

**CENTRO UNIVERSITÁRIO PRESIDENTE TANCREDO DE ALMEIDA NEVES –
UNIPTAN**

CURSO DE MEDICINA

SARA DETOMI TEIXEIRA

**DIAGNÓSTICO DE MELANOMA: REVISÃO ACERCA DA
COMPARAÇÃO ENTRE DERMATOLOGISTAS E INTELIGÊNCIA
ARTIFICIAL**

SÃO JOÃO DEL REI, JULHO 2020

SARA DETOMI TEIXEIRA

**DIAGNÓSTICO DE MELANOMA: REVISÃO ACERCA DA
COMPARAÇÃO ENTRE DERMATOLOGISTAS E INTELIGÊNCIA
ARTIFICIAL**

Trabalho de Conclusão do Curso,
apresentado para obtenção do grau de
médico no Curso de Medicina do Centro
Universitário Presidente Tancredo de
Almeida Neves, UNIPTAN.

Orientadores: Prof. MSc. Henrique Alvarenga da Silva
Prof. DSc Luiz Eduardo Canton Santos

SÃO JOÃO DEL REI, JULHO 2020

SARA DETOMI TEIXEIRA

**DIAGNÓSTICO DE MELANOMA: REVISÃO ACERCA DA COMPARAÇÃO ENTRE
DERMATOLOGISTAS E INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso
aprovado pela Banca Examinadora para
obtenção do Grau de médico, no Curso
de Medicina do Centro Universitário
Presidente Tancredo de Almeida Neves,
UNIPTAN.

São João Del Rei, 19 de julho de 2020.

BANCA EXAMINADORA

Prof. Henrique Alvarenga da Silva – Mestre – (Centro Universitário Presidente
Tancredo de Almeida Neves) – Orientador

Prof. Luiz Eduardo Canton Santos - Doutor - (Centro Universitário Presidente
Tancredo de Almeida Neves) – Orientador

Prof. Daniel Riani Gotardelo – Doutor – (Centro Universitário Presidente Tancredo de
Almeida Neves) – Orientador

RESUMO

Introdução: O melanoma atualmente é diagnosticado por meio do exame a olho nu e dermatoscopia, seguidos de determinação definitiva pelo exame anatomopatológico. Porém, a inteligência artificial possui enorme potencial na área, razão pela qual é necessário determinar seu estado da arte e suas possíveis contribuições para a medicina. **Metodologia:** Foi realizada revisão de pesquisas que compararam a acurácia de diagnósticos de melanoma realizados por dermatologistas e pela inteligência artificial. Ao total, 6 (seis) estudos originais foram incluídos, sendo 2 (dois) sobre a mesma pesquisa. **Resultados:** O número de profissionais, de variados níveis de formação, comparados à inteligência artificial variou de 8 a 157. A maioria dos treinamentos e testes dos estudos utilizou imagens provenientes *do International Skin Imaging Collaboration (ISIC)*. O número total de imagens utilizadas para o treinamento variou de 274 a 12.378 e para o teste de 65 a 808. Todos os artigos definiram a sensibilidade e a especificidade tanto dos algoritmos quanto dos profissionais. A sensibilidade média dos médicos variou entre 67,2% a 86,6% e a especificidade entre 58,7% a 72,65%. A sensibilidade média do algoritmo foi de 74,1% a 96% e a especificidade de 42,5% a 86,5%. **Discussão:** Todos os estudos mostraram que a inteligência artificial pode colaborar de forma eficaz no diagnóstico do melanoma, porém, não se pode olvidar das questões técnicas, éticas e legais que permeiam o assunto. **Conclusão:** A inteligência artificial e os dermatologistas, ambos detentores de potências e fraquezas, devem ser vistos como aliados. Afinal, nenhuma tecnologia é capaz de substituir a relação médico-paciente.

Palavras-chave: Dermatologia. Diagnóstico. Inteligência Artificial. Melanoma.

ABSTRACT

Introduction: Melanoma is currently diagnosed by examination with the naked eye and dermatoscopy, followed by definitive determination by anatomopathological examination. However, artificial intelligence has a huge potential in the area, that is why it is necessary to determine its state of the art and its possible contributions to medicine. **Methodology:** A review was made with studies that compared the accuracy of melanoma diagnoses performed by dermatologists and artificial intelligence. In total, 6 (six) original studies were included, 2 (two) on the same research. **Results:** The number of professionals, of varying levels of training, compared to artificial intelligence ranged from 8 to 157. Most of the training and tests of the studies used images from the International Skin Imaging Collaboration (ISIC). The total number of images used for training ranged from 274 to 12378 and for the test from 65 to 808. All articles defined the sensitivity and specificity of both the algorithms and the professionals. The average sensitivity of doctors varied between 67.2% to 86.6% and specificity between 58.7% to 72.65%. The average sensitivity of the algorithm was 74.1% to 96% and the specificity was 42.5% to 86.5%. **Discussion:** All studies have shown that artificial intelligence can collaborate effectively in the diagnosis of melanoma, however, cannot be forgotten that technical, ethical and legal issues that permeate the subject. **Conclusion:** Artificial intelligence and dermatologists, both possessing strengths and weaknesses, should be seen as allies. After all, no technology is capable of replacing the doctor-patient relationship.

Keywords: Dermatology. Diagnosis. Artificial intelligence. Melanoma.

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVOS	11
2.1	Objetivo principal	11
2.2	Objetivos específicos	11
3	ARTIGO	12
3.1	Introdução	13
3.2	Metodologia	14
3.2.1	Critérios de inclusão e exclusão de estudos	14
3.2.2	Estratégia de busca	14
3.2.3	Síntese e comparação dos estudos	14
3.3	Resultados	16
3.3.1	Profissionais	16
3.3.2	Imagens	17
3.2.3	Comparações dos diagnósticos	18
3.4	Discussão	21
3.5	Conclusão	24
3.6	Referências	25
4	REFERÊNCIAS	28

1 INTRODUÇÃO

O termo inteligência artificial (IA), cunhado por John McCarthy na ocasião da primeira conferência sobre o tema em 1956, no Dartmouth College, é hoje uma das tecnologias mais promissoras¹. Desde então esse campo de estudos cresceu de forma impressionante. Atualmente, usando o *deep learning*, uma versão refinada de redes neurais artificiais, é possível, com base em um conjunto de algoritmos desenvolvidos, fazer com que o computador não só identifique as imagens a que foi apresentado, como também aprenda, como autodidata, a reconhecer e analisar imagens nunca vistas, ou seja, quanto mais imagens analisa, melhor a acurácia da identificação².

Dentre todas as aplicações da IA, o reconhecimento de imagens por algoritmos inteligentes é uma das áreas mais populares nas últimas décadas. Dentre essas, destacam-se as pesquisas na área de análise de imagens médicas.

A área da radiologia (imagiologia) e da patologia, que trabalham fundamentalmente com extração da informação a partir de imagens, foram as primeiras a experimentarem o uso da IA em seus processos. O uso da IA se mostra também promissor na dermatologia, na qual o processo diagnóstico se baseia, principalmente, na análise de imagens de lesões de pele, e o melanoma tem se destacado como principal objeto de estudos nessa área.

O melanoma, tumor originário dos melanócitos, possui geralmente localização cutânea primária, podendo surgir em outras áreas como olhos, mucosas, meninges e outros³. Apesar de a etiopatogenia ainda não completamente esclarecida, é sabido que alguns fatores influenciam no estabelecimento e desenvolvimento da doença, como: genética, exposição solar, fototipo, número de nevos e síndrome do nevo displásico (SND)³. Mas, ainda que o fator mais importante, qual seja: exposição solar excessiva, seja muito conhecido, as pessoas não adotam medidas capazes de reduzir o risco de adquirir a doença⁴.

É considerado o mais maligno dentre os cânceres de pele e atinge milhares de pessoas em todo o mundo a cada ano. Em 2013, no Brasil, foram registradas 1.547 mortes em razão do melanoma⁵. Ainda, 6.260 novos casos da doença devem surgir

no país entre 2018 e 2019⁵.

Tendo em vista que pode rapidamente deixar de ser localizado para se tornar tumor invasivo, o diagnóstico precoce e preciso se torna imperioso para o bom prognóstico da doença.

Atualmente, além do exame a olho nu, para a investigação desta patologia, o dermatologista faz uso do dermatoscópio. A dermatoscopia é o exame da pele por meio de um sistema óptico ligado a uma fonte de luz que permite a magnificação e visualização em profundidade de estruturas, formas e cores⁶. Portanto, como permite acessar a estrutura abaixo da pele de forma não invasiva, é uma técnica que complementa e confirma o diagnóstico. Quando o médico, então, determina que a lesão é melanocítica, deve classificá-la em benigna, suspeita ou altamente suspeita⁷, em geral, utilizando a regra do ABCD (assimetria, irregularidade da borda, variação da cor, diâmetro maior que 0,6 cm), posteriormente, precisa fazer a biópsia, meio invasivo e muitas vezes demorado, para determinação do grau da lesão. Afinal, o exame anatomopatológico é insubstituível para confirmação diagnóstica³. Essa forma de diagnóstico, hoje considerada o “padrão-ouro”, é indiscutivelmente muito eficiente. No entanto, devemos nos atentar para subjetividade destes métodos, totalmente dependentes do conhecimento e experiência do profissional atuante.

Ilustrando esta problemática, Andrea Sboner⁸ observou o modelo de tomada de decisão para diagnóstico de melanoma por 6 (seis) dermatologistas, 3 (três) especialistas e 3 (três) não especialistas. Verificou-se, então, que não existia um consenso entre as diretrizes a serem utilizadas para o diagnóstico pelos médicos, o que reforça a ideia da presente e perigosa subjetividade em diagnósticos tão importantes.

Entretanto, o processo tradicional, com observação direta pelo médico, não é a única forma de diagnóstico. O *Machine Learning*, em palavras mais simples, consiste na técnica em que o pesquisador mostra ao computador várias imagens que correspondem a lesões de pele, dentre elas: o melanoma. Então, o sistema gradualmente aprende a classificar o que é ou não uma lesão maligna, até que em um determinado momento se torna capaz de classificar imagens nunca analisadas anteriormente por ele, o que permite dizer que aprendeu a forma de identificar as novas imagens submetidas à sua análise. Essas novas tecnologias, então, possuem grande capacidade para aumentar a objetividade diagnóstica, como também para garantir outros benefícios como celeridade, precisão, menor custo e outros. É

necessário, portanto, não só conhecer a inteligência artificial por ser método diagnóstico de enorme potencial, mas também por ser já ser realidade, ensejando a adaptação dos profissionais, visando sempre melhorar o atendimento do paciente.

2 OBJETIVOS

2.1 Objetivo Geral

Analisar o estado da arte da aplicação da inteligência artificial no diagnóstico do melanoma.

2.2 Objetivos Específicos

- a) Pesquisar artigos científicos que tiveram como objetivo o uso da inteligência artificial no diagnóstico do melanoma;
- b) Comparar resultados das pesquisas em termos de acurácia e sensibilidade conforme métodos utilizados;
- c) Identificar nas discussões se o uso da tecnologia foi capaz de superar dermatologistas experientes no tocante ao diagnóstico de melanoma;
- d) Determinar o estado da arte da aplicação da inteligência artificial no diagnóstico do melanoma;
- e) Analisar criticamente possíveis benefícios e malefícios do uso da tecnologia para pacientes e médicos.
- f) Identificar os desafios a serem enfrentados pela implantação da inteligência artificial como ferramenta diagnóstica.

3 ARTIGO

DIAGNÓSTICO DE MELANOMA: REVISÃO ACERCA DA COMPARAÇÃO ENTRE DERMATOLOGISTAS E INTELIGÊNCIA ARTIFICIAL

DIAGNOSIS OF MELANOMA: REVIEW OF THE COMPARISON BETWEEN DERMATOLOGISTS AND ARTIFICIAL INTELLIGENCE

Sara Detomi Teixeira,

CENTRO UNIVERSITÁRIO PRESIDENTE TANCREDO DE ALMEIDA NEVES

3.1 INTRODUÇÃO

O melanoma tem alta mortalidade, sendo “o número de mortes em decorrência dessa patologia três vezes maior que o de todas as demais neoplasias cutâneas em conjunto” ¹, principalmente em decorrência da capacidade metastática, razão pela qual o diagnóstico precoce e preciso se torna imperioso para o bom prognóstico da doença ². Com o exame a olho nu, a precisão no diagnóstico do melanoma é de apenas 65% a 80%, mas aumenta para 75% a 84% com o uso da dermatoscopia ³. A determinação definitiva do diagnóstico é finalmente obtida pelo exame anatomopatológico ¹, hoje considerado o “padrão-ouro”.

Atualmente, se uma lesão é classificada como melanocítica, deve ser categorizada em benigna, suspeita ou altamente suspeita ¹, em geral, utilizando a regra do ABCD (assimetria, irregularidade da borda, variação da cor, diâmetro maior que 0,6 cm). Entretanto, o processo tradicional, com observação direta pelo médico, não é a única forma de diagnóstico. No intuito de favorecer um diagnóstico precoce, soluções diagnósticas de melanoma utilizando inteligência artificial estão se tornando cada vez mais comuns ⁴⁻¹¹. Algoritmos que aprendem a classificar o que é ou não lesão maligna se encontram em constante evolução.

Em 2003, uma metanálise comprovou que a acurácia de algoritmos computadorizados era confiável e comparável a de dermatologistas ¹², e, em 2009, em outra metanálise, Rapjara et al. ⁴ concluíram que o diagnóstico do melanoma pela inteligência artificial e pelos especialistas com auxílio de dermatoscópios digitais era similar. Mais recentemente, em 2017, Esteva et al. ¹³, em estudo relevante na área, criaram algoritmos do tipo *Deep Learning Convolutional Neural Networks* (CNNs) capazes de desenvolver suas próprias regras diagnósticas e, ainda, obter a mesma acurácia dos especialistas ao classificar o câncer de pele.

Visto seu enorme potencial, convém, portanto, determinar o estado da arte das novas tecnologias, a fim de saber quais são as suas possíveis contribuições para a medicina. Este artigo apresenta, então, uma revisão de recentes pesquisas que compararam diagnósticos de melanoma realizados por dermatologistas e pela inteligência artificial.

3.2 METODOLOGIA

3.2.1 Critérios de inclusão e exclusão de estudos

Foram incluídos estudos originais, do tipo experimentais, que apresentassem resultados da comparação dos diagnósticos de melanomas realizados por dermatologistas e por algoritmos por meio da inteligência artificial. Os demais critérios de inclusão foram: data da publicação do estudo entre janeiro de 2015 e setembro de 2019, publicados em inglês, terem selecionado imagens de melanoma para o estudo, terem determinado a sensibilidade e especificidade de médicos e algoritmos com relação à classificação do melanoma. Foram excluídos estudos em que o foco do tipo de lesão dermatológica não era o melanoma, que não apresentavam desenvolvimento de um novo algoritmo e que não comparavam os resultados do algoritmo desenvolvido com médicos que atuam na área dermatológica.

3.2.2 Estratégia de busca

A pesquisa dos artigos foi realizada na base de dados PubMed, utilizando-se os termos *Artificial intelligence*, *Dermatologist*, *Melanoma*, *Machine learning* e foram encontrados 29 (vinte e nove) artigos. Foram pesquisadas também as referências bibliográficas dos artigos incluídos.

Realizou-se uma primeira avaliação, tendo por base os títulos e o resumo dos artigos e foram rejeitados aqueles que não preencheram os critérios de inclusão. Quando um estudo não pode ser incluído ou rejeitado com certeza, o texto completo foi analisado secundariamente. Ao total, 6 (seis) artigos foram selecionados para este trabalho.

3.2.3 Síntese e comparação dos estudos

Foi realizada síntese narrativa das pesquisas selecionadas, apresentadas segundo o número e as características dos profissionais submetidos ao estudo que atuam na área dermatológica. Dentre as características, optou-se por elencar informações como anos de experiência e cargo hierárquico dentro da

medicina, número de imagens e tipos de lesões nelas contidas nos treinamentos e testes realizados, as taxas de sensibilidade e de especificidade do diagnóstico dos profissionais quanto dos algoritmos encontradas pelos pesquisadores.

3.3 RESULTADOS

3.3.1 Profissionais

O número total de profissionais avaliados pelas pesquisas variou de 8 a 157. Todos os estudos relataram dados com relação aos anos de experiência prática na área dermatológica pelos profissionais submetidos e/ou seus níveis hierárquicos de formação.

No trabalho de Ferris et al.¹⁴, o diagnóstico assistido pelo computador foi acompanhado por 30 médicos que possuíam experiência em dermatologia, sendo eles 12 dermatologistas, 10 residentes e 8 clínicos que atuam na dermatologia.

Já, Marchetti et al., em um primeiro estudo, em 2018¹⁵, analisaram o diagnóstico realizado por 8 dermatologistas especialistas em câncer de pele, originados de 4 países, que tinham a média de experiência de 13 anos (3 a 31), e no ano seguinte¹⁶, esses mesmos dermatologistas foram incluídos em novo estudo, assim como mais 9 residentes da área, somando 17 profissionais.

Por outro lado, Haenssle et al.¹⁷, ampliaram o número de profissionais e desenvolveram a pesquisa com 58 dermatologistas, originados de 17 países, que foram classificados em 3 grupos conforme anos de experiência: 17 iniciantes (menos de 2 anos), 11 como treinados (2 a 5 anos) e 30 especialistas (mais de 5 anos).

Por fim, Brinker et al., em um primeiro momento, expandiram significativamente o número de profissionais a serem comparados com a performance de um algoritmo e analisaram as respostas de 157 dermatologistas, dentre eles 56,1% eram residentes e 43,9% eram especialistas, provenientes de 12 hospitais universitários alemães e cuja média de experiência era de 4 anos¹⁸. Em segundo momento, aplicaram questionários a 144 dermatologistas, de nove hospitais universitários alemães, 52 foram respondidos por dermatologistas certificados e 92 dermatologistas juniores¹⁹.

Tabela 1. Quantidade e qualificação de profissionais comparados à IA nas pesquisas

Autor, ano (referência)	Total	Dermatologistas	Residentes
Ferris et al., 2015 ¹⁴	30*	12	10
Marchetti et al., 2018 ¹⁵	8	8	
Marchetti et al., 2019 ¹⁶	17	8	9
Haenssle et al., 2018 ¹⁷	58	58**	
Brinker et al., 2019 ¹⁸	157	69	88
Brinker et al., 2019 ¹⁹	144	144***	

* 8 outros

** Dos 58, 17 iniciantes, 11 especializados e 30 especialistas

*** Dos 144, 52 certificados e 92 juniores

3.3.2 Imagens

Para que um algoritmo seja capaz de classificar novas imagens a ele apresentadas, deve ser treinado e alimentado com um banco de dados. A maioria dos treinamentos e testes nos estudos analisados utilizou imagens provenientes no *International Skin Imaging Collaboration* (ISIC), que, além de propor padrões de abordagem tanto tecnológica, quanto terminológica, das imagens dermatológicas, disponibiliza uma coleção pública de imagens de lesões para serem utilizadas no ensino e desenvolvimento de tecnologia ²⁰. Houve variação conforme o número imagens e tipos de lesões selecionados para cada etapa.

Marchetti et al., em 2018¹⁵, após selecionarem 1279 imagens deste banco de dados internacional, utilizaram 900 imagens para treinamento e 100 imagens para o teste, que continha 50 imagens de melanomas e 50 de lesões benignas. Em 2019 ¹⁶, com o crescimento do número de imagens no ISIC, escolheram 2750 imagens, das quais 2000 foram utilizadas no treinamento e 150 no teste, dentre estas: 50 melanomas, 50 nevos melanocíticos e 50 ceratoses seborreicas.

Brinker et al. treinaram a *Convolutional Neural Network* (CNN) com imagens de 1888 melanomas e 10490 nevos atípicos, já o teste baseou-se em 100 imagens (20 melanomas e 80 nevos atípicos) ¹⁸. Meses após, para treinamento da CNN usaram 4204 imagens do ISIC, ou seja, aproximadamente duas vezes menos em relação ao estudo anterior, e aumentaram o número de

imagens para o teste dos dermatologistas para 808, sendo 402 melanomas e 402 nevos¹⁹.

Dados sobre o número de imagens utilizadas para treinar a CNN não foram fornecidos por Haenssle et al.¹⁷. No entanto, fizeram a seleção de 300 imagens, incluindo 20% de melanomas e 80% de nevos melanocíticos benignos. Posteriormente, dois dermatologistas experientes selecionaram 100 imagens deste conjunto que foram usadas para testes da CNN em comparação com dermatologistas.

Ferris et al.¹⁴ optaram por maior variedade de tipos de lesões para o treinamento do algoritmo. As imagens malignas foram compostas por 105 melanomas, 29 nevos displásicos de alto grau, 23 carcinomas basocelulares e 3 carcinomas de células escamosas, enquanto dentre as benignas encontravam-se 93 lesões melanocíticas, 4 dermatofibromas, 14 ceratoses seborreicas e 3 angiomas. Para o teste do classificador de imagens de lesões melanocíticas foram usadas imagens de 39 melanomas, de 11 carcinomas basocelulares, 3 carcinomas de células escamosas e 120 de lesões benignas. E para o teste dos dermatologistas foram selecionadas 65 lesões, das quais 25 eram melanomas e 40 lesões benignas.

Tabela 2. Quantidade de imagens utilizadas no treinamento e nos testes

Autor, ano (referência)	Treinamento	Teste
Ferris et al., 2015 ¹⁴	274	238*
Marchetti et al., 2018 ¹⁵	900	100
Marchetti et al., 2019 ¹⁶	2000	150
Haenssle et al., 2018 ¹⁷	300	100
Brinker et al., 2019 ¹⁸	12378	100
Brinker et al., 2019 ¹⁹	4204	808

Nota: *173 para teste do classificador e 65 para teste dos dermatologistas

3.3.3 Comparações dos diagnósticos

Todos os artigos eleitos para análise definiram a sensibilidade e a especificidade tanto dos algoritmos quanto dos dermatologistas, dados que possibilitaram, então, a comparação.

No estudo de Ferris et al.¹⁴, o conjunto de 30 praticantes da dermatologia

alcançou a sensibilidade média para detecção de melanoma de 70,8% e especificidade de 58,7%. A sensibilidade dos clínicos foi de 80,5%, enquanto a dos residentes foi de 70,4% e dos especialistas foi de 64,7%. Já a especificidade dos clínicos foi 48,1%, dos residentes 59% e dos especialistas 65,4%. Para as mesmas imagens, o classificador de imagens de lesões dermatológicas atingiu sensibilidade de 96% e especificidade de 42,5%. E, diante desses dados, os autores concluíram que a tecnologia é capaz de auxiliar na decisão de biopsiar a lesão dermatológica.

Em 2018, Marchetti et al.¹⁵ concluíram que, ao classificarem as lesões em benignas ou malignas, a sensibilidade dos 8 dermatologistas participantes da pesquisa era de 82% e a especificidade de 59%. Então, tomando a sensibilidade como 82%, o algoritmo individual alcançou a especificidade de 62%, já o algoritmo de fusão de 76%. Neste estudo, os autores perceberam que a tecnologia excedeu apenas alguns dermatologistas submetidos à comparação. Em 2019, a sensibilidade dos 8 dermatologistas foi de 76% e a especificidade foi de 72,65%, enquanto a sensibilidade dos residentes foi de 56% e a especificidade de 76,3%. Por sua vez, utilizando 76% para sensibilidade, o algoritmo conseguiu a especificidade de 85%¹⁶.

Os 58 dermatologistas analisados por Haenssle et al.¹⁷, em um primeiro momento, quando a eles foram apresentadas apenas as imagens, obtiveram sensibilidade média de 86,6% e especificidade média de 71,3%. A sensibilidade e a especificidade foram 82,9% e 67,6% para iniciantes, 85,9% e 68,5% para treinados e 89,0% e 74,5% para especialistas, respectivamente. Após 4 semanas, quando às imagens foram adicionadas informações clínicas, a sensibilidade geral foi de 88,9% e a especificidade geral foi de 75,7%. A CNN, por outro lado, no teste com 100 imagens atingiu 95% de sensibilidade e 63,8% de especificidade e no com 300 alcançou 95% de sensibilidade e 80% de especificidade. Em suas conclusões, ainda, apontaram que a maioria dos dermatologistas foi superada pela inteligência artificial.

No que concerne à pesquisa de Brinker et al.¹⁸, a sensibilidade média dos 157 dermatologistas foi de 74,1% e a especificidade de 60%. Sendo a sensibilidade de 74,1%, a CNN alcançou 86,5% de especificidade. Já, tomando o valor de 60% para a especificidade, a sensibilidade foi de 87,5%. Já em estudo posterior¹⁹, esses mesmos autores determinaram que a sensibilidade e

especificidade médias dos 144 dermatologistas foi de 67,2% e 62,2%, respectivamente. Sendo a sensibilidade e especificidade dos médicos certificados, 63,2% e 65,2% e dos juniores, 68,9% e 58%. Por sua vez, a CNN obteve 82,3% de sensibilidade e 77,9% de especificidade. Após esta segunda pesquisa¹⁹, os autores afirmaram que foi a primeira vez que a CNN demonstrou ser significativamente superior aos dermatologistas, independente de sua experiência, com relação ao diagnóstico de melanoma.

Tabela 3. Dados de artigos incluídos na revisão de acordo com autor, ano de publicação, sensibilidade e especificidade de médicos e de algoritmos e conclusões principais

Autor ano (referência)	Sensibilidade médicos	Especificidade médicos	Sensibilidade algoritmo	Especificidade algoritmo
Ferris et al., 2015 ¹⁴	70,8%	58,7%	96%	42,5%
Marchetti et al., 2018 ¹⁵	82%	59%	82%	76%
Marchetti et al., 2019 ¹⁶	76%	72,6%	76%	85%
Haenssle et al., 2018 ¹⁷	86,6%	71,3%	95%	63,8%
Brinker et al., 2019 ¹⁸	74,1%	60%	74,1%	86,5%
Brinker et al., 2019 ¹⁹	67,2%	62,2%	82,3%	77,9%

3.4 DISCUSSÃO

A classificação de lesões possui dois objetivos primordiais: fornecer informações específicas e opções de tratamento e detectar precocemente o câncer de pele com sensibilidade e especificidade razoáveis²¹. Conforme se observa nos estudos analisados, para que esses objetivos sejam alcançados é almejado um diagnóstico cada vez mais preciso. Foi constatado que, ao longo dos anos, a tecnologia está se aperfeiçoando e, cada vez mais, se assemelhando a dermatologistas, e é nítido o avanço da inteligência artificial nesta esfera.

Os estudos variaram com relação ao número de profissionais e os seus anos de experiência prática na área dermatológica e/ou seus níveis hierárquicos de formação. Quando foram estabelecidas diferenças entre sensibilidade e especificidade entre profissionais de níveis hierárquicos diferentes, foi possível perceber, de modo geral, que a especificidade é maior dentre aqueles que possuem maior experiência^{14,17,19}. Já a sensibilidade tende a ser maior dentre os profissionais com menos experiência ou que não são especialistas^{14,19}, o que talvez se justifique pela menor confiança, então, em caso de dúvida classificam a lesão como maligna.

Conforme apontam Li et al.²², a performance da inteligência artificial quanto ao diagnóstico do melanoma é permeada por inúmeros problemas. Dentre eles, a insuficiência das bases de dados de imagens de doenças dermatológicas e a falta de compartilhamento entre hospitais é algo que influencia diretamente o treinamento das máquinas, que exigem a formação de “*big data*”.

Comprovando a ideia de que quanto mais dados fornecidos para alimentação do sistema de aprendizagem, maior precisão será alcançada, é possível reforçar o fato de que no teste de Haenssle et al.¹⁷, quando utilizadas 200 imagens a mais, houve aumento de 16,2% de especificidade, enquanto a sensibilidade foi mantida.

Visando a criação de padrões para imagens dermatológicas e a criação de um arquivo de acesso público para servir como recurso de imagens para aprendizado e desenvolvimento de tecnologias, foi fundado o *International Skin Imaging Collaboration: Melanoma Project*²⁰. E, de todos os estudos analisados,

quatro^{15,16,18,19} utilizaram imagens provenientes deste projeto.

O aumento dos dados a serem fornecidos permite a atenuação do impacto de variáveis como iluminação, zoom, cor, brilho, contraste e rotação de uma imagem, que poderiam alterar substancialmente a tomada de decisão pelas máquinas²³. Estes autores até mesmo sugerem, como possível solução, que a própria IA poderia ajudar a orientar ou automatizar a captura de imagens ou dados para que sejam encaixados em padrões e parâmetros predefinidos.

Ainda sobre a complexidade de captura de imagens, Mar et al.²⁴ comentam a dificuldade de se realizar um exame de toda a pele do paciente, visto que existem áreas, a exemplo do couro cabeludo e de superfícies sacrais, que são difíceis de serem fotografadas. É possível afirmar, então, que existe a questão da qualidade das imagens. Brinker et al.¹⁹, em último estudo, ao verificarem as imagens com resultados de biópsia, chegaram à conclusão de que a quantidade de imagens para treinamento não é tão determinante quanto à qualidade. Todas as demais pesquisas analisadas¹⁴⁻¹⁸ treinaram seus algoritmos com base em imagens confirmadas por decisões consensuais, colaborando, então, para o risco de a aprendizagem pela tecnologia se basear em alguns erros de julgamentos.

Ainda, apesar de para a dermatologia a inspeção ter alta relevância, a palpação pode ser essencial para a identificação de algumas lesões, inclusive de alguns melanomas²⁴. Ou seja, durante o exame clínico do paciente, o médico consegue obter mais informações para a tomada de decisão²¹. Então, a consideração de dados clínicos, como idade, sexo, histórico do paciente, histórico familiar e outras informações podem ajudar na obtenção de diagnóstico e prognóstico mais precisos²². Por exemplo, a aquisição de dados clínicos, conforme apresentado por Haensle et al.¹⁷, aumentou a performance de dermatologistas. Assim como poderiam aumentar a eficácia da classificação pela IA se forem adicionados aos métodos de aprendizado²¹.

Além das questões técnicas, o diagnóstico pela IA envolve questões éticas e legais²². Arcabouço jurídico-normativo regulamenta a responsabilidade civil do médico, caberia questionar, então, como se daria a responsabilização pelo erro do sistema e possível dano ao paciente.

Dentro deste contexto, também é preciso refletir acerca dos princípios norteadores da relação médico-paciente, como privacidade e confidencialidade,

que serão desafiados com a implantação da inteligência artificial. Afinal, a tecnologia geralmente depende de dados registrados em plataformas eletrônicas para que aprenda e consiga diagnosticar com maior precisão. Então, os profissionais podem ter dificuldades morais ao decidir as informações confidenciais que podem ser relatadas nos registros eletrônicos ²⁵. Ainda, é preciso atentar para o fato de que recentemente noticiaram que terceiros, como empregadores, serão capazes de acessar, controlar e monitorar essas informações, além do fato de as empresas de tecnologia não informarem sobre quem controla as escolhas e ações dos sistemas, sendo incerta a proteção da privacidade do paciente²⁵.

Essa situação, então, ensejará que profissionais das mais diversas áreas se esforcem para a criação de códigos de condutas que visem a salvaguardar direitos e princípios protetores tanto do paciente quanto do profissional submetidos à inteligência artificial.

Apesar dos inúmeros obstáculos a serem enfrentados, as vantagens a serem alcançadas pela inteligência artificial devem ser elucidadas. Sobre este ponto, Teixeira e Silva² as resumem e as ilustram de forma clara: “pode reduzir o tempo para o diagnóstico do melanoma, fator decisivo em seu prognóstico”; “pode facilitar o acesso ao diagnóstico em áreas remotas”; “pode reduzir custos”; “pode aumentar a produtividade e a confiabilidade dos diagnósticos médicos”.

No Brasil, poderia, inclusive, ser solução para a distribuição heterogênea de dermatologistas, visto que: “Em 2010, no Brasil, havia apenas 1 dermatologista para cada 23.000 habitantes, e apenas 9,1% dos municípios contavam com dermatologistas (SCHMITT; MIOT, 2014). Portanto, o acesso a especialistas, como dermatologistas, caros e escassos em muitas regiões, ainda é privilégio de poucos” ².

Tanto a inteligência artificial quanto os dermatologistas possuem vantagens e desvantagens. E o que se observa na grande maioria dos estudos atuais é a contraposição entre a atuação dos médicos e das máquinas em relação à precisão diagnóstica, razão, inclusive, da elaboração dessa pesquisa. No entanto, é preciso atentar para o poder do encontro de ambos. Em razão dos diagnósticos em equipe serem considerados mais precisos, Hekler et al.²¹ propuseram estudo pioneiro que demonstrou que quando as decisões dos dermatologistas e os resultados da tecnologia foram combinados, melhoraram

ainda mais o desempenho da inteligência artificial. E, além deste dado objetivo, em estudo qualitativo, elaborado por Nelson et al.²⁶, os próprios pacientes se demonstraram receptivos ao uso da inteligência artificial no diagnóstico de câncer de pele, se preservada a relação médico-paciente, ou seja, ao invés de substituir o médico, a tecnologia funcionaria como mais uma fonte de informações para a tomada de decisão do profissional. Afinal, “apesar de todo poder disruptivo de uma nova tecnologia, o médico, humano, será sempre insubstituível quando se tratar de cuidar do seu paciente e a tecnologia apenas é um instrumento que potencializa este cuidado”².

3.5 CONCLUSÃO

Máquinas não devem ser vistas como concorrentes, mas como complementares ao diagnóstico médico. Médicos, cientistas da computação, biomédicos e máquinas com inteligência artificial, devem trabalhar em conjunto, de forma multidisciplinar para que seja possível o diagnóstico cada vez mais precoce e eficaz.

Por mais que a análise da imagem da lesão possa um dia ser exclusividade da inteligência artificial, o dermatologista continuará tendo que analisar as informações geradas pelo computador, ou seja, caberá a ele o raciocínio clínico, quem decidirá a melhor conduta para aquele caso específico e quem acompanhará o paciente como pessoa humana.

3.6 REFERÊNCIAS

1. Azulay RD, Azulay DR, Azulay-Abulafia L. *Dermatologia*. 7a ed. Rio de Janeiro: Guanabara Koogan; 2017.
2. Teixeira SD, Silva HA. A inteligência artificial na dermatologia. In: Silva Neto BR, editor. *Ciências da Saúde: Da Teoria à Prática*. Atena Editora; 2019.
3. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, et al. Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. *J Med Internet Res [Internet]*2018;20(10):e11936.
4. Rajpara SM, Botello AP, Townend J, Ormerod AD. Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. *Br J Dermatol*. England; 2009 Sep;161(3):591–604.
5. Garcia Arroyo JL, Garcia Zapirain B. Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis. *Comput Biol Med*. United States; 2014 Jan;44:144–57.
6. Li L, Zhang Q, Ding Y, Jiang H, Thiers BH, Wang JZ. Automatic diagnosis of melanoma using machine learning methods on a spectroscopic system. *BMC Med Imaging*. England; 2014 Oct;14:36.
7. Saez A, Sanchez-Monedero J, Gutierrez PA, Hervas-Martinez C. Machine Learning Methods for Binary and Multiclass Classification of Melanoma Thickness From Dermoscopic Images. *IEEE Trans Med Imaging*. United States; 2016 Apr;35(4):1036–45.
8. Jaworek-Korjakowska J, Kleczek P. Automatic Classification of Specific Melanocytic Lesions Using Artificial Intelligence. *Biomed Res Int*. United States; 2016;2016:8934242.
9. Gareau DS, Correa da Rosa J, Yagerman S, Carucci JA, Gulati N, Hueto F, et al. Digital imaging biomarkers feed machine learning for melanoma screening. Vol. 26, *Experimental dermatology*. Denmark; 2017. p. 615–8.
10. Premaladha J, Ravichandran KS. Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms. *J Med Syst*. United States; 2016 Apr;40(4):96.
11. Safran T, Viezel-Mathieu A, Corban J, Kanevsky A, Thibaudeau S, Kanevsky J. Machine learning and melanoma: The future of screening. *J Am Acad Dermatol*. United States; 2018 Mar;78(3):620–1.

12. Rosado B, Menzies S, Harbauer A, Pehamberger H, Wolff K, Binder M, et al. Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. *Arch Dermatol*. United States; 2003 Mar;139(3):361–7; discussion 366.
13. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* [Internet]. Macmillan Publishers Limited, part of Springer Nature. All rights reserved.; 2017 Jan 25;542:115.
14. Ferris LK, Harkes JA, Gilbert B, Winger DG, Golubets K, Akilov O, et al. Computer-aided classification of melanocytic lesions using dermoscopic images. *J Am Acad Dermatol*. United States; 2015 Nov;73(5):769–76.
15. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kalloo A, et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *J Am Acad Dermatol*. United States; 2018 Feb;78(2):270–277.e1.
16. Marchetti MA, Liopyris K, Dusza SW, Codella NCF, Gutman DA, Helba B, et al. Computer Algorithms Show Potential for Improving Dermatologists Accuracy to Diagnose Cutaneous Melanoma; Results of ISIC 2017. *J Am Acad Dermatol* [Internet]. Elsevier; 2019 Jul 20; Available from: <https://doi.org/10.1016/j.jaad.2019.07.016>
17. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol Off J Eur Soc Med Oncol*. England; 2018 Aug;29(8):1836–42.
18. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer*. England; 2019 May;113:47–54.
19. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer*. England; 2019 Sep;119:11–7.
20. ISIC Archive. [Accessed November 2, 2019]; <https://isic-archive.com/>

21. Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, et al. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer*. England; 2019 Oct;120:114–21.
22. Li CX, Shen CB, Xue K, Shen X, Jing Y, Wang ZY, Xu F, Meng RS, Yu JB, Cui Y. Artificial intelligence in dermatology: past, present, and future. *Chin Med J* 2019;00:00–00. doi: 10.1097/CM9.0000000000000372
23. Novoa RA, Gevaert O, Ko JM. Marking the Path Toward Artificial Intelligence-Based Image Classification in Dermatology. *JAMA dermatology*. United States; 2019 Aug;
24. Mar VJ, Soyer HP. Artificial intelligence for melanoma diagnosis: how can we deliver on the promise? Vol. 29, *Annals of oncology : official journal of the European Society for Medical Oncology*. England; 2018. p. 1625–8.
25. Nabi J. How Bioethics Can Shape Artificial Intelligence and Machine Learning. *Hastings Cent Rep*. United States; 2018 Sep;48(5):10–3.
26. Nelson CA, Perez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. *JAMA dermatology*. 2020.

4 REFERÊNCIAS

1. Moor J. The Dartmouth College artificial intelligence conference: The next fifty years. *Ai Magazine*. 2006; 27(4):87-87. Available from: <https://doi.org/10.1609/aimag.v27i4.1911>
2. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA* [Internet]. 2016 Dec 13;316(22):2353–4. Available from: <https://doi.org/10.1001/jama.2016.17438>
3. Azulay RD, Azulay DR, Azulay-Abulafia L. *Dermatologia*. 7a ed. Rio de Janeiro: Guanabara Koogan; 2017.
4. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014 Nov 19;515:S109. Available from: <https://doi.org/10.1038/515S109a>
5. INCA. Estimativa 2018: Incidência de Câncer no Brasil. INCA. Rio de Janeiro. 2017. Available from: http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pele_melanoma/definicao+
6. Malvey J, Puig S, Argenziano G, Marghoob AA, Soyer HP. Dermoscopy report: proposal for standardization. Results of a consensus meeting of the International Dermoscopy Society. *J Am Acad Dermatol*. United States; 2007 Jul;57(1):84–95.
7. Azulay RD, Azulay DR. *Dermatologia*. 3a ed. Rio de Janeiro: Guanabara Koogan; 2004.
8. Sboner A, Aliferis CF. Modeling clinical judgment and implicit guideline compliance in the diagnosis of melanomas using machine learning. *AMIA. Annu Symp proceedings AMIA Symp*. United States; 2005;664–8.