relevante para que os especialistas tenham uma segunda opinião no diagnóstico da doença.

## 1.1 Objetivos

Este trabalho tem como objetivo geral desenvolver uma abordagem para a classificação de regiões em melanoma e não melanoma, utilizando descritores de textura baseado nos índices de diversidade filogenética.

Possuem como objetivos específicos:

1. Desenvolver e adaptar técnicas baseadas no comportamento filogenético para caracterização da textura das regiões;
2. Analisar a usabilidade dos descritores filogenéticos na categorização de imagens dermatológicas;
3. Construir uma metodologia para que os especialistas tenham uma segunda opinião no diagnóstico do melanoma.

## 1.2 Organização do Trabalho

Além da introdução, este trabalho está organizado em outros 5 Capítulos. O Capítulo 2 descreve todo o embasamento teórico para o entendimento da metodologia utilizada no trabalho. No Capítulo 3 são mostrados os principais trabalhos relacionados. No Capítulo 4 é abordado todo o desenvolvimento da pesquisa, tendo como início a aquisição

---

das imagens, extração de características e de textura baseado em índice de diversidade filogenética, classificação e por último a validação da metodologia proposta. O Capítulo 5 detalha os resultados obtidos com a metodologia proposta. Por fim, no capítulo 6 é apresentado a conclusão do trabalho proposto.

## 2 Referencial Teórico

Para melhor entendimento deste trabalho, faz-se necessário a descrição de conceitos relacionados ao tema. Este capítulo aborda todo o embasamento teórico utilizado na metodologia proposta para facilitar sua compreensão. Diante disto, serão abordados os seguintes tópicos: Melanoma, processamento digital de imagens, índices de diversidade filogenética, reconhecimento de padrões e métricas de validação.

### 2.1 Melanoma

O câncer de pele é uma doença que é provocada pelo crescimento anormal e descontrolado das células que compõem a pele. Essas células se dispõem formando camadas e, de acordo com as que forem afetadas, são definidos os diferentes tipos de câncer. Os mais comuns são os carcinomas basocelulares e os espinocelulares. O câncer de pele mais agressivo é conhecido como melanoma, sendo o mais raro e nocivo entre os outros tipos de cânceres de pele (INCA, 2018).

O câncer de pele é o tipo de câncer mais incidente em ambos os sexos no Brasil. Entretanto, as lesões são de fácil diagnóstico clínico e podem atingir índices de curas superiores a 95%, desde que diagnosticados nos estágios iniciais e tratados corretamente (GOMES, 2013).

O melanoma tem origem nos melanócitos, as células que produzem melanina, o pigmento que dá cor à pele. Normalmente, surge nas áreas do corpo mais expostas à radiação solar. Em estágios iniciais, o melanoma se desenvolve apenas na camada mais superficial da pele, o que facilita a remoção cirúrgica e a cura do tumor. Nos estágios mais avançados, a lesão é mais profunda e espessa, o que aumenta a chance de se espalhar para outros órgãos (metástase) e diminui as possibilidades de cura (COSTA et al., 2019).

Por isso, o diagnóstico precoce do melanoma é fundamental. Embora apresente pior prognóstico, avanços na medicina e o recente entendimento das mutações genéticas, que levam ao desenvolvimento dos melanomas, possibilitaram que pessoas com melanoma avançado hoje tenham aumento na sobrevida e na qualidade de vida (GUERRA et al., 2019).

Quando uma pessoa apresenta sinais de melanoma, o especialista analisa todo o histórico do paciente, observando os sintomas e fatores de risco que desencadeiam o melanoma. Para o diagnóstico do melanoma e determinar o grau de comprometimento da doença o profissional solicita alguns exames.

Um dos exames solicitados pelo especialista é o exame físico, o profissional observa a forma, tamanho, cor e textura das lesões e se há sangramento ou descamação. Em conjunto com o exame físico, alguns especialistas utilizam dermatoscopia para avaliar manchas de

pele, possuindo dois tipos, a dermatoscopia manual que utiliza o dermatoscópio para verificar o risco da lesão e a dermatoscopia digital que permite a análise de uma fotografia ampliada da lesão. Porém, o diagnóstico definitivo do melanoma só pode ser feito através da biópsia, que é feito através de amostras de pele enviadas ao patologista para serem analisadas, sendo este o profissional responsável pelo diagnóstico definitivo da doença, como afirma [BRAZIL \(2009\)](#).

## 2.2 Processamento Digital de Imagens

É visto que em [Gonzalez e Woods \(2009\)](#) a área de processamento digital de imagens (PDI) trata da manipulação de imagens por computador de modo que geralmente a entrada e a saída do processo sejam imagens. O interesse nos métodos de processamento digital de imagens provém de duas áreas principais de aplicação: Melhora das informações visuais para a interpretação humana e processamento de dados de imagens para armazenamento, transmissão e representação, considerando a percepção automática por máquinas.

Segundo [Knob \(2011\)](#) o PDI é dividido em 6 etapas, como ilustra a Figura 2. Sendo elas: Aquisição das imagens, pré-processamento, segmentação, representação e descrição, reconhecimento e interpretação.

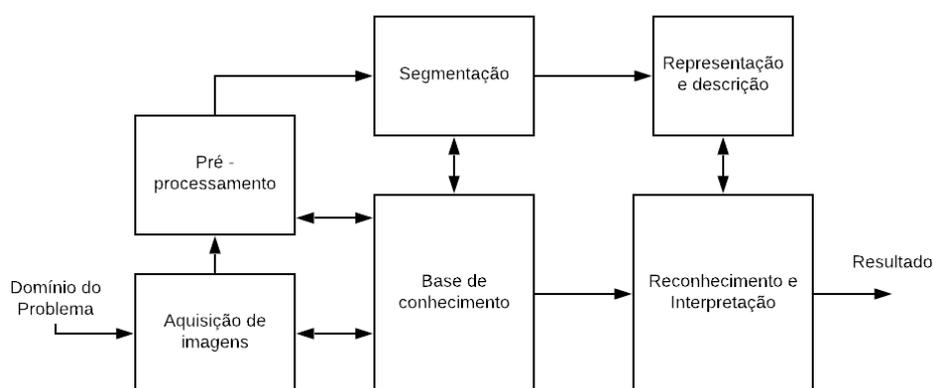


Figura 2: Etapas do processamento digital de imagens [Knob \(2011\)](#).

O fim de cada etapa do PDI é a entrada para a próxima etapa, no final de todas as etapas, pode ser gerado uma imagem digital ou não.

- A primeira etapa é conhecida como aquisição de imagens, onde o digitalizador converte uma imagem analógica para digital.
- A segunda etapa é o pré-processamento, tendo como objetivo melhorar a imagem, aumentando também os resultados das etapas seguintes. Pode ser usado técnicas como: remoção de ruídos, suaviza tons de cinza, dentre outros.

- A terceira etapa conhecida como segmentação, tem como objetivo separar a imagem em uma região de interesse. Sendo abordado de forma manual, semi-automática e automática.
- A quarta etapa é a representação e descrição, conhecida como extração de características. Com o objetivo de descrever e representar as imagens através de valores. Cria-se um vetor de característica definindo um padrão, com base nos valores obtidos de cada medida extraída.
- A quinta etapa é conhecida como base de conhecimento, onde é aplicado todo o embasamento teórico e prático na área do PDI, com o objetivo de aplicar as etapas necessárias para a elaboração da metodologia do estudo.
- A última etapa é conhecida como reconhecimento e interpretação, sendo parte do processamento que classifica os objetos a partir de informações encontradas na imagem, atribuindo um rótulo para o objeto baseado na informação fornecida pelos descritores.

## 2.3 Índices de Diversidade

A diversidade é um termo muito utilizado na área da ecologia. O seu objetivo é informar variedade de espécies presentes em uma comunidade ou área (OLIVEIRA et al., 2013).

Um índice de diversidade é a medida da heterogeneidade de uma comunidade. Ele é utilizado na ecologia para medir a biodiversidade em um ecossistema. Na economia, serve para medir a distribuição sobre os setores de atividades econômicas em uma região. De forma mais geral, os índices de diversidade podem ser utilizados para estimar a diversidade de uma população na qual cada membro pertence a um único grupo ou espécie (SOUSA et al., 2011).

Segundo Pianka (1994), para medir a diversidade necessita de dois componentes: *a riqueza das espécies*, que consiste no número de espécies encontradas em determinada região e a *abundância relativa*, que é o número de indivíduos de uma determinada espécie existentes numa dada área. As medidas de espécies são geralmente úteis para comparar padrões em diferentes áreas.

A forma mais básica de utilizar índices de diversidade em imagens, é quando a comunidade representa uma imagem ou região da mesma, as espécies sendo os níveis de cinza e os indivíduos sendo os pixels OLIVEIRA et al. (2013). A Tabela 1 relaciona a metodologia proposta e a biologia.

Tabela 1: Relação entre a metodologia e a biologia.

Metodologia Proposta	Biologia
Região de Interesse da Imagem	Comunidade
Níveis de Cinza da Imagem	Espécies
Píxels da Imagem	Indivíduos
Número de arestas entre duas Espécies	Distância Filogenética

### 2.3.1 Índice de Diversidade Filogenética

Diversidade filogenética é uma medida de uma comunidade que incorpora as relações filogenéticas das espécies (MAGURRAN, 2013). A premissa principal dessa medida é que a diversidade é maior em uma comunidade em que as espécies são filogeneticamente mais distintas, em outras palavras, pode-se afirmar que isso representaria uma imagem que possui muita variabilidade de níveis de cinza, ou seja, muitas espécies. A percepção dominante em ecologia evolutiva é que espécies coexistindo devem diferir significativamente e que a maior parte da variação entre espécies aparentadas é uma resposta adaptativa à competição no passado, quando os traços não diferiam (HARVEY; RAMBAUT, 2000). Esses índices consideram a quantidade de indivíduos, calculada a partir da arquitetura da árvore.

Uma árvore filogenética é uma representação gráfica, em forma de árvore, que apresenta relações evolutivas entre várias espécies ou outras entidades que possam ter um ancestral comum. Onde, as folhas representam os organismos e os nós internos representam os supostos ancestrais, as arestas da árvore denotam as relações evolutivas. A Figura 3 apresenta um exemplo de região de imagem onde acontece uma diversidade de espécies em relação aos indivíduos. Sendo possível o uso dos índices de diversidade filogenética.

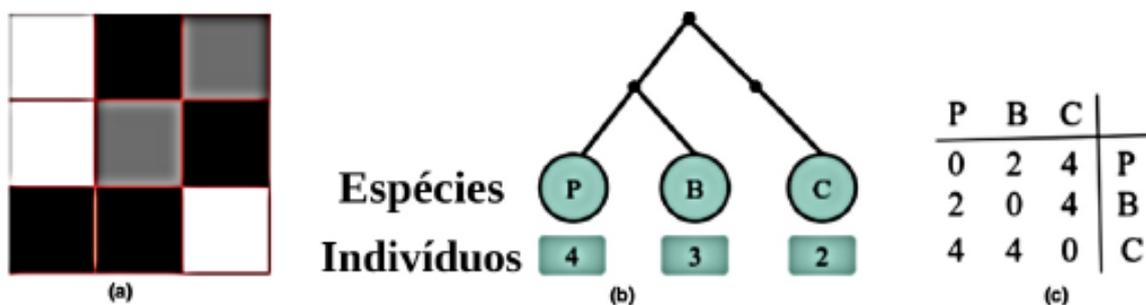


Figura 3: (a) Representação da região analisada (b)Árvore filogenética extraída (c) Matriz com distâncias levando em consideração a quantidade de arestas da árvore filogenética que as separa as espécies analisadas. Disponível em [Carvalho et al. \(2018\)](#).

Neste trabalho, são utilizados dois índices de diversidade filogenética para criação dos descritores de texturas, sendo eles: *Phylogenetic Diversity* (PD) e *Sum Of Phylogenetic Distances* (SPD).

O PD é o somatório dos comprimentos dos ramos da filogenia de cada espécie. Onde  $B$  é o número de ramificação da árvore,  $Li$  é o comprimento do ramo e  $Ai$  é a abundância média de espécies que ramo  $i$ . Representado pela Equação 2.1.

$$B * \frac{\sum_i^B LiAi}{\sum_i^B Ai}. \quad (2.1)$$

O SPD é a soma das distâncias filogenéticas entre cada par de espécies. Onde  $dmm$  é a distância entre as espécies  $m$  e  $n$ ;  $a_m$ , abundância de espécies  $m$ ;  $S$ , número de espécies no conjunto focal<sup>1</sup>. Como mostra a Equação 2.2.

$$SPD = \left(\frac{S(S-1)}{2}\right) * \frac{\sum \sum m < n^a d^{mn} m^a n}{\sum \sum m < n^a m^a n}. \quad (2.2)$$

## 2.4 Reconhecimento de Padrões

Segundo [Castro e Prado \(2002\)](#) o reconhecimento de padrões pode ser dividido em duas categorias: O estudo de seres humanos e organismos vivos com o objetivo de se estabelecer os modos pelos quais os mesmos desenvolvem e aprimoram suas capacidades de reconhecimento de padrões e o desenvolvimento de teorias e técnicas objetivando a construção de máquinas ou dispositivos capazes de apresentar características semelhantes às dos seres humanos em reconhecerem padrões.

A técnica de reconhecimento de padrões tem como objetivo classificar objetos tendo base um conhecimento prévio ou informações estatísticas extraídas dos padrões ([CASTRO; PRADO, 2002](#)).

O uso do reconhecimento de padrões é recorrente para discriminar e classificar possíveis alterações em imagens médicas, determinando a presença ou ausência da doença. Após a extração de características, geralmente utiliza-se o reconhecimento de padrões para detectar as possíveis alterações nas imagens e classificar as possíveis anomalias ou não da imagem que desencadeiam a doença. Foi utilizado a ferramenta WEKA para realizar a classificação da pesquisa proposta.

Weka é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Ele contém ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização ([WEKA, 2018](#)).

Para o reconhecimento de padrões desta metodologia, baseado em estudos sobre classificadores levando consideração seu desempenho, facilidade de adaptação nos experimentos, eficácia e junção com as ferramentas de extração de características em imagens. Embasado nestes critérios, foram escolhidos os seguintes classificadores: *Random Forest*, *Sequential Minimal Optimization* (SMO), *Multi Layer Perceptron* (MLP) e *Instance Based Learning* (IBK).

---

<sup>1</sup> Comunidade analisada

O *Random Forest* é um classificador composto por uma coleção de árvores, onde são aleatoriamente independentes e identicamente distribuídas e cada árvore vota na classe mais popular, sendo a escolhida para o objeto (OSHIRO, 2013).

O SMO é um dos principais métodos atuais para resolver o problema de programação quadrática que surge durante o treinamento de um SVM. Este método, em certa medida, pode diminuir o grau de dificuldade de um problema de programação quadrática através de estratégias de decomposição. Esses problemas são resolvidos analiticamente, permitindo que o SMO manipule um grande conjunto de treinamento e realize a distinção dos padrões entre os conjuntos de dados avaliados (ZENG ZHI-QIANG E YU, 2008).

O MLP é uma rede neural perceptron, formada por um conjunto de camadas. Tendo uma arquitetura de uma camada de entrada, que são várias unidades que recebem entradas no mundo real. Com uma ou duas camadas ocultas, elas podem aproximar virtualmente qualquer mapa de entrada-saída. E a camada de saída que retorna os resultados para o mundo real (PANCHAL GAURANG E GANATRA, 2011).

O IBK é um classificador baseado no aprendizado de instâncias, conhecido como o algoritmo do vizinho mais próximo, que possui grandes requisitos de armazenamento. Descrevem-se como os requisitos de armazenamento que podem ser reduzidos significativamente com, no máximo, pequenos sacrifícios na taxa de aprendizado e na precisão da classificação (HALL, 2000).

## 2.5 Métricas de Validação

Para avaliar o desempenho dos classificadores, geralmente são usadas métricas para validar a relevância da metodologia, onde são comumente utilizados as seguintes métricas: Acurácia (A), Sensibilidade (S) e Especificidade (E). Essas métricas fazem o uso da matriz de confusão, que faz a classificação correta e incorreta das classes, agrupando os resultados em quatro classes, respectivamente: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN). A Tabela 2 representa a matriz de confusão.

Tabela 2: Matriz de Confusão.

Resultado	Doença	
	Melanoma	Ausente
Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Acúrcia (A): É a proporção de predições corretas, sem levar em consideração o que é positivo e o que é negativo. Esta medida é altamente suscetível a desbalanceamentos do

conjunto de dados e pode facilmente induzir a uma conclusão errada sobre o desempenho do sistema. Representado pela Equação 2.3.

$$A = \frac{VP + VN}{VP + VN + FP + FN}. \quad (2.3)$$

Sensibilidade (S): É a proporção de verdadeiros positivos, isto é, a capacidade do sistema em prever corretamente a condição para casos que realmente a têm. Representado pela Equação 2.4.

$$S = \frac{VP}{VP + FN}. \quad (2.4)$$

Especificidade (E): É a proporção de verdadeiros negativos, isto é, a capacidade do sistema em prever corretamente a ausência da condição para casos que realmente não a têm. Representado pela Equação 2.5.

$$E = \frac{VN}{VN + FP}. \quad (2.5)$$

A análise ROC (*Receiver Operating Characteristic*) é uma ferramenta poderosa para medir e especificar problemas no desempenho do diagnóstico em medicina. Como visto em Braga (2001), a área sob curva (AUC) é caracterizado como uma boa forma de avaliar um sistema de classificação binária, onde o sistema de diagnóstico discrimina entre dois estados (especificidade e sensibilidade), verificando-se o melhor desempenho, alcançado pelo método avaliado, quando o valor da AUC for mais próximo de 1 (igual a 100%).

Para avaliar a concordância do experimento é usado o método estatístico de cálculo do Coeficiente Kappa (K). Sendo utilizado para avaliar a reprodutibilidade dos resultados na validação. Este resultado quanto mais próximo de 1, maior será a concordância do experimento. A Tabela 3 representa os níveis de concordância do K.

Tabela 3: Tabela de Concordância do Coeficiente Kappa.

Valor do Coeficiente Kappa	Nível de Concordância
<0	Não existe concordância
0 - 0.20	Concordância mínima
0.21 - 0.40	Concordância razoável
0.41 - 0.60	Concordância moderada
0.61 - 0.80	Concordância substancial
0.81 - 1.0	Concordância perfeita

No cálculo do  $K$ ,  $P_o$  é a taxa de aceitação relativa e  $P_e$  é a taxa hipotética da aceitação. Representado pela Equação 2.6.

$$K = \frac{P_o - P_e}{1 + P_e} \quad (2.6)$$

As métricas citadas acima, tem como finalidade estimar o desempenho da metodologia realizada neste trabalho, para determinar se os resultados obtidos foram relevantes ou

não e também apontar os pontos que foram positivos e negativos para que seja aplicado melhorias em trabalhos futuros para essa pesquisa.

## 3 Trabalhos Relacionados

A literatura especializada oferece estudos relacionados a classificação do melanoma, assim como técnicas de extração de características. Nesta seção será abordado um resumo de alguns desses trabalhos.

No trabalho de [Alencar \(2015\)](#) é desenvolvida uma metodologia para categorizar as imagens de pele em melanoma ou não, aplicando a limiarização de *otsu* para segmentar as imagens no canal azul do esquema RGB e extrair os contornos utilizando as funções *FindContours*, *ConvexHull* e *ConvexityDefects* da biblioteca OpenCv. Foi utilizado o classificador *Multi Layer Perceptron* para o reconhecimento de padrões, resultando em 86,6% de acurácia.

Na metodologia de [Rastgoo et al. \(2016\)](#) é proposto o uso de uma técnica que utiliza o *Random Forest* para reconhecer os padrões, juntamente com técnicas de extração de características que são, *Scale-Invariant Feature Transform* (SIFIT), Histogramas de Ângulos Opostos e Intensidade de RGB. O treinamento ocorreu através de validação cruzada com 10 divisões com a base de imagens ph2. Obtendo 100% de especificidade e 90,3% de sensibilidade.

No trabalho de [Moura et al. \(2017\)](#) apresenta uma metodologia para detecção automática do melanoma, através de um descritor híbrido para a extração de características, sendo eles, a regra ABCD, *Gray Level Run Length Matrix* (GLRLM), *Histograms of Oriented Gradients* (HOG), *Local Binary Pattern* (LBP), Tamura, *Box-Counting*. Onde foi utilizado o algoritmo Razão de Ganho de Informação (*Gain Ratio Information*) para criar o vetor de características e em seguida, reconhecimento de padrões em melanoma e não melanoma obtendo 95% de acurácia com o classificador MLP.

Na metodologia de [Li e Shen \(2018\)](#) é proposto uma abordagem para segmentar e classificar as imagens de lesões de pele de forma simultânea através da técnica *Lesion Indexing Network* (LIN). Através de 2 CNNs profundas, FCRNs-88, treinadas com diferentes conjuntos de treinamento, para produzir o resultado da segmentação e classificação. Usa-se a técnica *Lesion Indexing Calculation Unit* (LICU) para medir a importância do pixel na decisão da classificação. Resultando em 85,7% de acurácia, 96,6% em especificidade e 49,0% na sensibilidade.

No trabalho de [Bonafini e Perlin \(2018\)](#) mostra uma metodologia para o diagnóstico automático do melanoma, através das regras ABCDE (Assimetria, borda, cor, diâmetro e evolução) são extraídas as características e feito o reconhecimento de padrões com o classificador *Support Vector Machine* (SVM). Resultando em 84,5% de acurácia.

[Maia et al. \(2019\)](#) propõe uma metodologia para o diagnóstico do melanoma através das regras ABCD, com o uso da rede neural VGG16, *Inception* e *Resnet* para a extração de características. Realizando o reconhecimento de padrões com o classificador Random

Forest e obtendo uma acurácia de 92%, sensibilidade em 92,5% e especificidade de 91,8%.

A Tabela 4 apresenta o desempenho dos trabalhos relacionados, mencionando a metodologia aplicada, base de imagem e os resultados obtidos (acurácia, sensibilidade e especificidade).

Tabela 4: Trabalhos Relacionados.

Trabalho	Metodologia	Base	Acurácia (%)	Especificidade (%)	Sensibilidade (%)
(ALENCAR, 2015)	Contornos + MLP	PH2	86,6	-	-
(RASTGOO et al., 2016)	SIFIT + Random Forests	PH2	-	100	90,3
(MOURA et al., 2017)	Descritor Híbrido + MLP	PH2	95,5	95,4	95,5
(LI; SHEN, 2018)	LIN CNN	ISIC	85,7	96,6	49
(BONAFINI; PERLIN, 2018)	Regras ABCDE + SVM	ISIC	84,5	-	-
(MAIA et al., 2019)	Inception + Resnet + Random Forest	PH2	92	92,5	91,8

Os trabalhos relacionados acima apresentaram resultados promissores na detecção e diagnóstico do melanoma e lesões de pele, porém alguns apresentaram dificuldades na execução do método. Como nos trabalhos de Moura et al. (2017), Rastgoo et al. (2016) e Maia et al. (2019) que utilizaram uma base de imagens pequena composta por 200 imagens dermatológicas. Na metodologia de Alencar (2015) e Bonafini e Perlin (2018) poderia ter sido aplicado outras métricas para validar suas pesquisas. No projeto de Li e Shen (2018) poderia ser empregado técnicas para melhorar a sensibilidade.

A maioria dos trabalhos citados acima, utilizam descritores baseados em forma, seguindo as regras ABCD e descritores baseados em redes neurais, com isso, nesta pesquisa pretende-se utilizar uma nova abordagem com descritores de textura baseado em índices de diversidade filogenética para verificar a relevância da análise de textura em comparação aos descritores de forma e aos baseados em redes neurais, também aumentar a acurácia e melhorar os valores da sensibilidade e especificidade.

## 4 Método Proposto

A metodologia proposta neste trabalho teve como finalidade realizar a classificação das regions of interest (ROIs) em melanoma ou não. Compreendido em 4 etapas, na primeira, a aquisição do banco de imagens *International Skin Imaging Collaboration* (ISIC) (ISIC, 2016); na segunda, é realizado a extração de características e de textura, implementados neste trabalho com a biblioteca de manipulação de imagens OpenCv (PURI; JAIN, 2019); na terceira, utiliza-se algoritmos disponíveis na ferramenta WEKA (WEKA, 2018) para realizar a classificação das características extraídas; e por fim, é feita a validação dos resultados com as métricas apresentadas na Seção 2.5. A Figura 4 mostra o resumo de cada etapa desta metodologia.

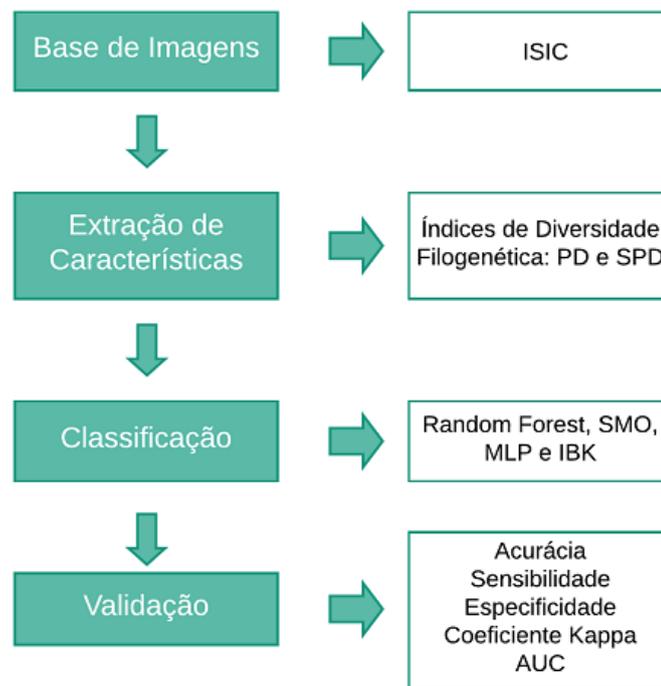


Figura 4: Metodologia Proposta.

### 4.1 Aquisição das Imagens

A ISIC é um esforço internacional para melhorar o diagnóstico do melanoma, patrocinado pela Sociedade Internacional de Imagem Digital da Pele (ISDIS). A base de imagens ISIC contém a maior coleção publicamente disponível de imagens dermatoscópicas (ISIC, 2016).

A base de imagens é composta por 900 imagens dermatológicas com as marcações feitas por especialistas. A partir dessas marcações são extraídas as regiões de interesse

(ROIs), gerando as imagens com as delimitações feitas pelos especialistas e as originais. Com 727 imagens não melanoma e 123 imagens melanoma.

## 4.2 Extração de Características

Após a aquisição das imagens e a extração das ROIs a partir das marcações feitas pelos especialistas, as imagens com as delimitações feitas pelos especialistas foram encaminhadas para a fase de extração de características de textura. Sendo aplicado a metodologia de extração de características de textura em imagens dermatológicas baseado em índices de diversidade filogenética.

Neste trabalho, são utilizados dois índices de diversidade filogenética para criação dos descritores de textura, sendo eles: A diversidade filogenética e a soma das distâncias filogenéticas. Onde, no final do processo de extração de características, esses índices descrevem as texturas das imagens e os atributos fornecidos buscam reconhecer as medidas de homogeneidade ou heterogeneidade da ROI. Servindo como base para categorizar em melanoma e não melanoma.

## 4.3 Classificação e Validação dos Resultados

Para a classificação, foi utilizado a ferramenta Weka, dentre os algoritmos presentes no Weka, foram selecionados: Radom Forest, SMO, MLP e IBK, por apresentarem bons resultados e pelo fato de serem rigorosos na presença de ruídos no processo de classificação, também utilizam descritores que se baseiam em comportamento dentro de comunidades, sendo uma boa escolha para a classificação da metodologia.

Utilizando os parâmetros com os valores padrão juntamente com a validação cruzada (*Cross Validation*) de *k-folds*, atribuindo o  $k = 10$ . Esse método tem como procedimento dividir os indivíduos em 10 grupos, separando o treino em 9 grupos e 1 grupo para teste, ocorrendo 10 cruzamentos alterando o grupo de teste, após realizar todos os cruzamentos é gerado uma média que atribue-se como resultado.

Por fim, após a etapa do reconhecimento de padrões é necessário a utilização de métricas para validar os resultados e analisar possíveis melhorias do mesmo. A metodologia proposta, utiliza métricas de validação recorrentes na área de processamento digital de imagens, sendo baseadas em estatísticas, respectivamente: Acurácia (A), Sensibilidade (S), Especificidade (E), Coeficiente Kappa (K) e Área Sobre a Curva (ROC).

## 5 Resultados e Discussões

Este capítulo mostra e discute os resultados obtidos da aplicação da metodologia proposta neste trabalho, obtendo o resultado da classificação das ROIs das imagens de pele em melanoma e não melanoma. Foram feitos testes dos descritores de características para imagens em níveis de cinza, do canal vermelho, do canal verde e do canal azul, com o intuito de verificar o potencial dos descritores de textura em cada um desses canais.

As Tabelas 5, 6, 7 e 8 são uma síntese dos principais testes feitos nesta metodologia, através de descritores de textura para a extração de características e utilizando os classificadores: IBK, MLP, *Random Forest* e SMO para o reconhecimento de padrões das imagens de pele.

A Tabela 5 demonstra os resultados dos testes dos descritores de textura extraídos para as imagens em níveis de cinza.

Tabela 5: Testes em níveis de Cinza.

Classificadores	Acurácia (%)	Especificidade (%)	Sensibilidade (%)	ROC	Kappa
<b>IBK</b>	<b>98,6</b>	<b>99,2</b>	<b>99,2</b>	<b>0,979</b>	<b>0,957</b>
MLP	92,6	99,2	92,3	0,805	0,732
Random Forest	97,6	98,6	98,5	0,996	0,924
SMO	82,1	99,9	81,9	0,537	0,113

Diante dos resultados apresentados na Tabela 5, o classificador SMO apresentou a menor taxa de predição, obtendo uma acurácia de 82%, especificidade em 99%, sensibilidade de 81,9%, ROC em 0,537 e K de 0,113. O resultado mais relevante foi obtido pelo classificador IBK, com porcentagem de 98% de acurácia, 99% em especificidade, 99% de sensibilidade, ROC em 0,979 e K de 0,957.

A Tabela 6 representa os resultados dos testes dos descritores de textura extraídos do canal vermelho.

Tabela 6: Testes com o canal vermelho (Red).

Classificadores	Acurácia (%)	Especificidade (%)	Sensibilidade (%)	ROC	Kappa
<b>IBK</b>	<b>99,8</b>	<b>99,9</b>	<b>100</b>	<b>0,999</b>	<b>0,996</b>
MLP	94,7	98,5	95,2	0,941	0,822
Random Forest	99,4	99,7	99,6	0,998	0,982
SMO	81,5	99,9	81,5	0,522	0,070

Entre os resultados obtidos da classificação mostrada na Tabela 6, o que obteve o pior resultado foi o classificador SMO, com 81% na acurácia, 99% em especificidade, 81% de sensibilidade, 0,522 para ROC e 0,070 para o K. A melhor predição foi obtida

pelo classificador IBK, alcançando 99% na acurácia, 99% para especificidade, 100% em sensibilidade, 0,999 na ROC e 0,996 no K.

A Tabela 7 apresenta os resultados dos testes dos descritores de textura extraídos do canal verde.

Tabela 7: Testes com o canal verde (Green).

Classificadores	Acurácia (%)	Especificidade (%)	Sensibilidade (%)	ROC	Kappa
<b>IBK</b>	<b>99,6</b>	<b>99,7</b>	<b>99,9</b>	<b>0,996</b>	<b>0,989</b>
MLP	94,8	97,4	96,3	0,968	0,832
Random Forest	98,7	99,3	99,2	0,996	0,960
SMO	80,6	99,9	80,8	0,499	-0,002

Diante dos resultados atingidos na Tabela 7, o resultado apresentado de menor relevância foi do classificador SMO, com 80% na acurácia, 99% de especificidade, 80% em sensibilidade, 0,499 para ROC e -0,002 para o K. O resultado mais promissor entre os classificadores foi o IBK, obtendo 99% de acurácia, 99% em especificidade, 99% na sensibilidade, 0,996 para ROC e 0,989 no K.

A Tabela 8 mostra os resultados dos testes dos descritores de textura extraídos do canal azul.

Tabela 8: Testes com o canal azul (Blue).

Classificadores	Acurácia (%)	Especificidade (%)	Sensibilidade (%)	ROC	Kappa
<b>IBK</b>	<b>99,3</b>	<b>99,4</b>	<b>99,7</b>	<b>0,991</b>	<b>0,978</b>
MLP	96,1	96,1	99,6	0,975	0,894
Random Forest	99,2	99,4	99,6	0,996	0,975
SMO	80,6	99,9	80,8	0,499	-0,002

Entre os resultados mostrados na Tabela 8, o classificador com a menor taxa preditiva foi o SMO, tendo 80% na acurácia, 99% de especificidade, 80% em sensibilidade, 0,499 na ROC e -0,002 no K. O de maior taxa preditiva em termos gerais foi o classificador IBK, porém, o classificador *Random Forest* obteve valores relevantes nas métricas de validação. O IBK com uma ROC de 0,991, K de 0,978 e o *Random Forest* uma ROC com 0,996, K de 0,975. O IBK com uma maior acurácia e sensibilidade, e a especificidade igual ao *Random Forest*.

Analisando os resultados apresentados nas Tabelas 5 a 8, com base em todos os experimentos, a Tabela 6 apresentou os melhores resultados, onde, a média de especificidade foi acima de 98%. Também obtendo o melhor resultado dentre todos os experimentos, com 99% de acurácia, 0,999 curva ROC e coeficiente kappa 0,996, com o uso do classificador IBK e os índices de diversidade filogenética, aplicado no canal vermelho (*RED*) atribuindo

os melhores resultados em relação aos testes para as imagens em níveis de cinza e nos canais verde e azul.

Entre todos os classificadores, o IBK mostrou os resultados mais relevantes por ser baseado em aprendizado de instâncias capaz de ter um maior número de acertos com descritores filogenéticos, com taxas de predições médias de 98% de acurácia em todos os canais, seguido pelo *Random Forest* que apresentou uma taxa média de 97% de acurácia em todos os canais, após esses, o MLP que apresentou uma taxa média de 92% de acurácia em todos os canais. Por fim, o que apresentou as piores predições o SMO, que apresentou uma taxa média de 80% de acurácia.

Com intuito de averiguar o potencial dos descritores de textura propostos neste trabalho, foram feitos testes com descritores comumente utilizados na literatura. O descritor escolhido foi *Gray-Level-Co-Occurrence Matrix* (GLCM). O GLCM é um método estatístico para descrição de texturas, capaz de armazenar a probabilidade de que dois valores de intensidade cinza estejam envolvidos por uma relação espacial. A partir da matriz de probabilidades, foram usadas as seguintes medidas para extrair as características: Contraste, dissimilaridade, homogeneidade, correlação, energia e o segundo momento angular (ASM). Os testes foram iguais ao da metodologia proposta, feitos com a mesma base de imagens. A Tabela 9 mostra os resultados dos testes do descritor GLCM.

Tabela 9: Teste com o descritor GLCM.

Classificadores	Acurácia (%)	Especificidade (%)	Sensibilidade (%)	ROC	Kappa
IBK	67,7	80,1	16,1	0,481	-0,037
<b>MLP</b>	<b>81,4</b>	<b>81,3</b>	<b>4,0</b>	<b>0,695</b>	<b>0,061</b>
Random Forest	77,0	80,9	21,0	0,590	0,014
SMO	80,7	100	-	0,500	0

Diante dos resultados apresentados na Tabela 9, o classificador MLP mostrou a melhor taxa preditiva com valores de 81,4 de acurácia, 81,3 para especificidade, 4,0 em sensibilidade, 0,695 na ROC e 0,061 no Kappa. O pior caso mostrou-se no classificador IBK, com taxa de 67,7 na acurácia, 80,1 em especificidade, 16,1 de sensibilidade, 0,481 na ROC e -0,037 no Kappa. Analisando os resultados obtidos com o descritor GLCM e o da metodologia proposta, feitos os testes com a mesma base de imagens, nota-se que a pesquisa proposta apresentou uma maior taxa preditiva nas métricas de validação, que determina o nível de relevância do estudo. O melhor resultado do GLCM apresentou 81,4 de acurácia, 81,3 para especificidade, 4,0 em sensibilidade, 0,695 na ROC e 0,061 no Kappa e a metodologia proposta apresentou 99% na acurácia, 99% para especificidade, 100% em sensibilidade, 0,999 na ROC e 0,996 no K.

A Tabela 10 apresenta uma comparação da metodologia deste trabalho com os trabalhos relacionados citados no Capítulo 3.

Tabela 10: Comparativo com os Trabalhos Relacionados.

Trabalho	Metodologia	Base	Acurácia (%)	Especificidade (%)	Sensibilidade (%)
(ALENCAR, 2015)	Contornos + MLP	PH2	86,6	-	-
(RASTGOO et al., 2016)	SIFIT + Random Forests	PH2	-	100	90,3
(MOURA et al., 2017)	Descritor Híbrido + MLP	PH2	95,5	95,4	95,5
(LI; SHEN, 2018)	LIN CNN	ISIC	85,7	96,6	49
(BONAFINI; PERLIN, 2018)	Regras ABCDE + SVM	ISIC	84,5	-	-
(MAIA et al., 2019)	Inception + Resnet + Random Forest	PH2	92	92,5	91,8
<b>Metodologia</b>	<b>Descritores Filogenéticos + IBK</b>	<b>ISIC</b>	<b>99,8</b>	<b>99,9</b>	<b>100</b>

A Tabela 10 faz um comparativo entre a metodologia proposta e os trabalhos relacionados, avaliando a acurácia, especificidade e sensibilidade. Como são bases de imagens diferentes deste trabalho, este comparativo visa apenas ilustrar as predições obtidas em cada metodologia. Onde, os resultados da metodologia proposta mostram-se promissores em comparação aos demais trabalhos. Na metodologia de Bonafini e Perlin (2018) foi obtido uma acurácia de 84,5%, sendo a menor taxa entre os trabalhos relacionados. O trabalho de Moura et al. (2017) obteve as melhores valores preditivos entre os trabalhos relacionados, tendo abaixo só a especificidade em relação ao trabalho de Rastgoo et al. (2016) com o valor de 100. Por fim, baseado nos valores das predições, a metodologia proposta apresentou resultados satisfatórios diante da bibliografia.

## 6 Conclusão

Com o aumento da incidência de doenças, os especialistas necessitam cada vez mais de mecanismos para uma segunda opinião em seus diagnósticos de doenças. O uso de sistemas computadozizados tornou-se recorrente no auxílio do diagnóstico das doenças, por serem precisos em seus resultados.

Este trabalho apresentou uma metodologia para classificação em melanoma ou não, com o uso de índices de diversidade filogenética em imagens, acoplado com o reconhecimento de padrões, capazes de discriminar e classificar, em imagens, as anomalias nas regiões suspeitas que desencadeiam o melanoma, atribuindo uma segunda opinião para os especialistas.

Os índices de diversidade filogenética mostraram-se eficientes na categorização de regiões em melanoma e não melanoma. Foram obtidos resultados promissores nas métricas de validação, com taxas de acertos de 99%.

Como análise nos resultados obtidos, os descritores baseados em textura mostraram resultados promissores na classificação do melanoma, sendo possível proporcionar aos especialistas uma segunda opinião no diagnóstico do melanoma.

Pretende-se, como trabalhos futuros:

- Aumentar as amostras de melanoma e não melanoma utilizando mais bases de imagens;
- Aplicar um método de segmentação nas imagens de pele;
- Aplicar os índices filogenéticos em outras lesões de pele.

## Referências

- ALENCAR, F. E. S. Desenvolvimento de um sistema para classificação automática de imagens dermatoscópicas para dispositivos móveis. *Universidade do Estado do Rio Grande do Norte, Biblioteca Central Orlando Teixeira, BR-RN*, 2015. Citado 3 vezes nas páginas 24, 25 e 31.
- AZEVEDO-MARQUES, P. M. de. Diagnóstico auxiliado por computador na radiologia. *Radiologia Brasileira*, SciELO Brasil, v. 34, n. 5, p. 285–293, 2001. Citado na página 14.
- BAKOS, L. Melanoma cutâneo: estudos de base populacional no brasil cutaneous melanoma: population-based studies in brazil. *An bras dermatol*, SciELO Brasil, v. 81, n. 5, p. 402, 2006. Citado na página 13.
- BONAFINI, B. L.; PERLIN, H. A. Um estudo sobre reconhecimento de padrões aplicado a detecção de câncer do tipo melanoma maligno. *Revista Científica Interdisciplinar INTERLOGOS*, v. 3, n. 1, p. 137–154, 2018. Citado 3 vezes nas páginas 24, 25 e 31.
- BRAGA, A. Curvas roc: aspectos funcionais e aplicações. 2001. Citado na página 22.
- BRAZIL, S. M. I. Melanoma cutâneo no brasil. *Arquivos Catarinenses de Medicina*, v. 38, n. Suplemento 01, p. 14, 2009. Citado na página 17.
- CARVALHO, E. D. et al. Method of differentiation of benign and malignant masses in digital mammograms using texture analysis based on phylogenetic diversity. *Computers & Electrical Engineering*, Elsevier, v. 67, p. 210–222, 2018. Citado 2 vezes nas páginas 8 e 19.
- CASTRO, A. A. M. D.; PRADO, P. P. L. D. Algoritmos para reconhecimento de padrões. *Revista Ciências Exatas*, v. 8, n. 2002, 2002. Citado na página 20.
- COSTA, E. S. et al. Análise da mortalidade e distribuição de neoplasia cutânea em sergipe no período de 2010 a 2015. In: *Congresso Nacional de Enfermagem-CONENF*. [S.l.: s.n.], 2019. v. 1, n. 1. Citado na página 16.
- GOMES. *Cancer de pele*. 2013. <https://www.minhavidade.com.br/saude/temas/cancer-de-pele/>. [Online; accessed 19-November-2017]. Citado na página 16.
- GONZALEZ, R. C.; WOODS, R. C. *Processamento digital de imagens*. [S.l.]: Pearson Educación, 2009. Citado na página 17.
- GUERRA, R. L. et al. Custo-utilidade de terapias-alvo comparadas à dacarbazina para o tratamento de primeira linha do melanoma avançado não-cirúrgico e metastático no sistema único de saúde do brasil. *Value in Health Regional Issues*, Elsevier, v. 20, p. 103–109, 2019. Citado na página 16.
- HALL, M. A. Correlation-based feature selection of discrete and numeric class machine learning. University of Waikato, Department of Computer Science, 2000. Citado na página 21.

- HARVEY, P. H.; RAMBAUT, A. Comparative analyses for adaptive radiations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, The Royal Society, v. 355, n. 1403, p. 1599–1605, 2000. Citado na página 19.
- INCA. *Instituto Nacional do Câncer*. 2018. <http://www2.inca.gov.br/wps/wcm/connect/inca/portal/home/>. [Online; accessed 19-November-2017]. Citado 2 vezes nas páginas 13 e 16.
- ISIC. *COLIVAC*. 2016. <https://challenge.kitware.com/#challenge/560d7856cad3a57cfde481ba>. [Online; accessed 19-November-2018]. Citado na página 26.
- KNOB, A. H. Aplicação do processamento de imagens digitais para análise da anisotropia da massa de grãos. 2011. Citado 2 vezes nas páginas 8 e 17.
- LI, Y.; SHEN, L. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 2, p. 556, 2018. Citado 3 vezes nas páginas 24, 25 e 31.
- MAGURRAN, A. E. *Measuring biological diversity*. [S.l.]: John Wiley & Sons, 2013. Citado na página 19.
- MAIA, L. B. et al. Aprendizagem profunda aplicada ao diagnóstico de melanoma. Universidade Federal do Maranhão, 2019. Citado 3 vezes nas páginas 24, 25 e 31.
- MOURA, N. et al. Proposta de um descritor híbrido para aprimoramento da identificação automática de melanoma. In: SBC. *17<sup>o</sup> Workshop de Informática Médica (WIM 2017)*. [S.l.], 2017. v. 17, n. 1/2017. Citado 3 vezes nas páginas 24, 25 e 31.
- OLIVEIRA, F. S. S. d. et al. Classificação de tecidos da mama em massa e não-massa usando índice de diversidade taxonômico e máquina de vetores de suporte. Universidade Federal do Maranhão, 2013. Citado na página 18.
- OLIVEIRA, R. B. Método de detecção e classificação de lesões de pele em imagens digitais a partir do modelo chan-vese e máquina de vetor de suporte. Universidade Estadual Paulista (UNESP), 2012. Citado na página 13.
- OSHIRO, T. M. *Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 21.
- PANCHAL GAURANG E GANATRA, A. e. K. Y. e. P.-D. Análise comportamental de perceptrons multicamadas com múltiplos neurônios ocultos e camadas ocultas. *Jornal Internacional de Teoria e Engenharia de Computadores*, v. 3, n. 2, 2011. Citado na página 21.
- PIANKA, E. *Evolutionary Ecology. 5a Edição*. [S.l.]: Editora Harper Collins College Publishers, 1994. Citado na página 18.
- PURI, R.; JAIN, V. Barcode detection using opencv-python. *Science*, v. 4, n. 1, p. 97–99, 2019. Citado na página 26.

- RASTGOO, M. et al. Classification of melanoma lesions using sparse coded features and random forests. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2016: Computer-Aided Diagnosis*. [S.l.], 2016. v. 9785, p. 97850C. Citado 3 vezes nas páginas 24, 25 e 31.
- SAMIS, T. *doença de pele*. 2006. <https://revistas.pucsp.br/index.php/RFCMS/article/view/74/pdf/>. [Online; accessed 1-September-2006]. Citado na página 13.
- SOARES, H. B. Análise e classificação de imagens de lesões da pele por atributos de cor, forma e textura utilizando máquina de vetor de suporte. Universidade Federal do Rio Grande do Norte, 2008. Citado na página 14.
- SOUSA, U. S. et al. Classificação de massas na mama a partir de imagens mamográficas usando índice de diversidade de shannon-wiener. Universidade Federal do Maranhão, 2011. Citado na página 18.
- WEKA. *WEKA*. 2018. <https://www.cs.waikato.ac.nz/ml/weka/>. [Online; accessed 19-November-2017]. Citado 2 vezes nas páginas 20 e 26.
- ZENG ZHI-QIANG E YU, H.-B. e. X. H.-R. e. X. Y.-Q. e. G. J. Máquinas de vetores de suporte de treinamento rápido usando otimização mínima sequencial paralela. In: . [S.l.: s.n.], 2008. v. 1, p. 997–1001. Citado na página 21.



**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA  
“JOSÉ ALBANO DE MACEDO”**

**Identificação do Tipo de Documento**

- ( ) Tese  
( ) Dissertação  
 Monografia  
( ) Artigo

Eu, Wellington João da Silva Sousa Júnior,  
autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de  
02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar,  
gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação  
Análise filogenética aplicada a exploração de textura de imagens  
de pele para auxílio no diagnóstico do melanoma  
de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título  
de divulgação da produção científica gerada pela Universidade.

Picos-PI 02 de julho de 2019.

Wellington João da Silva Sousa Júnior  
Assinatura

Wellington João da Silva Sousa Júnior  
Assinatura