



Reimaginando a eliminação da hanseníase com análise de IA de uma combinação de imagens de lesões cutâneas com dados demográficos e clínicos

Barbieri, a *,¹ Kevin Yixi Xu, b *,¹ Lucy Setian, c Paulo Thiago Souza-Santos, a Anusua Trivedi, b Jim Cristofono, b Ricardo Bhering, a Raquel R White, b Anna M Sales, a Geralyn Miller, b Jose Augusto C Nery, a Michael Sharman, b Richard Bumann, b Shun Zhang, b Mohamad Goldust, d, e Euzenir N Sarno, a Fareed Mirza, c Arielle Cavaliero, c Sander Timmer, b Elena Bonfiglioli, b Cairns Smith, f David Scollard, g Alexander A. Navarini, d Ann Aerts, c Juan Lavista Ferres, b *, e Milton O Moraes *

^a Laboratório de Hanseníase Instituto Oswaldo Cruz, Fundação C, ~ao Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brasil

^b Microsoft, One Microsoft Way, One Microsoft Way, Redmond, WA, Estados Unidos

^c Fundação Novartis, Basileia, Suíça

^d Universidade de Basileia, Basileia, Suíça

^e Departamento de Dermatologia, University Medical Center Mainz, Mainz, Alemanha

^f Universidade de Aberdeen, Aberdeen, Escócia

^g Aposentado, Wilbraham, MA, Estados Unidos

Resumo

Contexto A hanseníase é uma doença infecciosa que afeta principalmente populações carentes. Embora tenha sido em grande parte eliminada, ainda são diagnosticados anualmente cerca de 200.000 novos pacientes. Na ausência de um teste de diagnóstico, o diagnóstico clínico é muitas vezes tardio, podendo levar a danos neurológicos irreversíveis e ao estigma resultante, bem como à transmissão continuada. A aceleração do diagnóstico poderia contribuir significativamente para o avanço da eliminação global da lepra. A tecnologia impulsionada pela Inteligência Digital e Artificial (IA) demonstrou potencial para aumentar as capacidades dos profissionais de saúde na realização de diagnósticos mais rápidos e precisos, especialmente quando se utilizam imagens, como nas áreas de dermatologia ou oftalmologia. Isso nos fez iniciar a busca por um assistente de diagnóstico de hanseníase baseado em IA, baseado em imagens de pele.

Métodos Aqui descrevemos a precisão de um assistente de diagnóstico de hanseníase baseado em imagens habilitado para IA, chamado AI4Le-prosy, com base em uma combinação de imagens de pele e dados clínicos, coletados seguindo um processo padronizado. Em um centro nacional de referência em hanseníase no Brasil, foram incluídos 222 pacientes com hanseníase ou outras condições dermatológicas, e as 1.229 imagens de pele coletadas e 585 conjuntos de metadados são armazenados em um conjunto de dados de código aberto para outros pesquisadores explorarem.

Resultados Usamos esse conjunto de dados para testar se um algoritmo de IA baseado em CNN poderia contribuir para o diagnóstico da hanseníase e empregamos três modelos de IA, testando imagens e metadados de forma independente e em combinação. A modelagem por IA indicou que os sinais clínicos mais importantes são perda de sensibilidade térmica, nódulos e pápulas, parestesia nos pés, número de lesões e sexo, mas também superfície descamativa e prurido que foram negativamente associados à hanseníase. O uso da regressão logística com rede elástica proporcionou alta acurácia de classificação (90%) e área sob a curva (AUC) de 96,46% para o diagnóstico de hanseníase.

Interpretação A validação futura destes modelos está em andamento, reunindo conjuntos de dados maiores de populações de diferentes tipos de pele e coletando imagens com câmeras de smartphones para imitar configurações do mundo real. Esperamos que os resultados da nossa investigação conduzam a soluções clínicas que ajudem a acelerar a eliminação global da lepra.

Financiamento Este estudo foi parcialmente financiado pela Fundação Novartis e pela Microsoft (contribuição em espécie).

Copyright 2022 Os Autores. Publicado pela Elsevier Ltd. Este é um artigo de acesso aberto sob a licença CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Palavras-chave: Hanseníase; Inteligência artificial; IA; Diagnóstico baseado em imagens; Dermatologia; Lesões de pele; AI4lepra

The Lancet Regional
Saúde - Américas

2022;9: 100192

Publicado on-line em 3 de fevereiro

ano de

2022 <https://doi.org/10.1016/j.lana.2022.100192>

* Autores correspondentes.

Endereços de e-mail: jlavista@microsoft.com (JL Ferres), milton.moraes@fiocruz.br (MO Moraes).

¹ Co-primeiros autores.

Pesquisa em contexto

Evidências antes do estudo

De acordo com as Diretrizes de Hanseníase da OMS (2018) o diagnóstico deve ser baseado na avaliação clínica, idealmente com assistência de esfregaços de pele ou biópsias de pele ou nervos exame histopatológico. Um dos seguintes sinais cardinais devem ser detectados, como a presença das micobactérias, através de testes de coloração ácido-resistente esfregaços cutâneos ou presença de manchas periféricas espessadas nervosismo. A maioria dos métodos laboratoriais disponíveis exibe baixa sensibilidade, enquanto qPCR, que apresenta maior sensibilidade, ainda precisa de reprodutibilidade e padronização com boas práticas de fabricação. Além disso, todos os métodos, desde esfregaços cutâneos até biópsias, são invasivos e trabalhoso, exigindo infraestrutura que não está disponível em muitas configurações. O diagnóstico precoce é fundamental para a hanseníase controle, mas o atraso no diagnóstico tem impedido novas conquistas no sentido da redução da incidência.

Tecnologias digitais de saúde, como aplicativos baseados em IA poderia melhorar o diagnóstico da hanseníase, uma vez que foi aplicado com sucesso no diagnóstico de melanoma e outras condições de pele.

Valor acrescentado deste estudo

De 2018 a 2020, protocolo de pesquisa para coleta de imagens de alta resolução, de pacientes com hanseníase e outras doenças dermatológicas, juntamente com aspectos clínicos e dados demográficos foram implementados na FIOCRUZ clínica, Rio de Janeiro, Brasil. As imagens foram armazenadas, rotuladas e algoritmos de inteligência artificial foram testados. Dados mostram que a precisão do modelo de IA para detectar hanseníase previamente identificada por dermatologistas atingiu 90% e uma área sob a curva (AUC) de 96,46% quando a rede elástica foi utilizada regressão logística. Aqui fornecemos uma prova de conceito de que a IA pode ser efetivamente usada para determinar a probabilidade de lepra.

Implicações de todas as evidências disponíveis

Com validação independente adicional, este modelo de IA pode ser o primeiro passo para aliviar os desafios em torno doenças de pele. Este modelo de IA pode levar a uma abordagem não invasiva abordagem, usando um conjunto específico de preditores, que é fácil implementar em telefones celulares para estimar a probabilidade de hanseníase. No futuro, esta abordagem poderá ajudar profissionais de saúde no encaminhamento de pacientes suspeitos para centros de referência, quando necessário, o que por sua vez proporciona maior precisão nas estratégias de saúde pública no controle da carga de doenças.

Introdução

A hanseníase é uma doença tropical negligenciada (DTN) para a qual O diagnóstico é muitas vezes tardio porque os sintomas desaparecem entre dois meses a 20 anos para aparecer e doença a progressão é lenta.¹ Embora a análise histológica e

qPCR de biópsias de pele pode auxiliar no diagnóstico de hanseníase, assim como Coloração de Ziehl-Neelsen para detecção de Mycobacterium leprae em esfregaços cutâneos de lóbulos das orelhas,² atualmente não há nenhum teste diagnóstico considerado padrão-ouro para lepra. Até o momento, o diagnóstico da hanseníase permanece clínico.¹ A diminuição da experiência em hanseníase entre os profissionais de saúde trabalhadores, e o atraso na procura de cuidados de saúde causado pelo estigma associado à hanseníase, muitas vezes resulta em atraso no diagnóstico, quando o dano neurológico já ocorreu, levando a incapacidades irreversíveis.

Embora a ampla disponibilidade de terapia multimedicamentosa (PQT) gratuita tenha assegurado uma redução de 99% do carga global de lepra, ainda existem até 200.000 pacientes recém-diagnosticados anualmente, e novas ferramentas são necessário para percorrer os últimos quilômetros rumo à eliminação da hanseníase,³ definida como transmissão zero. Uma estratégia recentemente validada é o rastreamento de contatos seguido de profilaxia pós-exposição (PEP) para pessoas de contato de pacientes com hanseníase, recomendado pela OMS após a implementação bem-sucedida da iniciativa PEP para Lepra.⁴

A apresentação clínica da hanseníase depende de uma imunidade específica do paciente. A Organização Mundial da Saúde (OMS) classifica a doença em paucibacilar (PB) e multibacilar (MB), para pessoas que apresentam de 1 a 5 ou mais de 6 lesões cutâneas, respectivamente.⁵ Em comparação com uma classificação padrão para fins de pesquisa,⁶ o grupo PB de hanseníase inclui tuberculose tuberculóide (TT) e pacientes borderline tuberculóides (BT), enquanto o MB grupo inclui pacientes com doença limitrofe (BB), limitrofe-virchowiano (BL) e virchowiano (LL) lepra. Diagnóstico diferencial para hanseníase PB inclui manifestações dermatológicas, como aquelas causada por pitiríase alba, sífilis, psoríase, granuloma anulare e sarcoidose, enquanto a hanseníase MB e especificamente as formas LL precisam ser diferenciadas de doenças como linfoma, neurofibromatose e xantoma.

Como demonstrado anteriormente, a tecnologia digital pode acelerar o diagnóstico da hanseníase, permitindo a saúde da linha de frente trabalhadores enviem imagens da pele do celular a um dermatologista de referência para orientação e encaminhamento.⁷ Aumentando as evidências em torno do aprendizado de máquina permitindo diagnósticos baseados em imagens mais rápidos e precisos em disciplinas como radiologia, patologia e dermatologia, nos motivou a desenvolver um Artificial "Acelerador de diagnóstico" orientado por inteligência (IA) para lepra, AI4Lepra, usando uma combinação de pele imagens, sintomas clínicos e relatados.

A maior parte das evidências de diagnóstico baseadas em IA em dermatologia vem do melanoma, e algoritmos como redes neurais convolucionais profundas (CNN) forneceram precisão comparável à dos dermatologistas, na diferenciação de lesões malignas de lesões benignas.⁸ Neurais profundas redes demonstraram exceder os especialistas na diferenciação de melanoma de mimetizadores benignos, como nevos e ceratoses seborreicas em imagens dermatoscópicas.⁹ Conjuntos de dados de treinamento maiores e avanços no algoritmo

o desenvolvimento aumentou ainda mais seu desempenho,¹⁰ ao mesmo tempo em que aumentou o desempenho humano em lesões onde os médicos relataram baixa confiança no diagnóstico.¹¹

Este artigo descreve o processo de desenvolvimento de um modelo de IA que analisa lesões de pele, metadados e informações desidentificadas do paciente para determinar a probabilidade de hanseníase, explicando a aquisição e compilação de imagens em um banco de dados de código aberto de imagens de alta resolução, sua análise CNN e combinação com dados clínicos. Além disso, descrevemos a metodologia para armazenamento de imagens de pele e sua rotulagem como lepra ou lesões semelhantes à lepra. Nossas imagens desidentificadas incluem uma diversidade de manifestações clínicas e tipos de pele e estão disponíveis abertamente em um repositório de dados para uso posterior por pesquisadores e profissionais de saúde. Combinando imagens, dados clínicos e demográficos de pacientes com hanseníase confirmada e pacientes com lesões cutâneas semelhantes à hanseníase, treinamos algoritmos de IA para diferenciar a hanseníase de outras condições. Até onde sabemos, este é o primeiro grande banco de dados e imagens de código aberto disponível para hanseníase e modelo de IA para suspeita de hanseníase. Nossos resultados indicam que a modelagem de IA pode funcionar como uma ferramenta poderosa para acelerar e aumentar a precisão da suspeita de hanseníase para confirmação adicional do diagnóstico.

Métodos

Ética

A coleta de dados e imagens foi realizada no ambulatório de hanseníase da Fundação Oswaldo Cruz, no Rio de Janeiro, centro de referência nacional que recebe diversos pacientes e seus familiares de todo o Brasil.

Entrevistas com pacientes, consultas clínicas e coletas de imagens seguiram processos operacionais padrão. O estudo foi aprovado pelo Comitê de Ética e Pesquisa da Fundação Oswaldo Cruz CAAE:

38053314.2.0000.5248, número: 976.330 y 03/10/2015.

Todos os participantes do estudo forneceram consentimento informado e puderam solicitar esclarecimentos sobre a pesquisa ou seus riscos e benefícios a qualquer momento. Imagens de lesões cutâneas foram desidentificadas e fotos de acompanhamento puderam ser tiradas para avaliar a melhora ou progressão da doença.

Desenho e fluxo de trabalho do

estudo Os critérios de inclusão e exclusão de pacientes e o fluxo de trabalho geral até a análise dos dados de IA estão ilustrados na Figura 1 Complementar. Na primeira consulta, foram examinados os pacientes que apresentavam lesões cutâneas e aqueles diagnosticados com "lesões semelhantes à hanseníase" (mácula), placas ou nódulos) eram elegíveis para inscrição. Esses pacientes receberam informações sobre a pesquisa, preencheram o consentimento informado, bem como informações gerais de saúde e demográficas, juntamente com possíveis sintomas. A fotografia foi tirada e, em seguida, biópsias de pele e esfregaços de pele dos lóbulos das orelhas foram usados para confirmar (ou excluir) a hanseníase antes de convidar os pacientes para participar do

diagnóstico. O diagnóstico de hanseníase foi definido somente após os resultados da histologia e da coloração de Ziehl-Neelsen dos esfregaços cutâneos para detecção de características morfológicas do tecido hanseniano ou *Mycobacterium leprae*, respectivamente. O qPCR só foi realizado quando foram encontrados resultados equivocados na análise histológica.² O diagnóstico final só foi estabelecido na terceira consulta (até 30 dias após a segunda visita), pois os resultados laboratoriais estariam então disponíveis e permitiriam aos dermatologistas comunicar o diagnóstico e iniciar o tratamento adequado. O diagnóstico de hanseníase seguiu a classificação operacional da OMS e de Ridley e Jopling⁶ (Figuras Complementares 2 e 3). Os pacientes confirmados com hanseníase foram tratados com PQT, de acordo com as diretrizes nacionais do Ministério da Saúde e da OMS.⁵

Os dados sociodemográficos dos pacientes foram registrados no sistema de informação de saúde local padrão, incluindo uma etiqueta (nome e número de registro) para identificar os pacientes do estudo. As imagens das lesões cutâneas foram armazenadas em um computador utilizado exclusivamente para o estudo e em um dispositivo de backup, e rotuladas com o anotador de imagens VGG.¹² As avaliações das imagens seguiram os requisitos mínimos da International Skin Imaging Collaboration (ISIC), incluindo cor de fundo, iluminação, campo de visão, foco/profundidade de campo, resolução, escala e calibração de cores.

Detalhes sobre o processo passo a passo para captura e armazenamento de imagens estão descritos na Figura 4 Complementar. No geral, foram tiradas até três imagens de cada lesão cutânea: uma foto panorâmica para identificar a parte do corpo onde a lesão estava localizada, uma foto em close e outra foto da borda da lesão, incluindo a pele normal circundante. O DOI do repositório pode ser acessado em: <https://doi.org/10.35078/1PSIEL>. O código também pode ser visto aqui: <https://github.com/micro-soft/leprosy-skin-lesion-ai-análise>.

Para manter as características originais, as imagens foram armazenadas em formato raw e foram geradas extensões .jpeg para transferência para o Microsoft AzureTM. Metadados específicos do estudo descrevendo dados demográficos, informações clínicas e detalhes de lesões cutâneas foram inseridos no Epi infoTM ¹³ e as imagens foram carregadas na nuvem para desenvolver os algoritmos de IA.

Critérios de inclusão e exclusão Foram

incluídos todos os pacientes com lesões cutâneas semelhantes à hanseníase e com idade superior a seis anos, exceto aqueles que não forneceram consentimento informado ou que foram diagnosticados com HIV ou tuberculose. O consentimento dos pais foi necessário para menores de 18 anos e foram excluídos os pacientes que não compareceram para o diagnóstico final na visita 3.

Exame clínico e coleta de dados

Os dermatologistas avaliaram os critérios de inclusão e exclusão antes de convidar os pacientes para participar do

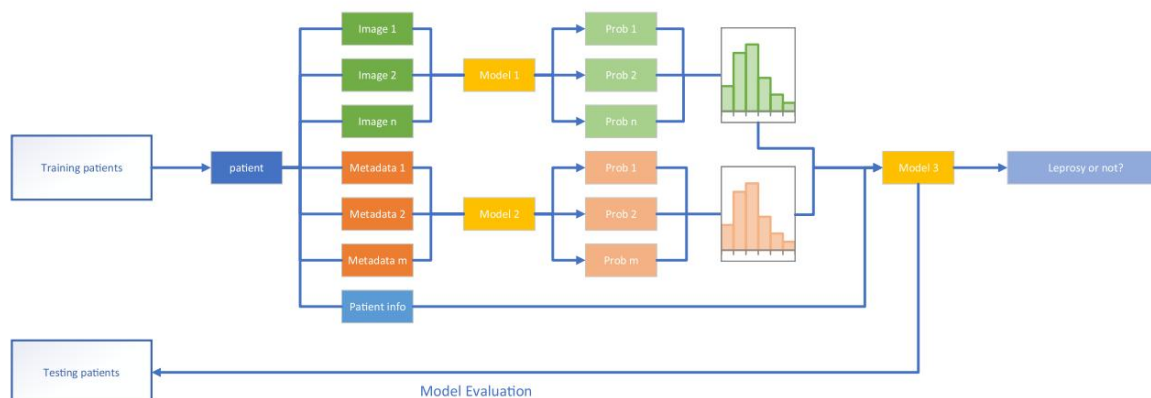


Figura 1. Visão geral da modelagem de dados. Dos 228 pacientes recrutados, 222 pacientes foram finalmente incluídos na análise dos dados. Imagens (modelo 1) ou metadados (modelo 2) de 182 pacientes foram usados para treinar os algoritmos em um conjunto de dados de treinamento, enquanto 40 pacientes foram separados como um grupo de teste independente usado apenas para validação no modelo 3. Os resultados dos modelos 1 e 2 foram os histogramas que alimentaram os cálculos de precisão e área sob a curva (AUC).

estudo, preencheu os formulários do estudo e vinculou cada imagem de lesão cutânea aos metadados correspondentes do paciente.

Durante a consulta, os dermatologistas avaliaram sintomas nos locais das lesões, como dor, prurido, perda de sensibilidade ou hiperestesia, e verificaram a presença de parestesia nas mãos e nos pés. O espessamento dos nervos periféricos também foi avaliado por exame dermatoneurológico. Além disso, os dermatologistas avaliaram se havia infiltração cutânea difusa, perda de sobrancelhas, liquenificação ou descamação na superfície da lesão. Foram registradas a cor (eritematosa, hipo ou hiperpigmentada) e a localização corporal (membros superiores, membros inferiores ou tronco) das lesões cutâneas, bem como seu tipo (mácula, pápula, placa, nódulo, úlcera, vesícula ou bolha).).

O diâmetro das lesões cutâneas foi medido e para lesões ≥ 4 cm, a temperatura foi avaliada na lesão e na área de pele saudável adjacente, bem como contralateralmente. Por fim, foi verificada a sensibilidade (térmica, dolorosa e tátil) para aquelas lesões cutâneas maiores que 1,5 cm, resultando em nódulos, por exemplo, não sendo avaliados quanto à temperatura ou sensibilidade. A sensibilidade nas lesões cutâneas da hanseníase foi avaliada por meio de teste clínico padrão realizado durante o exame dermatológico. Para avaliar a sensibilidade térmica foram utilizados tubos contendo água morna e fria, um bastão pontiagudo ou agulha foi utilizado para avaliar a sensação de dor e a sensação tátil foi avaliada com um pedaço de algodão. Tabelas demográficas e descritivas foram construídas para todas as variáveis, utilizando estatística contínua ou discreta para comparar variáveis demográficas e de metadados.

Modelagem de dados

Como os modelos preditivos que usamos exigem dimensões de entrada iguais, desenvolvemos um modelo de duas etapas no nível do paciente, primeiro prevendo a probabilidade de hanseníase com base na imagem da lesão cutânea (Modelo 1) ou nos metadados

(Modelo 2). Cada modelo produziu uma probabilidade de hanseníase para cada imagem ou conjunto de metadados, conforme mostra a [Figura 1](#). Dado que os pacientes poderiam ter múltiplas lesões ou registros de metadados, combinamos os resultados de ambos os modelos por paciente em um histograma, para representar as probabilidades previstas. Por fim, o Modelo 3 foi treinado para combinar a análise feita na primeira etapa com as informações do paciente. Esta última etapa estabeleceu a probabilidade global combinando os histogramas dos Modelos 1 e 2, com informações do paciente.

Pré-processamento de

dados Quando alguma das variáveis teve poucas observações, elas foram agrupadas com a variável de ocorrência mais frequente. Por exemplo, combinamos "Nódulo" e "Pápula" em um novo grupo para a variável "tipo". Se as lesões fossem muito pequenas, o diâmetro e as temperaturas não eram medidos e o diâmetro seria imputado como 1 cm, enquanto a temperatura seria substituída pela mediana das temperaturas do paciente. Se nenhuma temperatura pudesse ser medida para um paciente, seria usada a mediana das temperaturas de todos os pacientes. O valor "ausente" tornou-se uma categoria adicional para perda sensorial de lesões que eram pequenas demais para medir a função sensorial. Os preditores que consistiam principalmente em um único valor foram eliminados.

Papel da fonte de

financiamento A Fundação Novartis estabeleceu o conceito de AI4leprosy, forneceu informações técnicas na fase de concepção da iniciativa e garantiu a coordenação geral da iniciativa, embora não estivesse envolvida na análise de dados. A Microsoft forneceu experiência em ciência de dados e crédito em espécie na nuvem para o desenvolvimento do modelo.

Resultados

Um total de 228 pacientes diagnosticados com hanseníase ou outras manifestações dermatológicas foram recrutadas. Três pacientes que não aderiram ao agendamento de consultas e não confirmaram o diagnóstico (visita 3) foram excluído. Idade, sexo e características clínicas gerais de os 225 pacientes incluídos são apresentados na Tabela Suplementar 1. Devido à ausência de imagens em close, três pacientes foram excluídos, resultando em um total final de 222 pacientes do estudo (Figura 5 suplementar). Desses, um foram coletadas 582 lesões cutâneas e 1.226 imagens, carregado e armazenado na nuvem. Perda de sobrancelhas ou cílios, infiltração difusa da pele, nervos aumentados e parestesias nos pés foram mais comuns em pacientes com hanseníase confirmada, enquanto a liquenificação e a descamação da superfície das lesões cutâneas foram observadas com mais frequência em pacientes com outras condições dermatológicas (Suplemento Tabela 1). Todos os modelos foram avaliados comparando a probabilidade estimada de hanseníase com o anterior diagnósticos feitos por dermatologistas.

Algoritmos de aprendizagem

Primeiro dividimos os 222 pacientes em dois grupos. Um grupo de 40 pacientes selecionados aleatoriamente, foi usado como testar pacientes para validar os modelos finais em nível de paciente. Os 182 pacientes restantes foram utilizados para ambos os modelos treinamento e seleção. Para cada experimento do Modelo 1 ou 2, usamos validação cruzada de 5 vezes para avaliar o desempenho de cada algoritmo (Figura 6 suplementar).

Para o Modelo 1, combinamos as seguintes configurações (i) a arquitetura de rede neural: Inception-v4 ou

ResNet-50; (ii) estratégia de ajuste: ajuste tudo (ajuste o modelo de rede neural completo) ou congelar (treinar apenas a camada de saída); (iii) tipo de imagem de entrada: apenas close-up ou todas imagens; (iv) otimizador: descida gradiente estocástica. A o ResNet-50 ajustado usando imagens em close deu a melhor precisão de validação cruzada (ACC), bem como a área abaixo da curva, AUC (Tabela 1). Este foi o algoritmo final para treinar o Modelo 1, que gerou a probabilidade histograma para alimentar o Modelo 3, conforme mostrado na Figura 1. O o desempenho do ResNet-50 foi superior (66,6% precisão e 74,56% para AUC), embora menor precisão e a AUC foi alcançada quando tanto o close-up quanto outros imagens foram incluídas (Tabela 1).

Para o Modelo 2, testamos os três métodos de aprendizado de máquina métodos, regressão logística de rede elástica (LR), XGBoost (XGB), florestas aleatórias (RF) (Tabela 2). Temperaturas e os diâmetros são demorados para serem medidos na prática. Assim, avaliamos o impacto da remoção desses dois recursos criando um subconjunto de recursos sem temperaturas e diâmetros. Os dados demonstraram que a regressão logística de rede elástica usando as características do subconjunto alcançou a pontuação de AUC mais alta (88,6%). Esse foi o algoritmo final para treinar o Modelo 2, que gerou o histograma de probabilidade para alimentar o Modelo 3.

Removendo temperatura e diâmetro da pele lesões influenciaram apenas minimamente o desempenho de os algoritmos. Para interpretar o modelo, utilizamos regressão logística de rede elástica com cruzamento repetido de 10 vezes. validação, no conjunto de dados completo, incluindo o recursos de subconjunto. O modelo seleciona as seguintes variáveis: tipo, cor, local, perda sensorial, térmica, tátil, prurido, hiperestesia e assintomático, conforme mostrado

Média ACC e AUC		ResNet-50		Início-v4	
		Sintonizar tudo	Congelar	Sintonizar tudo	Congelar
Toda	ACC (SD)*	0,6138 (0,040)	0,5723 (0,051)	0,5828 (0,070)	0,5212 (0,050)
	AUC (SD)	0,6760 (0,057)	0,6003 (0,094)	0,6144 (0,106)	0,5487 (0,045)
Fechar-se	ACC (SD)	0,6660 (0,099)	0,5790 (0,064)	0,5834 (0,092)	0,5661 (0,059)
	AUC (SD)	0,7456 (0,113)	0,6542 (0,104)	0,6590 (0,099)	0,5919 (0,089)

Tabela 1: A comparação de desempenho para o Modelo 1 (somente imagens) usando arquiteturas de rede neural ResNET-50 e Inception-v4.

* ACC – precisão; AUC-área sob a curva; DP ÿ desvio padrão.

Pontuação média	Regressão de rede elástica (LR)		XGBoost (XGB)		Floresta aleatória (RF)	
	Completo	Subconjunto	Completo	Subconjunto	Completo	Subconjunto
ACC (SD)	0,813 (0,058)	0,817 (0,06)	0,808 (0,086)	0,818 (0,075)	0,779 (0,088)	0,818 (0,073)
AUC (SD)	0,881 (0,082)	0,880 (0,080)	0,849 (0,086)	0,846 (0,092)	0,836 (0,071)	0,863 (0,090)
Sensibilidade (SD)	0,841 (0,118)	0,845 (0,115)	0,818 (0,067)	0,85 (0,092)	0,795 (0,129)	0,845 (0,090)
Especificidade (SD)	0,791 (0,173)	0,794 (0,174)	0,79 (0,177)	0,784 (0,155)	0,763 (0,177)	0,789 (0,158)

Tabela 2: Comparação de desempenho do Modelo 2 utilizando características extraídas da forma da lesão para regressão logística Elastic-net (LR), métodos de aprendizado de máquina XGBoost (XGB) e florestas aleatórias (RF). Para a análise completa dos dados, os dados incluíram 15 preditores coletados na avaliação clínica conforme descrito na seção de métodos. Para análise de subconjunto, as características de temperatura e diâmetro foram excluídas.

Section 1				Section 2				Section 3				Section 4			
Item	Value	Item	Value	Item	Value	Item	Value	Item	Value	Item	Value	Item	Value		
1	10	1	20	1	30	1	40	1	50	1	60	1	70		
2	20	2	30	2	40	2	50	2	60	2	70	2	80		
3	30	3	40	3	50	3	60	3	70	3	80	3	90		
4	40	4	50	4	60	4	70	4	80	4	90	4	100		
5	50	5	60	5	70	5	80	5	90	5	100	5	110		
6	60	6	70	6	80	6	90	6	100	6	110	6	120		
7	70	7	80	7	90	7	100	7	110	7	120	7	130		
8	80	8	90	8	100	8	110	8	120	8	130	8	140		
9	90	9	100	9	110	9	120	9	130	9	140	9	150		
10	100	10	110	10	120	10	130	10	140	10	150	10	160		
11	110	11	120	11	130	11	140	11	150	11	160	11	170		
12	120	12	130	12	140	12	150	12	160	12	170	12	180		
13	130	13	140	13	150	13	160	13	170	13	180	13	190		
14	140	14	150	14	160	14	170	14	180	14	190	14	200		
15	150	15	160	15	170	15	180	15	190	15	200	15	210		
16	160	16	170	16	180	16	190	16	200	16	210	16	220		
17	170	17	180	17	190	17	200	17	210	17	220	17	230		
18	180	18	190	18	200	18	210	18	220	18	230	18	240		
19	190	19	200	19	210	19	220	19	230	19	240	19	250		
20	200	20	210	20	220	20	230	20	240	20	250	20	260		
21	210	21	220	21	230	21	240	21	250	21	260	21	270		
22	220	22	230	22	240	22	250	22	260	22	270	22	280		
23	230	23	240	23	250	23	260	23	270	23	280	23	290		
24	240	24	250	24	260	24	270	24	280	24	290	24	300		
25	250	25	260	25	270	25	280	25	290	25	300	25	310		
26	260	26	270	26	280	26	290	26	300	26	310	26	320		
27	270	27	280	27	290	27	300	27	310	27	320	27	330		
28	280	28	290	28	300	28	310	28	320	28	330	28	340		
29	290	29	300	29	310	29	320	29	330	29	340	29	350		
30	300	30	310	30	320	30	330	30	340	30	350	30	360		
31	310	31	320	31	330	31	340	31	350	31	360	31	370		
32	320	32	330	32	340	32	350	32	360	32	370	32	380		
33	330	33	340	33	350	33	360	33	370	33	380	33	390		
34	340	34	350	34	360	34	370	34	380	34	390	34	400		
35	350	35	360	35	370	35	380	35	390	35	400	35	410		
36	360	36	370	36	380	36	390	36	400	36	410	36	420		
37	370	37	380	37	390	37	400	37	410	37	420	37	430		
38	380	38	390	38	400	38	410	38	420	38	430	38	440		
39	390	39	400	39	410	39	420	39	430	39	440	39	450		
40	400	40	410	40	420	40	430	40	440	40	450	40	460		
41	410	41	420	41	430	41	440	41	450	41	460	41	470		
42	420	42	430	42	440	42	450	42	460	42	470	42	480		
43	430	43	440	43	450	43	460	43	470	43	480	43	490		
44	440	44	450	44	460	44	470	44	480	44	490	44	500		
45	450	45	460	45	470	45	480	45	490	45	500	45	510		
46	460	46	470	46	480	46	490	46	500	46	510	46	520		
47	470	47	480	47	490	47	500	47	510	47	520	47	530		
48	480	48	490	48	500	48	510	48	520	48	530	48	540		
49	490	49	500	49	510	49	520	49	530	49	540	49	550		
50	500	50	510	50	520	50	530	50	540	50	550	50	560		
51	510	51	520	51	530	51	540	51	550	51	560	51	570		
52	520	52	530	52	540	52	550	52	560	52	570	52	580		
53	530	53	540	53	550	53	560	53	570	53	580	53	590		
54	540	54	550	54	560	54	570	54	580	54	590	54	600		

na Tabela 3, cujas estimativas de OR estão resumidas em Tabela Suplementar 2.

Os resultados deste modelo estão alinhados com observações clínicas, dado que, por exemplo, a perda sensorial é uma característica típica da hanseníase ou que as lesões da hanseníase raramente causar prurido.

Para o Modelo 3, os outros dois modelos foram retrainados com o conjunto completo de 182 pacientes, utilizando ResNet-50, respectivamente algoritmos de regressão logística de rede elástica. Esses dois foram os melhores algoritmos selecionados por 5 vezes validação cruzada na primeira etapa. Então treinamos o modelo 3 utilizando todos os 182 pacientes por regressão logística de rede elástica, XGBoost e Random Forest. Os modelos finais em nível de paciente foram validados nos 40 pacientes testados. Random Forest, usando apenas informações do paciente, forneceu a AUC mais alta - 98,74% (Tabela 3). Para XGBoost e Random Forest, há poucos benefícios em incluir Saídas do Modelo 1 e Modelo 2, além do paciente Informação. No entanto, para a regressão logística de rede elástica, a inclusão dos produtos do Modelo 2 aumenta a AUC em comparação com o modelo construído apenas com base nas informações do paciente (Tabela 3).

XGBoost e Random Forest são melhores na captura de relações não lineares do que a regressão logística de rede elástica, e esta pode ser a razão pela qual a inclusão de

Os resultados do modelo 2 aumentam a AUC para regressão logística de rede elástica, mas não para XGBoost ou Random Forest.

Além disso, uma dimensão de entrada maior poderia levar a uma superestimação da precisão do algoritmo, especialmente para modelos complexos de aprendizado de máquina. Tão superequipado como os modelos tendem a produzir resultados excessivamente otimistas, o que pode exagerar seu poder preditivo. Finalmente, campos sobrepostos existiam entre os metadados e as informações do paciente como perda sensorial ou prurido. Embora os valores fossem não são iguais, caracterizaram aspectos semelhantes: por exemplo, 'nódulo' seria registrado tanto na lesão documentada e as informações do paciente. Quanto aos modelos construídos sobre um único recurso de dados, usar apenas os resultados do Modelo 1 produz uma AUC moderada em torno de 70%, enquanto usar resultados do Modelo 2 ou informações do paciente fornece um modelo forte com uma AUC > 90%.

Conforme mostrado na [Tabela 2](#), Random Forest usando pacientes a informação por si só atinge a AUC mais alta nos pacientes testados. As dez características mais importantes que identificamos foram: perda de sensibilidade térmica, nódulos e pápulas, parestesia nos pés, número de lesões, sexo, superfície descamativa, prurido, tronco, ausência de sintomas na lesão cutânea e infiltração difusa ([Figura 2](#)). A importância do recurso de permutação foi medida pela diminuição do modelo precisão ao alterar os valores do recurso. Curiosamente, as estimativas do odds ratio (OR) usando o modelo 2 por regressão logística de rede elástica com informações clínicas confirmou algumas dessas variáveis exibindo os maiores valores de OR (Tabela Suplementar 2).

Devido ao pequeno tamanho da amostra, optamos por também usar regressão logística de rede elástica com resultados do Modelo 2 e informações do paciente. Isto ofereceu um alto

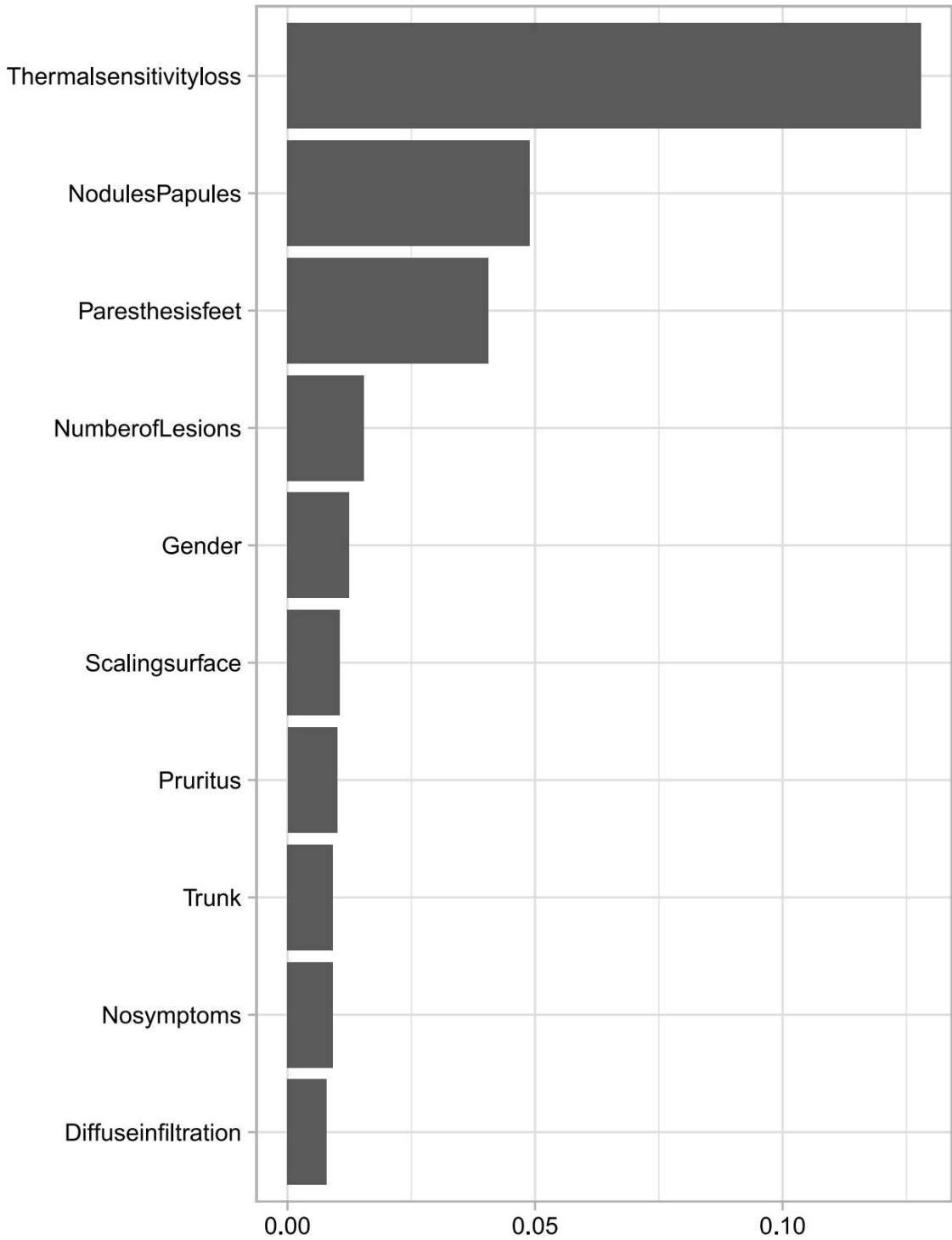


Figura 2. Valor do recurso de permutação, representando os recursos mais importantes após o treinamento do modelo 3.

precisão de classificação (90%) e AUC (96,46%) e foi mais simples e interpretável que XGBoost ou Random Forest. Para definir o modelo final, utilizamos regressão logística de rede elástica com validação cruzada repetida de 10 vezes no conjunto de dados completo, incluindo os pacientes de teste, que está resumido na Tabela Suplementar 3.

As 10 principais características descritas na [Figura 2](#) foram selecionadas por regressão logística, exceto nódulos, pápulas, prurido e ausência de sintomas. As três características restantes foram selecionadas pelo Modelo 2 e o histograma de probabilidade foi utilizado por regressão logística.

Assim, tanto as florestas aleatórias quanto a regressão logística de rede elástica têm uma preferência semelhante em relação às variáveis que utilizam.

Discussão

Nosso estudo foi projetado para fornecer um conjunto de dados de código aberto de imagens de pele e dados clínicos de alta qualidade para avaliar a viabilidade e precisão do modelo de análise de Hanseníase baseado em IA com base em imagens de pele e sintomas clínicos. Não temos conhecimento de outros conjuntos de dados existentes e esperamos que nosso trabalho possa ser uma inspiração para o treinamento adicional de modelos de IA em dermatologia. Como diversas outras DTN apresentam manifestações cutâneas, o modelo aqui descrito poderia ser um exemplo para replicação e auxiliar outras pesquisas sobre DTN, que podem ser doenças debilitantes e muitas vezes afetam as populações mais pobres do mundo. Entendemos que a segurança e a privacidade dos dados são uma grande preocupação e, por isso, definimos um protocolo que evita o reconhecimento individual. Embora forneçamos acesso ao conjunto de dados, ele será restrito e disponibilizado somente mediante cadastro e validação do usuário.

Os resultados de nossa pesquisa indicam que os modelos de probabilidade foram capazes de reconhecer a Hanseníase com alta precisão (96,4%), especialmente quando combinaram os resultados do modelo de regressão Elastic-net 2 e as informações do paciente. Outro assistente de diagnóstico dermatológico baseado em CNN (AIDDA) oferece 89,4% de precisão para o diagnóstico de psoríase e 92,57% para dermatite atópica e eczema.

A disponibilidade da AIDDA em um aplicativo para smartphone provou aumentar significativamente a precisão do diagnóstico para profissionais de saúde menos especializados.¹⁴ Com base nas evidências da pesquisa da AIDDA, na próxima fase da pesquisa sobre AI4Leprosy, treinaremos o algoritmo coletando imagens e dados por meio de um aplicativo para smartphone, para ajudar a validar e melhorar ainda mais o modelo em imagens de qualidade inferior e, em última análise, imitar melhor as configurações do mundo real. Prevê-se que esta implementação aconteça em ambientes rurais e urbanos remotos em ambientes asiáticos e africanos, com o objectivo de remediar qualquer viés de selecção que teria sido introduzido no nosso estudo através da recolha de dados num centro de referência em Hanseníase. Treinar o modelo em ambientes de linha de frente e em outros tipos de pele será essencial, embora o Brasil tenha uma população bastante diversificada que permitiu a coleta de imagens de diversos tipos de pele.

Um estudo anterior da Índia também testou o diagnóstico de Hanseníase baseado em imagens da pele com IA,¹⁵ no entanto, não foi possível rastrear a origem das imagens e verificar os rótulos do conjunto de dados. Como é extremamente comum pacientes com Hanseníase apresentarem múltiplas lesões cutâneas, ignorar esse fato poderia fazer com que o modelo tentasse memorizar padrões de pele e fornecesse previsões de resultados excessivamente otimistas. Um protocolo rigoroso de coleta e processamento de imagens é um pré-requisito para ajudar a prevenir a introdução de vieses sistemáticos nos

algoritmos.⁸ Não é de surpreender que um sinal fundamental para a Hanseníase, como a perda de sensibilidade, tenha contribuído significativamente para o algoritmo de IA. Esse sintoma, somado às características de infiltração difusa da pele, nervos dilatados, parestesia e hiperestesia nas mãos ou pés e perda de sobrancelhas ou cílios, foram ideais para construir um modelo logístico alinhado às observações clínicas e fornecer uma estimativa precisa. Além disso, superfícies de liquenificação e descamação também foram incorporadas ao modelo, como sinais que são principalmente indicativos de lesões não hanseníase. A presença de lesões cutâneas em mais de um local aumentou significativamente a probabilidade predita de Hanseníase.

Por outro lado, nosso estudo demonstra que o modelo apresenta mais dificuldades na detecção da Hanseníase a partir de imagens de pele do que a partir de metadados ou informações do paciente. Tipo, cor e sensibilidade da lesão, por exemplo, são descrições introduzidas por especialistas humanos, que sabem que essas características são úteis no diagnóstico da Hanseníase. Se ~~documentamos~~ o modelo exclusivamente com imagens e não com descrições clínicas, o algoritmo teria que aprender isso sozinho. Contrariamente às nossas observações, outros demonstraram correlações bastante boas entre dados dermatológicos e características visuais das lesões cutâneas.¹⁶ Embora entendamos que a coleta de dados clínicos pode ser desafiadora e exigiria treinamento, desenvolvemos um aplicativo para ajudar os coletores de dados utilizarem adequadamente o algoritmo, aumentando ainda mais sua aplicabilidade. Este aplicativo não pretende ser uma ferramenta de diagnóstico, mas sim uma ferramenta que ajuda o médico ou profissional de saúde a identificar lesões suspeitas de Hanseníase e, como tal, pode acelerar o encaminhamento aos profissionais de saúde para um diagnóstico adequado.

Um modelo só funcionará bem quando dados adicionais de novos pacientes de múltiplas origens geográficas forem usados para continuar a treiná-lo e melhorá-lo. Isso é possível quando a posterior coleta de dados segue um protocolo semelhante ao utilizado para o treinamento do modelo, aprimorado por ferramentas digitais de coleta de dados para evitar erros manuais. A transferibilidade do modelo em diferentes populações deve ser considerada.¹⁷ Em 2018, Han et al.¹⁸ treinaram o seu modelo de lesão cutânea numa população asiática e testaram-no numa população europeia. Isso resultou em uma precisão de 55,7% nas 10 classes do Dermfit,¹⁸ que foi significativamente menor do que a de outros modelos treinados e testados no Dermfit (81%).¹⁹ Essa diminuição na precisão pode ser atribuída ao diferenças nas manifestações cutâneas entre as populações ou a falta de

transferibilidade em recursos aprendidos entre conjuntos de dados, devido a protocolos de aquisição de imagens. Por esse motivo, será fundamental treinar o modelo AI4leprosy em populações múltiplas e diversas.

Colaboradores

O conceito do AI4Leprosy foi estabelecido por AA, MOM, JLF, RRB, AMS, AT, ENS, AC, ST, FM, EB e GM e seu desenvolvimento completo foi coordenado por LS, JLF, GM, RB, EB e JC. A coleta de dados e imagens foi realizada por RRB, RB, PTSS, AMS, JACN, ES. Os experimentos foram realizados por RB1, ST, RRB, PTSS, YX, RB2 e MS, enquanto a curadoria de dados foi realizada por RRB, PTSS e YX; a análise e modelagem dos dados foram feitas por YX, MOM, JLF, RRB, PTSS e MS. A análise dos resultados e o desenho global do protocolo foram realizados por LS, MG, AAN, JLF com contribuições de KW, JC, AT, CS, DS, AC, AA. A perícia em hanseníase foi fornecida por CS, DS, AC e AA. O rascunho completo foi desenvolvido por RRB, YX, MOM, JLF e AA. Todos os autores revisaram e aprovaram a versão final.

Financiamento Este estudo foi financiado pela Fundação Novartis, com experiência em espécie da Microsoft AI4Health.

Declaração de interesses

Não declaramos interesses conflitantes.

Agradecimentos

Expressamos nossa gratidão a John Kahan da Microsoft por defender e impulsionar esta parceria desde o início e por seu apoio incomparável a toda a iniciativa. Agradecemos a Johannes Boch da Fundação Novartis por apoiar a coordenação desta iniciativa em vários momentos. E somos gratos a todos os pacientes pela disposição em participar do estudo.

Materiais complementares

O material complementar associado a este artigo pode ser encontrado na versão online em [doi:10.1016/j.lana.2022.100192](https://doi.org/10.1016/j.lana.2022.100192).

Referências

- Cooreman E, Gillini L, Pemmaraju V, et al. Diretrizes para diagnóstico, tratamento e prevenção da hanseníase. *Órgão Mundial de Cura*. 2018;1.
- Barbieri RR, Manta FSN, Moreira SJM, et al. Reação em cadeia da polimerase quantitativa no diagnóstico da hanseníase paucibacilar: estudo de acompanhamento. *PLoS Negl Trop Dis*. 2019;13(3). <https://doi.org/10.1371/journal.pntd.0007147>.
- OMS. Registro epidemiológico semanal. Atualização global sobre hanseníase, 2018: avançando em direção à hanseníase. *Wkly Epidemiol Rec*. 2019;94.
- Richardus JH, Tiwari A, Barth-Jaeggi T, et al. Profilaxia pós-exposição à hanseníase com rifampicina em dose única (LPEP): um programa de viabilidade internacional. *Lanceta Glob Cura*. 2021;9(1):e81–e90. [https://doi.org/10.1016/S2214-109X\(20\)30396-X](https://doi.org/10.1016/S2214-109X(20)30396-X).
- Comitê de Especialistas em Hanseníase da OMS (1997: Genebra, Suíça) e Organização Mundial da Saúde. Comitê de Especialistas em Hanseníase da OMS: Sétimo Relatório. Organização Mundial de Saúde; 1998. <https://apps.who.int/iris/handle/10665/42060>.
- Ridley DS, Jopling WH. Classificação da hanseníase de acordo com a imunidade. Um sistema de cinco grupos. *Int J Lepr Outros Mycobact Dis*. 1966;34(3).
- Mieras LF, Taal AT, Post EB, Ndeve AGZ, van Hees CLM. O desenvolvimento de uma aplicação móvel para apoiar os profissionais de saúde periféricos no diagnóstico e tratamento de pessoas com doenças de pele em locais com poucos recursos. *Trop Med Infect Dis*. 2018;3(3). <https://doi.org/10.3390/tropicalmed3030102>.
- Esteve A, Kuprel B, Novoa RA, et al. Classificação do câncer de pele em nível de dermatologista com redes neurais profundas. *Natureza*. 2017;542 (7639):115–118. <https://doi.org/10.1038/nature21056>.
- Brinker TJ, Hekler A, Enk AH, et al. Uma rede neural convolucional treinada com imagens dermatoscópicas teve desempenho semelhante ao de 145 dermatologistas em uma tarefa de classificação de imagens de melanoma clínico. *Eur J Câncer*. 2019;111. <https://doi.org/10.1016/j.ejca.2019.02.005>.
- Marchetti MA, Liopyris K, Dusza SW, et al. Algoritmos de computador mostram potencial para melhorar a precisão dos dermatologistas no diagnóstico de melanoma cutâneo: resultados da colaboração internacional de imagens da pele 2017. *J Am Acad Dermatol*. 2020;82(3). <https://doi.org/10.1016/j.jaad.2019.07.016>.
- Marchetti MA, Codella NCF, Dusza SW, et al. Resultados do simpósio internacional de colaboração internacional em imagens da pele de 2016 sobre o desafio da imagem biomédica: comparação da precisão de algoritmos de computador com dermatologistas para o diagnóstico de melanoma a partir de imagens dermatoscópicas. *J Am Acad Dermatol*. 2018;78. <https://doi.org/10.1016/j.jaad.2017.08.016>.
- Dutta A, Zisserman A. O software de anotação VIA para imagens, áudio e vídeo. In: *Anais do MM 2019 - 27ª Conferência Internacional ACM sobre Multimídia*. 2019. <https://doi.org/10.1145/3343031.3350535>.
- Dean AG, Amer TG, Sunki G, et al. Epi InfoTM, um programa de banco de dados e estatísticas para profissionais de saúde pública. 2011.
- Wu H, Yin H, Chen H, et al. Uma abordagem de aprendizagem profunda baseada em imagens para diagnóstico automatizado de doenças inflamatórias da pele. *Ann Transl Med*. 2020;8(9). <https://doi.org/10.21037/atm.2020.04.39>.
- Baweja HS, Parhar T. Reconhecimento de lesões de lepra usando redes neurais convolucionais. In: *Anais da Conferência Internacional sobre Aprendizagem de Máquina e Cibernética*. 1, 2016. <https://doi.org/10.1109/ICMLC.2016.7860891>.
- Yang J, Sun X, Liang J, Rosin PL. Diagnóstico clínico de lesões cutâneas utilizando representações inspiradas em critérios dermatologistas. In: *Anais da Conferência da IEEE Computer Society sobre Visão Computacional e Reconhecimento de Padrões*. 2018. <https://doi.org/10.1109/CVPR.2018.00137>.
- Kawahara J, Hamarneh G. Diagnóstico visual de doenças dermatológicas: desempenho humano e de máquina. *arXiv*. Publicado on-line em 2019.
- Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classificação das imagens clínicas para tumores cutâneos benignos e malignos utilizando algoritmo de aprendizagem profunda. *J Invest Dermatol*. 2018;138(7). <https://doi.org/10.1016/j.jid.2018.01.028>.
- Kawahara J, Bentaieb A, Hamarneh G. Recursos profundos para classificar lesões cutâneas. In: *Anais do Simpósio Internacional de Imagens Biomédicas*. 2016. <https://doi.org/10.1109/ISBI.2016.7493528>.